

# Open chromatin structures regulate the efficiencies of pre-RC formation and replication initiation in Epstein-Barr virus

Peer Papior,<sup>1</sup> José M. Arteaga-Salas,<sup>2</sup> Thomas Günther,<sup>3</sup> Adam Grundhoff,<sup>3</sup> and Aloys Schepers<sup>1</sup>

<sup>1</sup>DNA Replication and Epigenetics group, Research Unit Gene Vectors, Helmholtz Zentrum München, German Research Center for Environmental Health, 81377 München, Germany

<sup>2</sup>Adolf Butenandt Institut, Ludwig Maximilians University, 80336 München Germany

<sup>3</sup>Heinrich Pette Institute, Leibniz Institute for Experimental Virology, 20251 Hamburg Germany

Whether or not metazoan replication initiates at random or specific but flexible sites is an unsolved question. The lack of sequence specificity in origin recognition complex (ORC) DNA binding complicates genome-scale chromatin immunoprecipitation (ChIP)-based studies. Epstein-Barr virus (EBV) persists as chromatinized minichromosomes that are replicated by the host replication machinery. We used EBV to investigate the link between zones of pre-replication complex (pre-RC) assembly, replication initiation, and micrococcal nuclease (MNase) sensitivity at different cell cycle stages in a genome-wide fashion. The dyad symmetry element

(DS) of EBV's latent origin, a well-established and very efficient pre-RC assembly region, served as an internal control. We identified 64 pre-RC zones that correlate spatially with 57 short nascent strand (SNS) zones. MNase experiments revealed that pre-RC and SNS zones were linked to regions of increased MNase sensitivity, which is a marker of origin strength. Interestingly, although spatially correlated, pre-RC and SNS zones were characterized by different features. We propose that pre-RCs are formed at flexible but distinct sites, from which only a few are activated per single genome and cell cycle.

## Introduction

Eukaryotic cells initiate their genome duplication from hundreds to several tens of thousands of sites, called replication origins. The genomic organization of replication origins is very different across the eukaryotic kingdom (Aladjem et al., 2006). In yeast, origins are mainly defined by DNA sequence. *Saccharomyces cerevisiae* replication origins are located in ~150-bp-long autonomous replicating sequences characterized by an 11-bp AT-rich consensus motif. *Schizosaccharomyces pombe* origins are 500–1,000-bp-long AT-rich sequences, which lack a consensus sequence but support autonomous replication (Aladjem, 2007). Both yeast species feature an excess of origin sites, and the local chromatin structure limits the number of active origins to ~400 (Breier et al., 2004). In multicellular eukaryotes, origins are defined independently of sequence, and various approaches

to identify essential features of origins have led to ambiguous results (Schepers and Papior, 2010). In humans, replication starts from an estimated 30,000 origins. The mode of origin recognition and activation is characterized by its flexibility and plasticity, allowing an adequate response to environmental constraints and diverse demands during differentiation (Aladjem, 2007).

Despite differences in origin definition, the principles of origin recognition are highly conserved from yeast to human. The first step is always the binding of the origin recognition complex (ORC) that acts as an interactive platform for the subsequent assembly of pre-replication complexes (pre-RCs) during the G1 phase of the cell cycle. Pre-RC formation is characterized by the reiterative loading of the minichromosome maintenance complex (Mcm2-7) that requires the help of two auxiliary proteins, Cdc6 and Cdt1 (Sivaprasad et al., 2006). The DNA binding features of ORC reflect the plasticity of origin recognition.

P. Papior and J.M. Arteaga-Salas contributed equally to this paper.

Correspondence to Aloys Schepers: schepers@helmholtz-muenchen.de

Abbreviations used in this paper: ChIP, chromatin immunoprecipitation; DS, dyad symmetry element; EBV, Epstein-Barr virus; FR, family of repeats; HMM, hidden Markov model; MR, MNase resistance; MSR, MNase-sensitive region; ORC, origin recognition complex; pre-RC, pre-replication complex; SNS, short nascent strand; TSS, transcription start site.

© 2012 Papior et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

Supplemental Material can be found at:  
<http://jcb.rupress.org/content/suppl/2012/08/09/jcb.201109105.DC1.html>

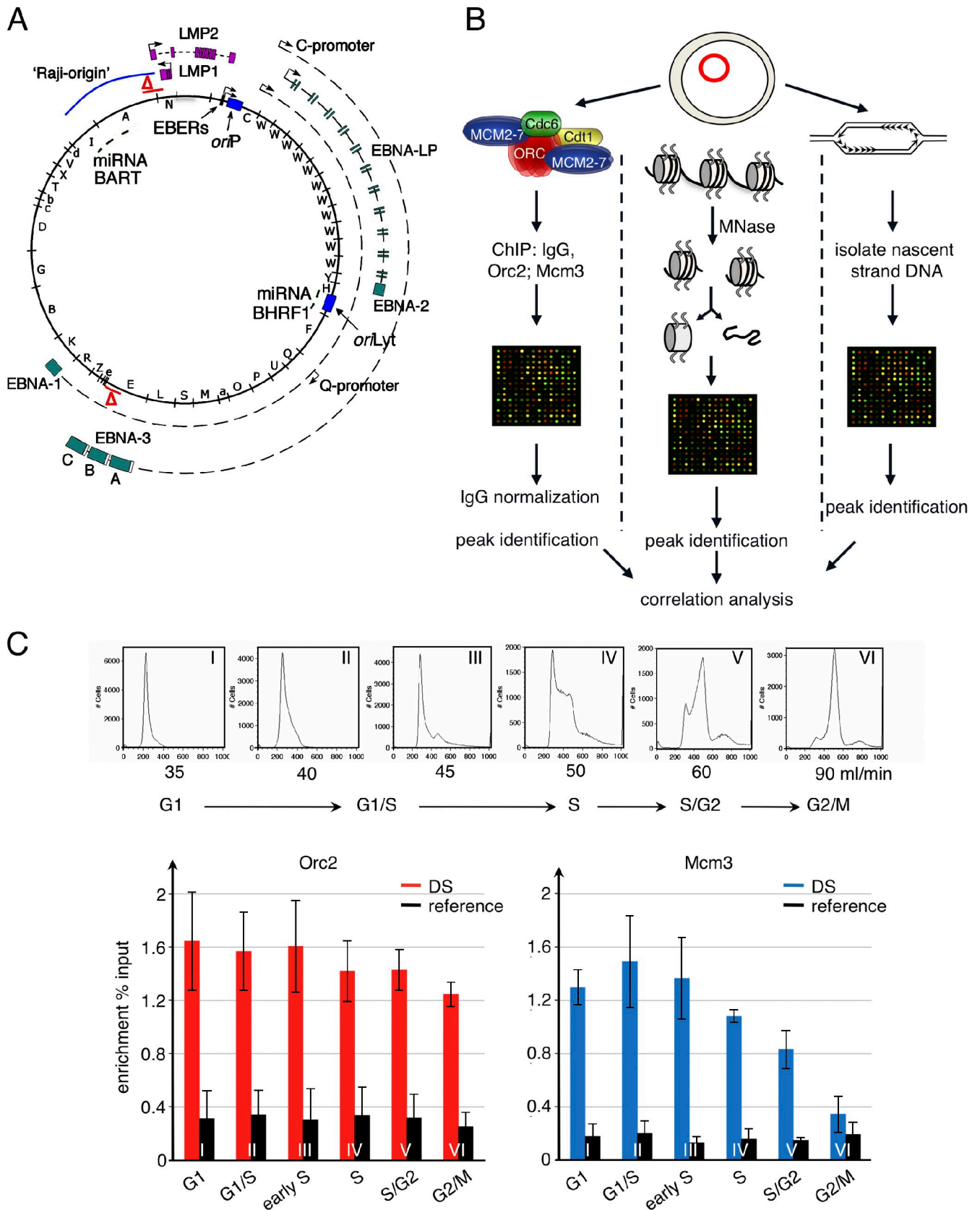


Figure 1. **Scheme of the EBV genome and experimental design for the analyses of pre-RC and SNS zones as well as mapping of MR profiles.** (A) Scheme of the circular EBV genome. In addition to the latent origins oriP (blue box) and the 14-kbp "Raji origin" (blue line), the lytic origin (oriLyt) is shown. The latent EBV nuclear antigens 1, 2, 3A-C, EBNA-LP genes (turquoise), and LMP1 and -2 (purple) are depicted, including their transcripts and promoters. The EBNA1 and -2 and the miRNAs regions (BART and BARF) are indicated (green lines). The Raji genome harbors two deletions (red Δ, nt 86,000–89,000 and 163,978–166,635). These regions do not produce array signals in comparison to the reference strain type I used for the design of the EBV microarray.

Although *S. cerevisiae* ORC (ScORC) recognizes origin-specific sequences, *S. pombe* ORC (SpORC) targets AT-rich DNA regions via an AT-hook extension of the SpOrc4 subunit (Aladjem et al., 2006; Masai et al., 2010). *Drosophila melanogaster* ORC (DmORC) has some bias for polyA tracts, whereas human ORC binds to DNA without any marked preference for distinct sequences (Vashee et al., 2003; Schaarschmidt et al., 2004; Balasov et al., 2007). ORC localizes to MNase-sensitive regions (MSRs), which are flanked by positioned nucleosomes (Berbenetz et al., 2010; Eaton et al., 2010; MacAlpine et al., 2010). In higher eukaryotic systems, additional features such as DNA topology, histone modifications, and chromatin structures might contribute to pre-RC binding and origin activation (Thomae et al., 2008; Méchali, 2010). For example, it has been postulated that pre-RCs assemble in zones of increased MNase sensitivity at the dihydrofolate reductase (DHFR) initiation region (Lubelsky et al., 2011). Genome-scale studies in human and mouse cells using short nascent strand (SNS) DNA as readout suggest that strong origins are often located in promoter regions, particularly transcription start sites (TSS), and map to CpG islands (Cadoret et al., 2008; Sequeira-Mendes et al., 2009; Cayrou et al., 2011). However, the high plasticity of ORC-DNA binding in human and other metazoan cells still hampers our understanding of origin formation and selection (Gilbert, 2010; Schepers and Papior, 2010).

In this study, we used Epstein-Barr virus (EBV) as a model to study the relationship between sites of pre-RC formation, origin activation, and nucleosome dynamics at origins in the background of human cells. EBV infects human B cells and establishes a persistent latent infection. The viral genome is maintained autonomously in proliferating cells and replicates once per cell cycle during S phase in synchrony with the host's chromosomal DNA (Adams, 1987; Yates and Guan, 1991). The latent origin, oriP, is the only cis-acting element required to sustain the autonomous state of the EBV genome (Yates et al., 1984). OriP is bound by the viral transactivator EBNA1. OriP was discovered due to its ability to support replication of plasmids, and it was believed that EBV's latent DNA replication initiates only at oriP. 2D gel analyses suggested that DNA synthesis frequently initiates outside oriP (Little and Schildkraut, 1995). Single molecule analyses demonstrated that initiation of DNA replication occurs at many sites across the viral genome, although only one or very few initiation events per genome occur in any given S phase (Norio and Schildkraut, 2001, 2004). OriP consists of two EBNA1 binding arrays: the family of repeats (FR) and the dyad symmetry element (DS). FR tethers the EBV genome to human chromosomes, thus ensuring stable retention (Marechal et al., 1999; Wu et al., 2000, 2002; Sears et al., 2003, 2004). DS is the origin element. The replication function of DS is based on EBNA1's ability to interact directly with ORC (Schepers et al., 2001).

This interaction allows a highly efficient assembly of pre-RCs at or near DS (Chaudhuri et al., 2001; Dhar et al., 2001; Schepers et al., 2001; Ritzi et al., 2003).

As mentioned in the previous paragraph, and even given the current high-throughput approaches, the complexity of mammalian genomes and the intrinsic flexibility in origin selection precluded these studies. Using EBV as a model system, we circumvented these problems by investigating the autonomous viral genome that, in many aspects, mimics a cellular chromosome. The advantage of this system is that the EBV genome is small enough to capture the entire molecule at high resolution on microarrays, yet large enough to allow formation of complex chromatin patterns and multiple replication origins. EBV has several advantages to study the replication initiation process in human cells. (a) Like all  $\gamma$ -herpesviruses, EBV persists as fully chromatinized genome that is exclusively replicated by the host replication machinery. This makes EBV an ideal reductionist model system to study replication across the entire genome. (b) DS can be used as an internal control site. DS has the unique advantage of being a well-characterized, highly specific and very efficient pre-RC site. (c) The high copy number of EBV genomes in combination with its small genome facilitates genome-wide experiments. We performed a comparative analysis of different pre-RC components, SNS mapping, and the pattern of mononucleosomes isolated at different stages of the cell cycle. Microarray analyses revealed highly similar DNA binding profiles for Orc2 and Mcm3, allowing the identification of 64 pre-RC zones in the EBV genome. We asked to what extent SNS and pre-RC zones coincide and whether these processes are characterized by MNase sensitivity patterns. Finally, we investigated a potential correlation of pre-RC assembly and replication initiation with nucleotide composition or proximity with TSSs. Our data demonstrates that pre-RC and SNS zones correlate spatially and are generally linked with regions of increased MNase sensitivity, though with distinct differences.

## Results

To study the parameters determining origin selection and activation in human cells, a comprehensive survey was performed using the EBV genome of the Burkitt's lymphoma cell line Raji (Fig. 1 A). Using a custom-made 6-bp resolution tiling array of the EBV genome, the relationship between zones of pre-RC formation, replication initiation, and nucleosome dynamics at origins were analyzed at high resolution. We chromatin-immunoprecipitated (ChIP) Orc2 and Mcm3 as members of the pre-RC from G1 cells and compared the array data with zones of actual initiation by measuring SNS DNA; we also compared them to mononucleosomal DNA isolated from cell cycle-fractionated chromatin, determining MNase-sensitive and -resistant regions (Fig. 1 B).

(B) Chart of the experimental set up to map pre-RC zones (left), MNase profiles (central), and SNS DNA (right). (C) Cell cycle phases of logarithmically growing Raji cells were separated by centrifugal elutriation. The DNA content of the different fractions was determined by FACS analysis (top, I-VI). The FACS profiles of one out of three experiments are shown. The quality of coprecipitated DNA was determined by quantitative PCR. The histograms show the mean values of three independent Orc2 (bottom left) and Mcm3 (bottom right) immunoprecipitations. The enrichments of Orc2 (red bars) and Mcm3 enrichments (blue bars) at the DS region are shown. The black bars indicate the enrichments of Orc2 and Mcm3 to a reference site. Error bars indicate mean  $\pm$  SEM.

## Genome-wide localization of Orc2 and Mcm3

To identify pre-RC zones, we cell cycle–fractionized cells using centrifugal elutriation (Ritzi et al., 2003) and performed ChIP with Orc2- and Mcm3-specific antibodies (Fig. 1 C). Orc2 binding to DS is cell cycle independent, whereas Mcm3 binding is clearly cell cycle regulated (Ritzi et al., 2003). The reference near oriLyt shows reduced amounts of Orc2 and Mcm3. Three biological replicates of Orc2- and Mcm3-specific precipitations and IgG controls of G1 chromatin (fraction II in Fig. 1 C) were hybridized against input DNA to the tiling array and analyzed (see Fig. S1 A for experiments designed to control potential biases introduced by linear amplification of ChIP material). The mean values of three independent Orc2/Mcm3 (Cy5) and input (Cy3) log<sub>2</sub> ratios were normalized against the IgG/input log<sub>2</sub> ratios. A sliding window of 150 bp was used to smooth the signal, and we then identified ChIP-enriched sites using a hidden Markov model (HMM; see Materials and methods). As expected, both Orc2 and Mcm3 show the most prominent enrichment at DS (Fig. S1 B). However, in addition to DS, many reproducible albeit less pronounced signals were observed across the EBV genome.

To determine the best possible resolution and to differentiate between background and true signals, we used several criteria. First, we considered the influence of the fragment length of the input DNA in the resolution of microarrays. Fig. S1 D simulates the resolution of an isolated binding site (top left) or of two neighboring binding sites (top right) with a uniform fragment population of 700 bp (see Materials and methods for the deduction of the formula for signal calculation). The simulated profile of a single signal has the shape of a triangle centered at the binding site with a width of twice the fragment length. Thus the fragment length has no influence on the resolution of a single signal per se, but may affect the separation of neighboring signals. When two binding sites are separated by less than the fragment length, their peaks will not be resolved, and appear as a trapezoid. The fragmentation process of a ChIP experiment, however, generates a population of fragments with varying lengths. Fig. S1 C shows the length distribution for one of our ChIP experiments. Fragments of ~700 bp are the most abundant. Fig. S1 D also shows signal simulations assuming the fragment length distribution shown in Fig. S1 C. Although the presence of large fragments broadens the total width of the peaks, the contribution of smaller fragments increases the overall resolution (Fig. S1 D, bottom left). As a consequence, it is possible to resolve individual binding sites within a distance of less than the mean (or median) fragment length (Fig. S1 D, bottom right). How well these peaks are resolved will ultimately depend on baseline fluorescence and noise levels. In summary, these simulations suggest that the resolution could in fact be higher than the mean fragment length.

Second, to minimize false positives within the obtained Orc2 and Mcm3 zones, we included for further analyses all peaks with a width of ≥400 bp (Fig. 2 A, see Materials and methods). Because peaks ≥400 bp are not representative for a specific site, we use the term “zone” for a region of adjacent probes with elevated signals. Note that this definition is different

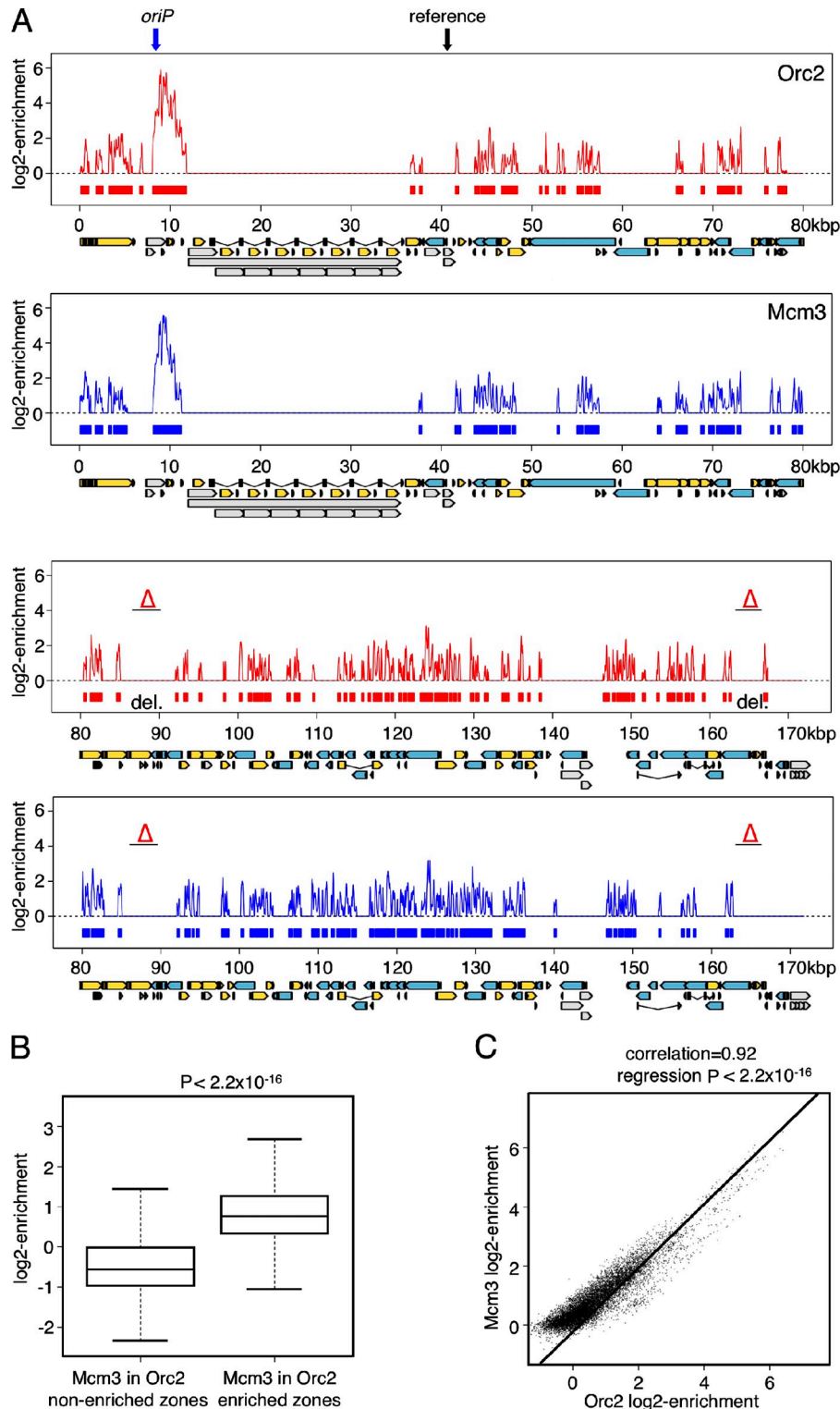
from a “replication initiation zone” describing a large region with delocalized initiation (Dijkwel et al., 1991). The Orc2 and Mcm3 profiles are highly similar, and Mcm3 log<sub>2</sub> ratios at Orc2 enriched zones have a significantly higher mean than at Orc2 nonenriched zones (Fig. 2 B;  $P < 2.2 \times 10^{-16}$ , one-sided Student’s *t* test). A linear regression of Orc2 and Mcm3 log<sub>2</sub> ratios at pre-RC zones confirmed a significant fit (Fig. 2 C;  $P < 2.2 \times 10^{-16}$ , regression *F*-test) and a high correlation (Pearson correlation = 0.92) between the enrichments. These results suggest that it is appropriate to combine Orc2 and Mcm3 log<sub>2</sub> ratios to define pre-RC enrichments. However, because Mcm3 but not Orc2 is essential for initiation once pre-RCs are formed, to define pre-RC zones we included zones with probes enriched not only with Mcm3 and Orc2 but also with Mcm3 only. From the identified 64 pre-RC zones, 55 are enriched in at least 5% of their width with both Mcm3 and Orc2, and 9 are Mcm3-only zones (Fig. 3 A). Detailed information about the location and composition of the zones is given in Table S1.

## Relationship between pre-RC zones and MSRs

Increasing evidence suggests that defined chromatin structures contribute to the definition of origins and that increased MNase sensitivity is one conserved feature of eukaryotic origins (Berbenetz et al., 2010; Eaton et al., 2010; Gilbert, 2010; Lubelsky et al., 2011; Givens et al., 2012; Xu et al., 2012). To confirm whether the positions of origins correlate with increased MNase accessibility, we generated MNase profiles of the EBV genome. Because we speculated that the MNase profile at origins might change dynamically during the cell cycle, we first isolated mononucleosomal DNA from MNase digested G1 chromatin (Fig. S2 A). A common misconception is to interpret MNase sensitivity as equivalent to nucleosome depletion. Yet, MNase sensitivity can also be produced by other factors (e.g., nonhistone proteins). Also, regions of extended high MNase protection are not digested to mononucleosomes and appear similar to MSRs. In this study, we define an MSR as a region of at least 150 bp in which all probes have a negative input/MNase ratio, which is indicative of increased MNase sensitivity (Fig. 3 B; see Materials and methods). A box plot shows that the MNase resistance (MR) was significantly lower in pre-RC enriched zones than in nonenriched zones (Fig. S2 B;  $P < 2.2 \times 10^{-16}$ , one sided Student’s *t* test). 81.3% of the pre-RCs are located in MSRs (mean probability of MSR being located in a random pre-RC = 0.0173). To evaluate whether the locations of probes in pre-RC zones and MSRs are independent, we used a two-way contingency table. A  $\chi^2$  test rejected the null hypothesis ( $P < 2.2 \times 10^{-16}$ ; Fig. S2 C). These results confirm the relationship between pre-RC zones and MSRs. Furthermore, the anti-correlation of the ChIP and MNase profiles clearly suggests that the ChIP signals are not systematic random noise.

## Pre-RC assembly and replication initiation zones correlate

The observation of 64 potential pre-RCs called into question how many of them can function as replication initiation sites.



**Figure 2. Orc2 and Mcm3 localize at highly similar locations and correlate with MSRs.** (A) Orc2 and Mcm3 ChIP experiments were performed with fraction II of the elutriated Raji cells. Orc2 (red) and Mcm3 (blue) enriched zones (width  $\geq 400$  bp) are plotted as a function of the EBV genome. Solid lines indicate the log<sub>2</sub> enrichments at the identified zones and the rectangles below indicate the width of each zone. EBV genes encoded in the upper strand (yellow boxes) and the lower strand (blue) are shown; repetitive elements are shown in gray. OriP and the reference site as well as the two Raji deletions (del.) are indicated. (B) A box plot analysis of Mcm3 log<sub>2</sub> enrichment in Orc2 enriched and nonenriched zones indicates a significant difference between the mean signals. (C) A linear regression of Mcm3- and Orc2-enriched probes in pre-RC zones confirms a significant relationship between the enrichments.

To identify active initiation sites, we isolated SNS DNA from asynchronous cells using alkaline gel electrophoresis (Kamath and Leffak, 2001). On average, we obtained  $<10$  ng SNS DNA from  $10^8$  cells, which is in the range of the expected amount (Cadoret et al., 2008). Nonproliferating cells did not yield any SNS DNA (not depicted). We prepared two independent samples, which were quality controlled by quantitative PCR at the hypoxanthine-guanine phosphoribosyltransferase (HPRT)

origin and at reference regions (Fig. S3 A; Cohen et al., 2002). SNS preparations were amplified and hybridized against genomic input DNA. As for ChIP DNA, we used Southern blot analysis with EBV-specific probes to monitor the length of viral DNA fragments (Fig. S1 C). To identify SNS-enriched zones we used the criteria described in Materials and methods. Most importantly, to avoid an overlap with Okazaki fragments, we omitted all potential SNS zones with a width of  $<400$  bp.

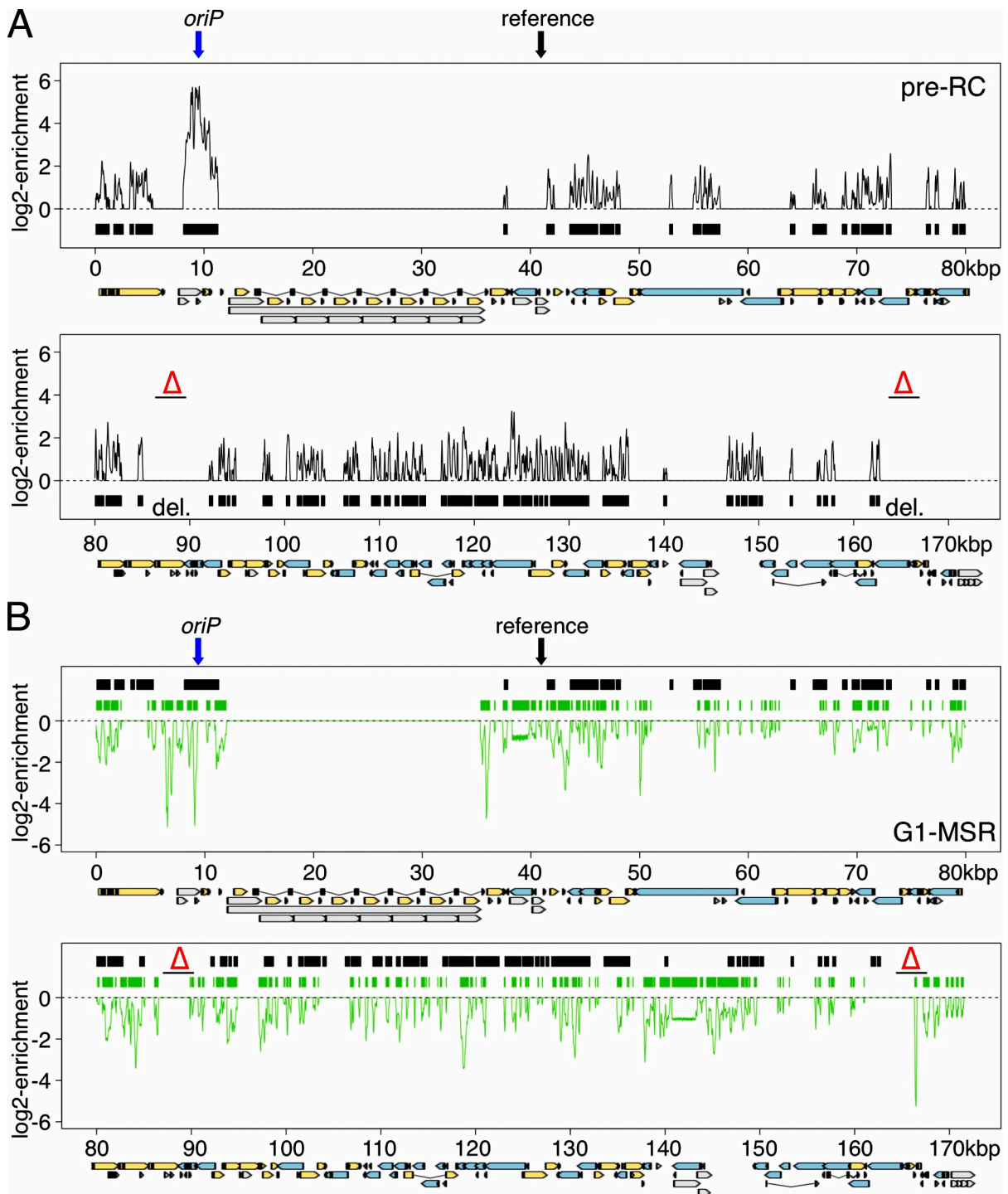


Figure 3. **Relationship between pre-RC zones and MSRs.** (A) 64 overlapping Orc2 and Mcm3 zones and Mcm3-only zones  $\geq 400$  bp were defined as pre-RC zones (black boxes). (B) MSRs (green), defined as regions  $\geq 150$  bp with negative MR versus genomic input ratios  $< 1$ , overlap with pre-RC zones (black).

Thus we identified 57 distinct potential SNS-enriched zones (Fig. 4 A). It was immediately obvious that replication initiates at many regions of the EBV genome and that DS is not the most prominent initiation zone. The region between the W repeats and nt 65,000 is relatively lacking in initiation zones, which is in line with studies from the Schildkraut laboratory (Fig. S3 B; Norio and Schildkraut, 2001, 2004).

One paradigm of the replication initiation model is that initiation occurs at or near pre-RC sites. Our observation that SNS log<sub>2</sub> ratios at pre-RC enriched zones have a significantly higher mean than at pre-RC nonenriched zones supports this hypothesis (Fig. 4 B;  $P < 2.2 \times 10^{-16}$ , one-sided Student's *t* test). 46 out of 57 SNS zones (81%) overlap with at least 5% of their width with a pre-RC zone (Table 1). Several arguments

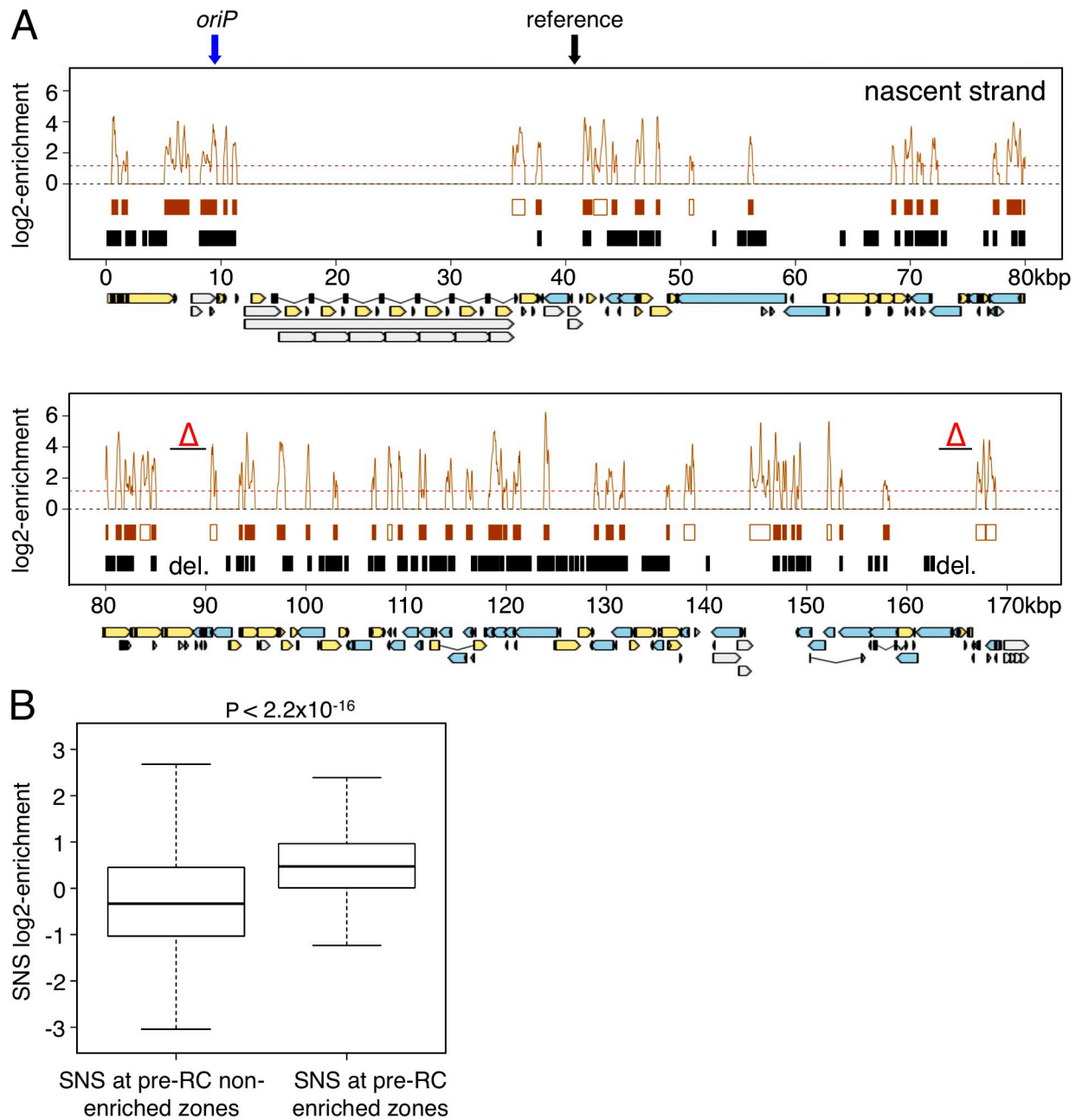


Figure 4. **Relationship between pre-RC and SNS zones.** (A) Enriched SNS zones (width  $\geq 400$  bp). SNS zones overlapping with at least 5% of their width with a pre-RC zone are indicated by the brown boxes. The nine nonoverlapping SNS zones are depicted as open boxes. The red line indicates the mean log<sub>2</sub> enrichment of the weakest SNS zone. The EBV map is the same as in Fig. 2. (B) A box plot of SNS log<sub>2</sub> enrichment at pre-RC enriched and nonenriched zones confirms a significant difference between the mean signals.

Table 1. **The majority of SNS zones overlap with at least 5% of their width with a pre-RC zone, a relationship that is also found in topSNSs**

Zone type	SNS	pre-RC
Total	57	64
Overlapping	46 (80.7%)	43 (67.2%)
Nonoverlapping	11 (19.3%)	21 (32.8%)
Top 30% enriched zones	17	19
Overlapping	14 (82.4%)	12 (63.2%)
Nonoverlapping	3 (17.6%)	7 (36.8%)

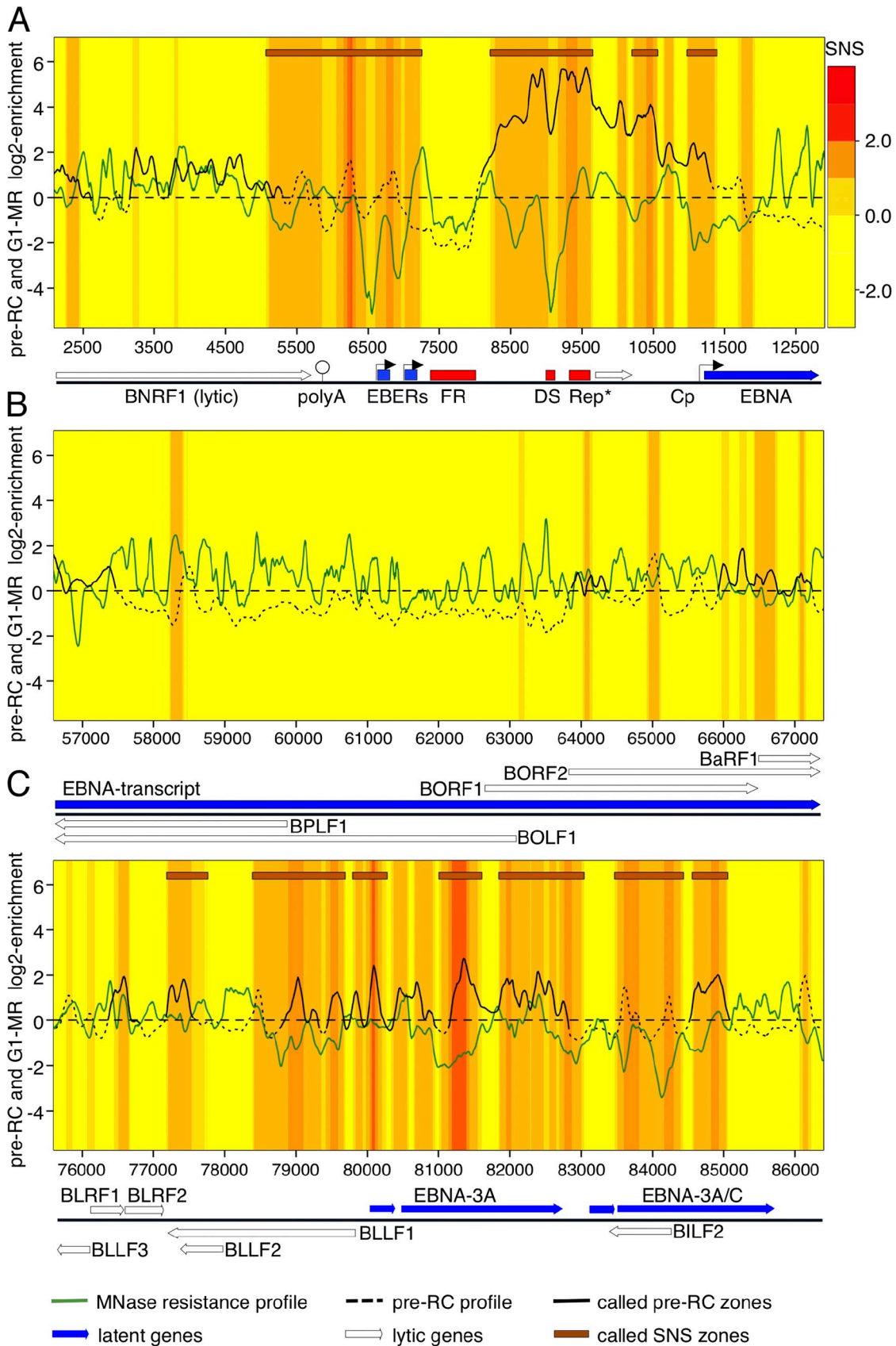


Figure 5. Heat map at three different EBV regions. (A–C) SNS heat map with pre-RC (black line) and G1-MNase profiles (green line). The SNS log<sub>2</sub> enrichment efficiency at an enlargement of the oriP region (A) and two exemplary regions (B and C) is shown. The SNS values are presented as heat maps, with red and yellow indicating high and low initiation activity, respectively. Dashed black lines indicate all pre-RC log<sub>2</sub> enrichments, whereas solid black lines represent only those pre-RC signals passing our filter for pre-RC zones. True SNS zones are marked by brown rectangles above the graph. The positions



might explain why 19% of the identified SNS zones do not overlap. First, our stringent criteria in defining pre-RC zones might exclude some true positive zones. For example, reducing the cut-off size for pre-RCs to 300 bp increases the number of potential pre-RC zones from 64 to 79, and the overlap between pre-RC and SNS zones raises the number from 81% to 89% (not depicted). Second, SNS zones not overlapping with a pre-RC zone are located in extended MSRs, and the majority have pre-RC signals within a short distance (Fig. S3 C). This suggests that SNS zones are spatially linked with pre-RC zones, although they are not located at identical sites. A list of all 57 SNS zones, as well as their mean and maximum peak intensities, is given in [Table S5](#). [Tables S2 and S6](#) contain detailed information about SNSs not overlapping with pre-RCs and vice versa.

These data suggests that many pre-RCs might also function as initiation sites, although not all potential origins are necessarily used in every EBV genome and cell cycle. We next examined a potential link between the mean efficiencies of pre-RC assembly and origin activation. To verify this, we compared SNS and pre-RC log<sub>2</sub> enrichments at SNS zones with a linear regression (Fig. S3 D). The regression provides a significant fit ( $P < 2.2 \times 10^{-16}$ , regression *F*-test), but the overall correlation of 0.27 is low. Because strong origins need to be efficient in both pre-RC assembly and initiation, we examined a potential correlation between both activities. Table 1 shows that 82.4% of the 30% strongest SNS zones (topSNS,  $n = 17$ ) overlap by at least 5% of their width with a pre-RC zone, the majority overlapping with one of the 30% strongest pre-RC zones (top-pre-RC,  $n = 19$ ). When analyzing top-pre-RCs and topSNSs in more detail, a relationship between these became obvious. Using a two-way contingency table, we tested the null hypothesis that the locations of probes in topSNSs and top-pre-RCs are independent. A  $\chi^2$  test rejected the null hypothesis ( $P = 5.6 \times 10^{-16}$ ; Fig. S3 E). We conclude that a significant relationship between top-pre-RCs and topSNSs exists. However, this association does not extend to all pre-RC and SNS zones. At present it is unclear which parameters determine the relationship. A list of all top-pre-RCs and topSNSs, including their mean and maximum peak intensities, is given in [Tables S3 and S7](#).

Our previous data demonstrated that DS is flanked by positioned nucleosomes (Zhou et al., 2005). To analyze the relationship between pre-RC assembly, MNase sensitivity, and initiation efficiency, we aligned these features using heat maps (Fig. 5). Fig. 5 A shows oriP, a multifunctional region, in which transcriptional activity, pre-RC assembly, replication initiation activity, and MNase sensitivity are spatially and functionally linked. Both oriP elements, FR and DS, are constantly bound by the EBV-transactivator EBNA1 and represent MSRs flanked by MNase-resistant regions (MRRs). Interestingly, both SNS and pre-RC signals peak not at DS but in the neighboring regions, confirming that EBNA1 targets ORC to a broad area (Schepers et al., 2001; Ritz et al., 2003). The pre-RC zone at oriP is flanked on one side by FR and on the other side by the C promoter

(nt 7,690–11,800). This zone contains three SNS zones, which suggests that multiple ORC molecules might bind to this region. The region between nt 5,100 and 7,250 represents one extended SNS and includes the noncoding EBER transcripts (Minaurovits et al., 1992). In Raji cells, only EBER1 is transcribed at a high level (Pratt et al., 2009). The promoter regions of both EBER genes are MNase sensitive and are characterized by the presence of +1 nucleosomes. Two pre-RC enrichments localized at the EBER promoters did not qualify as enriched zones because of our stringent scoring conditions. This example demonstrates that the criteria chosen to eliminate false positive signals and to efficiently reduce background noise come at the expense of sensitivity and might also eliminate true positive signals. Fig. 5 (B and C) shows two additional selected regions. The region between nt 57,000 and 67,000 displays three weak pre-RCs, which indicates that not every potential pre-RC zone is used as an initiation site (Fig. 5 B). The region between nt 76,000 and 86,000 has multiple pre-RC zones overlapping with SNS zones, which are preferentially located in MSRs; this suggests that replication initiation and increased MNase sensitivity are linked (Fig. 5 C).

#### The MNase sensitivity at pre-RC zones is dynamic over the cell cycle

Different studies demonstrate that origins are located in MSR (Berbenetz et al., 2010; Eaton et al., 2010; Gilbert, 2010; Lubelsky et al., 2011). To explore a potential MNase sensitivity at origins, we aligned and plotted the mean mononucleosome log<sub>2</sub> enrichments of G1 cells in a  $\pm 1,000$  bp window surrounding the maximum peak of the 64 pre-RCs (Fig. 6 A, panel 1; see Materials and methods). The alignment of all pre-RCs indicates only a moderate MNase sensitivity during G1. The standard deviation of the mean profiles confirms this analysis (Fig. S4 A). As control, we also aligned the  $\pm 1,000$ -bp neighborhood of 250 randomly selected positions across the EBV genome (Fig. 6 A, panel 2). Next, we examined whether the extent of MNase sensitivity is linked to the efficiency of pre-RC formation. The alignments of the 30% least prominent pre-RCs (bot-pre-RC,  $n = 19$ ; [Table S4](#)) and the top-pre-RCs indicate only small differences in MNase sensitivity at pre-RCs in G1 phase chromatin (Fig. 6 A, panels 3 and 4; and Fig. S4 A).

Pre-RC formation is limited to the G1 phase of the cell cycle, and pre-RCs are disassembled after origin firing. Therefore, we determined whether the MNase sensitivity at pre-RCs changes over the cell cycle. Fig. 4 B shows mean pre-RC and MR profiles, now also including the S- and G2/M-MR (S phase, fraction IV; G2/M, fraction VI of Fig. 1 C). In contrast to G2/M and G1 cells, we observed a significant increase in MNase accessibility at pre-RC zones during S phase, whereas on average the MR at pre-RC flanking regions do not change over the cell cycle (Fig. 6 B, left). Top-pre-RCs display pronounced MNase sensitivity during S phase, whereas this link is not obvious in bot-pre-RCs (Fig. 6 B, center and right; see Fig. S4 B for standard deviations).

of the oriP elements FR, DS, and Rep\* (red boxes), the C promoter, and the RNA Pol III transcribed EBER1 and 2 (arrows) are indicated. Latent (blue arrows) and silent lytic genes (white arrows) are depicted.

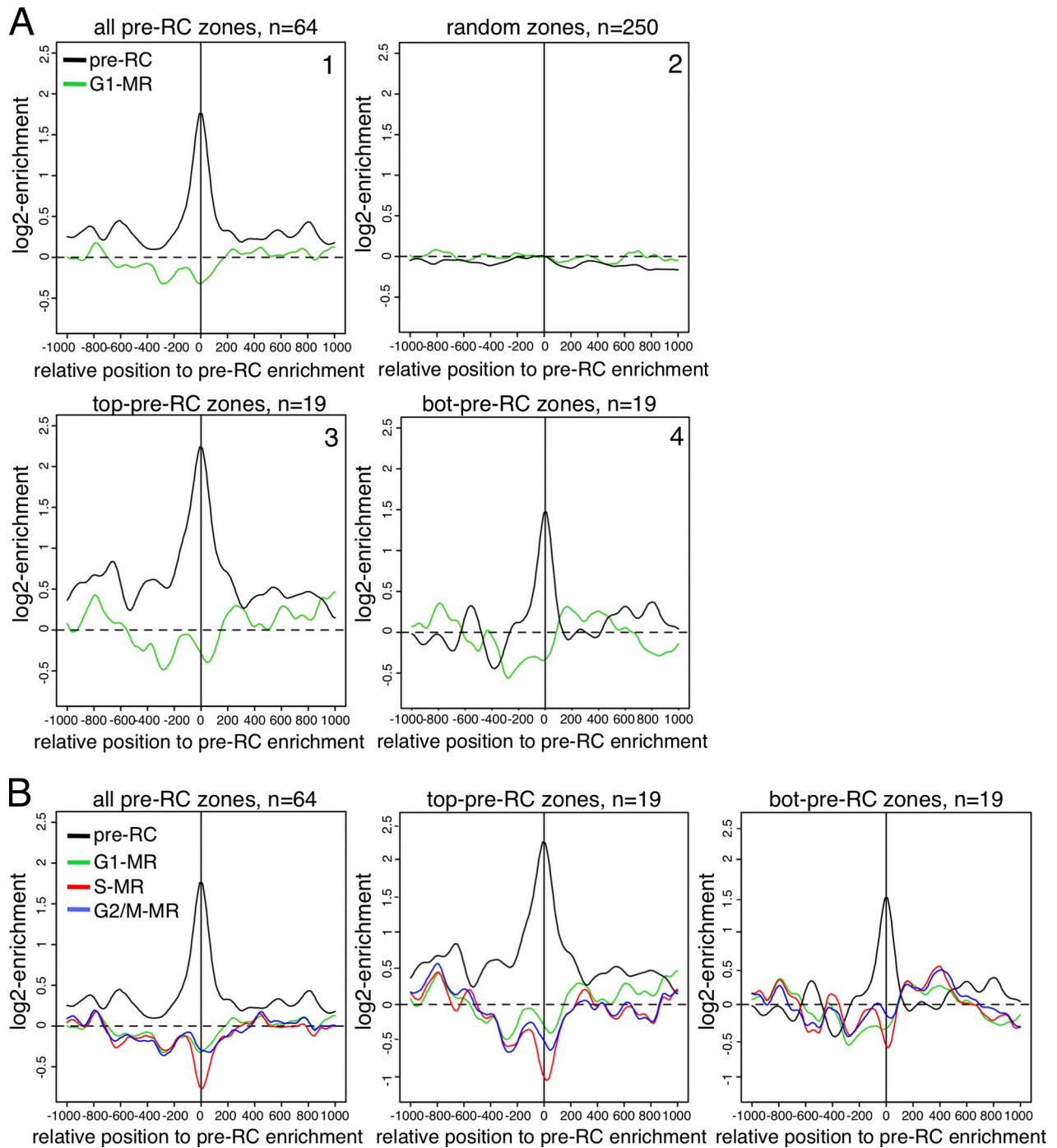


Figure 6. **Mean MR profiles at pre-RCs are dynamic over the cell cycle.** (A) Mean profile of pre-RC (black) and G1 phase MR log<sub>2</sub> enrichments (green) in a  $\pm 1,000$ -bp window centered at: (1) the maximum peak of all pre-RCs (top left), (2) 250 randomly chosen locations (top right), (3) the maximum peak of the top-pre-RCs (bottom left), and (4) the maximum peak of the bot-pre-RCs (bottom right). (B) Mean profile of pre-RC (black), G1 phase MR (green), S phase MR (red), and G2/M phase MR (blue) log<sub>2</sub> enrichments. The log<sub>2</sub> enrichments in a  $\pm 1,000$ -bp window are centered at the maximum peaks of pre-RCs as in A.

It is possible that pre-RCs protect DNA against MNase digestion, an effect that is lost when pre-RCs and ORC are disassembled in human cells after origin activation. The increased MNase sensitivity is S phase specific, whereas the average profile of the G2/M fraction is similar to the G1 fraction. It is important to note that the increased MNase sensitivity does not necessarily mean that nucleosomes are evicted, but that structural changes might occur that expose DNA, thus increasing the accessibility.

#### Efficiency of replication initiation correlates with MNase sensitivity

Pre-RC formation and replication initiation are independent processes that occur in different cell cycle phases. Although most SNS and pre-RC zones overlap, they are linked to different cell cycle phases, which might result in different chromatin states (Tables 1, S1, and S5). Therefore, we next analyzed the mean MR profiles of SNS zones and their standard deviations

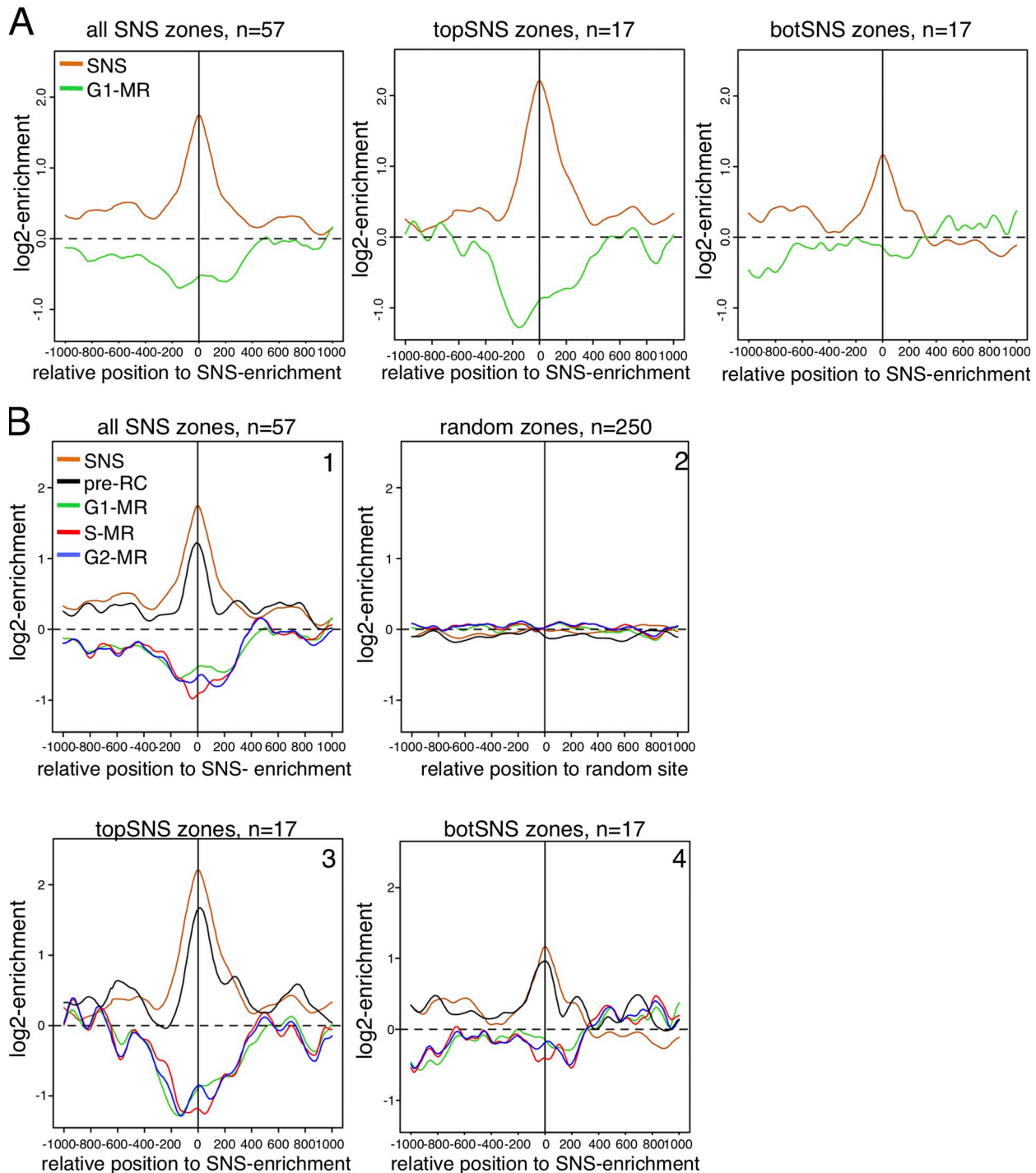


Figure 7. **Origin activity is linked to increased MNase sensitivity.** (A) Mean profile of SNS (brown) and G1-MR log<sub>2</sub> enrichments (green) in a  $\pm 1,000$ -bp window centered at the maximum peak of all SNSs (left), the topSNSs (center), and the botSNSs (right). (B) Mean profile of SNS (brown), pre-RC (black), G1-MR (green), S-MR (red), and G2/M-MR (blue) log<sub>2</sub> enrichments in a  $\pm 1,000$ -bp window centered at the maximum peak of the SNS zones as described in A.

(Fig. 7 A and Fig. S4 B). In G1 cells, SNSs are characterized by increased MNase sensitivity, whereas the topSNSs are characterized by a pronounced MNase sensitivity. A decreased sensitivity is observed within the 30% least prominent SNS zones (botSNSs,  $n = 17$ ; Table S8). This finding is in line with a recent report by Lantermann et al. (2010), who suggested a correlation between origin strength and MSRs for *S. pombe* origins. In contrast to the situation at pre-RC zones, no cell cycle dependence

was evident in the MNase sensitivity profiles at SNS zones. A pronounced MNase sensitivity was particularly evident in all phases of the cell cycle for the topSNSs (Fig. 7 B, panel 3). This relationship is missing in botSNSs (Fig. 7 B, panel 4). Randomly selected control sites show no regular pattern (Fig. 7 B, panel 2). We conclude from these analyses that both pre-RC and SNS zones show different features with respect to MNase sensitivity. Although pre-RCs are characterized by

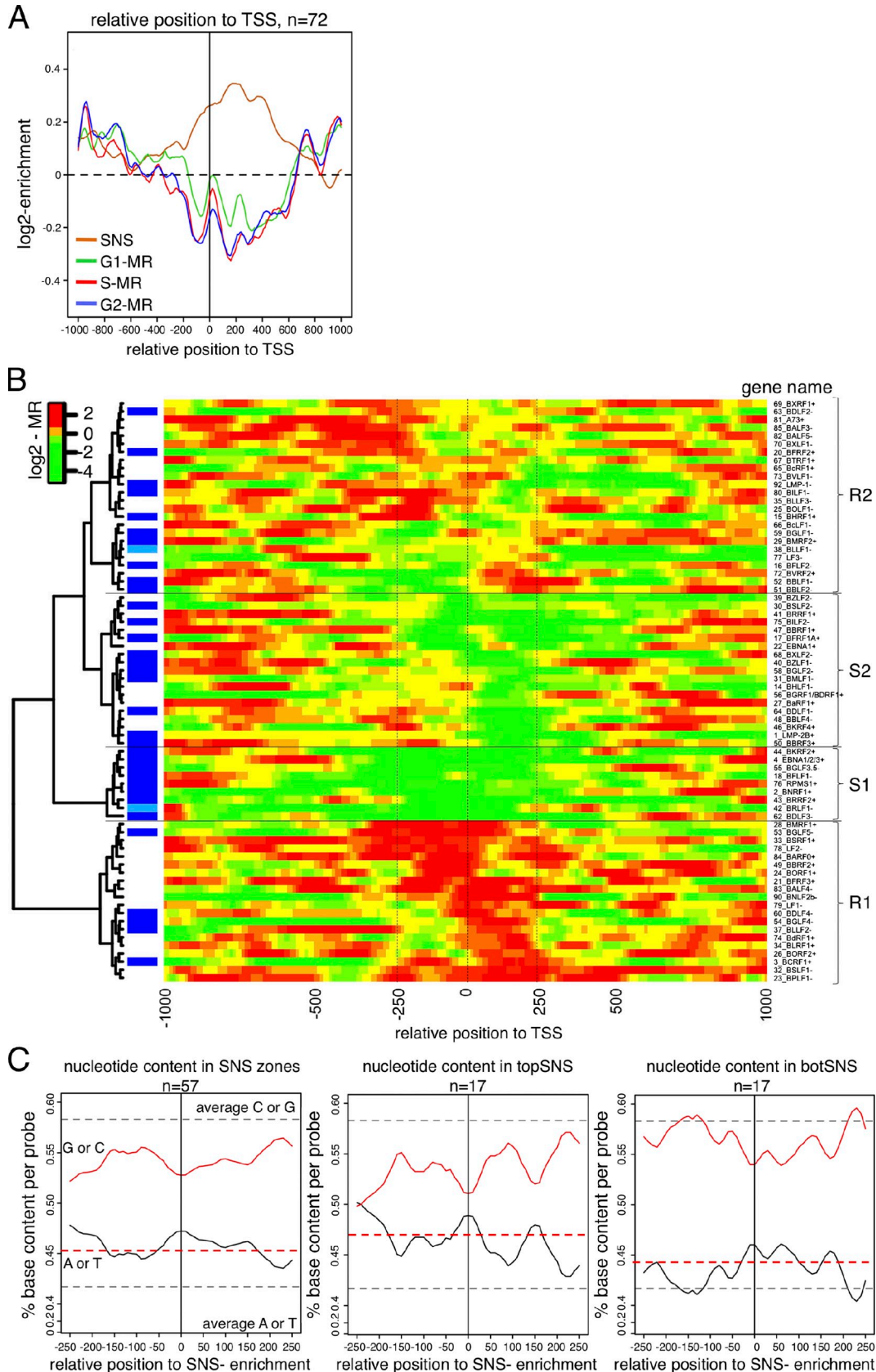


Figure 8. **Nucleosome occupancy at TSSs, and local nucleotide composition at pre-RC and SNS zones.** (A) Profiles of SNS (brown), G1-MR (green), S-MR (red), and G2/M-MR (blue) log<sub>2</sub> enrichments in a ±1,000-bp window centered at 72 TSS. (B) Heat map of G1-MR log<sub>2</sub> enrichments in a ±1,000-bp window centered at the 72 TSSs. The red color indicates high nucleosome resistance, whereas green indicates low. The IDs and names of the 72 genes are shown on the right axis. The dendrogram was obtained with a hierarchical cluster analysis based on the neighborhood [TSS - 250 bp, TSS + 250 bp]. Based on the dendrogram, we define four clusters of TSS: R1 (resistance, n = 22), S1 (sensitive, n = 9), S2 (n = 3), and R2 (n = 21). SNSs located in

dynamic profiles, the efficiency of origin activation is clearly linked with the degree of sensitivity, reflecting an open chromatin state.

### Replication initiation at EBV promoter regions

Recent genome-wide studies in different systems show a link between TSS and replication origins (Cadoret et al., 2008; Sequeira-Mendes et al., 2009; Eaton et al., 2010; Karnani et al., 2010; Cayrou et al., 2011). In comparison to the human genome, the EBV genome is very gene dense and comprises ~100 genes within 170 kbp. Most of these genes are efficiently silenced in latently infected cells. A recent study using the Raji cell line indicated RNA polymerase II binding only at the EBER regions, the DS/Cp domain, the BART miRNA region, and the LMP promoters (Holdorf et al., 2011). To study the relationship between replication SNS zones and TSSs, we generated mean enrichment profiles of SNSs and cell cycle MR profiles aligned at the TSSs (Fig. 8 A and Materials and methods). We omitted from the analysis those genes with a distance of <500 bp between their TSSs, which resulted in 72 TSSs used for analyses (Table S9). On average, promoters were found to exhibit MR just upstream of the TSSs. These indicate cell cycle independence, and were positioned +1 and +2 nucleosomes within the gene body, with the +1 nucleosome peaking at TSS + 20 bp, and the +2 nucleosome peaking at TSS + 220 bp.

Fig. 8 A shows that the mean replication initiation activity is high in the region [TSS, TSS + 500 bp], peaking in an MSR in the gene body. The mean nucleosome phasing is similar to promoter regions with an elongating or stalled RNA Pol II (Schones et al., 2008). In total, 37 SNS zones are located in the region [TSS, TSS + 500 bp] of the 72 analyzed TSSs (see Table S9). 33 regions [TSS, TSS + 500 bp] have one SNS and two [TSS, TSS + 500 bp] have two SNSs. In comparison to the genome mean, these regions show an ~2.5-fold higher density of SNSs. Most of the analyzed TSSs represent silent promoters, which are reactivated in the productive cycle. These lytic genes are expressed in a sequential order and are accordingly classified as early or late genes. We hypothesized that a correlation exists between the MNase profile of these classes and replication initiation. To investigate this, we performed a cluster analysis of the 72 promoters according to their MNase sensitivity in the 500-bp region [TSS - 250 bp, TSS + 250 bp] (Fig. 8 B; see Materials and methods, "Cluster analysis and heat map generation"). Generally, two major groups can be defined. The majority of late lytic genes (28 out of 43; subgroups R1 and R2) represent genes with high MR. In contrast, the latent genes, the miRNA regions, and genes preferentially expressed in the early lytic phase (28 TSSs: 5 latent and miRNA, 13 early lytic) are characterized by increased MNase sensitivity (subgroups S1 and S2). The cluster analysis revealed that 71.4% of the TSSs in the S groups contain SNSs, whereas only 38.6% of TSSs in the

R groups have an SNS (Table S10). None of the five origins within R1 belong to the topSNSs, whereas five of the 10 S1-SNSs are topSNSs. These results suggest that TSSs with an open chromatin structure are more frequently associated with SNSs, especially with topSNSs, than they are associated with a more closed chromatin state. Active transcription is not a prerequisite for this association. Our finding of two different "gene expression classes" is in accordance with studies of epigenetic modifications in the Kaposi's sarcoma-associated herpesvirus (Günther and Grundhoff, 2010; Toth et al., 2010). These studies revealed that early genes tend to be more enriched, with chromatin marks that usually correlate with active transcription, whereas late genes are more enriched with repressive histone modifications. We conclude that herpesvirus genes destined for rapid expression upon reactivation preserve an open chromatin state during latency. Our data strongly suggest that the prime determinant of pre-RC formation and initiation is not transcriptional activity as such, but rather an open and dynamic local chromatin structure.

### Nucleotide preferences at pre-RC and SNS zones

Previous in vitro ORC binding and origin mapping experiments show that metazoan ORC does not display any sequence preference. Recent meta-analysis of replication origins in *Drosophila melanogaster* corroborated that the primary sequence together with active chromatin features contributes to ORC binding, although to a low degree (MacAlpine et al., 2010; Eaton et al., 2011). Cayrou et al. (2011) reported that *D. melanogaster* and mouse origins are characterized by GC-rich motifs. We investigated the nucleotide composition and the occurrence of dinucleotide motifs in a  $\pm 250$ -bp window surrounding the highest peaks of pre-RC and SNS zones. Table 2 shows that pre-RCs assemble without any nucleotide preference relative to the genome wide mean; we observed only very minor differences between top- and bot-pre-RC zones. We observed slight significant differences in the A/G/T content between top- and bot-pre-RCs, in particular minor advantages of CG and G stretches (Tables 3 and 4).

Table 2 also indicates that origin activation is moderately affected by the nucleotide composition. We observed an elevated A/T content at SNS zones in relation to genome mean, which is more pronounced at topSNS than at botSNS. The EBV genome has an A/T content of 41.7%, whereas the SNS zones display a mean A/T content of 45.6%, with topSNSs having a mean of 46.6%. Fig. 8 C visualizes the preference for A/T-rich sequences at SNS zones by plotting the mean nucleotide content in a  $\pm 250$ -bp window centered at their maximum peak, which confirms the increased A/T frequency at topSNSs. The analysis of AT dinucleotide pairs indicates a slight overrepresentation of any A/T pair at topSNSs in relation to genome mean. Conversely, we observed a slight bias in disfavor of C/G pairs.

[TSS, TSS + 500 bp] are indicated in blue. Light blue highlights those TSSs with two SNSs. (C) Mean profile of the nucleotide base content in a  $\pm 250$ -bp window centered at the maximum peak of: all SNSs (left), topSNSs (center), and botSNSs (right). The nucleotide content is depicted in red for G or C and black for A or T, respectively. The gray dashed horizontal lines in each of the three panels represent the mean GC (top) or mean AT (bottom) content in the EBV genome. The red dashed line represents the mean AT content at the SNS zones.

Table 2. Base composition in SNS and pre-RC zones

Zone type	As	Ts	Gs	Cs	CG	GC	CC	GG	GC or CG	TA	AT	AA	TT	AT or TA
All pre-RC	0.206	0.232	0.284	0.277	0.050	0.073	0.083	0.086	0.124	0.034	0.043	0.044	0.053	0.077
Top-pre-RC	0.204	0.218	0.294	0.285	0.055	0.078	0.089	0.090	0.134	0.032	0.040	0.042	0.046	0.073
Bot-pre-RC	0.220	0.235	0.270	0.275	0.043	0.070	0.081	0.078	0.113	0.035	0.046	0.051	0.055	0.080
All SNS	0.221	0.235	0.283	0.262	0.043	0.069	0.080	0.088	0.112	0.039	0.048	0.051	0.057	0.087
TopSNS	0.223	0.243	0.276	0.257	0.042	0.066	0.082	0.081	0.108	0.041	0.051	0.050	0.066	0.092
BotSNS	0.218	0.224	0.296	0.262	0.046	0.073	0.076	0.094	0.119	0.032	0.045	0.051	0.050	0.077
Genome mean	0.203	0.214	0.301	0.282	0.054	0.079	0.090	0.101	0.133	0.032	0.041	0.043	0.048	0.074

Summary of nucleotide features within enriched pre-RC and SNS zones. The numbers indicate the percentages of single nucleotides, dinucleotide pairs (AT, TA, GC, and CG), and of single dinucleotide pairs. The top and bottom values give the percentages of the strongest and weakest enrichment zones ( $n = 19$  for pre-RC;  $n = 17$  for SNS).

In summary the initiation process is moderately favored by A/T-rich stretches, independent from specific primary sequence motifs, whereas no correlation between the efficiency of pre-RC assembly and the underlying sequence can be detected. It is important to note that this relationship does not have any predictive power to explain why origins are placed where they are.

## Discussion

Significant progress has been made in understanding the features controlling DNA replication in the context of chromatin in mammals. However, mechanisms regulating the efficiencies of pre-RC formation and origin firing are still a conundrum. By analyzing pre-RC and SNS zones, as well as mononucleosome profiles from different cell cycle stages, we show that pre-RCs are characterized by an S phase-specific MNase sensitivity, and that the efficiency of origin activation correlates with increased MNase sensitivity. Given that latent EBV replication is akin to that of host cell DNA in nearly every aspect studied to date, there is every reason to believe that the findings of our study are extendable to mammalian chromatin.

The replicon paradigm that guided the search for replication origins for many years does not reflect origin selection and activation in metazoan cells (Rhind, 2006; Hamlin et al., 2008). In contrast to *S. cerevisiae*, which nearly follows the replicon model, metazoan pre-RCs are established at flexible sites in each genome. In frog embryos, the plasticity is extreme and suggests a random origin pattern (Harland and Laskey, 1980; Hyrien and Méchali, 1993). The flexibility in pre-RC formation has implications on ChIP experiments and makes the identification of binding sites very difficult: signals are diluted, and reliable parameters to allow for a clear distinction between enriched binding sites and background signals are missing (Gilbert, 2010; Hamlin et al., 2010; Schepers and Papior, 2010). In addition to the advantages described previously, our study of the parameters supporting pre-RC formation in the latent EBV replication system has the unique advantage of using a well-characterized, highly specific, and efficient pre-RC site at DS that serves as internal positive control. We detect many Orc2- and Mcm3-enriched sites throughout the EBV genome, which exhibit a

very high correlation between binding sites and efficiencies. To reduce background noise, we performed three independent experiments, which were normalized against IgG controls. The resulting Orc2 and Mcm3 profiles were highly similar, which allowed us to combine both profiles to one pre-RC profile. To eliminate false positive signals, we chose a cut-off width of 400 bp for the identified enriched zones, although the fragment distribution might have allowed a higher resolution. The resulting 64 pre-RC zones correlate with increased MNase sensitivity, providing further evidence that these signals are true positive pre-RC zones and not random noise caused by antibody or hybridization artifacts. Pre-RCs are distributed over the entire EBV genome. Some regions contain clusters of assembly sites, whereas other regions are relatively sparse in pre-RC zones. We conclude that pre-RC formation occurs at multiple places of the EBV genome, with DS being the dominant assembly site. Furthermore, not the full contingent but rather only a small subset of these sites are used per individual genome and cell cycle.

Nucleosomes limit the accessibility of DNA for binding partners, and increasing evidence suggests that nucleosome organization might be one defining parameter of replication origins (Berbenetz et al., 2010; Eaton et al., 2010; MacAlpine et al., 2010; Lubelsky et al., 2011; Givens et al., 2012; Xu et al., 2012). Open chromatin structures are often found at transcriptionally active regions. Also, chromatin remodeling complexes mobilize nucleosomes to allow origin formation (Collins et al., 2002; MacAlpine et al., 2004; Zhou et al., 2005; Cadoret et al., 2008; Sugimoto et al., 2008; Sequeira-Mendes et al., 2009). Here, we performed the first comparative genome-wide analysis between pre-RC and SNS zones and MR profiles generated at different stages of the cell cycle. We found that pre-RCs are characterized by a dynamic MNase pattern, which exhibits an increased sensitivity during S phase (Fig. 6). In an analogy to the extended pre-RC-specific DNaseI footprint in *S. cerevisiae*, it is conceivable that pre-RCs also protect mammalian origin DNA in G1 (Diffley et al., 1994). The increased MNase sensitivity during S phase is in line with previous findings that human ORC dissociates after origin firing, which is likely to result in increased enzymatic accessibility (Gerhardt et al., 2006; Siddiqui and Stillman, 2007). In G2/M phase the MNase profile at pre-RCs is similar to the G1 profile. This observation might

Table 3. Deviations versus genome mean (p-value, one sided Student's t-test)

Zone type	As	Ts	Gs	Cs	CG	GC	CC	GG	GC or CG	TA	AT	AA	TT	AT or TA
All pre-RC	0.23	3.5 × 10 <sup>-9</sup>	3.6 × 10 <sup>-6</sup>	0.07	0.01	1.1 × 10 <sup>-4</sup>	1.2 × 10 <sup>-3</sup>	2.2 × 10 <sup>-9</sup>	1.9 × 10 <sup>-4</sup>	0.13	0.06	0.22	2.7 × 10 <sup>-3</sup>	0.05
Top-pre-RC	0.46	0.21	0.16	0.40	0.26	0.38	0.33	0.01	0.44	0.47	0.34	0.44	0.26	0.39
Bot-pre-RC	2.6 × 10 <sup>-3</sup>	2.9 × 10 <sup>-4</sup>	2.9 × 10 <sup>-7</sup>	0.18	2.5 × 10 <sup>-5</sup>	1.6 × 10 <sup>-3</sup>	0.02	4.0 × 10 <sup>-8</sup>	5.0 × 10 <sup>-5</sup>	0.15	0.04	0.01	0.02	0.05
All SNS	1.8 × 10 <sup>-7</sup>	4.5 × 10 <sup>-8</sup>	3.4 × 10 <sup>-6</sup>	8.2 × 10 <sup>-8</sup>	2 × 10 <sup>-12</sup>	1.2 × 10 <sup>-9</sup>	1.4 × 10 <sup>-5</sup>	2.3 × 10 <sup>-7</sup>	5 × 10 <sup>-14</sup>	1.7 × 10 <sup>-6</sup>	3.8 × 10 <sup>-6</sup>	2.4 × 10 <sup>-5</sup>	1.4 × 10 <sup>-5</sup>	5.8 × 10 <sup>-8</sup>
TopSNS	6.0 × 10 <sup>-4</sup>	3.5 × 10 <sup>-5</sup>	4.0 × 10 <sup>-4</sup>	4.4 × 10 <sup>-4</sup>	3.0 × 10 <sup>-7</sup>	4.2 × 10 <sup>-6</sup>	0.03	1.8 × 10 <sup>-5</sup>	2.8 × 10 <sup>-8</sup>	1.8 × 10 <sup>-4</sup>	1.1 × 10 <sup>-4</sup>	0.01	1.8 × 10 <sup>-5</sup>	1.1 × 10 <sup>-5</sup>
BotSNS	0.02	0.07	0.27	1.7 × 10 <sup>-3</sup>	0.01	0.04	5.1 × 10 <sup>-4</sup>	0.08	0.01	0.38	0.09	0.02	0.28	0.24

Nucleotide feature deviations between pre-RC or SNS zones in relation to the genome mean. The p-values shown indicate the significance of the deviation between each nucleotide feature in Table 2 and the genome mean.

be explained by a rebinding of ORC. However, the reassembly of pre-RCs is not completed in the G2/M fraction. Alternatively, structural changes exposing origin DNA might explain the cell cycle-dependent MNase sensitivity of origin DNA. Comparing the most and least efficient pre-RCs, we found a more pronounced sensitivity at top-pre-RCs than at bot-pre-RCs. In *S. cerevisiae*, pre-RCs are characterized by positioned nucleosomes. As these origins have an orientation, the mean size of a MSR is dependent on the alignment to the T-rich strand (Field et al., 2008; Berbenetz et al., 2010; Eaton et al., 2010). However, a limitation of our system is the small number of origins detected in the EBV genome. This results in a relatively low sample size for any statistical analyses, and thus in high variance, limiting any conclusions regarding the mean flanking nucleosome positions and the existence of an orientation in these origins.

Pre-RC assembly and origin activation are temporally separated but functionally linked events. To detect initiation sites, we isolated SNS DNA by an enzyme-free method and found that >80% of SNS and pre-RC zones overlap. When taking into account that the majority of the nonoverlapping SNS zones are located in the direct neighborhood of pre-RC zones, the spatial correlation increases to >90%. We do not observe a 100% overlap because: (a) Experimentally, we do not have a single-nucleotide resolution in our ChIP and SNS experiments; and (b) the definition of pre-RC and SNS zones for our analyses is most likely not perfect, and has some intrinsic fuzziness. Also, we might exclude true positive zones as well as include false positive signals. Lubelsky et al. (2011) have also observed the spatial separation of origin recognition and replication initiation, where pre-RCs and SNSs do not align perfectly.

Origin recognition at pre-RCs and replication initiation at SNSs are reflected in different features. First, pre-RC zones

are characterized by a cell cycle-dependent MNase profile, whereas SNS zones appear as cell cycle-independent MSRs. The efficiency of origin activation clearly correlates with the degree of MNase sensitivity. Second, our findings indicate that the initiation efficiency is moderately influenced by the underlying sequence. Our comparative analysis indicates that A/T-rich tracks are preferentially found at topSNSs. An increased A/T content thermodynamically destabilizes the DNA duplex, thus facilitating base unpairing, an event that is part of the initiation process, but not of pre-RC assembly. Furthermore, A/T-rich elements, particularly homopolymeric poly(dA:dT), are less favorable for nucleosome formation (Segal et al., 2006; Segal and Widom, 2009), which might explain the relationship between A/T content and SNS. Currently, no experimental data exist that describe how the EBV sequence influences nucleosome positioning. In contrast to our findings, Cayrou et al. (2011) found that SNSs correlate with GC richness and CpG islands, whereas we observe a bias toward AT-rich elements. This could either be explained by the different model organisms analyzed or by the different experimental methods used to isolate SNS DNA. A general feature of SNS DNA is the very low copy number, which makes all methods sensitive for contaminations or biases introduced during the experimental process. For example, λ-exonuclease might induce a bias toward GC-rich DNA. However, Karnani et al. (2010) compared this enzymatic method with the enzyme-independent immunoprecipitation of newly BrdU-labeled DNA without any apparent differences in terms of AT content. Further experiments are essential to clarify the strengths and limitations of the individual methods.

Our observations suggest a two-step model to explain the plasticity of origin formation and selection in human cells.

Table 4. Deviations, top versus bottom zones (p-value, one sided Student's t-test)

Zone type	As	Ts	Gs	Cs	CG	GC	CC	GG	GC or CG	TA	AT	AA	TT	AT or TA
pre-RC	0.02	0.02	1.9 × 10 <sup>-3</sup>	0.2	2.7 × 10 <sup>-4</sup>	0.03	0.13	0.02	1.1 × 10 <sup>-3</sup>	0.21	0.05	0.03	0.02	0.07
SNS	0.24	0.02	0.02	0.26	0.12	0.04	0.23	0.02	0.04	1.8 × 10 <sup>-3</sup>	0.04	0.47	1.4 × 10 <sup>-3</sup>	3.3 × 10 <sup>-3</sup>

Nucleotide feature deviations between top and bottom pre-RC or SNS zones. The p-values shown indicate the significance of the deviation between top and bottom pre-RC or SNS zones.

In the first step, a limited number of pre-RCs are assembled independent of sequence. At present it is unclear which mechanisms exist to limit this number; however, we propose that the efficiency is linked to the local chromatin structure and its ability to mobilize nucleosomes. It is very unlikely that each potential pre-RC is used in each cell cycle for complex formation because the copy number of initiation proteins is too low (Wong et al., 2011). The excess of pre-RCs in relation to SNSs and the relative ratios between the efficiencies of pre-RC assembly at DS and other sites corroborate this data. Assuming that a pre-RC is formed at the DS region in every cell cycle, the mean efficiency of a non-DS pre-RC in the EBV genome is on average 5.98 times weaker than at DS (peak<sub>max</sub> at DS,  $2^{3.34}$ ; peak<sub>average</sub>,  $2^{0.76}$ ). This means that only 15–20% of potential pre-RC sites are used per genome and cell cycle for pre-RC formation. In a second step, a subset of pre-RCs is activated to initiate replication. SMARD data shows that only 1–3 origins are activated per EBV genome, which suggests that the origin activation efficiency is in the range of 10–20%. This model explains the discrepancy between the observed plasticity of initiation sites, the limited number of pre-RCs present in each cell, and the even lower number of initiation events. With this, the Jesuit model (“many are called, but few are chosen”) functions at two temporally separated levels (DePamphilis, 1993, 1996).

The genome-wide mapping of pre-RC proteins and its correlation with replication initiation sites and MSRs provides new insights into our understanding of how replication origins are organized in mammalian cells. Our study demonstrates that a ChIP analysis of pre-RC components is technically possible; however, it requires very careful controls (i.e., sufficient replicates, various pre-RC proteins, IgG-controls) and considerations in the selection of threshold levels for enriched zone width. The high copy number of the EBV genome might have facilitated our analyses. Strong origins are characterized by efficient pre-RC assembly and replication initiation processes. However, to be a weak origin, only one of these processes needs to be inefficient. DS is a perfect example of a strong pre-RC site which may function as an internal control site, but which at the same time represents only a weak initiation site. DNA accessibility and nucleosome mobility are likely to contribute to efficient pre-RC formation, whereas initiation efficiency is influenced by additional parameters such as the A/T content. Our study may help to unravel the conflict between the strict replicon model and an entirely stochastic origin pattern (Rhind, 2006).

## Materials and methods

### Cell culture

Raji and DG75 cells were grown in suspension culture with RPMI medium, supplemented with 10% FCS at 37°C in 5% CO<sub>2</sub>.

### Centrifugal elutriation and flow cytometry

Centrifugal elutriation (J6-MC centrifuge; Beckman Coulter) was used to separate the different cell cycle phases. For ChIP experiments,  $5 \times 10^9$  logarithmically growing Raji cells were washed with PBS and resuspended in 50 ml RPMI supplemented with 1% FCS, 1 mM EDTA, and 0.25 U/ml DNase I (Roche). Cells were injected into a JE-5.0 rotor (Beckman Coulter) with a large separation chamber at 1,500 rpm and a flow rate of 30 ml/min controlled with a Masterflex pump (Cole-Palmer). The rotor speed was kept constant and 400-ml fractions were collected at increasing flow

rates (35–100 ml/min). Individual fractions were counted and processed for the ChIP assay as described in the next section.

For flow cytometry,  $10^6$  cells were washed once with PBS, resuspended in 1 ml 80% ethanol/20% PBS, and incubated for 1 h on ice. Fixed cells were washed twice with PBS, and 900  $\mu$ l PBS supplemented with 200 U RNase was added. After 15 min of incubation on ice, 100  $\mu$ l of propidium iodide stain was added (5  $\mu$ g/ml PI and 50 mM EDTA in PBS). Samples were kept on ice until the DNA content was determined using a FACS Calibur (BD).

### Chromatin preparation and ChIP experiments

$10^8$  cells of the corresponding cell cycle fractions were centrifuged (1,200 rpm, 10 min) and washed twice with ice-cold PBS. The pellet was resuspended in 20 ml of PBS (room temperature). Cells were fixed for 10 min at room temperature by adding 20 ml of a freshly prepared 2% (vol/vol) formaldehyde solution. Adding 1.25 M glycine to a final concentration of 125 mM stopped the reaction. Cells were immediately transferred to ice and incubated for 5 min. After centrifugation, the cells were washed twice with ice-cold PBS and subsequently lysed in 10 ml ice-cold ChIP lysis buffer 1 (50 mM HEPES-KOH, pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% [vol/vol] glycerol, 0.5% [vol/vol] NP-40, 0.25% Triton X-100, and a freshly added 1 $\times$  protease inhibitor cocktail [Roche]). After incubation for 10 min at 4°C, nuclei were precipitated by centrifugation (1,500 rpm) and solved in 5 ml of ice-cold ChIP lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and freshly added 1 $\times$  protease inhibitor cocktail). After incubation and centrifugation (1,500 rpm, 10 min, 4°C), the chromatin was resuspended in 5 ml of ice-cold ChIP lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5% N-lauryl sarcosine, 0.1% sodium deoxycholate, and 1 $\times$  protease inhibitor cocktail). Acid-washed glass beads (212–300  $\mu$ m) were added, and the cross-linked chromatin was fragmented by sonication (8  $\times$  30 s at 35% power output, with pulses set to 1 s on/1 s off) in an ice-water bath using a sonicator (250-D; Branson). Micrococcal nuclease (MNase) digestion was performed (10 U MNase/ml and 4 mM CaCl<sub>2</sub>) for 10 min at 37°C. Adding 40 mM EGTA stopped the reaction. The combination of DNA shearing by sonication and DNA digestion by MNase resulted in a mean fragment size of the bulk genomic DNA of 200–800 bp. Triton X-100 was added to a final concentration of 0.5% (vol/vol), and the lysate was centrifuged (13,200 rpm, 5 min, 4°C). The chromatin extract was quantified and diluted to 1 mg/ml with ChIP lysis buffer 3. Pre-clearing of the lysate was performed with 100  $\mu$ l protein A or G Sepharose beads (pre-absorbed with PBS/0.5% [wt/vol] BSA) for 2 h at 4°C. Pre-cleared extracts were incubated with 10  $\mu$ g affinity purified antibodies overnight at 4°C (rabbit Orc2 [Schepers et al., 2001], rabbit Mcm3 [Ritzi et al., 2003], and rabbit anti-IgG [Dianova]). 50  $\mu$ l of blocked protein A or G Sepharose beads were added and incubated for 4 h. The antibody–protein–DNA complexes were collected by centrifugation (1,500 rpm, 2 min, 4°C) and washed twice with 10 ml RIPA (1 mM EDTA, 150 mM NaCl, 0.1% SDS, 0.5% DOC, and 1% NP-40), 10 ml LiCl (250 mM LiCl, 0.1% SDS, 0.5% DOC, 1% NP-40, and 50 mM Tris, pH 8.0), and 10 ml TE, pH 8.0, respectively. Sepharose beads were transferred to 1.5-ml reaction tubes, and the protein–DNA complexes were eluted twice for 10 min using 100  $\mu$ l ChIP elution buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 1% SDS) at 65°C under constant agitation. Beads were removed (1,500 rpm, 2 min, room temperature) and the supernatant was incubated for 2 h at 37°C adding 5  $\mu$ g DNase-free RNase A. The cross-link was reversed by incubation for 12 h with 80  $\mu$ g Proteinase K at 56°C. Input DNA (10% of the chromatin material used for ChIP) was prepared in parallel to the ChIP samples. RNase, Proteinase K treatment, and reversion of the cross-link were performed as described previously. Co-precipitated and input DNA were purified via NucleoSpin Extract II (Machery-Nagel) according to manufacturer’s instructions. The DNA was eluted in 22  $\mu$ l of elution buffer.

### Quantitative real-time PCR

Real-time PCR was performed with the LightCycler (Roche) according to the manufacturer’s instructions. The provided FastStart Reaction Mix was supplemented with MgCl<sub>2</sub> to a final concentration of 2 mM. The amplification of PCR products was monitored on-line and usually stopped after 40 cycles. The following settings were used: 10 min at 95°C, cycles with 1 s at 95°C, 10 s at 62°C, and 20 s at 72°C. The sequences of the primers used are shown in Table S11.

### Whole genome amplification

Co-precipitated DNA was amplified before microarray hybridization using WGA II (Sigma-Aldrich) according to the manufacturer’s instructions. 10  $\mu$ l of



the purified CHIP DNA and 100 ng of purified input DNA were used for amplification. The PCR was performed in a thermocycler (Mastercycler personal; Eppendorf) using 15 cycles. The amplified DNA was purified via NucleoSpin Extract II according to manufacturer's instructions.

#### Array design and control experiments

EBV samples were hybridized on a custom-made  $2 \times 105,000$  array (Agilent Technologies) covering both strands of the EBV genome (EBV strain type I; GenBank/EMBL/DBJ accession no. NC\_007605) in overlapping, melting temperature-optimized 60 mers. For normalization purposes, the array also contained 60 mers of the Adenovirus 5 (Ad5) genome. Probes were tiled every 12 bp along the upper and lower strand, with strand-specific probes being shifted 6 bp relative to each other, resulting in an overall resolution of 6 nucleotides. The Raji genome harbors two deletions located at nt 86,000–89,000 and 163,978–166,635, respectively. The EBV genome contains a series of W repeats as internal repeats. Six copies of the W repeats were plotted onto the array, but the repetitive regions typically vary from genome to genome, and thus could introduce undesired background noise to our analyses. Therefore, we discarded these repeats from all analyses.

To minimize errors introduced by the antibodies and the unspecific hybridization of cellular DNA to the EBV genome, we performed two control experiments. First, we hybridized the IgG control to identify probes that are specifically recognized by IgG (Fig. S5 A). No overlap with the pre-RC-specific signals was observed. Second, we analyzed the Orc2 pattern of the EBV-negative cell line DG75 hybridized against the EBV tiling array (Fig. S5 B).

#### Microarray hybridization

Before hybridization, the concentrations and absorbance ratios of coimmunoprecipitated DNAs were recorded for all DNA samples using a spectrophotometer (ND-1000 UV-VIS; NanoDrop). For microarray hybridizations, only high-quality DNA samples with an A260/A280 ratio of 1.8–2.0 and a A260/A230 ratio  $>1.0$  were used. 500 ng of DNA from each sample was subjected to restriction digestion with a combination of AluI and RsaI. The digested DNA samples were directly labeled with exo-Klenow polymerase and random primers by using Cyanine-5 dUTP for the experimental samples and Cyanine-3 dUTP for the reference samples (Genomic DNA Enzymatic Labeling kit; Agilent Technologies). After purification, the DNA concentrations and Cyanine-5 and Cyanine-3 dye concentrations (pmol/ $\mu$ l) were recorded for all labeled samples.

After clean-up and quantification of labeled DNA, each Cyanine-5-labeled experimental sample was combined with a corresponding Cyanine-3-labeled reference sample. Human Cot-1 DNA was added to block the repetitive sequences in the genomic DNA. The combined samples were prehybridized and prepared for two-color based hybridization (Oligo aCGH Hybridization kit; Agilent Technologies). Each combination of experimental and reference DNA samples was hybridized at 65°C for 40 h on custom-made EBV-specific microarrays ( $2 \times 105,000$  format). Microarrays were washed with increasing stringency using Oligo aCGH wash buffers (Agilent Technologies) followed by drying with acetonitrile. Before scanning, each slide was washed in a drying and stabilization solution (Agilent Technologies) to stabilize the fluorescence for future scans. Fluorescent signal intensities for both dyes were detected on an Agilent DNA Microarray Scanner using Scan Control A8.4.1 Software (Agilent Technologies). Images were extracted using Feature Extraction 10.5.1.1 Software (Agilent Technologies). After a one-pass scan, an additional scan was performed by applying eXtended Dynamic Range (XDR) at 10% as well as 100% laser capacity. Primary array analyses and data normalization were performed on a GenePix Personal 4100A scanner using GenePix Pro 6.0 software (Axon Instruments).

#### SNS analysis

$2 \times 10^7$  cells were washed with PBS and resuspended in PBS with 10% glycerol (Kamath and Lefak, 2001). Cells were lysed for 10 min in slots of a 1.2% alkaline agarose gel (50 mM NaOH and 1 mM EDTA). DNA was separated by electrophoresis overnight (low melting temperature agarose; Biozym). After neutralization with 1 $\times$  TAE (40 mM Tris, pH 8.0, 20 mM acetic acid, and 1 mM EDTA) for 45 min, the lane containing DNA size markers was separated from the rest of the gel and visualized by ethidium bromide (EtBr) for 15 min. The size of SNS DNA in the unstained gel was determined by comparing it with the EtBr-stained DNA size marker. Subsequently, SNS fragments of 800–1,500 nt were extracted from the gel using the QIAquick gel extraction kit. SNS abundance was measured by quantitative real-time PCR using primer pairs for the human HPRT locus (Cohen et al., 2002). The concentration of purified SNS-DNA was determined

using a Quant-iT dsDNA high-sensitivity assay kit according to the manufacturer's instructions. After amplification, Southern blot analysis with an EBV-specific probe was performed to determine the length of EBV-specific nascent strand DNA. Fragments of 100–2,500 bp were detected, indicating that the size of the marker in the alkaline gel does not correspond with the actual length of the isolated SNS fragments.

#### Mononucleosome preparation

For each sample,  $10^7$  cells were harvested, washed with PBS, and resuspended in 4 ml hypotonic buffer A (10 mM Hepes-KOH, pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.34 M sucrose, 10% glycerol, 1 mM DTT, and 1 $\times$  protease inhibitor mix). Cells were lysed by adding 0.04% Triton X-100 and incubated for 10 min on ice. Samples were centrifuged (4 min, 1,300 g, 4°C) to separate soluble cytosolic and nucleosolic proteins from chromatin. Nuclei were washed in 5 ml of ice-cold buffer A supplemented with 200 mM NaCl. After centrifugation (5 min, 1,300 g, 4°C), nuclei were carefully resuspended in 1 ml MNase digestion buffer (10 mM Hepes-KOH, pH 7.6, 120 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, 10% (vol/vol) glycerol, 1 mM DTT, and 1 $\times$  protease inhibitor mix). MNase digestion was performed by adding 30 U MNase and incubating for 3 min at 37°C. The reaction was stopped by adding 40  $\mu$ l EGTA (0.5 M) on ice. RNA was removed by incubation with 20 U RNase for 2 h at 37°C, then subsequently incubated with Proteinase K at 56°C over night (80  $\mu$ g/ml). Mononucleosomes were isolated from a 1.2% agarose TAE gel and purified via NucleoSpin Extract II according to the manufacturer's instructions.

#### Southern blotting

500 ng of sonicated and MNase-digested CHIP DNA or nascent strand DNA were separated on a 1.0% TAE gel and transferred to membrane (Immobilon Ny+; EMD Millipore). After prehybridization in Church buffer (0.5 M sodium phosphate, 7% SDS, and 1 mM EDTA, pH 7.2) for 3 h at 65°C, labeled and denatured EBV-specific probes (nt 9,170–9,347, nt 37,794–37,970, and nt 50,012–50,269) were added and hybridized for 16 h at 65°C. After extensive washings with 2 $\times$  SSC, 0.1% SDS and 0.5 $\times$  SSC, 0.1% SDS, the data were obtained by digital imaging.

#### Bioinformatical methods

**Software.** All numerical and statistical analyses were done using R (<http://www.r-project.org>). Additionally, we used the packages *limma* (<http://www.bioconductor.org/packages/2.8/bioc/html/limma.html>), *affy* (<http://www.bioconductor.org/packages/release/bioc/html/affy.html>), *gplots* (<http://cran.r-project.org/web/packages/gplots/>) and *tileHMM* (<http://cran.r-project.org/web/packages/tileHMM/>). All functions were used with default parameters unless stated otherwise.

**Orc2 and Mcm3 signal enrichment and IgG-normalization.** Primary array analysis of enriched (Cy5) and input (Cy3) DNA as well as data normalization was performed on a GenePix Personal 4100A scanner using GenePix Pro 6.0 software (Axon Instruments). Enriched (Cy5) and input (Cy3) signal intensities were converted into log<sub>2</sub> enrichment ratios (log<sub>2</sub>(Cy5/Cy3)) individually for each biological replicate. Using these ratios, we then separately normalized the Orc2 and Mcm3 log<sub>2</sub> ratios with IgG using the *lmFit* function of the *limma* package. Finally, we scaled the resulting normalized log<sub>2</sub> enrichments (i.e., so that the set of log<sub>2</sub> enrichments have a mean of 0 and a standard deviation of 1) separately for each concentration to allow for an appropriate comparison.

**Orc2 and Mcm3 enriched zones calculation.** Enrichment zones for Mcm3 and Orc2 were then calculated with the IgG-normalized ratios using the *tileHMM* package (Humburg et al., 2008). The package identifies ChIP-enriched regions using a two-state HMM with *t* distributions. The initial parameters of the HMM were obtained by adjusting the default parameters for the probe length (probe region) and the mean size of the DNA fragments to the values 60 and 800, respectively, which are appropriate choices for our analyses. For the estimation of transition probabilities, we assigned the "enriched" state to those probes with log<sub>2</sub> ratios  $> 0$ , and the "non-enriched" state to those probes with log<sub>2</sub> ratios  $\leq 0$ . Log<sub>2</sub> ratios in repetitive and dilution regions in the EBV genome were excluded from the analysis.

Before obtaining the enriched zones, we smoothed the data with a moving mean in overlapping windows with a size of 150 bp. Thus, after parameter optimization with the Viterbi and EM algorithms, *tileHMM* suggested 140 enriched zones for Mcm3 and 174 enriched zones for Orc2. We observed that several of these zones had short widths (minimum zone width = 114 bp), and thus were suspicious of including mostly background signals. To avoid including false positive-enriched zones in further analyses, we discarded all enriched zones with widths  $<400$  bp. Thus, we obtained

64 enriched zones for Mcm3 (mean width = 650 bp) and 76 enriched zones for Orc2 (mean width = 570 bp). 55 of the 64 zones were enriched for both Mcm3 and Orc2.

**Estimation of single pre-RC log<sub>2</sub> ratios and enriched zones.** With the normalized Orc2 and Mcm3 log<sub>2</sub> ratios, we then estimated a single log<sub>2</sub> ratio for pre-RC using the *lmFit* function of the *limma* package. This single estimate was then scaled (as described previously) to allow for appropriate comparisons with other concentrations. To obtain pre-RC enriched zones, we observed the overlap of the previously identified Mcm3 and Orc2 enriched zones on a probe-by-probe basis, and defined them as regions with length  $\geq 400$  bp, composed of probes enriched with both Orc2 and Mcm3 or with Mcm3 only. This resulted in the identification of 64 pre-RC enriched zones.

**SNS DNA signal enrichment.** Signal intensities for the SNS arrays were converted into log<sub>2</sub> ratios for each biological replicate as described previously. Using the mean of both replicates, we obtained a single estimated SNS signal. This signal was then smoothed with a moving mean in overlapping windows of size 200 bp, and scaled similarly as for Orc2, Mcm3, and pre-RC.

**SNS enriched zone calculation.** Using the SNS log<sub>2</sub> ratios, we followed three criteria to identify SNS enrichment zones: (1) the probes must have a positive log<sub>2</sub> ratio, (2) ratios in the bottom quintile of all positive ratios qualified as an estimate of background noise and were thus removed from further analyses, and (3) a minimum of 58 consecutive probes that fulfill criteria 1 and 2 must be present to be considered an SNS-enriched zone (i.e., minimum length of SNS enriched zone = 402 bp).

The number of 58 consecutive probes used to estimate the minimal length of SNS-enriched zones was chosen to avoid an overlap with Okazaki fragments. This resulted in the identification of 57 SNS enriched zones.

**SNS heat map generation.** The heat maps displayed in Fig. 3 C were generated using the *image* function in R.

**G1 MSR estimation.** To estimate the sensitive regions of MNase-digested G1 chromatin, we first converted the signal intensities of the nucleosome G1 array into log<sub>2</sub> ratios, and then scaled these ratios as described previously. We then smoothed the data with a moving mean in overlapping windows with a size of 100 bp. The MSRs were estimated according to three criteria: (1) the probes must have a negative log<sub>2</sub> ratio, (2) ratios in the bottom decile of all negative ratios qualified as an estimate of background noise and were thus removed from further analyses, and (3) a minimum of 16 adjacent probes that fulfill criteria 1 and 2 must be present to be considered an MSR (i.e., minimum length of a MSR = 150 bp). The number of 16 consecutive probes used to estimate the minimal length of a MSR was chosen based on the typical length of a nucleosome.

**S and G2/M MSR signal enrichment.** Signal intensities for the nucleosome arrays (S and G2/M phases) were converted to log<sub>2</sub> enrichment ratios, scaled to allow for proper comparisons, and smoothed with a moving mean in overlapping windows of size 100 bp.

**Mean pre-RC, SNS, and MR log<sub>2</sub> enrichment profiles.** The pre-RC, SNS, and MR profiles (Fig. 4, A–C; and Fig. 5, A and B) were created by averaging the pre-RC, SNS, G1 phase MR, S phase MR, and G2/M phase MR log<sub>2</sub> ratios in a  $\pm 1,000$  bp neighborhood centered at the maximum peak of the pre-RC or SNS enriched zones (all, top 30%, or bottom 30%). The profiles were calculated using sliding windows 50 bp in size, sliding the window in steps of 10 bp. The mean log<sub>2</sub> enrichments at each step were obtained with a 5% trimmed mean to avoid the effects of outliers.

The random profiles were obtained as described previously, except that the neighborhood was centered at 250 randomly selected positions with a uniform distribution across the EBV genome (excluding repetitive and dilution regions).

**Standard deviation for mean profiles.** Because of the heterogeneity of the data (in some cases caused by small sample sizes), we also obtained the standard deviation of the mean profiles described previously (Fig. S4, A and B). The dotted lines in each panel represent the mean  $\pm 1$  standard deviation for each step of the sliding window.

**Mean profiles at TSSs.** The mean log<sub>2</sub> enrichment profiles centered at the TSSs were obtained according to the description given previously. We first centered the  $\pm 1,000$ -bp neighborhood at the selected TSSs, and then obtained mean log<sub>2</sub> enrichments of SNS, G1 phase MR, G2/M phase MR, and S phase MR with a 5% trimmed mean. The profiles were calculated using sliding windows 50 bp in size, sliding the window in steps of 10 bp.

**Cluster analysis and heat map generation.** The hierarchical cluster analysis on the 72 TSSs was obtained using the Ward's minimum variance method, and is based on the G1 phase MR in the region [TSS – 250 bp, TSS + 250 bp].

The heat map shown in Fig. 6 B was produced using the *gplots* package, and it represents the G1 phase MR in the  $\pm 1,000$  bp neighborhood centered at each TSS. The red color indicates high nucleosome occupancy, whereas green indicates low. The IDs of the 72 TSSs are shown on the right axis of the heat map, and the dendrogram obtained with the hierarchical cluster analysis is shown on the left.

**Nucleotide base composition at pre-RC and SNS zones.** The mean nucleotide composition profiles were produced by centering a  $\pm 250$ -bp neighborhood at the maximum peak of the SNS enrichment zones (all, top, or bottom), and then averaging the percentage of nucleotide composition of the probes comprising each zone. The profiles were calculated using sliding windows 50 bp in size, sliding the window in steps of 10 bp.

**Note on p-values and box plots.** All box plots were plotted without outliers. In addition, several p-values in the figures and supplementary figures display the value “ $P < 2.2 \times 10^{-16}$ .” The reason for the frequent repetition of this value is because of the fact that the lowest possible p-value that the software obtains (for numerical reasons) is precisely  $2.2 \times 10^{-16}$ .

**Calculation of simulated microarray signals.** Resolution of fragments with a uniform length: To simulate the resolution of microarray signals, we considered a hypothetical genome/chromosome of length  $l$  that contains several discrete sites ( $b \in B = \{b_1, b_2, \dots, b_x\}$ ), which can be used to isolate subgenomic fragments. We assume that the fragmentation process results in the generation of subgenomic fragments that are uniformly distributed along the genome. Hence, within a fragment population of length  $l$ , the coverage of each nucleotide position is identical, and there are an identical number of fragments that begin or end at this position (this is only strictly true for nucleotides that are at least 1 nucleotide away from either end of the parental molecule). Assuming that the presence of a single selection site is sufficient to retain a given fragment, the selection process isolates all fragments that contain at least one site  $b \in B$ . We refer to this population as  $F_B$ . For each probe  $p$ , the pool of fragments that contributes to its signal is represented by the subpopulation of  $F_B$  that contains the target sequence for  $p$ . The maximum possible array signal is observed when probe position and a binding site coincide, and thus 100% of fragments that contain  $p$  are retained. Hence, a relative and normalized signal for each probe  $p$  is calculated by dividing the number of selected fragments that contain  $p$  by the total number of fragments that contain  $p$  (see also Fig. S1 E), such that:

$$s_{norm}(p) = \frac{|F_B \cap F_p|}{|F_p|},$$

where  $F_B$  is the set of fragments retained during selection,  $F_p$  is the set of fragments that contain  $p$ ,  $F_B \cap F_p$  is the intersection of  $F_B$  and  $F_p$ , and thus the set of fragments containing  $p$  as well as at least one selection site  $b \in B$ , and  $|\dots|$  denotes the cardinality, i.e., total number of elements/fragments within a given set.

Resolution of fragments with varying length: in ChIP experiments, the fragmentation process results in the generation of a population of fragments of varying length  $l$ . Under nonsaturating conditions, the signal of each probe is represented by the sum of the normalized contribution of each fragment pool. To calculate a normalized signal value, we introduce a normalization factor  $f(l)$ , represented by the frequency distribution of sequence coverage by individual fragment length pools before selection. For example, if the population consists of fragment lengths  $l_1$ ,  $l_2$ , and  $l_3$  that account for 80%, 19%, and 1% of sequence coverage, their frequency values are 0.8, 0.19, and 0.01, respectively. We are not considering potential competition between probes for longer subgenomic fragments because the Agilent protocol includes a step that generates shorter fragments of 50–200 bp immediately before hybridization. Furthermore, fragments are labeled by synthesis using Cy5- and Cy3-labeled nucleotides, and thus the amount of label is directly proportional to the fragment length. Therefore, if fragments of different lengths equally cover a given probe, competition by adjacent probes for longer fragments is compensated for by the higher fluorescence of those fragments. Based on these considerations:

$$s_{norm}(p) = \sum_{l=l_{min}}^{l_{max}} f(l) \times \frac{|F_B^l \cap F_p^l|}{|F_p^l|},$$

where  $l$  is the length of subgenomic fragment population,  $l_{min}$  and  $l_{max}$  are the lower and upper bounds, respectively, of  $l$ ,  $f(l)$  is the coverage frequency for fragments of length  $l$ ,  $F_B^l$  is the selected set of fragments of length  $l$ ,  $F_p^l$  is the set of fragments of length  $l$  that contain  $p$ ,  $F_B^l \cap F_p^l$  is the

intersection of  $F_B^l$  and  $F_p^l$ , and thus the set of fragments containing  $p$  as well as at least one selection site  $b \in B$ , and  $|\dots|$  denotes the cardinality, i.e., the total number of elements/fragments within a given set.

### Online supplemental material

Fig. S1 A illustrates a small bias introduced by the whole genome amplification step. Fig. S1 B shows genome-wide profiles for Orc2, Mcm3, G1-MNase, and pre-RC. Fig. 1 C shows Southern blot experiments to determine the fragment length distribution of the ChIP-input DNA and the SNS DNA, and a quantitative analysis of the fragment length distribution of the ChIP DNA. Fig. S1 (D and E) shows figures simulating the influence of the fragment length on the resolution of microarray data. Fig. S2 shows a control agarose gel of an MNase digest (A), box plots of G1-MR at pre-RC and SNS-enriched zones (B), and a  $\chi^2$  test demonstrating that pre-RC zones are not independent of MSRs. Fig. S3 A shows an experiment to control the quality of isolated SNS DNA at the published HPRT locus (Cohen et al., 2002). Fig. S3 B shows the replication initiation activity within the EBV genome in 5-kbp steps to facilitate the comparison with the SMARD results (Norio and Schildkraut, 2001, 2004). Fig. S3 C shows that the majority of SNS zones that are not overlapping with at least 5% of their width with a pre-RC zone have nearby pre-RC zones. Fig. S3 D shows a linear regression between the mean SNS and pre-RC log<sub>2</sub> enrichments at SNS enriched zones. Only a minor correlation between both enrichments can be detected. In contrast to this finding, a  $\chi^2$  test on a two-way contingency table revealed a significant relationship between top-pre-RCs and topSNSs (Fig. S3 E). Fig. S4 shows means and standard deviations of the mean profiles shown in Fig. 4 and Fig. 5. Fig. S5 shows box plots of different IgG antibodies hybridized on the EBV array (A), and a control experiment in the EBV-negative cell line DG75 (B). Tables S1–S8 contain lists of all pre-RC and SNS zones, their locations, sizes, and efficiencies. Table S9 lists all TSS in the EBV genome and their overlap with SNS zones. Table S10 shows the cluster information of SNS zones located in [TSS – 250 bp, TSS + 250 bp]. Table S11 lists all primer pairs used for quantitative PCR experiments. Online supplemental material is available at <http://www.jcb.org/cgi/content/full/jcb.201109105/DC1>.

We thank P. Becker, T. Straub, and P. Korber for helpful discussions, and C.E. Mayer and M. Ravasz for critical reading of the manuscript. The microarray hybridization experiments were performed by IMG M Laboratories GmbH, Martinsried.

This work was supported by Deutsche Forschungsgemeinschaft (DFG) grants SFB646 and SFB/TR05 (A. Schepers) and by the Hamburger Stiftung zur Förderung der Krebsbekämpfung (A. Grundhoff). The Heinrich Pette Institute is a member of the Leibniz Gemeinschaft and is supported by the City of Hamburg and the Federal Ministry of Health. The authors declare that they have no conflicts of interest.

Submitted: 21 September 2011

Accepted: 13 July 2012

## References

Adams, A. 1987. Replication of latent Epstein-Barr virus genomes in Raji cells. *J. Virol.* 61:1743–1746.

Aladjem, M.I. 2007. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.* 8:588–600. <http://dx.doi.org/10.1038/nrg2143>

Aladjem, M.I., A. Falaschi, and D. Kowalski. 2006. Eukaryotic DNA replication origins. In *DNA Replication and Human Diseases*. M.L. DePamphilis, editor. Cold Spring Harbor Press, Cold Spring Harbor, New York. 31–63.

Balaso, M., R.P. Huijbrechts, and I. Chesnokov. 2007. Role of the Orc6 protein in origin recognition complex-dependent DNA binding and replication in *Drosophila melanogaster*. *Mol. Cell. Biol.* 27:3143–3153. <http://dx.doi.org/10.1128/MCB.02382-06>

Berbenetz, N.M., C. Nislow, and G.W. Brown. 2010. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.* 6:e1001092. <http://dx.doi.org/10.1371/journal.pgen.1001092>

Breier, A.M., S. Chatterji, and N.R. Cozzarelli. 2004. Prediction of *Saccharomyces cerevisiae* replication origins. *Genome Biol.* 5:R22. <http://dx.doi.org/10.1186/gb-2004-5-4-r22>

Cadoret, J.C., F. Meisch, V. Hassan-Zadeh, I. Luyten, C. Guillet, L. Duret, H. Quesneville, and M.N. Prioleau. 2008. Genome-wide studies highlight indirect links between human replication origins and gene regulation.

*Proc. Natl. Acad. Sci. USA.* 105:15837–15842. <http://dx.doi.org/10.1073/pnas.0805208105>

Cayrou, C., P. Coulombe, A. Vigneron, S. Stanojic, O. Ganier, I. Peiffer, E. Rivals, A. Puy, S. Laurent-Chabaliere, R. Desprat, and M. Méchali. 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* 21:1438–1449. <http://dx.doi.org/10.1101/gr.121830.111>

Chaudhuri, B., H. Xu, I. Todorov, A. Dutta, and J.L. Yates. 2001. Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc. Natl. Acad. Sci. USA.* 98:10085–10089. <http://dx.doi.org/10.1073/pnas.181347998>

Cohen, S.M., B.P. Brylawski, M. Cordeiro-Stone, and D.G. Kaufman. 2002. Mapping of an origin of DNA replication near the transcriptional promoter of the human HPRT gene. *J. Cell. Biochem.* 85:346–356. <http://dx.doi.org/10.1002/jcb.10136>

Collins, N., R.A. Poot, I. Kukimoto, C. García-Jiménez, G. Dellaire, and P.D. Varga-Weisz. 2002. An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.* 32:627–632. <http://dx.doi.org/10.1038/ng1046>

DePamphilis, M.L. 1993. Eukaryotic DNA replication: anatomy of an origin. *Annu. Rev. Biochem.* 62:29–63. <http://dx.doi.org/10.1146/annurev.bi.62.070193.000333>

DePamphilis, M.L. 1996. Origins of DNA Replication. In *DNA Replication in Eukaryotic Cells*. M.L. DePamphilis, editor. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 45–86.

Dhar, S.K., K. Yoshida, Y. Machida, P. Khaira, B. Chaudhuri, J.A. Wohlschlegel, M. Leffak, J. Yates, and A. Dutta. 2001. Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell.* 106:287–296. [http://dx.doi.org/10.1016/S0092-8674\(01\)00458-5](http://dx.doi.org/10.1016/S0092-8674(01)00458-5)

Diffley, J.F.X., J.H. Cocker, S.J. Dowell, and A. Rowley. 1994. Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell.* 78:303–316. [http://dx.doi.org/10.1016/0092-8674\(94\)90299-2](http://dx.doi.org/10.1016/0092-8674(94)90299-2)

Dijkwel, P.A., J.P. Vaughn, and J.L. Hamlin. 1991. Mapping of replication initiation sites in mammalian genomes by two-dimensional gel analysis: stabilization and enrichment of replication intermediates by isolation on the nuclear matrix. *Mol. Cell. Biol.* 11:3850–3859.

Eaton, M.L., K. Galani, S. Kang, S.P. Bell, and D.M. MacAlpine. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24:748–753. <http://dx.doi.org/10.1101/gad.1913210>

Eaton, M.L., J.A. Prinz, H.K. MacAlpine, G. Tretyakov, P.V. Kharchenko, and D.M. MacAlpine. 2011. Chromatin signatures of the *Drosophila* replication program. *Genome Res.* 21:164–174. <http://dx.doi.org/10.1101/gr.116038.110>

Field, Y., N. Kaplan, Y. Fondufe-Mittendorf, I.K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4:e1000216. <http://dx.doi.org/10.1371/journal.pcbi.1000216>

Gerhardt, J., S. Jafar, M.P. Spindler, E. Ott, and A. Schepers. 2006. Identification of new human origins of DNA replication by an origin-trapping assay. *Mol. Cell. Biol.* 26:7731–7746. <http://dx.doi.org/10.1128/MCB.01392-06>

Gilbert, D.M. 2010. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.* 11:673–684. <http://dx.doi.org/10.1038/nrg2830>

Givens, R.M., W.K. Lai, J.M. Rizzo, J.E. Bard, P.A. Mieczkowski, J. Leatherwood, J.A. Huberman, and M.J. Buck. 2012. Chromatin architectures at fission yeast transcriptional promoters and replication origins. *Nucleic Acids Res.* In press.

Günther, T., and A. Grundhoff. 2010. The epigenetic landscape of latent Kaposi sarcoma-associated herpesvirus genomes. *PLoS Pathog.* 6:e1000935. <http://dx.doi.org/10.1371/journal.ppat.1000935>

Hamlin, J.L., L.D. Mesner, O. Lar, R. Torres, S.V. Chodaparambil, and L. Wang. 2008. A revisionist replicon model for higher eukaryotic genomes. *J. Cell. Biochem.* 105:321–329. <http://dx.doi.org/10.1002/jcb.21828>

Hamlin, J.L., L.D. Mesner, and P.A. Dijkwel. 2010. A winding road to origin discovery. *Chromosome Res.* 18:45–61. <http://dx.doi.org/10.1007/s10577-009-9089-z>

Harland, R.M., and R.A. Laskey. 1980. Regulated replication of DNA micro-injected into eggs of *Xenopus laevis*. *Cell.* 21:761–771. [http://dx.doi.org/10.1016/0092-8674\(80\)90439-0](http://dx.doi.org/10.1016/0092-8674(80)90439-0)

Holdorf, M.M., S.B. Cooper, K.R. Yamamoto, and J.J. Miranda. 2011. Occupancy of chromatin organizers in the Epstein-Barr virus genome. *Virology.* 415:1–5. <http://dx.doi.org/10.1016/j.viro.2011.04.004>

Humburg, P., D. Bulger, and G. Stone. 2008. Parameter estimation for robust HMM analysis of ChIP-chip data. *BMC Bioinformatics.* 9:343. <http://dx.doi.org/10.1186/1471-2105-9-343>

Hyrien, O., and M. Méchali. 1993. Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of *Xenopus* early embryos. *EMBO J.* 12:4511–4520.

- Kamath, S., and M. Leffak. 2001. Multiple sites of replication initiation in the human beta-globin gene locus. *Nucleic Acids Res.* 29:809–817. <http://dx.doi.org/10.1093/nar/29.3.809>
- Karnani, N., C.M. Taylor, A. Malhotra, and A. Dutta. 2010. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell.* 21:393–404. <http://dx.doi.org/10.1091/mbc.E09-08-0707>
- Lantermann, A.B., T. Straub, A. Strålfors, G.C. Yuan, K. Ekwall, and P. Korber. 2010. *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.* 17:251–257. <http://dx.doi.org/10.1038/nsmb.1741>
- Little, R.D., and C.L. Schildkraut. 1995. Initiation of latent DNA replication in the Epstein-Barr virus genome can occur at sites other than the genetically defined origin. *Mol. Cell. Biol.* 15:2893–2903.
- Lubelsky, Y., T. Sasaki, M.A. Kuipers, I. Lucas, M.M. Le Beau, S. Carignon, M. Debatisse, J.A. Prinz, J.H. Dennis, and D.M. Gilbert. 2011. Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. *Nucleic Acids Res.* 39:3141–3155. <http://dx.doi.org/10.1093/nar/gkq1276>
- MacAlpine, D.M., H.K. Rodríguez, and S.P. Bell. 2004. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* 18:3094–3105. <http://dx.doi.org/10.1101/gad.1246404>
- MacAlpine, H.K., R. Gordán, S.K. Powell, A.J. Hartemink, and D.M. MacAlpine. 2010. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* 20:201–211. <http://dx.doi.org/10.1101/gr.097873.109>
- Marechal, V., A. Dehee, R. Chikhi-Brachet, T. Piolot, M. Coppey-Moisán, and J.C. Nicolas. 1999. Mapping EBNA-1 domains involved in binding to metaphase chromosomes. *J. Virol.* 73:4385–4392.
- Masai, H., S. Matsumoto, Z. You, N. Yoshizawa-Sugata, and M. Oda. 2010. Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* 79:89–130. <http://dx.doi.org/10.1146/annurev.biochem.052308.103205>
- Méchali, M. 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* 11:728–738. <http://dx.doi.org/10.1038/nrm2976>
- Minarovits, J., L.F. Hu, Z. Marcsek, S. Minarovits-Kormuta, G. Klein, and I. Ernberg. 1992. RNA polymerase III-transcribed EBER 1 and 2 transcription units are expressed and hypomethylated in the major Epstein-Barr virus-carrying cell types. *J. Gen. Virol.* 73:1687–1692. <http://dx.doi.org/10.1099/0022-1317-73-7-1687>
- Norio, P., and C.L. Schildkraut. 2001. Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science.* 294:2361–2364. <http://dx.doi.org/10.1126/science.1064603>
- Norio, P., and C.L. Schildkraut. 2004. Plasticity of DNA replication initiation in Epstein-Barr virus episomes. *PLoS Biol.* 2:e152. <http://dx.doi.org/10.1371/journal.pbio.0020152>
- Pratt, Z.L., M. Kuzembayeva, S. Sengupta, and B. Sugden. 2009. The microRNAs of Epstein-Barr Virus are expressed at dramatically differing levels among cell lines. *Virology.* 386:387–397. <http://dx.doi.org/10.1016/j.virol.2009.01.006>
- Rhind, N. 2006. DNA replication timing: random thoughts about origin firing. *Nat. Cell Biol.* 8:1313–1316. <http://dx.doi.org/10.1038/ncb1206-1313>
- Ritzi, M., K. Tillack, J. Gerhardt, E. Ott, S. Humme, E. Kremmer, W. Hammerschmidt, and A. Schepers. 2003. Complex protein-DNA dynamics at the latent origin of DNA replication of Epstein-Barr virus. *J. Cell Sci.* 116:3971–3984. <http://dx.doi.org/10.1242/jcs.00708>
- Schaarschmidt, D., J. Baltin, I.M. Stehle, H.J. Lipps, and R. Knippers. 2004. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J.* 23:191–201. <http://dx.doi.org/10.1038/sj.emboj.7600029>
- Schepers, A., and P. Papior. 2010. Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Res.* 18:63–77. <http://dx.doi.org/10.1007/s10577-009-9087-1>
- Schepers, A., M. Ritzi, K. Bousset, E. Kremmer, J.L. Yates, J. Harwood, J.F. Diffley, and W. Hammerschmidt. 2001. Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J.* 20:4588–4602. <http://dx.doi.org/10.1093/emboj/20.16.4588>
- Schones, D.E., K. Cui, S. Cuddapah, T.Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 132:887–898. <http://dx.doi.org/10.1016/j.cell.2008.02.022>
- Sears, J., J. Kolman, G.M. Wahl, and A. Aiyar. 2003. Metaphase chromosome tethering is necessary for the DNA synthesis and maintenance of oriP plasmids but is insufficient for transcription activation by Epstein-Barr nuclear antigen 1. *J. Virol.* 77:11767–11780. <http://dx.doi.org/10.1128/JVI.77.21.11767-11780.2003>
- Sears, J., M. Ujihara, S. Wong, C. Ott, J. Middeldorp, and A. Aiyar. 2004. The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. *J. Virol.* 78:11487–11505. <http://dx.doi.org/10.1128/JVI.78.21.11487-11505.2004>
- Segal, E., and J. Widom. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19:65–71. <http://dx.doi.org/10.1016/j.sbi.2009.01.004>
- Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. *Nature.* 442:772–778. <http://dx.doi.org/10.1038/nature04979>
- Sequeira-Mendes, J., R. Díaz-Uriarte, A. Apedaile, D. Huntley, N. Brockdorff, and M. Gómez. 2009. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* 5:e1000446. <http://dx.doi.org/10.1371/journal.pgen.1000446>
- Siddiqui, K., and B. Stillman. 2007. ATP-dependent assembly of the human origin recognition complex. *J. Biol. Chem.* 282:32370–32383. <http://dx.doi.org/10.1074/jbc.M705905200>
- Sivaprasad, U., A. Dutta, and S.P. Bell. 2006. Assembly of Pre-replication Complexes. In *DNA Replication and Human Disease*. M.L. DePamphilis, editor. Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 63–88.
- Sugimoto, N., I. Kitabayashi, S. Osano, Y. Tatsumi, T. Yugawa, M. Narisawa-Saito, A. Matsukage, T. Kiyono, and M. Fujita. 2008. Identification of novel human Cdt1-binding proteins by a proteomics approach: proteolytic regulation by APC/CCdh1. *Mol. Biol. Cell.* 19:1007–1021. <http://dx.doi.org/10.1091/mbc.E07-09-0859>
- Thomae, A.W., D. Pich, J. Brocher, M.P. Spindler, C. Berens, R. Hock, W. Hammerschmidt, and A. Schepers. 2008. Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proc. Natl. Acad. Sci. USA.* 105:1692–1697. <http://dx.doi.org/10.1073/pnas.0707260105>
- Toth, Z., D.T. Maglinte, S.H. Lee, H.R. Lee, L.Y. Wong, K.F. Brulois, S. Lee, J.D. Buckley, P.W. Laird, V.E. Marquez, and J.U. Jung. 2010. Epigenetic analysis of KSHV latent and lytic genomes. *PLoS Pathog.* 6:e1001013. <http://dx.doi.org/10.1371/journal.ppat.1001013>
- Vashee, S., C. Cvetec, W. Lu, P. Simanek, T.J. Kelly, and J.C. Walter. 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* 17:1894–1908. <http://dx.doi.org/10.1101/gad.1084203>
- Wong, P.G., S.L. Winter, E. Zaika, T.V. Cao, U. Oguz, J.M. Koomen, J.L. Hamlin, and M.G. Alexandrow. 2011. Cdc45 limits replicon usage from a low density of preRCs in mammalian cells. *PLoS ONE.* 6:e17533. <http://dx.doi.org/10.1371/journal.pone.0017533>
- Wu, H., D.F. Ceccarelli, and L. Frappier. 2000. The DNA segregation mechanism of Epstein-Barr virus nuclear antigen 1. *EMBO Rep.* 1:140–144. <http://dx.doi.org/10.1093/embo-reports/kvd026>
- Wu, H., P. Kapoor, and L. Frappier. 2002. Separation of the DNA replication, segregation, and transcriptional activation functions of Epstein-Barr nuclear antigen 1. *J. Virol.* 76:2480–2490. <http://dx.doi.org/10.1128/jvi.76.5.2480-2490.2002>
- Xu, J., Y. Yanagisawa, A.M. Tsankov, C. Hart, K. Aoki, N. Kommajosyula, K.E. Steinmann, J. Bochicchio, C. Russ, A. Regev, et al. 2012. Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.* 13:R27. <http://dx.doi.org/10.1186/gb-2012-13-4-r27>
- Yates, J., N. Warren, D. Reisman, and B. Sugden. 1984. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. *Proc. Natl. Acad. Sci. USA.* 81:3806–3810. <http://dx.doi.org/10.1073/pnas.81.12.3806>
- Yates, J.L., and N. Guan. 1991. Epstein-Barr virus-derived plasmids replicate only once per cell cycle and are not amplified after entry into cells. *J. Virol.* 65:483–488.
- Zhou, J., C.M. Chau, Z. Deng, R. Shiekhhattar, M.P. Spindler, A. Schepers, and P.M. Lieberman. 2005. Cell cycle regulation of chromatin at an origin of DNA replication. *EMBO J.* 24:1406–1417. <http://dx.doi.org/10.1038/sj.emboj.7600609>