HAL
open science

Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations OPEN

Arach Goldar, A. Arneodo, B. Audit, Françoise Argoul, A. Rappailles, G. Guilbaud, Nataliya Petryk, Malik Kahli, Olivier Hyrien

To cite this version:

Arach Goldar, A. Arneodo, B. Audit, Françoise Argoul, A. Rappailles, et al.. Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations OPEN. Scientific Reports, 2016, 6 (22), pp.22469 (1-8). 10.1038/srep22469 . hal-01412574

HAL Id: hal-01412574

https://hal.science/hal-01412574v1

Submitted on 4 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SCIENTIFIC REPORTS

OPEN

Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations

Received: 26 June 2015
Accepted: 16 February 2016
Published: 03 March 2016

A. Goldar[1], A. Arneodo[2,3], B. Audit[2,3], F. Argoul[2,3], A. Rappailles[4,5], G. Guilbaud[4,6], N. Petryk[4,7], M. Kahli[4] & O. Hyrien[4]

We propose a non-local model of DNA replication that takes into account the observed uncertainty on the position and time of replication initiation in eukaryote cell populations. By picturing replication initiation as a two-state system and considering all possible transition configurations, and by taking into account the chromatin's fractal dimension, we derive an analytical expression for the rate of replication initiation. This model predicts with no free parameter the temporal profiles of initiation rate, replication fork density and fraction of replicated DNA, in quantitative agreement with corresponding experimental data from both *S. cerevisiae* and human cells and provides a quantitative estimate of initiation site redundancy. This study shows that, to a large extent, the program that regulates the dynamics of eukaryotic DNA replication is a collective phenomenon that emerges from the stochastic nature of replication origins initiation.

At the heart of genetic transmission, DNA duplication mechanisms are conserved among eukaryotes[1]. The core of the eukaryal replicative helicase, the MCM2-7 complex, is loaded around DNA in the form of an inactive head-to-head double hexamer (dh-MCM2-7) during the first phase (G1) of the proliferative cell cycle. During the following DNA synthetic (S) phase, a complex reaction, involving several replication factors, activates a fraction of dh-MCM2-7 to form a pair of divergent replication forks that unwind and replicate DNA until they meet with convergent forks assembled at adjacent initiation sites[1–4]. Initiation sites are called replication origins. Inactive dh-MCM2-7 at the start of S phase correspond to potential origins[5–9]. These may become activated later in S phase, or may be unloaded (inactivated) by progressing forks. The mechanisms that determine the location of potential and activated origins remain elusive[10,11]. While in *S. cerevisiae*, a unicellular eukaryote, origins are defined by a conserved DNA sequence motif[2], in metazoans no conserved sequence pattern is detected. However, in all eukaryotes the number of potential origins is higher than the number of fired ones[5]. The duration of S phase is finite and the DNA replication process must be completed within a reliable time. This constraint led to the assumption that origins firing is under the control of a deterministic program that regulates their rate and the spatio-temporal pattern of firing[12,13]. Recent experimental and theoretical works[14–17] challenged this view and suggested that a stochastic firing of randomly distributed potential origins could also meet the temporal constraint imposed by the cell cycle as long as the rate of origin firing increases as S phase progresses[5,16].

The majority of available mathematical and numerical models of DNA replication are founded on an analogy with a one-dimensional crystallization and growth process (KJMA model)[18]. This analogy allows to model replication dynamics by analyzing snapshots of the system to infer its evolution; this model describes the system's changes of states but not its evolution[16–22]. Due to the atomistic and geometric nature of the KJMA model in its simplest form, the exact position of fired origins must be defined (localized) to describe the replication dynamics of surrounding regions and the effect of the origin firing propagates along the DNA via the emanating replication

[1]Ibitec-S, CEA, Gif-sur-Yvette, France. [2]Université de Lyon, F-69000 Lyon, France. [3]Laboratoire de Physique, Ecole Normale Supérieure de Lyon, CNRS UMR5672, F-69007 Lyon, France. [4]Institut de Biologie de l'Ecole Normale Supérieure (IBENS) CNRS UMR8197, Inserm U1024, 75005 Paris, France. [5]Institut Pasteur, 75015 Paris, France. [6]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK. [7]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. Correspondence and requests for materials should be addressed to A.G. (email: arach.goldar@cea.fr)

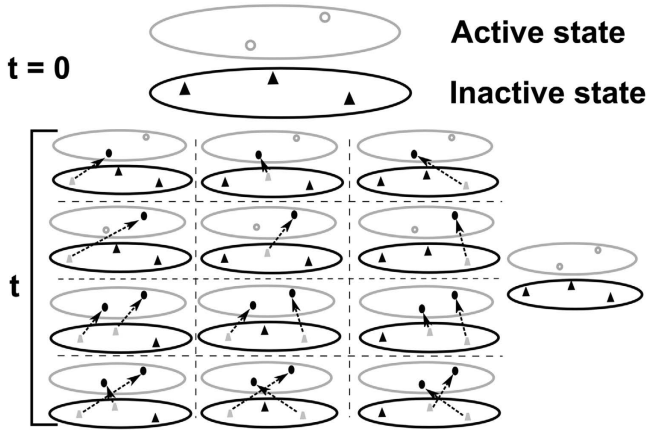

Figure 1. Non localized model of origin firing. m_0 dh-MCM2-7 complexes fill initially ($t = 0$) all positions of inactive state (filled black triangles). The active state is empty (O_{total} open grey circles). At time t, these complexes can transit from inactive state to active state with individual probability rate ψ (dashed arrows). By the end of this process, one or several potential origins (filled grey triangles) have fired (filled black circles). However, as potential origins are indistinguishable and the position and the time of fired origins are variable from cell to cell, one cannot designate precisely which potential origin corresponds to which fired origin and one must consider all possible configurations.

forks. Furthermore, in its simplest form, the KJMA model assumes the independence of firing among individual origins and in an arbitrary manner a temporal distribution of origin firing. Thus, these models of DNA replication are adequate to describe the replication process locally but cannot explain how it is influenced by the global compact conformation of the genome. In an effort to link the global conformation of the chromatin to the dynamic of replication, Gauthier & Bechhoefer[23] developed a model that reproduces the temporal profile of the rate of origin firing by assuming (i) a sequence of three- and one-dimensional replication origin search process for a replication initiation trans-acting factor, and (ii) the independence of firing among individual origins. Along the same line, to reproduce the experimental profile of the rate of origin firing, Goldar *et al.*[19] have assumed that this rate is regulated by the density of replication forks. The predictions of these models rely on the mechanistic ingredients used to describe the temporal changes of the rate of origin firing per time and per length of unreplicated DNA ($I(t)$).

In this paper, we explicitly introduce the unlocalized character of origin firing by picturing the firing process as a transition probability between two possible states (fired or not fired) for any potential origin. We describe the kinetics of replication by using a formal analogy between origin firing in a cell population and scattering process in inhomogeneous media[24]. This point of view does not require to know the exact position or distribution of replication origins along the genome and takes into account the effect of the compact conformation of the genome on the rate of origin firing. Our approach is thus complementary to existing ones: the KJMA model describes the state of a system, our non-local modeling describes the process that leads to the observed state. By considering the experimental observations that (i) dh-MCM2-7 act as potential origins (m_0), (ii) a stochastic process governs their firing and (iii) by the end of S phase a finite number of potential origins (O_{total}) have fired, we predict the temporal profile of the population-averaged number of fired origins as S phase progresses. The outcome of the developed model (i) is in good agreement with experimental observations of parameters describing the kinetics of DNA replication, (ii) confirms the adequacy of equilibrium globule picture to describe the budding yeast chromatin conformation and (iii) can be used to discriminate between two possible pictures to describe the chromatin conformation in human cells.

Results

As biological observations are performed in a cell population, the genomic positions of potential or fired origins and their firing times are not univocally defined[25]. This leads us to assume that at the start of S phase, there exists a cloud of potential origins (m_0 dh-MCM2-7 loaded on DNA during G1 phase) in each cell. Once the S phase starts, some of them transit from this inactive state to an active state (origin firing) until the end of the S phase where O_{total} origins are supposed to have fired, leaving $m_0 - O_{total}$ dh-MCM2-7 in inactive state. We call $k(t)$ the rate of transition (per potential origin, per time unit and per cell) between the inactive and active state. For simplification, we do not distinguish between the loaded but inactive dh-MCM2-7 state and the unloaded state. Lygeros *et al.*[26] have previously used the transition probability theory to model replication process in *S. pombe*. By distinguishing loaded but inactive potential origins and unloaded origins, these authors have defined 6 possible states for a potential origin. Here, we model the firing process distinguishing only between fired and non-fired origins. In this 2-state description, the rate of origin firing per cell is equal to the rate of transition, times the number of dh-MCM2-7 that are in inactive state, times the number of free locations in the active state as schematized in Fig. 1:

$$\frac{dO(t)}{dt} = k(t)(O_{total} - O(t))(m_0 - O(t)). \tag{1}$$

The non local picture of origin firing in a cell population depicted in Fig. 1 implies that each potential origin has the possibility to explore all available configurations in the active state before filling one of them. In other words, two observed fired origins at different times of S phase in a cell population can originate from a common potential origin. Therefore, the calculation of $O(t)$ is formally similar to the determination of the scattering amplitude in an inhomogeneous media (Supplementary material section 1), where the scattered intensity from two distinct scatterers can originate from a common scatterer[24]. Using this formal analogy and following Matsson's treatment of the ligand target interaction[27], the proportion of origin firing per cell at time $t\left(\rho(t) = \frac{O(t)}{m_0}\right)$ is represented as a Bethe-Salpeter like ladder graph which after summation yields:

$$\rho(t) \approx \frac{O_{total}}{2m_0}\left(\sum_{\nu=0}^{\infty} C^{\nu}\psi(t)^{\nu} - 1\right) \approx \frac{O_{total}}{2m_0}\left(\frac{1}{1 - C\psi(t)} - 1\right), \tag{2}$$

where $C = 2m_0/(m_0 + O_{total})$, and $\psi(t)(\ll 1)$ corresponds to the transition probability of an isolated dh-MCM2-7 that is associated in this picture with forward scattering amplitude (see Supplementary material for detailed derivation of Eq. (2)). Note that while $\psi(t)$ represents the probability of origin firing of an isolated potential origin (low density behavior of the system, meaning no interaction among fired origins), $\rho(t)$ corresponds to the probability of origin firing considering the cellular context (high density behavior, meaning interaction among fired origins). Direct insertion of Eq. (2) into Eq. (1) together with the change of variable $\phi(t) = (2m_0\rho(t)/O_{total} + 1)$ leads to the following compact evolution equation for the observed dynamics of origin firing per cell (Supplementary material section 2):

$$\frac{d\phi}{dt} = \frac{k'(t)}{a}[a^2 - (a^2 - b^2)\phi^2(t)], \tag{3}$$

where

$$k'(t) = m_0 k(t), \quad a = \frac{m_0 + O_{total}}{2m_0}, \quad b = \sqrt{\frac{O_{total}}{m_0}}. \tag{4}$$

As $\rho(t)$ is a probability, its values should be always positive, therefore only the forward solution of Eq. (3) has a physical meaning. Using the initial condition that at the start of S phase no origin has fired, we obtain the following general solution:

$$\rho(t) = \frac{b^2}{2}\left(\frac{a}{c}\tanh\left(c\int_0^t k'(t')dt' + \tanh^{-1}\left(\frac{c}{a}\right)\right) - 1\right), \tag{5}$$

where $c = \sqrt{a^2 - b^2}$.

Rate of transition $k(t)$. $k(t)$ represents the population-averaged transition rate between the inactive and active states per potential origin. The firing of an origin requires that trans-acting replication factors, that diffuse in the volume defined by the compacted genome (chromatin), find and activate one of the inactive dh-MCM2-7 complexes[25] that are able to freely diffuse on DNA[28]. Assuming that dh-MCM2-7 complexes are uniformly distributed along the genome, the radius of the volume explored by a dh-MCM2-7 scales as $R(t) \propto t^{\frac{1}{2}}$. Therefore, the probability $P_0(t)$ to find at time t a dh-MCM2-7 complex in the nuclear subspace filled by the chromatin is $P_0(t) = \left(\frac{R_0}{R(t)}\right)^{d_f} \propto t^{-\frac{d_f}{2}}$, where R_0 is the characteristic size of the dh-MCM2-7 and d_f is the chromatin's fractal dimension. The probability to find a trans-acting factor in the fractal structure of chromatin[29] at time t is proportional to $P_1(t) \propto t^{-\frac{d_f}{d_w}}$, where d_w is the fractal dimension of the trans-acting replication factor's random walk[30,31]. Hence, the probability that in an elementary volume at time t a trans-acting factor meets a dh-MCM2-7 is $S(t) = P_0(t)P_1(t) \propto t^{-d_f\left(\frac{1}{d_w}+\frac{1}{2}\right)}$. Since the spatial distribution of both dh-MCM2-7 and trans-acting factors are not homogeneous in the volume of the nucleus, the transport process that leads to the encounter between these two actors cannot be neglected. Thus, the rate of transition from inactive to active sites is no longer a time constant k_0 equal to the population averaged probability of origin firing per potential origin and per cell, but it has to be normalized by a fraction of the total number of dh-MCM2-7 and trans-acting factor encounters during the time t: $\int_0^t S(t')dt'$. This leads to the following time dependence of the transition rate:

$$k(t) = k_0 t^{d_f\left(\frac{1}{d_w}+\frac{1}{2}\right)-1}. \tag{6}$$

Fraction of replicated DNA: $f_{DNA}(t)$. To calculate $f_{DNA}(t)$, we use the analogy between DNA replication and one-dimensional nucleation and growth phenomena[18]. In this analogy, the firing of a potential origin corresponds to a nucleation event and the propagation of divergent replication forks at constant velocity v to a growth event. Following Avrami[32], we consider the genome at an instant t, and assume that $O(t)$ origins have already fired. The probability that, at time t, a particular locus of the genome is not covered by a particular replicon is $1 - \frac{2vt}{L_u(t)}$, where $L_u(t)$ is the length of the unreplicated genome. So the probability that it is not covered by any $O(t)$

replicons is $\left(1-\frac{2vt}{L_u(t)}\right)^{O(t)}$. Assuming that $2vt \ll L_u(t)$, this probability becomes $\left(1-\frac{2vt}{L_u(t)}\right)^{O(t)} \sim \exp\left(\frac{-2O(t)vt}{L_u(t)}\right)$. Finally, the probability that a locus is covered at time t, is just the fraction of replicated DNA:

$$f_{DNA}(t) = 1 - \exp(-\theta_{ext}(t)), \tag{7}$$

where $\theta_{ext}(t) = \frac{2O(t)vt}{L_u(t)}$. As firing of origins is an asynchronous phenomenon, in reality $\theta_{ext}(t) = \sum_i^{O(t)} \frac{2v}{L_u(t)}(t-t_i)$, where i is an index running over all fired origins. Each origin i fires and starts growing at t_i. We change the discrete sum on i to a continuous integral over time (t) and considering that $L_u(t) = L(1-f_{DNA}(t))$, we get for $\theta_{ext}(t)$:

$$\theta_{ext}(t) = \frac{2vm_0}{L}\int_0^t \frac{d\rho(t')}{dt'}\frac{t-t'}{1-f_{DNA}(t')}dt', \tag{8}$$

where L is the size of the genome.

Rate of origin firing per unreplicated length of DNA ($I(t)$) and fork density ($N_f(t)$). $I(t)$ is defined as the number of fired origins per unit of time and per unit of length of unreplicated DNA:

$$I(t) = \frac{d\rho(t)}{dt}\frac{m_0}{L(1-f_{DNA}(t))}. \tag{9}$$

It is interesting to note the similarity between Eq. (9) and the expression of $I(t)$ derived by Gauthier and Bechhoefer (Eq. (6) in ref. 23). Both expressions of $I(t)$ are obtained assuming that the trans-acting replication factor diffuses in the volume defined by the chromatin. However while here we consider the collective rate of origin firing $\frac{d\rho(t)}{dt}$, in ref. 23, the authors assume that the origins fire independently (see Supplementary material section 3, last paragraph for more discussion). Following the expression of domain (replication bubble) density calculated by Yang *et al.*[33], the density of replication forks is obtained under the following integral form:

$$N_f(t) = 2\int_0^t I(t')dt' \exp\left(-2v\int_0^t\int_0^{t'} I(t'')dt'dt''\right). \tag{10}$$

Then by introducing Eqs (5) and (6) into Eqs (7–10), we show that the dynamics of $f_{DNA}(t)$, $I(t)$ and $N_f(t)$ during the S phase can be completely characterized by the knowledge of 7 measurable parameters: m_0, O_{total}, d_f, d_w, k_0, v and L.

Recent technological developments have facilitated the access to the replication dynamics of *S. cerevisiae* and *H. sapiens* and provide some reliable quantitative estimates of our model parameters. *S. cerevisiae* has a genome of length $L^{S.c} = 12 \times 10^3 kb$ while the size of the haploid human genome is ~280 times larger ($L^H = 3200 \times 10^3 kb$). The number of dh-MCM2-7 complexes per cell has been estimated experimentally both in *S. cerevisiae* ($m_0^{S.c} = 322$)[34,35] and human HeLa cells ($m_0^H = 6.8 - 8.5 \times 10^4$)[36]. In *S. cerevisiae* on average $O_{total}^{S.c} = 168 \pm 20$ origins are referenced to fire systematically during a single S phase per cell[37,38]. In contrast the number of systematically fired origins in a human cell population is rather poorly known. Recent single-molecule[15] and genome-wide[3,39] studies estimated that on average between $O_{total}^H = 3$ to 9.2×10^4 origins fire per cell cycle. The speed of fork progression was measured experimentally in *S. cerevisiae*[40] as $v^{S.c} = 1.68\,kb.\,min^{-1}$ and was deduced from single-molecule and genome-wide replication timing studies of replicating HeLa cells[15] to range between $v^H = 0.8$ and $3.5\,kb.min^{-1}$. The geometrical fractal dimension d_f and the dynamic fractal dimension d_w can be combined to define the spectral dimension[41] $d_s = 2d_f/d_w$. The spectral dimension characterizes the power-law decay of the intra-chain contact probability of a polymer as $P_c \propto s^{-\alpha}$, where s is the number of monomers along the chain[41,42] and $\alpha = \frac{d_s}{2}$. From the experimentally measured distribution of the frequency of intra-chromosomal contact points, one can extract d_s. In the case of *S. cerevisiae*, it was experimentally measured that[43] $\alpha^{S.c} = d_s^{S.c}/2 = 3/2$. As the conformation of the chromatin inside the yeast nucleus can be reasonably considered to be an equilibrium globule[44], hence $d_f^{S.c} = 3$ and so $d_w^{S.c} = 2$ (normal diffusion). In HeLa, the observed the intra-chromosome contact probability was observed to be inversely proportional to the distance between the contact points[45], $\alpha^H = d_s^H/2 = 1$. In HeLa, two different models for chromatin organization inside the nucleus were proposed. The first and historical interpretation is to consider that the chromatin fiber is self-organized into a long-lived, non-equilibrium unknotted conformation allowing easy opening and closing of chromosomal regions over large distances in the nucleus[45]; this interpretation leads to model the chromatin as a "crumple" or fractal globule[44,45]. Following this model, as it is independently measured that in HeLa cells[46,47] $d_w^H = 2.6$ (subdiffusion), we conclude that $d_f^H = 2.6$ (see Dissussion). The second alternative interpretation is based on the recent analysis of Hi-C data in different human cell types by Boulos *et al.*[48]. By combining an integrative analysis of epigenetic maps and Hi-C data, these authors have shown that the 3D equilibrium globule model with $d_f = 3$ and $d_w = 2$ provides a comprehensive description of the Hi-C contact probability power-law exponent $\alpha = \frac{d_f}{d_w} = \frac{3}{2}$ observed in (i) embryonic stem cells as the signature of an accessible and permissive genome structure possibly shaped by pluripotency factors[49], and (ii) somatic cells between gene rich, early replicating euchromatin pairs of loci confirming that active chromatin in differentiated cell lines is preferentially positioned in the nuclear interior[49,50]. Importantly, Boulos *et al.*[48] have further shown that Hi-C contact probability exponent $\alpha \leq 1$ is indeed observed in differentiated cells between gene poor, late replicating heterochromatin pairs of loci as an indicator of the confining of this lamina-associated heterochromatin to the nucleus

periphery[49,50], consistent with the prediction of the 2D equilibrium globule model $d_f = 2$, $d_w \geq 2\left(\alpha = \frac{d_f}{d_w} \leq 1\right)$. Using this interpretation, we propose that the observed replication signals result from the superposition of the replication dynamics influenced by a 3-D and 2-D equilibrium globule organization of the chromatin fibre. To find the proportion of each signals, we follow the interpretation of Boulos *et al.*[48] and assume that the signal from 2-D equilibrium globule organization of the chromatin represents only 38% of the total signal, representing the amount of chromatin that interacts with the lamina in a constitutive manner[51].

Now, using Eq. (5) and the boundary condition that by the end of S phase (t_{end}) $\rho(t_{end}) = \frac{O_{total}}{m_0}$, we obtain

$$k_0 = \frac{2}{m_0}\left[\frac{d_f\left(\frac{1}{d_w} + \frac{1}{2}\right)}{1 - b^2} t_{end}^{-d_f\left(\frac{1}{d_w}+\frac{1}{2}\right)} \tanh^{-1}\left(\frac{1 - b^4}{2b^2 - (1 - b^2)^2}\right)\right]. \tag{11}$$

As during S phase, origins are fired in a continuous and irreversible manner[52], and only once per cell cycle[4], then $0 \leq k_0 \leq +\infty$ and from Eq. (10), we find the following boundaries to $\rho(t_{end})$:

$$\frac{1}{2} < \frac{O_{total}}{m_0} < 1. \tag{12}$$

This inequality is verified for *S. cerevisiae* $\left(\frac{O_{total}^{S.c}}{m_0^{S.c}} = 0.52\right)$, indicating that $\rho(t_{end})$ almost saturates the lower bound in Eq. (12). This observation turns out to be also valid for HeLa cells where the comparison of our model predictions with the replication dynamical data (see below) also selects an origin redundancy $m_0/O_{total} \simeq 2$ with $v^H = 1.1 \pm 0.3\,kb.\ min^{-1}$, $m_0^H = 7 \times 10^4$ and $O_{total}^H = 3.8 \times 10^4 \left(\frac{O_{total}^H}{m_0^H} = 0.54\right)$. Knowing that $t_{end}^{S.c} = 42\,min$[53] and $t_{end}^H = 480\,min$[15], we get $k_0^{S.c} = 7.7 \times 10^{-7}\,min^{-1}$ and $k_0^H = 1.1 \times 10^{-10}\,min^{-1}$.

Our 7 model parameters being fixed, we use Eqs (5–10) to numerically calculate $f_{DNA}(t)$, the flow cytometry (Facs) profiles, $I(t)$ and $N_f(t)$ for both *S. cerevisiae* and HeLa and compare the obtained theoretical profiles to recent experimental data reported in refs 15 and 53 respectively. As shown in Fig. 2, the agreement between theory and experiment is very good.

Discussion

The success of this analysis sheds light particularly on two aspects of DNA replication. First, we explicitly link the rate of origin firing to the global conformation of chromatin and to the diffusion of replication factors inside the nucleus. We find that for both considered organisms, the spectral dimension $d_s \geq 2$, suggesting that origin firing is only transiently regulated by the random encounter of a transacting factor and a dh-MCM2-7 complex[54]. Furthermore, in both cases the encounter probability $S(t)$ decreases faster than t^{-1}, a behavior that is representative of non compact exploration diffusion process (*i.e.* the number of sites explored by the transacting factor is smaller than the number of sites present in the volume defined by the chromatin)[30]. This is not surprising as only the encounter of a transacting factor with an inactive dh-MCM2-7 that is still bounded to a non replicated region of the genome can lead to the transition of the latter to the active state. Second, the irreversibility of replication process involves that the number of fired origins should at least represents half of the potential origins per cell. Note that our model further suggests that if during the S phase less than half of potential origins are used, the rate of transition $k(t)$ would have a dissipative component (k_0 become a complex number) inducing that by the end of S phase all the genome would not be replicated. The results reported in Fig. 2(b,b') provide a quantitative estimate of origin redundancy[8,9] in a single cell to $m_0/O_{total} \simeq 2$. We propose that the finite length of S phase applies an evolutive pressure that fixes m_0/O_{total}.

Profiles of $I(t)$, $N_f(t)$ and $f_{DNA}(t)$ are sensitive to the origin usage, but the shape of $I(t)$ and $N_f(t)$ are particularly sensitive to d_f and d_w in both *S. cerevisiae* and Hela (Supplementary material section 4). Importantly, our analysis confirms that the conformation of the chromatin in budding yeast can be represented as an equilibrium globule[44] in three dimensions ($d_f = 3$, $d_w = 2$) (Fig. 3(a–d)), consistent with the observed power-law decay of the intra-chromosome contact probability[43] with exponent $\alpha^{S.c} = \frac{d_s^{S.c}}{2} = \frac{d_f}{d_w} = \frac{3}{2}$. The scarcity of experimental replication data in Hela cells makes these data less selective for the estimate of d_f^H and d_w^H in human (Fig. 3(a'–d')). This explains that rather equal agreement of the replication data was obtained in Fig. 2(a'–c') with both the fractal globule model[44,45] and the 3D-2D equilibrium globule model[48]. The consistency in human somatic cells between replication data and the compartmentalization of the genome into an early replicating 3D equilibrium globule euchromatin organization in the nucleus interior and a late replicating 2D equilibrium globule heterochromatin confined at the nuclear envelop requires further investigation of new experimental data.

To conclude, the existing models of DNA replication[16–18,21,22] require an *a priori* knowledge of spatio-temporal map of origin firing, and the variability of the latter is treated as a small deviation from their population averaged values. Here, we explicitly consider the variability on the position and firing time distribution of origins in a cell population[14,25] and use a non-local treatment to calculate their rate of firing. This allows us to develop an effective description of DNA replication dynamics using a physical analogy between origin firing and scattering phenomena in an inhomogeneous medium. One of the outcome of such a description is that this dynamics is self-referential. The self-reference arises because we consider the replication process in a cell population, demonstrating that the temporal pattern of DNA replication is emergent and not predefined as in the KJMA theory. Furthermore, the distributed nature of our analysis (Supplementary material, section 3), allows (i) linking the

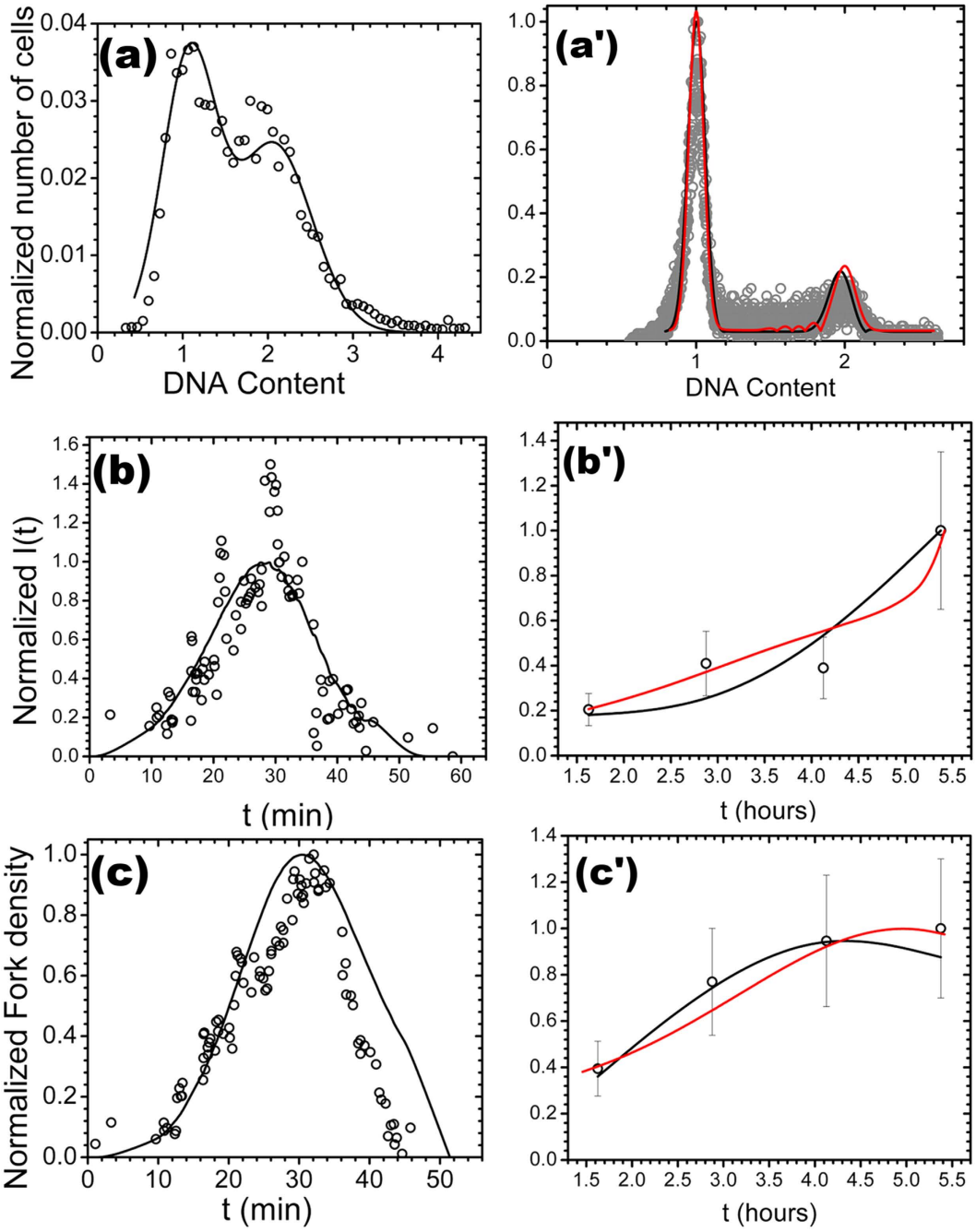

Figure 2. The open circles are experimental data and the solid lines are the calculated profiles. 3D equilibrium globule model of chromatin (black curve; $d_f = 3$, $d_w = 2$) for *S. cerevisiae* (data from Ma *et al.*[53]): (**a**) Facs profile calculated from $f_{DNA}(t)$[53] ($C = 0.94$, $P < 10^{-10}$); (**b**) $I(t)$ ($C = 0.87$, $P < 10^{-10}$); (**c**) $N_f(t)$ ($C = 0.89$, $P < 10^{-10}$). Fractal globule model of chromatin (black curve; $d_f = 2.6$, $d_w = 2.6$) for HeLa (data from Guilbaud *et al.*[15]): (**a'**) Facs profile ($C = 0.94$, $P < 10^{-10}$); (**b'**) $I(t)$ ($C = 0.94$, $P = 5.5 \times 10^{-12}$); (**c'**) $N_f(t)$, ($C = 0.96$, $P = 3.5 \times 10^{-2}$). 3D-2D equilibrium globule organization model of chromatin (red curve; $d_f = 3$ (62%), 2 (38%), $d_w = 2$ see text) for HeLa (data from Guilbaud *et al.*[15]): (**a'**) Facs profile ($C = 0.97$, $P < 10^{-10}$); (**b'**) $I(t)$ ($C = 0.96$, $P = 1.5 \times 10^{-2}$); (**c'**) $N_f(t)$, ($C = 0.93$, $P = 6.3 \times 10^{-2}$). $C = cov\left(X_{theo}X_{exp}\right)/(\sigma_{theo}\sigma_{exp})$ is the Pearson linear correlation coefficient.

kinetics of observed proportion of fired origins (Eq. (5)) to the averaged rate of single origin firing by taking into account the global topology of the genome inside the cell nucleus through its spectral dimension and (ii) ties the individual probability of firing of an origin to their collective observed probability (Eq. (2)). Therefore this model is a macroscopic model that overpasses the detailed molecular mechanisms necessary to the firing of an origin and retains the minimal necessary steps of DNA replication that are conserved among eukaryotes.

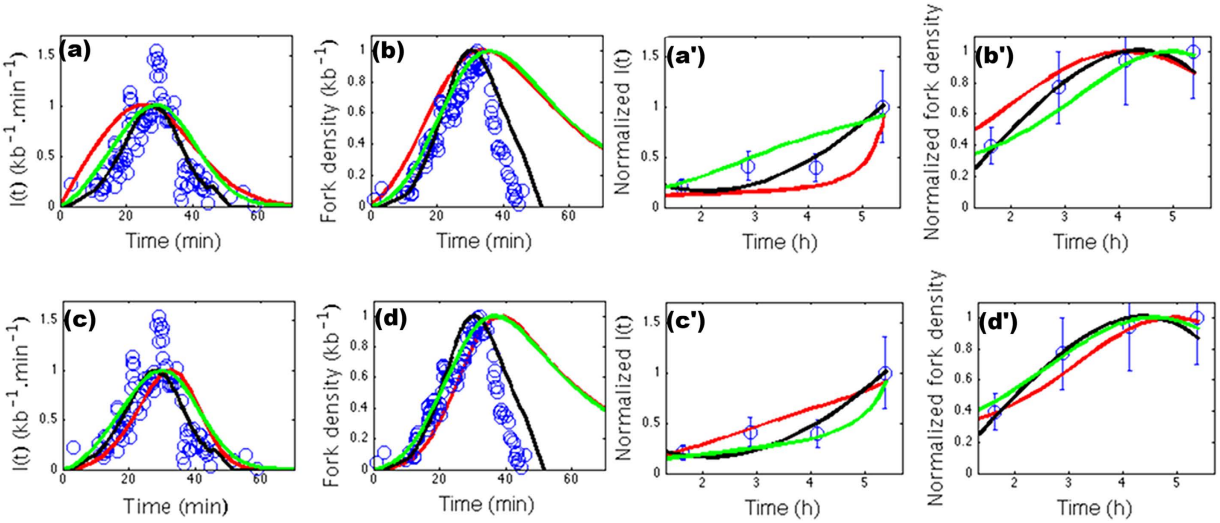

Figure 3. The open circles are experimental data and the solid lines are the calculated profiles. *S. cerevisiae* (data from Ma *et al.*[53]): (**a**) $I(t)$ and (**b**) $N_f(t)$ for different values of the chromatin fractal dimension: $d_f = 2$ (red), 3 (black) and 2.5 (green). (**c**) $I(t)$ and (**d**) $N_f(t)$ for different values of the dynamics fractal dimension: $d_w = 1.5$ (red), 2 (black) and 2.5 (green). HeLa (data from Guilbaud *et al.*[15]): (**a**) $I(t)$ and (**b**) $N_f(t)$ for different values of the chromatin fractal dimension: $d_f = 2$ (red), 2.6 (black) and 3 (green). (**c**) $I(t)$ and (**d**) $N_f(t)$ for different values of the dynamics fractal dimension: $d_w = 2$ (red), 2.6 (black) and 3 (green).

References

1. DePamphilis, M. L. & Bell, S. D. *Genome Duplication: concepts, mechanisms, evolution and disease* (Garland Science, New York, 2011).
2. Newlon, C. S. & Theis, J. F. The structure and function of yeast ARS elements. *Curr. Opin. Genet. Dev.* **3,** 752–758 (1993).
3. Mesner, L. D. *et al.* Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res.* **23,** 1774–1788 (2013).
4. Siddiqui, K., On, K. F. & Diffley, J. F. X. Regulating *DNA replication in eukarya*. Cold Spring Harb. Perspect. Biol. **5,** a012930 (2013).
5. Hyrien, O., Marheineke, K. & Goldar, A. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *Bioessays* **25,** 116–125 (2003).
6. Edwards, M. C. *et al.* MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in xenopus egg extracts. *J. Biol. Chem.* **277,** 33049–33057 (2002).
7. Harvey, K. J. *et al.* CpG methylation of DNA restricts prereplication complex assembly in xenopus egg extracts. *Mol. Cell. Biol.* **23,** 6769–6779 (2003).
8. Lucas, I., Chevrier-Miller, M., Sogo, J. M. & Hyrien, O. Mechanisms ensuring rapid and complete DNA replication despite random initiation in xenopus early embryos. *J. Mol. Biol.* **296,** 769–786 (2000).
9. Blow, J. J. & Ge, X. Q. A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep.* **10,** 406–412 (2009).
10. Hyrien, O. *et al.* From simple bacterial and archaeal replicons to replication N/U-domains. *J. Mol. Biol.* **425,** 4673–4689 (2013).
11. Renard-Guillet, C., Kanoh, Y., Shirahige, K. & Masai, H. Temporal and spatial regulation of eukaryotic DNA replication: from regulated initiation to genome-scale timing program. *Semin. Cell Dev. Biol.* **30,** 110–120 (2014).
12. Diffley, J. F. Replication conrol: choreographing replication origins. *Curr. Biol.* **8,** R771–R773 (1998).
13. Raghuraman, M. K. *et al.* Replication dynamics of the yeast genome. *Science* **a,** 115–121 (2001).
14. Czajkowsky, D. M., Liu, J., Hamlin, J. L. & Shao, Z. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J. Mol. Biol.* **375,** 12–19 (2008).
15. Guilbaud, G. *et al.* Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* **7,** e1002322 (2011).
16. Yang, S. C.-H., Rhind, N. & Bechhoefer, J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.* **6,** 404 (2010).
17. Retkute, R., Nieduszynski, C. A. & de Moura, A. Dynamics of DNA replication in yeast. *Phys. Rev. Lett.* **107,** 068103 (2011).
18. Herrick, J., Jun, S., Bechhoefer, J. & Bensimon, A. Kinetic model of DNA replication in eukaryotic organisms. *J. Mol. Biol.* **320,** 741–750 (2002).
19. Goldar, A., Labit, H., Marheineke, K. & Hyrien, O. A dynamic stochastic model for DNA replication initiation in early embryos. *PLoS One* **3,** e2919 (2008).
20. Goldar, A., Marsolier-Kergoat, M.-C. & Hyrien, O. Universal temporal profile of replication origin activation in eukaryotes. *PLoS One* **4,** e5899 (2009).
21. de Moura, A. P. S., Retkute, R., Hawkins, M. & Nieduszynski, C. A. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.* **38,** 5623–5633 (2010).
22. Baker, A., Audit, B., Yang, S. C.-H., Bechhoefer, J. & Arneodo, A. Inferring where and when replication initiates from genome-wide replication timing data. *Phys. Rev. Lett.* **108,** 268101 (2012).
23. Gauthier, M. G. & Bechhoefer, J. Control of DNA replication by anomalous reaction-diffusion kinetics. *Phys. Rev. Lett.* **102,** 158104 (2009).
24. Ishimaru, A. *Wave Propagation and Scattering in Random Media* (Academic Press, New York, 1978).
25. Aparicio, O. M. Location, location, location: it's all in the timing for replication origins. *Genes Dev.* **27,** 117–128 (2013).
26. Lygeros, J. *et al.* Stochastic hybrid modeling of DNA replication across a complete genome. *Proc. Natl. Acad. Sci. USA* **105,** 12295–12300 (2008).
27. Matsson, L. Response theory for non-stationary ligand-receptor interaction and a solution to the growth signal firing problem. *J. Theor. Biol.* **180,** 93–104 (1996).

28. Remus, D. *et al.* Concerted loading of MCM2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139,** 719–730 (2009).
29. Arneodo, A. *et al.* Multi-scale coding of genomic information: From DNA sequence to genome structure and function. *Phys. Rep.* **498,** 45–188 (2011).
30. De Gennes, P. Kinetics of diffusion controlled processes in dense polymer systems. I. Nonentangled regimes. *J. Chem. Phys.* **76,** 3316 (1982).
31. Kopelman, R. Rate processes on fractals: Theory, simulations, and experiments. *J. Stat. Phys*. **42,** 185–200 (1986).
32. Avrami, M. Kinetics of phase change. I. General theory. *J. Chem. Phys.* **7,** 1103–1112 (1939).
33. Yang, S. C.-H. & Bechhoefer, J. How xenopus laevis embryos replicate reliably: investigating the random-completion problem. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **78,** 041917 (2008).
34. Wyrick, J. J. *et al.* Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science* **294,** 2357–2360 (2001).
35. Xu, W., Aparicio, J. G., Aparicio, O. M. & Tavar, S. Genome-wide mapping of ORC and MCM2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. *BMC Genomics* **7,** 276 (2006).
36. Wong, P. G. *et al.* CDC45 limits replicon usage from a low density of preRCS in mammalian cells. *PLoS One* **6,** e17533 (2011).
37. MacAlpine, D. M. & Bell, S. P. A genomic view of eukaryotic DNA replication. *Chromosome Res.* **13,** 309–326 (2005).
38. Hawkins, M. *et al.* High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep.* **5,** 1132–1141 (2013).
39. Picard, F. *et al.* The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet.* **10,** e1004282 (2014).
40. Sekedat, M. D. *et al.* Gins motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* **6,** 353 (2010).
41. Rammal, G., R. & Toulouse. Random walks on fractal structures and percolation clusters. *J. Phys. Lett.* **44,** L13–L22 (1983).
42. Havlin, S. & Ben-Avraham, D. Diffusion in disordered media. *Advances in Physics* **36,** 695–798 (1987).
43. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465,** 363–367 (2010).
44. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19,** 37–51 (2011).
45. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
46. Weiss, M., Elsner, M., Kartberg, F. & Nilsson, T. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys. J.* **87,** 3518–3524 (2004).
47. Banks, D. S. & Fradin, C. Anomalous diffusion of proteins due to molecular crowding. *Biophys. J.* **89,** 2960–2971 (2005).
48. Boulos, R. E., Drillon, G., Argoul, F., Arneodo, A. & Audit, B. Structural organization of human replication timing domains. *FEBS Lett.* **589,** 2944–2957 (2015).
49. Julienne, H., Audit, B. & Arneodo, A. Embryonic stem cell specific "master" replication origins at the heart of the loss of pluripotency. *PLoS Comput. Biol.* **11,** e1003969 (2015).
50. Julienne, H., Zoufir, A., Audit, B. & Arneodo, A. Human genome replication proceeds through four chromatin states. *PLoS Comput. Biol.* **9,** e1003233 (2013).
51. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23,** 270–280 (2013).
52. DePamphilis, M. L. *DNA Replication and Human Disease* (Cold Spring Harbor Laboratory Press, 2006).
53. Ma, E., Hyrien, O. & Goldar, A. Do replication forks control late origin firing in saccharomyces cerevisiae? *Nucleic Acids Res.* **40,** 2010–2019 (2012).
54. Hwang, S., Yun, C.-K., Lee, D.-S., Kahng, B. & Kim, D. Spectral dimensions of hierarchical scale-free networks with weighted shortcuts. *Phys. Rev. E* **82,** 056110 (2010).

Acknowledgements

This work was supported by ANR REFOPOL, ANR10 BLANC1615, ANR-10-LABX-54 MEMO LIFE ANR-11-IDEX-0001-02 PSL* Research University and EDFV4_104.Rev.

Author Contributions

A.G. developed the model. A.G., A.A., B.A., F.A., A.R., G.G., N.P., M.K. and O.H. discussed the correctness and the biological relevance of the model and analyzed the data. A.G., A.A. and O.H. wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Goldar, A. *et al.* Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations. *Sci. Rep.* **6**, 22469; doi: 10.1038/srep22469 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/