DNA Replication Origin Interference Increases the Spacing between Initiation Events in Human Cells[□]

Ronald Lebofsky,^{*†} Roland Heilig,[‡] Max Sonnleitner,[§] Jean Weissenbach,[‡] and Aaron Bensimon^{*}

^{*}Unité de Stabilité des Génomes, Institut Pasteur, 75724, Paris, France; [‡]Genoscope, Centre National de Séquençage, 91000, Evry, France; and [§]Upper Austrian Research, Zentrum für Biommedizinische Nanotechnologie, 4020, Linz, Austria

Submitted April 11, 2006; Revised September 8, 2006; Accepted September 20, 2006
Monitoring Editor: A. Gregory Matera

Mammalian DNA replication origins localize to sites that range from base pairs to tens of kilobases. A regular distribution of initiations in individual cell cycles suggests that only a limited number of these numerous potential start sites are converted into activated origins. Origin interference can silence redundant origins; however, it is currently unknown whether interference participates in spacing functional human initiation events. By using a novel hybridization strategy, genomic Morse code, on single combed DNA molecules from primary keratinocytes, we report the initiation sites present on 1.5 Mb of human chromosome 14q11.2. We confirm that initiation zones are widespread in human cells, map to intergenic regions, and contain sequence motifs found at other mammalian initiation zones. Origins used per cell cycle are less abundant than the potential sites of initiation, and their limited use increases the spacing between initiation events. Between-zone interference decreases in proportion to the distance from the active origin, whereas within-zone interference is 100% efficient. These results identify a hierarchical organization of origin activity in human cells. Functional origins govern the probability that nearby origins will fire in the context of multiple potential start sites of DNA replication, and this is mediated by origin interference.

INTRODUCTION

Eukaryotic cells have a limited amount of time, defined by the length of S phase, to duplicate their genomes. This is achieved by synthesizing DNA at replication forks, which extend from multiple sites of initiation. Because fork speed is not scaled according to S-phase length, regulating the frequency of initiation along each respective chromosome is required to prevent unreplicated gaps before the onset of mitosis (Hand and Tamm, 1973; Edenberg and Huberman, 1975).

Although there are exceptions, the common view is that somatic mammalian origins fire at 50- to 300-kilobase (kb) intervals (Edenberg and Huberman, 1975; Berezney *et al.*, 2000). This suggests that Metazoa do possess a mechanism to evenly distribute initiation events. Placing strong replicator sequences at regular distances is one such mechanism that is used by the budding yeast, *Saccharomyces cerevisiae* (Newlon *et al.*, 1991; Shirahige *et al.*, 1993). In higher eukaryotes, genetic elements play a role in origin activation; however, they are not sufficient by themselves to drive initiation (Gilbert, 2004). Furthermore, although some Metazoan origins localize to well-circumscribed sites of a few base pairs, a large number localize to more disperse initiation zones ranging up to tens of kbs (DePamphilis, 1999).

This raises the problem of how to achieve a regular distribution of activated origins from a range of potential sites that possess low intrinsic efficiency.

One method to regulate origin activity is to change the probability that it will be replicated passively. As an elongating fork from an origin neighbor mediates this suppression, this form of origin deactivation has been termed “origin interference” (Brewer and Fangman, 1993). Most of our understanding concerning origin interference has been provided by work in *S. cerevisiae*. In budding yeast, there are many more assembled pre-replicative complexes (pre-RCs) than those that are either needed or used to complete replication (Dershowitz and Newlon, 1993; Raghuraman *et al.*, 2001; Wyrick *et al.*, 2001; Pasero *et al.*, 2002). Analysis of origin efficiency on yeast chromosomes III and VI, revealed that origins are used between 5 and 90% of cell cycles (Friedman *et al.*, 1997; Yamashita *et al.*, 1997; Poloumienko *et al.*, 2001). Licensed origins are inefficient owing to their scheduled timing late in S phase or relatively late compared with other origins in the vicinity (Santocanale and Diffley, 1996; Vujcic *et al.*, 1999). As a consequence, these competent origins are replicated passively by forks that elongate from flanking initiation sites (Santocanale *et al.*, 1999).

According to the above-mentioned explanation for origin interference, origin neighbors must be preprogrammed in G₁ to fire at different times during S phase (Raghuraman *et al.*, 1997). This requirement, however, may not be satisfied in higher eukaryotes, where 1) timing control is exerted over extended regions of ~100 kb (MacAlpine *et al.*, 2004; Norio *et al.*, 2005) and 2) origins situated next to each other fire simultaneously in clusters (Berezney *et al.*, 2000). In contrast, interference between yeast origins can still occur even though they have been programmed to fire simultaneously

This article was published online ahead of print in *MBC in Press* (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E06-04-0298>) on September 27, 2006.

[□] The online version of this article contains supplemental material at *MBC Online* (<http://www.molbiolcell.org>).

[†] Present address: Harvard Medical School, Boston, MA 02115.

Address correspondence to: Aaron Bensimon (abensim@pasteur.fr).

(Brewer and Fangman, 1993; Dubey *et al.*, 1994). Indeed, at the amplified AMPD2 locus of Chinese hamster ovary (CHO) cells, significant predefined timing differences between nearby origins was not observed (Anglana *et al.*, 2003). Nevertheless coactivation of adjacent origins at well-defined base pairs locations was blocked (Anglana *et al.*, 2003). Which mode of origin interference applies to broad initiation zones in human cells remains to be determined.

To understand how regular initiation intervals are achieved in human cells and whether origin interference contributes to this process, we queried a 1.5-Mb region of human chromosome 14q11.2 from primary keratinocytes for origin activity. A single molecule approach exploiting molecular combing technology was chosen for the following reasons. First, sufficient origin firing events can be obtained to position all the potential start sites of DNA replication in a particular cell type. Second, we could determine which origins single cells use in individual S phases and their activation timing with respect to each other. This is required to ascertain the spatiotemporal distribution of initiation events. These data were combined to evaluate whether origins that have already fired regulate downstream potential initiation site use. We found that origins self-regulate one another according to a hierarchy established by the active origin, which is selected stochastically without predefined timing preferences. Furthermore, origin interference yields conserved initiation event spacing. The reasons for and the mechanisms used to implement human origin interference are discussed.

MATERIALS AND METHODS

DNA Preparation

Normal human primary keratinocytes were derived from skin biopsies. Cells were cultured in standard keratinocyte-SFM (Invitrogen, Carlsbad, CA). Thereafter, nascent DNA was labeled with 5-iodo-2'-deoxyuridine (IdU) and 5-chloro-2'-deoxyuridine (CldU) for 20 min each as described previously (Lebofsky and Bensimon, 2005). DNA extraction and combing were done according to standard protocols.

Hybridization and Fluorescent Revelation

The 25 GMC probes were produced by long-range polymerase chain reaction (PCR) by using primer pairs listed in Supplemental Table S1. To help amplify 5- to 7-kb probes, TaKaRa LA Taq was used (Takara Bio, Gennevilliers, France). Bacterial artificial chromosomes that served as templates for PCR products are found in Supplemental Table S2. Probes were pooled at a final concentration of ~20 ng/ μ l according to their symbol (a, b, c, and d). Biotinylation of probes was achieved by random priming (Invitrogen) the four symbols separately. For individual slide assays, ~250 ng of each biotinylated probe was combined with 10 μ g of Human cot-1 DNA (Invitrogen). After that, hybridization on combed DNA conformed to published methods (Lebofsky and Bensimon, 2005). The immunofluorescent steps to detect probes, IdU and CldU, were as follows: 1) Alexa 488-conjugated streptavidin (Invitrogen), mouse anti-5-bromo-2'-deoxyuridine (BrdU) (BD Biosciences, San Jose, CA), and rat anti-bromodeoxyuridine (Harlan Serlab, Crawley, United Kingdom); 2) biotin-conjugated rabbit anti-streptavidin (Rockland, Gilbertsville, PA), Alexa 350-conjugated goat anti-mouse (Invitrogen), and Texas Red-conjugated donkey anti-rat (Jackson ImmunoResearch Laboratories, West Grove, PA); 3) Alexa 488-conjugated streptavidin (Invitrogen), and Alexa 350-conjugated donkey anti-goat (Invitrogen). Antibody incubations, washes, and slide mounting were performed as reported previously (Lebofsky and Bensimon, 2005).

Image Acquisition

Half of the images were captured with a Zeiss Axioplan 2 microscope equipped with an HQ charge-coupled device camera (Photometrics, Tucson, AZ). The other half was acquired using the Cytoscout high-throughput scanning device (Upper Austrian Research, Linz, Austria). Background fluorescent dots were removed using Photoshop (Adobe Systems, Mountain View, CA) to highlight the molecule of interest.

Sequence Analysis

The 03/2006 NCBI Build 36.1 version of the human genome produced by the International Human Genome Consortium was used as a reference for all sequence analyses. Known protein encoding genes were annotated from the

National Center for Biotechnology Information mRNA reference sequences collection (RefSeq). Gene expression was assessed using the available data from the Genomics Institute of the Novartis Research Foundation (Su *et al.*, 2002). Base composition was analyzed using the Artemis software package (<http://www.sanger.ac.uk/Software/Artemis>). To find AT-rich regions, a 500-base pair sliding window was used.

RESULTS

Genomic Morse Code (GMC) Hybridization Strategy

As a first step toward analyzing which origins are active and silent within a given cell cycle, all potential initiation sites within a region were obtained. For this purpose, we used a single molecule approach based on molecular combing (Bensimon *et al.*, 1994). Asynchronous human primary keratinocytes displaying a normal karyotype were given 20-min pulses of IdU followed by CldU. DNA from this cell population was extracted and combed. The incorporation of the BrdU analogues along newly synthesized DNA was visualized by immunological detection of IdU in blue and CldU in red. This experimental paradigm gives rise to three types of signals from which the start site of bidirectional replication can be inferred (Figure 1A; for a detailed description, see Anglana *et al.*, 2003; Lebofsky and Bensimon, 2005). By combining BrdU revelation with fluorescence in situ hybridization, initiation can be attributed to specific sequence tracts wherever replication tracts overlap with probe signals.

Conventionally, probe pairs of either different color (Figure 1B, i) or different size (Figure 1B, ii) are used to visualize a genomic region on the slide. DNA breakage, however, limits the distance they can cover. Thus, walking down a chromosome by using this hybridization strategy is a time-consuming process because experiments increase proportionally to the number of probe pairs. Ideally, one could use several colors and/or probes of varying lengths to cover a large region; however, these strategies are not suitable owing to spectral overlap and nonspecific hybridization of repetitive sequences, respectively.

The first part of our unique solution came with the realization that gaps of different size provide the same information as probes of different color or size. In the example provided in Figure 1B, iii, gap 1 is defined by one probe set and gap 2 is defined by another probe set. Also, the gap size between the two probe sets is distinct from gaps 1 and 2. Gap 3 becomes useful during DNA breakage. With its help, the molecule can still be oriented even though the complete set of probes is not visualized (Figure 1B, iv and v). As gaps provide positional information, their numbers are no longer limited, i.e., spectral overlap and repetitive sequences during hybridization are no longer an issue. By using gaps of different sizes, a GMC covering ~1.5 Mb in human chromosome 14q11.2 was generated (Figure 1C). The entire GMC was hybridized in individual assays, and all probes were detected in green. Before molecular combing, DNA manipulation causes the fibers to break in random locations. Fiber size, however, was frequently sufficient to permit the visualization of multiple symbols on individual molecules. In contrast, because of fiber breakage, occasionally only a few of the probes from a symbol were detectable. Origins were mapped whenever replication tracks denoting initiation colocalized with a decodable set of GMC probes (Figure 1D). Thus, the novel hybridization strategy, GMC, allowed origin mapping over a large region in a limited number of experiments.

Initiation Mapping on 1.5 Mb of Human Chromosome 14q11.2

Using this experimental paradigm, we detected 307 initiation events on 232 single DNA molecules in the GMC region.

Figure 1. DNA replication initiation mapping on 1.5 Mb in human chromosome 14q11.2. (A) Three types of replication signals on combed DNA that indicate an origin. Initiation occurs before the labeling periods, during the IdU pulse (blue) and during the CldU pulse (red) giving rise to the signals shown in i, ii, and iii, respectively. In all three cases, the midpoint of the tracks is assumed to be the site of initiation. (B) Hybridization strategies on combed DNA. Two probes of equal length but detected with different colors (i) or two probes of different length detected with the same color (ii) are hybridized to visualize a genomic region of interest. Alternatively, gaps between probe sets can be used to provide the same information. In iii, four short probes are hybridized giving rise to three informative gaps. Gap 3 allows the molecule to be oriented during breakage (iv and v). DNA breaks are denoted by a pair of vertical solid lines. (C) GMC covering 1.356 Mb in human chromosome 14q11.2. The linear patterns of the first four symbols in Morse code, a, b, c, and d, are provided. GMC comprises these four symbols, each symbol represented by a collection of probes. Probes are shown in green. Coding gaps are short gap and long gap. Start gap and end gap were included to help orient symbols when DNA breaks. Symbols are separated by space gaps. Probe and gap sizes in kb are given above each, respectively. (D) Examples of observed initiation events in the GMC region. White arrows indicate the initiation site. For fibers 2, 7, 8, and 13, GMC is still decidable, even though all probes pertaining to a symbol are not present. Initiation events flanking the symbols were mapped when one of the adjacent symbols was decoded (molecules 1, 5, 11, and 14) or space gap information was available (molecules 4 and 10). Bar, 100 kb.

Data clustering was carried out to objectively establish zones of preferential initiation. First, we created a hierarchical clustering tree (Duda *et al.*, 2001). To achieve the best partition, the spread of data within clusters should be minimized and the separation between clusters should be maximized. These two features are called within variance (W) and between variance (B) (Figure 2A). Hence, the desired cluster set must have small W and big B or maximal values of $B - W$. When the data were divided into 9, 22, and 45 clusters, relatively high $B - W$ values were obtained (Figure 2A). Figure 2B shows how initiation events are partitioned according to these cluster sets. Dividing the data set into 45 clusters yields the narrowest regions of initiation. We con-

sidered these 45 clusters to represent individual initiation zones (Figure 2C). Some of these clusters contained very few initiation events, which may have been due to background noise. Therefore, only clusters with greater than three initiation events were used for subsequent analyses. In this way, 38 initiation zones were identified and their sizes varied between 2.6 kb (minimum; min.) and 21.6 kb (maximum; max.) with an average of 13.5 ± 5.2 kb. These values fall within the range of other initiation zones reported for mammalian cells (DePamphilis, 1999).

Using combing technology, initiation sites are positioned at the midpoint of the two bidirectional replication signals (Figure 1A). Importantly, heterogeneity in the speed of the

Figure 2. Initiation zone identification by cluster analysis. (A) Defining ideal clusters. Equations for within variance (W) and between variance (B) are shown in the inset. For W , N is the number of clusters and V_i is the variance of cluster i . For B , C_i is the centroid of cluster i and $\bar{C}$ is the mean of all the centroids in a cluster set. Plotting $B - W$ as a function of the number of clusters revealed maximal values when the data were divided into 9 (blue circle), 22 (green circle), and 45 (red circle) groups. (B) The breakdown of one cluster into its components when 9 (blue line), 22 (green line), and 45 (red line) partitions are applied to the data. Regions underneath the red lines represent initiation zones, which are illustrated by the white boxes above the molecules. White arrows indicate the initiation site. Bar, 100 kb. (C) Distribution of initiation zones in the GMC region. The horizontal red lines denote zone size and position. Vertical white lines designate positions of all the 307 initiation sites mapped. Zones that contain fewer than four data points are marked by a horizontal gray line instead of a red line and are not considered in subsequent analyses. A horizontal black line indicates the cluster in B. Bar, 100 kb.

two outgoing forks would artificially yield initiation zones even though discrete origin sites of 1–2 kb underlie the initiation events. To exclude this possibility, fork speed in the GMC region was analyzed. Fork speed was calculated by dividing the length of individual blue or red segments by the time of labeling for the corresponding IdU or CldU pulses (20 min). On average, forks traveled at a rate of 1.24 ± 0.32 kb/min ($n = 572$). Indeed, the high SD indicates that fork speed was heterogeneous in the GMC region; however, as mentioned previously, accuracy of assigned origin position specifically depends on the fork speed difference of the two outgoing forks from a single initiation event and not on the global fork speed variability. To analyze bidirectional fork speed variation from a single origin, fork speeds on either side of an initiation event were compared with each other. The 184 fork speed pairs were found to be highly correlated ($R = 0.884$; $p < 0.001$),

and this correlation was found irrespective of the distance between the two outgoing forks. Furthermore, only 2.1% (4/189) of the pairs displayed $>30\%$ fork speed differences. Assume that two outgoing forks travel at the mean 1.24 kb/min fork speed and possess 30% speed differences. This would translate into an error of 7.4 kb when attempting to position an initiation site ($(1.24 \text{ kb/min} \times 20 \text{ min}) - (1.24 \text{ kb/min} \times 70\% \times 20 \text{ min})$). The theoretical 7.4-kb error is approximately one-half the size of the average 13.5-kb initiation zone. Furthermore, considering that almost no two pairs of forks showed 30% fork speed differences, this error is a high estimate. Therefore, our finding that initiation takes place in zones is unlikely to be a consequence of errors in determining origin position. Alternatively, because of this error, we cannot exclude the possibility that discrete initiation sites reside at the smaller initiation zones (<7 kb). In the future, these re-

Figure 3. Initiation zone correlation with gene location and sequence. Horizontal red lines denote initiation zones. Initiation events that contributed to initiation zone mapping are shown as vertical white lines. Blue rectangles illustrate gene position and size. The gene name is given above the corresponding blue rectangle. Initiation zones that lie entirely in intergenic regions look white. Light gray, dark gray, and black initiation zones represent zones that reside in 90, 50–90, and <50% intergenic regions, respectively. Sequence elements underlying the initiation zones were analyzed as follows: AT, 500-base pair elements with AT content >70%; AG, at least 50 base pairs with AG content greater than 98%; (TTA)₄₋₅; A₃₋₄, five A₃₋₄ stretches each separated by 10 base pairs. The numbers of these elements found within an initiation zone are provided below each initiation zone. Bracketed AT numbers indicate AT-rich 500-bp elements with >65% AT content. Bar, 100 kb.

regions can be probed with other higher resolution techniques (Todorovic *et al.*, 2005) to resolve this ambiguity.

To ascertain the correlation between initiation zones and genes, we positioned all known genes in the region (Figure 3). Of the 38 initiation zones, 25 (66%) were situated entirely in the intergenic regions. Eight (21%) and three (8%) initiation zones had >90 and 50% intergenic sequences, respectively. Only two (5%) zones contained >50% gene-encoding sequences. The expression of 36 of the 46 genes found in this locus were previously profiled using identical tissue, namely, skin (Su *et al.*, 2002). Although all 36 genes in the locus were expressed at low or barely detectable levels, none were highly expressed in this cell type. In summary, the majority of the initiation zone sequence reported in this study map to intergenic regions, and this was not related to high expression of genes found in the locus.

We then searched for sequence motifs that had been previously reported for other Metazoan initiation zones (Aladjem and Fanning, 2004). Motifs analyzed included AT-rich elements, AG-rich tracks (TTA)₄₋₅, and A₃₋₄ positioned at 10 nucleotide intervals (Figure 3). Twenty-nine of the 38 (76%) initiation zones possessed at least one and up to 11 elements with >70% AT content. Of the remaining nine zones, eight contained at least one element with >65% AT content. Twelve initiation zones (32%) contained AG-rich elements, five (13%) contained the (TTA)₄₋₅ motif, and four (11%) contained the A₃₋₄ motif. We could find no common ar-

rangement of sequence motifs giving rise to an initiation zone. Nevertheless, our data confirm the presence of sequence motifs, which reside at other mammalian initiation zones.

Spatiotemporal Analysis of Activated Origin Neighbors

We next turned our attention toward how initiation zones were distributed relative to one another. Measuring distances between zone centroids revealed an interzone average of 40.6 ± 20.7 kb (min. = 14.3 kb; max. = 93.1 kb). This was surprising considering that interorigin distances in mammalian cells generally range between 100 and 150 kb (Berezney *et al.*, 2000). The discrepancy can be explained if only a subset of zones is activated per cell cycle. To explore this possibility, we analyzed the spacing between multiple initiations on individual fibers (Figure 4A). Because of the single molecule level of our analysis, these origins correspond to those that are actually used by one cell in one S phase. DNA breakage prevented the visualization of flanking origins for 173 of the 307 initiation events observed. The remaining cases were observed in the presence of an active origin neighbor (134/307). The two nearest and the two furthest functional origins were separated by 31.4 kb and 390.8 kb, respectively. Interestingly, the mean interorigin distance was calculated as 113 ± 66.4 kb (Figure 4B). In comparison with the interzone distance (~40 kb), this result

Figure 4. Spatiotemporal analysis of functional origins. (A) Replication signals that provide interorigin distances (X). In i, the replication tracks from two initiation sites remain separate. In ii and iii, oncoming forks merge during the IdU and CldU pulses, respectively. (B) Histogram showing the frequency of measured interorigin distances. (C) Examples of molecules with at least two initiation events in a and b (i) and c and d (ii). White arrows indicate the initiation site. Initiation zones are marked by horizontal red lines. For individual molecules, the initiation zone from which an origin fires is indicated by a white box. Dark boxes designate silent initiation zones. Bar, 100 kb.

suggests that, on average, only one origin fires from out of three potential zones in a given cell cycle (Figure 4C).

We next considered whether the drop-off of small and large interorigin distances (Figure 4C) were due to the experimental limitations of combed technology. In the first instance, resolution limits could have affected our ability to detect two closely spaced origins. The resolution limit of combed DNA is 1–5 kb. Therefore, detection of linear segments by using this technique requires 1–5 kb of labeled DNA. Assuming a lower resolution limit of 5 kb, the closest that two theoretical origins could be distinguished would be ~15 kb. In this case, both origins would be required to fire at the end of either the first (IdU) or second (CldU) pulse. For the former pulse, a 5-kb red CldU label would split the two 5-kb blue IdU labels where the origins had fired. For the latter, a 5-kb unlabeled or nonfluorescent DNA fragment would separate the two 5-kb red CldU signals where the origins had fired. Fifteen kilobases is significantly smaller

than the distance between the two closest origins detected (31.4 kb). Therefore, the drop-off of short interorigin distances is probably not a consequence of combing resolution limits. With respect to large interorigin distances, DNA breakage could have precluded our ability to detect distant origins. To address this possibility, the lengths of the fibers with multiple origins were measured. The mean fiber length was calculated as 347.9 ± 92.1 kb. The two furthest origins were separated by 390.8 kb; however, large interorigins started to drop-off at ~175 kb (Figure 4B), which is significantly shorter than the mean fiber length. Therefore, at least in the 400-kb range, our inability to observe a substantial number of interorigin distances beyond 175 kb is not likely to be because of fiber breakage. Conversely, for origins that are spaced further apart, i.e., >500 kb, the current size of combed molecules would be insufficient for their simultaneous detection. Therefore, interorigin distances of this greater magnitude would remain unaccounted for.

To investigate whether origins from specific zones reproducibly fired early or later with respect to one another, activation times were examined. Based on the type of replication signals indicating an origin (Figure 1A), the time of initiation with respect to the labeling periods could be attributed. This applies to origins that fired during either the IdU or CldU pulses, which were 20 min each in duration (Figure 1A, ii and iii). This equally applies to origins that fired before the labeling period, provided that the outgoing forks could be visualized by their incorporation of the modified nucleotides (Figure 1A, i). For this to occur, the time of origin activation could not precede the IdU/CldU pulses by more than 20 min on average (for example, see the second molecule in Figure 4C, i). Therefore, the window of analysis covers ~60 min in total, comprising 20 min before the pulses, 20 min during the IdU pulse, and 20 min during the CldU pulse. We could find no timing preferences for any of the 38 initiation zones. Furthermore, adjacent origins did not fire at the same time (Figure 4C). It should be noted that multiple initiations on individual fibers were detectable within the 60 min afforded by our experimental paradigm. Therefore, although precise synchrony between initiation events was not observed, the timing differences between any two activated origins are limited to ~1 h.

Because activation times between adjacent origins were slightly staggered, potential origins in the unreplicated regions between two oncoming forks might still have been activated at some later time. Origins firing from these regions would yield lower interorigin distances. Indeed, occasionally, retarded origins between two previously activated origins were detected, which reduced the interorigin distance (for an example, see the second and fourth origin in molecule 7; Figure 4C, ii). The majority of adjacent origins are considered to fire within 30 min of each other (Berezney *et al.*, 2000). Because our window of analysis is 60 min (see above paragraph), almost all origins within a cluster are predicted to be activated. Therefore, although it is possible that retarded origins could fire thereby reducing the interorigin distances measured, it is unlikely that this phenomenon would significantly alter the mean interorigin distance reported here.

Fork Extension across Potential Initiation Sites

Up until now, replication tracks have been used only for the purpose of inferring their start site or initiation. Their bidirectional extension into the surrounding region, however, provides another important piece of data. Signals originating from one initiation zone that overlap a flanking zone implies for the latter the prior passage of a replication fork and removal of an origin's license. This renders the passively replicated zone refractory from firing at some later time in S phase. Insofar as all potential origins in human cells are licensed as they are in yeast (Santocanale and Diffley, 1996), this observation provides evidence for origin interference (Figure 5A). For forks that extend partway into an initiation zone, zones were only considered as suppressed if the centroid was reached. We used signals from elongating forks to analyze how far from an active origin interference occurs (Figure 5B).

In total, 528 initiation zones were found to be suppressed. Their distance from the initiation site did not significantly differ when labels representing centromeric moving forks were considered (56.5 ± 37.7 kb; $n = 266$) versus labels representing telomeric moving forks (55.1 ± 38.4 kb, $n = 262$; Figure 5C). Consequently, for the following analyses, origin interference mediated by forks moving in both directions was combined. The closest and farthest suppressed

Figure 5. Origin interference based on fork extension. (A) Forks elongating from the active origin in zone (v) cover the region bounded by the vertical line pairs. The termination of the leftward moving fork is observed (inverted open triangles). Initiation zones i and ii are not interfered with, because the fork from origin (v) does not extend to its boundaries. Zones iii, iv, and vi are suppressed as the fork passively replicates their entire lengths. The rightward moving fork penetrates zone (vii), but it does not reach its centroid (black dot). This zone is not included in the origin interference data. Gray rectangles designate initiation zones and dark gray rectangle designate suppressed initiation zones. (B) Examples of molecules that display origin interference. The initiation zones relevant to this figure are illustrated by the horizontal red lines. The white box marks the initiation zone from which origins fire (white arrows). Dark boxes indicate initiation zones that are suppressed because of fork extension. Bar, 100 kb. (C) Histogram showing the frequency of distances between an initiation event and zones interfered with. Zones that were suppressed by centromeric and telomeric moving forks are represented by negative and positive values, respectively.

zones were located 7 and 284.6 kb from the functional origin, respectively. On average, replication tracks from active origins overlapped with zones 55.8 ± 38 kb away. This distance is significantly shorter than the average length of combed DNA (see above commentary for interorigin distances). Therefore, the drop-off in frequency of suppressed zones at greater distances is probably not because of fiber breakage. These data translate into the following. The zone situated immediately next to an initiation site was suppressed 314 times. The second zone was suppressed 137 times, and subsequent zones after that (three or greater) were suppressed 77 times. Together, these data based on 528 suppressed initiation zones suggest that origin interference extends for the most part over one to two flanking initiation zones.

In addition to between-zone interference, we also analyzed within-zone interference. Forks from an active origin extended beyond the boundaries of its own initiation zone 100% of the time (for examples, see Figure 5B). If this form of interference is

robust, the probability of more than one origin firing per initiation zone in any given cell cycle should be low. To carry out this analysis depends on our ability to discriminate short replication tracts representative of closely spaced origins in relatively small initiation zones. The maximal resolution of linear fluorescent segments on combed DNA is 1–5 kb. This complicates the visualization of multiple initiations in zones smaller than the average of 13.5 kb. For the larger initiation zones, however, observation of several origins is not limited by the resolution of molecular combing. Regardless of initiation zone size, two or more origins were never observed to fire from within the same initiation zone in individual S phases. Therefore, in contrast to the between-zone interference that decreases with distance from the initiation event, within-zone interference is extremely efficient and does not depend on the distance forks have to travel.

DISCUSSION

In conclusion, we have mapped all detectable initiation zones throughout a 1.5-Mb region in human chromosome 14q11.2 by using a single molecule approach (Figures 1 and 2). Primarily, initiation zones are situated in the intergenic regions, even though none of the genes are highly expressed in the cell type used for this study (Figure 3). Common sequence motifs underlying other mammalian origins were also found in the reported initiation zones (Figure 3). We have also demonstrated that only a fraction of the initiation zones are actually used in individual cell cycles (Figure 4). Last, limited origin activation yields regular interorigin firing distances (Figure 4). Accordingly, we show for the first time that conserved initiation event spacing is maintained in the context of broad mammalian initiation zones.

A high potential to active origin ratio has been described in yeast and CHO cells (Raghuraman *et al.*, 2001; Wyrick *et al.*, 2001; Pasero *et al.*, 2002; Anglana *et al.*, 2003). Our data suggest that this ratio is a conserved feature in human cells. This raises an important question: Why is origin redundancy a recurrent theme in eukaryotic cells? Deleting several origins on one arm of a yeast chromosome had negligible effects on genome stability (Dershowitz and Newlon, 1993). This would suggest that so many origins are not necessary. More recently, however, it was shown that preventing the full complement of assembled pre-RC resulted in chromosomal rearrangements (Lengronne and Schwob, 2002; Tanaka and Diffley, 2002). Although the reason for this is unknown, several proposals converge on the idea that an excess of potential origins provides a safety net in the event of perturbed DNA replication (Schwob, 2004). First, if a fork is blocked, it can be converted into a substrate for recombination (Rothstein *et al.*, 2000). Activation of a downstream “extra” origin gives rise to an oncoming fork. This fork merges with the blocked fork thereby rescuing it from recombination. Second, if some origins fail to fire, cells may undergo mitosis with unreplicated DNA. This fragment will break when the centromeres are pulled apart. An oversupply of potential origins reduces the likelihood of this happening. Last, optimal cell cycle arrest by the S-phase checkpoint requires a sufficient number of forks (Shimada *et al.*, 2002). Forks are lost when an attempt to initiate fails. The firing of a backup origin generates two additional forks to compensate, thus rendering the checkpoint operational. Clearly, further work is needed to evaluate which of these models is applicable.

Origin interference has been invoked as a mechanism to explain how a high potential to active origin ratio is achieved in eukaryotes. It involves the removal of pre-RCs, which represent licensed origins, by forks progressing from

earlier activated origins (Brewer and Fangman, 1993). Origin interference has been observed in yeast, *Xenopus*, and CHO cells (Brewer and Fangman, 1993; Dubey *et al.*, 1994; Lucas *et al.*, 2000; Anglana *et al.*, 2003). Here, we show for the first time that origin interference plays a significant role in modulating origin function in human cells, and moreover, that this occurs in the context of initiation zones (Figure 5). Before molecular combing, DNA is deproteinated. Therefore, it was not possible to observe which of the initiation zones contained licensed origins. Indeed, passively replicated zones, which were interpreted as suppressed, may simply not have been licensed to begin with. Future work will assay pre-RC assembly among initiation zones. This will allow us to determine whether origin interference occurs according to the canonical definition of the term.

In yeast, origin interference can occur between origins that have been programmed to fire either at similar or different times (Brewer and Fangman, 1993; Dubey *et al.*, 1994; Lucas and Raghuraman, 2003). In agreement with work performed in CHO cells (Anglana *et al.*, 2003), we did not find any strong programmed timing differences for adjacent origins (Figure 4). Therefore, our data suggest that origin interference occurs between two origins with similar activation times in the context of initiation zones in human cells.

We observed that between-zone interference gradually decreases with distance from the active origin (Figure 5). If the probability of origin firing is low due to limited initiation factors (Walter and Newport, 1997), the origin interference reported here may be an indirect outcome of this low probability, and, consequently, a passive phenomenon. In contrast, if origin-firing probabilities are high, origin interference must be actively regulated. For example, checkpoint proteins that are present at unperturbed elongating forks might suppress distal origins from firing. This would actively increase the chance that delayed origins are passively replicated and therefore suppressed (Marheineke and Hyrien, 2004; Shechter *et al.*, 2004; Sorensen *et al.*, 2004; Syljuasen *et al.*, 2005). Future research will reveal which of these models is responsible for between-zone interference.

Recently, a mathematical study proposed that only potential origins 11 kb apart can be sequestered together in a replication focus and therefore activated simultaneously (Jun *et al.*, 2004). This restriction is determined by the persistence length of DNA, which limits DNA bending. Persistence length may explain within-zone interference: DNA stiffness prevents two potential initiation sites from one zone to be concentrated within a replication focus, thus preventing their simultaneous activation. The robustness of a mechanism based on the physical properties of DNA could produce the high efficiency of within-zone interference reported here.

The mechanism of origin interference within and among mammalian initiation zones depends upon the molecular determinants that underlie these regions. During licensing, multiple minichromosome maintenance (MCM) complexes spread away from pre-RCs (Ritzi *et al.*, 1998; Edwards *et al.*, 2002). It has been suggested that origins firing at one of these MCM sites explain the presence of initiation zones in mammalian cells (Hyrien *et al.*, 2003; Blow and Dutta, 2005; Cvetic and Walter, 2005). Accordingly, the initiation zones reported here (Figure 2) may arise due to reiterative MCM loading. Interestingly, the sequence motifs underlying the initiation zones were diverse and variably arranged (Figure 3), which may contribute to complex initiation factor recruitment. Furthermore, although the majority of initiation zones fall within intergenic regions, this restriction was not due to high levels of gene expression (Figure 3). Determining the molecular machinery and interactions responsible for mammalian initiation

zones will help us understand how human origin interference is executed and initiation event spacing is regulated.

ACKNOWLEDGMENTS

We thank Claude Lalou for work in preparing the keratinocytes used in this study. R.L. is supported by the Natural Sciences and Engineering Council of Canada and the Pasteur-Weizmann Foundation.

REFERENCES

- Aladjem, M. I., and Fanning, E. (2004). The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep.* 5, 686–691.
- Anglana, M., Apiou, F., Bensimon, A., and Debatisse, M. (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* 114, 385–394.
- Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., and Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science* 265, 2096–2098.
- Berezney, R., Dubey, D. D., and Huberman, J. A. (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* 108, 471–484.
- Blow, J. J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. *Nat. Rev. Mol. Cell Biol.* 6, 476–486.
- Brewer, B. J., and Fangman, W. L. (1993). Initiation at closely spaced replication origins in a yeast chromosome. *Science* 262, 1728–1731.
- Cvetic, C., and Walter, J. C. (2005). Eukaryotic origins of DNA replication: could you please be more specific? *Semin. Cell Dev. Biol.* 16, 343–353.
- DePamphilis, M. L. (1999). Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5–16.
- Dershowitz, A., and Newlon, C. S. (1993). The effect on chromosome stability of deleting replication origins. *Mol. Cell. Biol.* 13, 391–398.
- Dubey, D. D., Zhu, J., Carlson, D. L., Sharma, K., and Huberman, J. A. (1994). Three ARS elements contribute to the *ura4* replication origin region in the fission yeast, *Schizosaccharomyces pombe*. *EMBO J.* 13, 3638–3647.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*, Wiley: New York.
- Edenberg, H. J., and Huberman, J. A. (1975). Eukaryotic chromosome replication. *Annu. Rev. Genet.* 9, 245–284.
- Edwards, M. C., Tutter, A. V., Cvetic, C., Gilbert, C. H., Prokhorova, T. A., and Walter, J. C. (2002). MCM2–7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *J. Biol. Chem.* 277, 33049–33057.
- Friedman, K. L., Brewer, B. J., and Fangman, W. L. (1997). Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 667–678.
- Gilbert, D. M. (2004). In search of the holy replicator. *Nat. Rev. Mol. Cell Biol.* 5, 848–855.
- Hand, R., and Tamm, I. (1973). DNA replication: direction and rate of chain growth in mammalian cells. *J. Cell Biol.* 58, 410–418.
- Hyrien, O., Marheineke, K., and Goldar, A. (2003). Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *Bioessays* 25, 116–125.
- Jun, S., Herrick, J., Bensimon, A., and Bechhoefer, J. (2004). Persistence length of chromatin determines origin spacing in *Xenopus* early-embryo DNA replication: quantitative comparisons between theory and experiment. *Cell Cycle* 3, 223–229.
- Lebofsky, R., and Bensimon, A. (2005). DNA replication origin plasticity and perturbed fork progression in human inverted repeats. *Mol. Cell. Biol.* 25, 6789–6797.
- Lengronne, A., and Schwob, E. (2002). The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G(1). *Mol. Cell.* 9, 1067–1078.
- Lucas, I., Chevrier-Miller, M., Sogo, J. M., and Hyrien, O. (2000). Mechanisms ensuring rapid and complete DNA replication despite random initiation in *Xenopus* early embryos. *J. Mol. Biol.* 296, 769–786.
- Lucas, I. A., and Raghuraman, M. K. (2003). The dynamics of chromosome replication in yeast. *Curr. Top. Dev. Biol.* 55, 1–73.
- MacAlpine, D. M., Rodriguez, H. K., and Bell, S. P. (2004). Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* 18, 3094–3105.
- Marheineke, K., and Hyrien, O. (2004). Control of replication origin density and firing time in *Xenopus* egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint. *J. Biol. Chem.* 279, 28071–28081.
- Newlon, C. S., et al. (1991). Analysis of a circular derivative of *Saccharomyces cerevisiae* chromosome III: a physical map and identification and location of ARS elements. *Genetics* 129, 343–357.
- Norio, P., Kosiyatrakul, S., Yang, Q., Guan, Z., Brown, N. M., Thomas, S., Riblet, R., and Schildkraut, C. L. (2005). Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol. Cell* 20, 575–587.
- Pasero, P., Bensimon, A., and Schwob, E. (2002). Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev.* 16, 2479–2484.
- Poloumienko, A., Dershowitz, A., De, J., and Newlon, C. S. (2001). Completion of replication map of *Saccharomyces cerevisiae* chromosome III. *Mol. Biol. Cell* 12, 3317–3327.
- Raghuraman, M. K., Brewer, B. J., and Fangman, W. L. (1997). Cell cycle-dependent establishment of a late replication program. *Science* 276, 806–809.
- Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., and Fangman, W. L. (2001). Replication dynamics of the yeast genome. *Science* 294, 115–121.
- Ritzi, M., Baack, M., Musahl, C., Romanowski, P., Laskey, R. A., and Knippers, R. (1998). Human minichromosome maintenance proteins and human origin recognition complex 2 protein on chromatin. *J. Biol. Chem.* 273, 24543–24549.
- Rothstein, R., Michel, B., and Gangloff, S. (2000). Replication fork pausing and recombination or “gimme a break”. *Genes Dev.* 14, 1–10.
- Santocanale, C., and Diffley, J. F. (1996). ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J.* 15, 6671–6679.
- Santocanale, C., Sharma, K., and Diffley, J. F. X. (1999). Activation of dormant origins of DNA replication in budding yeast. *Genes Dev.* 13, 2360–2364.
- Schwob, E. (2004). Flexibility and governance in eukaryotic DNA replication. *Curr. Opin. Microbiol.* 7, 680–690.
- Shechter, D., Costanzo, V., and Gautier, J. (2004). ATR and ATM regulate the timing of DNA replication origin firing. *Nat. Cell Biol.* 6, 648–655.
- Shimada, K., Pasero, P., and Gasser, S. M. (2002). ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. *Genes Dev.* 16, 3236–3252.
- Shirahige, K., Iwasaki, T., Rashid, M. B., Ogasawara, N., and Yoshikawa, H. (1993). Location and characterization of autonomously replicating sequences from chromosome VI of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 13, 5043–5056.
- Sorensen, C. S., Syljuasen, R. G., Lukas, J., and Bartek, J. (2004). ATR, Claspin and the Rad9-Rad1-Hus1 complex regulate Chk1 and Cdc25A in the absence of DNA damage. *Cell Cycle* 3, 941–945.
- Su, A. I., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* 99, 4465–4470.
- Syljuasen, R. G., Sorensen, C. S., Hansen, L. T., Fugger, K., Lundin, C., Johansson, F., Helleday, T., Sehested, M., Lukas, J., and Bartek, J. (2005). Inhibition of human Chk1 causes increased initiation of DNA replication, phosphorylation of ATR targets, and DNA breakage. *Mol. Cell. Biol.* 25, 3553–3562.
- Tanaka, S., and Diffley, J. F. (2002). Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. *Genes Dev.* 16, 2639–2649.
- Todorovic, V., Giadrossi, S., Pelizon, C., Mendoza-Maldonado, R., Masai, H., and Giacca, M. (2005). Human origins of DNA replication selected from a library of nascent DNA. *Mol. Cell* 19, 567–575.
- Vujcic, M., Miller, C. A., and Kowalski, D. (1999). Activation of silent replication origins at autonomously replicating sequence elements near the HML locus in budding yeast. *Mol. Cell. Biol.* 19, 6098–6109.
- Walter, J., and Newport, J. W. (1997). Regulation of replicon size in *Xenopus* egg extracts. *Science* 275, 993–995.
- Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., and Aparicio, O. M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360.
- Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., and Shirahige, K. (1997). The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 655–665.