## Replication timing is regulated by the number of MCMs loaded at origins

Shankar P. Das,<sup>1</sup> Tyler Borrman,<sup>1</sup> Victor W.T. Liu,<sup>1</sup> Scott C.-H. Yang,<sup>2,3</sup> John Bechhoefer, $^2$  and Nicholas Rhind<sup>1</sup>



Replication timing is a crucial aspect of genome regulation that is strongly correlated with chromatin structure, gene expression, DNA repair, and genome evolution. Replication timing is determined by the timing of replication origin firing, which involves activation of MCM helicase complexes loaded at replication origins. Nonetheless, how the timing of such origin firing is regulated remains mysterious. Here, we show that the number of MCMs loaded at origins regulates replication timing. We show for the first time in vivo that multiple MCMs are loaded at origins. Because early origins have more MCMs loaded, they are, on average, more likely to fire early in S phase. Our results provide a mechanistic explanation for the observed heterogeneity in origin firing and help to explain how defined replication timing profiles emerge from stochastic origin firing. These results establish a framework in which further mechanistic studies on replication timing, such as the strong effect of heterochromatin, can be pursued.

[Supplemental material is available for this article.]

 $T_{\text{max}} = T_{\text{max}} \sum_{i=1}^{N} \sum_{i=1}^{N} \sigma_i \sigma_i$ metabolism that correlates with correlates with  $\mathcal{U}$ late, chromatin structure, generally  $\mathbb{P}N$  repair, and cellu- $\frac{1}{2}$  differentiation (Goren and Cedar 2003; Gilbert et al. 2010).  $R = \frac{1}{2}$  is determined by the timing is determined by the timing of replication of replication or  $\frac{1}{2}$  $\mathcal{L} = \mathcal{L} \left( \begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix} \right)$ .  $\mathbf{P} \left( \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right)$ cell cycle, the Origin Recognition Complex (ORC) binds origin  $L$ and loads the ring-shaped MCM replicative helication  $\mathbb{M}$  $(B_2 \hspace{1cm} 10^{-1})$ . Activity of the MCM complex initiates in the MCM complex in the MCM complex in the MC replication and thus determines replication timing. Nonetheless,  $\mathcal{F}_{\text{max}}^{\text{W}}$  origin first origin first is regulated in the such original is uncertainty is uncertainty in the such as  $\mathcal{F}_{\text{max}}$ Although the average replication times of origins, as  $\mathcal{A}$  $\mathbf{u}_1 = \mathbf{v}_1 \mathbf{v}_2 = \mathbf{u}_2 \mathbf{v}_3 = \mathbf{u}_3 \mathbf{v}_4 = \mathbf{u}_3 \mathbf{v}_5 = \mathbf{u}_4 \mathbf{v}_5 = \mathbf{u}_5 \mathbf{v}_6 = \mathbf{u}_5 \mathbf{v}_7 = \mathbf{u}_6 \mathbf{v}_7 = \mathbf{u}_7 \mathbf{v}_8 = \mathbf{u}_7 \mathbf{v}_7 = \mathbf{u}_7 \mathbf{v}_8 = \mathbf{u}_7 \mathbf{v}_7 = \mathbf{u}_7 \mathbf{v}_8 = \mathbf{u}_7 \mathbf{v}_8 = \math$  $\mathcal{L}(\mathbf{R}) = \mathbf{q}(\mathbf{R})$  individual cells are heterogeneous (Rhind and Gilbert and Gil  $(0^*)$ . In fact, single-molecule studies in both budding and final fission buddhn  $y_{\alpha}$  that shown that origin first first (Patel et al. 2006;  $\mathcal{C} \times \mathcal{C}$ kowsky et al. 2008). Nonetheless, if individual origins fire storage stock  $\mathcal{C}$ chastically with a characteristic efficiency, they will exhibit reproducible average first average first origins for  $\mathcal{L}$ earlier, on average (Bechhoefer and Rhind 2012). Therefore, understand  $\mathbf{0}^*$ standing the timing of original first variables understanding  $\mathcal{L}(\mathcal{N})$ regulates the efficiency of  $\mathcal{L}$  origin first fir One strong influence on replication timing is heterochroma-

tin (Rhind and Gilbert 2013). In the budding year of the state  $\mathcal{L}_{\mathbf{r}}$  $m_{\rm{max}}$  meric heterochromatin is both necessary and sufficient to delay and  $\mathcal{F}_A = \mathcal{F}_A$  first is regulated in  $\mathcal{F}_B = \mathcal{F}_A = \mathcal{F}_B$ . This effect is regulated in by Rift-dependent records records to protein protein phosphatase I, which phosphatase I, which phosphatase I, which is a set of protein particle of  $\mathcal{P}_n$ may antagonize origination by the cyclin- and  $\mathcal{L}_{\text{max}}$ dent kinases (Lian et al. 2011; Davé et al. 2014; Hiraga et al. 2014; Hiraga et al. 2014; Hiraga et al. 2014;  $\mathbb{M}_{\alpha}$  is an extended  $\mathbb{C}$ ,  $\mathbb$  $t_1$  and  $t_2$  and  $t_3$  and  $\mathbf{R}$  are timing of timing o



that replication times is regulated by  $\mathcal{F}_t$ the number of  $M^{\text{th}}$  local departure at each  $\alpha$  $\mathcal{I}$  .

Results



et al. 2015), as does comparing our MCM Chip-seq data to other mass published estimates of origin timing and  $\alpha$  and  $\alpha$  origin times  $\alpha$ . r 0.0, P  $^{\circ}$   $^$  $\mathbb{E}_{\mathbf{X}}(t)$  al. 2013). From the set of the relationships  $\mathbb{E}_{\mathbf{X}}(t)$ ative number of  $\mathcal{A}$  regions during  $\mathcal{A}$  regions during G1 regions during G1 regions during G1 reg- $U_{\rm{max}} = 1$  , their first  $\mathbb{R} \bullet$ . However, the results of results of results of  $\mathcal{L}_{\mathbf{a}}$  becomes at originative number of  $\mathcal{L}_{\mathbf{a}}$  at origins; they do number of  $\mathcal{L}_{\mathbf{a}}$ not distinguish between  $\mathcal{N}$  in which early origins have multipliers have multipliers have multipliers have multipliers have multipliers  $\mathcal{N}$ tiple MCMs ( $Y_{\rm eff}$  et al. 2010) and the  $Y_{\rm eff}$  in which early origins  $Y_{\rm eff}$  in which early origins  $Y_{\rm eff}$ have a single MCM complex loaded and late origins have substantial orig chiometric MCM loading  $\frac{1}{2}$  (defined by  $\frac{1}{2}$  al. 2010;  $\frac{1}{2}$  and  $\frac{1}{2}$  $0^{22}$   $\frac{1}{2}$   $\frac$ 

 $T = \frac{1}{2} \sqrt{1 + \frac{1}{2} \left( \frac{1}{2} \right)^2 + \frac{1}{2} \left( \frac{1}{2} \right)^2}$  , and  $T = \frac{1}{2} \sqrt{1 + \frac{1}{2} \left( \frac{1}{2} \right)^2 + \frac{1}{2} \left( \frac{1}{2} \right)^2}$  $i,j,k$  test test test the second prediction  $\mathcal{L}$  the second prediction—that early origins  $\mathcal{L}$ have multiple  $\mathcal{N}$  and  $\mathcal{N}$  and  $\mathcal{N}$  as single-original single-or purification strategy. We engineer the three contributions in the three contributions  $T_A$  plasmid affinition system (United States system system (United States system  $200$  and  $25$  and  $25$  and  $25$  and  $25$  single binding single binding single  $25$  $\mathbf{P} \mathbf{N}$  binding protein  $\mathbf{N}$  binds to its 10-bp recognition binds of  $\mathbf{N}$  $\mathbb{S}^N$ site with sub-nanomolar affinity (Elerod-Erickson and Pabo 1999), as a internal control. We have  $\mu$  or  $\alpha$  with the MCM2 and  $\alpha$ epitope and expressed and  $H_1$ -tagged  $\mathbb{R}$  (Supplemental Fig.  $S_{\alpha}$  ( $\lambda$ ), we then purified TALO8 plasmids containing different or- $\mathcal{I}$ igins from G1-arrested cells and determined by Western by Western by Western by Western blotting by Western blotting and determined by Western blotting and determined by Western blotting and determined by Western  $\mathcal{N}$  many  $\mathcal{N}$  many  $\mathcal{N}$  complexes were loaded in viewe to the total in viewe  $Z_1(x) = \int_{\mathbb{R}^2} L(x) \sqrt{2\pi} \int_{\mathbb{R}^2} \left( \frac{L(x)}{2\pi} + \frac{L(x)}{2\$ 

naturally explain the observation that slowers the rate of  $\mathbb{R}^n$  $\mathcal{L}_{\mathbf{z}}(t) = \mathbf{0} - \mathbf{0} - \mathbf{0}$  , which if  $\mathcal{N}_{\mathbf{z}}$  and  $\mathbf{z} = \mathbf{0} - \mathbf{1}$  all  $\mathcal{N}_{\mathbf{z}}$  and  $\mathbf{z} = \mathbf{0}$ 2007; Rhind 2008; Koren et al. 2010). If a rate limiting factor re- $\mathcal{L}_{\mathcal{A}} = \mathcal{I}(\mathcal{W}_{\mathcal{A}}^{\mathcal{A}}, \mathcal{A}_{\mathcal{A}}^{\mathcal{A}})$  with the replication for  $\mathcal{S}_{\mathcal{A}}$  $\mathbb{R}$  and  $\mathbb{R}$  (iii) Nurse 2009; Mantiero et al. 2011; Tanaka and 2011; Tanaka et al. 2011; Gindin et al. 2014),  $e^{i\theta}$  is  $e^{i\theta}$  in  $e^{i\theta}$  would be unable to unabl first previously in  $\mathcal{S}_{\alpha}$  in an extension for  $c \in \mathcal{C}$ ing times to  $f_{\bullet}$  progression ( $\bullet$  00).

Another advantage of the multiple-MCM model is the multiple-MCM model is that it explanets how events that including the impact of  $\mathcal{P}_{\text{max}}$  is defined by the impact of  $\mathcal{P}_{\text{max}}$  $f(x,y,z) = \frac{1}{2} \int_{-\infty}^{\infty} f(x,y) \, dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x,y) \, dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x,y) \, dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x,y) \, dx$  $\alpha$  affinition of  $\alpha$  affects the time during  $\alpha$  affects the time-timental action ac tivation (Wu  $Q_0$ ),  $\mathbb{R}^n$  and  $\mathbb{$  $\mathcal{N}_\text{C}$  is only active during G1, it has been unclear how  $\mathcal{N}_\text{C}$  or  $\mathcal{N}_\text{C}$  . We give  $t_1$  and a few decomposition during S phase after  $S$ activated by CDK activity. The loading of  $m<sub>1</sub>$ early of G1 Origins provides a memory of G1 ORC activity and stable tablishes the replication times the replication times  $T$ 

 $A$  although our results suggest that the number of  $M$ ed at origins affect origins affect original factors contribute as  $\mathcal{F}_{\mathcal{F}}$  $W$ , in particular chromatin structure (Rhind and Gilbert 2013).  $T_{\rm eff}$  correlation we see between  $T_{\rm eff}$   $T_{\rm eff}$  in Fig. signal and replica-sequences  $T_{\rm eff}$  $t_1 = 1/\sqrt{0}$ ,  $\int_{0}^{1} L^{2}$ ,  $\int_{0}^{1}$  (see Eq. i. s.  $\int_{0}^{1}$   $\int_{0}^{N}$  s = Eq. i.  $\int_{0}^{N}$  s  $M_{\rm H}$  number  $\frac{1}{2}$  significant but not exclusive role in regular role lating origin timing. Furthermore, we see a similar level of correlation between other published MCM  $\mathcal{L}$  and  $\mathcal{L}$ -sequence  $\mathcal{L}$  $t = \frac{1}{\sqrt{2}} \int_{\mathbb{R}^2} [1 + \sqrt{2} \cos \theta] \sin \theta \, d\theta$   $\mathbb{H}^{k}$  institution et al. 2013;  $\mathbb{H}^{k}$  et al. 2013; Belsky et al. 2013;  $g_{\alpha}(t)$  the result is result is robust to experimental details. The lack of corresponding to  $\alpha$ relation seen with order  $\mathcal{E}$  and  $\mathcal{E}$  all  $\mathcal{E}$  sets (Wyrick et al.)  $2001$ ,  $2006$  is presented to the lower dynamic range to the lower dynamic range  $\frac{1}{2}$  $\overline{\phantom{a}}$ of those data sets.

If  $\sum_{i=1}^{N} x_i \overline{y_i} \overline{y_i} = \overline{y_i} \overline{y_i}$  by a reported to be affected by a finite b the Rpd3 historic deach deach control  $\mathcal{L}^{\bullet}$  historic binding protein, the KU telomere binding protein, the  $\mathcal{L}^{\bullet}$  $t_{\rm max}$  the Federal transcription factor, or the  $C_{\rm max}$  $\mathcal{F}_{\bullet}^{\mathbf{V}}$  to be correlation in proves to between  $\mathcal{F}_{\bullet}^{\mathbf{V}}$  or  $\mathbf{0}$ .  $T_{\rm eff}$  , and the these factors, all of which affect chromating  $\alpha$  of which affect chromating  $\alpha$ structure, modify original first origin first origin first origin for  $\mathcal{L}$ also suggest that other as yet unidentified factors as yet unidentified factors and  $\alpha$ portant role in role in regulation times originate times or  $R$  . Cheromating originates the chromatin structure  $R$ could affect origin effect origin effect origin licensing in the origin licensing in the origin licensing origin licensing order or  $\alpha$ and  $\mathbb{R}^N$  and  $\mathbb{R}^N$  binding,  $\mathbb{R}^N$  binding,  $\mathbb{R}^N$  activities. McM activation.  $\mathcal{H}_{\text{max}}$  the factor of  $\mathcal{H}_{\text{max}}$  large numbers of  $\mathcal{H}_{\text{max}}$  loaded numbers of  $\mathcal{H}_{\text{max}}$ at late-replication telometers (Fig. 1A) suggests that heterochroma-replication telometers that heterochromatin can delay the first of load MCMs, perhaps by counteraction of loaded  $M$  $i$  and  $j$  of  $i$  of  $r$  and  $i$  activity activity of  $\left(\begin{array}{ccc} 1 & a & c & 0 \\ c & c & a & c \end{array}\right)$ ;  $\ldots$   $\ldots$   $\ldots$   $0^{\bullet}$ ).

 $T_{\text{total}}$  , and  $T_{\text{total}}$  are multi-multi ple McMs may be located after located after located after located after location  $\mathcal{P}_n$  $t_1$  locally loaded in the nucleosome-free region next to  $\mathcal{L}$ and  $(\mathbf{0}^*)$ , for multiple  $\mathbf{0}$ ,  $\mathbf{0}^*$  to be localed,  $\mathbf{0}^*$  would would be localed,  $\mathbf{0}^*$  would be localed to be localed,  $\mathbf{0}^*$  would be a set of  $\mathbf{0}^*$ . need to  $t^{\mathcal{W}_\ell}$  from this loading site into surrounding site into surrounding characteristic into surrounding  $\mathcal{W}_\ell$ matrix. Such diffusion has been observed on characteristic templates in frog equation (Edwards et al. 2002). Function  $\ell$ recent high-resolution  $\mathcal{L}$  at budding  $\mathcal{L}$  mapping of  $\mathcal{L}$  at budding  $\mathcal{L}$  at budding years or  $\mathcal{L}$  $\overline{I}$  in  $\overline{I}$  if  $\overline{I}$  is the  $\overline{I}$  grad preference with  $\overline{N}_{\alpha}$  in  $\overline{I}$  is the sum of  $\overline{I}$ flanking nucleosomes (Belsky et al. 2015). Although  $L_1$ the strongest MCM  $\frac{1}{2}$  signal tends to be either at the  $\frac{1}{2}$  or  $\frac{$  $n_{\text{max}}$ nucleosome, individual original actions show signal actions show signal actions  $\mathbb{R}$ ing nucleosomes, allowing the possibility that multiple  $\mathcal{N}$  the possibility that multiple  $\mathcal{N}$ could associate with multiple nucleosomes in a heterogeneous in a het manner.

 $T_{\rm tot}$  , as to as to why more  $\mathcal{P}_{\rm tot}^{\rm W}$  as to  $\mathcal{P}_{\rm tot}$  as to why more  $\mathcal{P}_{\rm tot}$ loaded at one origin than another. A simple hypother. A simple hypothesis is that early original have higher-after-affinity ORC binding sites, so that  $\mathcal{L}(\mathbf{r},\mathbf{r})$ spends more of G1 phase bound at the those sites and can load more sites and  $\mathbb{E}[\mathbf{M}^{\mathbf{p}}(f_{\mathbf{q}},\mathbf{r}_{\mathbf{q}}), \mathbf{M}_{\mathbf{p}}]$  is supported by a modest, but significant si icant, correlation between  $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$   $\mathcal{L}$  $(n-0.45)$  (Supplemental Fig. S<sub>1</sub>C; Eaton et al. 2010). However,  $t_{\rm eff}$  affinity of  $\sim$   $\sim$  0.000 for  $\sim$  0.000 for  $\sim$  0.000 for  $\sim$  0.000 for the more than  $\sim$ simply the local original origin sequence, because for a significant number of a significant number of a significant number of  $\mathcal{L}$ of origins, in violence is affected by local chromatin structure is affected by local chromatin structure.  $t_{\rm orb}$  , as  $t_{\rm orb}$  as well as the direct affinity of  $\Delta t_{\rm orb}$  the origin sequence  $(T_{H}-1.20)$ .  $T_{H}$   $T_{H}$ ing regulators of originating effects of original structure, may be structured and chromatin structure, may be also affect the number of  $\overline{M}$  localizations.  $R_{\rm eff} = 1000$  recent recent suggests and may be conceptually similar mechanism may be conceptually similar m

regulate origin timing in  $\mathbb{R}^n$  in the cells (Gindin et al. 2014; Rhindin et al. 2014; Rhi  $[0^{\bullet}, 0^{\bullet}]$ . In human cells, the density of DNAse I hypersensitive sites in the density of  $\mathbb{R}$ is an excellent predictor of replication times, with regions dense  $\mathcal{L}$ in DNAse I hypersensitive sites replication of  $\mathcal{L}_A$  ,  $\mathcal{L}_B$  ,  $\mathcal{L}_B$  ,  $\mathcal{L}_C$  ,  $\mathcal{L}_C$  ,  $\mathcal{L}_C$  ,  $\mathcal{L}_C$  $(0^\bullet, 0^\bullet)$ . Moreover, a model that uses  $M_{\rm{max}}$  is defined that uses  $M_{\rm{max}}$ as a proxy for licensed origins, which are first are fired by a hypothetical rate limiting activation for  $\mathbb{R}^n$  and  $\mathbb{R}^n$  activities  $\mathbb{R}^n$  and  $\mathbb{R}^n$ predicts developmentally regulated replication times of  $\overline{P}$ ly, the model does not require that early first have a higher probability of the higher density of  $f(x)$  $\overline{a}$  of origins in early firing regions is sufficient to increase the chance the chance the chance the chance the chance three chances in the chance of the chance three chances in the chance of the chance of the chanc  $\mathcal{L}$  such regions replicating early. Thus, in this model, the effect of effect of effect of effect of effect of effect of  $\mathcal{L}$ 







## **origins Replication timing is regulated by the number of MCMs loaded at**

- $S$  P.D , T B , T.L, .
- $\widetilde{A}$  . The most recent vecent version at  $10.1101/$  . 195305.115 *Genome Res.* 2015 25: 1886-1892 originally published online September 10, 2015  $\boldsymbol{G}$





## **<http://genome.cshlp.org/subscriptions>** To subscribe to *Genome Research* go to:

2015 D.; P. C. S. H. L. P.