nature structural & molecular biology

Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks

Celina Costas[1,7], Maria de la Paz Sanchez[1,6,7], Hume Stroud[2,7], Yanchun Yu[3], Juan Carlos Oliveros[4], Suhua Feng[5], Alberto Benguria[4], Irene López-Vidriero[4], Xiaoyu Zhang[3], Roberto Solano[4], Steven E Jacobsen[2,5] & Crisanto Gutierrez[1]

Genome integrity requires faithful chromosome duplication. Origins of replication, the genomic sites at which DNA replication initiates, are scattered throughout the genome. Their mapping at a genomic scale in multicellular organisms has been challenging. In this study we profiled origins in *Arabidopsis thaliana* by high-throughput sequencing of newly synthesized DNA and identified ~1,500 putative origins genome-wide. This was supported by chromatin immunoprecipitation and microarray (ChIP-chip) experiments to identify ORC1- and CDC6-binding sites. We validated origin activity independently by measuring the abundance of nascent DNA strands. The midpoints of most *A. thaliana* origin regions are preferentially located within the 5′ half of genes, enriched in G+C, histone H2A.Z, H3K4me2, H3K4me3 and H4K5ac, and depleted in H3K4me1 and H3K9me2. Our data help clarify the epigenetic specification of DNA replication origins in *A. thaliana* and have implications for other eukaryotes.

Faithful duplication of the genetic material is crucial in maintaining genomic integrity. DNA replication in eukaryotic cells initiates at multiple sites, known as replication origins, which are scattered throughout the genome[1–3]. The number of origins ranges from hundreds to thousands depending on cell type and/or physiological state[3]. One of the key steps in understanding the function of replication origins is determining whether and how they are specified in the genome. In *Saccharomyces cerevisiae*, a strict sequence-dependent specification occurs whereby the origin recognition complex (ORC) recognizes an 11-base-pair (bp) sequence to define the site of each active replication origin[4,5]. This mechanism seems to be unique because a consensus sequence has not been found in other organisms. For example, in *Schizosaccharomyces pombe*, although origins are associated with (A+T)-rich stretches, they are not specified by a known consensus DNA sequence[6,7].

The identification of the molecular nature of replication origins in multicellular organisms has been challenging, and only a handful of origins have been analyzed[2,8–10]. The large genome size of multicellular eukaryotes, their different developmental strategies and the diversity of proliferating cell populations have made it difficult to determine origin specification, function and spatiotemporal regulation at a genomic scale[3]. Local epigenetic modifications can further affect origin selection and usage, for example, replication timing[3,11–13]. Although attempts to obtain genome-wide maps of replication origins in mammalian cells have been reported[14–17], the molecular features defining replication origins

in higher eukaryotes and, in particular, their links to epigenetic modifications still remain largely unknown.

In this study, we identified replication origins, analyzed their organization and defined their epigenetic signatures at a high-resolution genome-wide scale in the plant *A. thaliana*. Its compact genome (~125 Mb, ~28,000 protein coding genes), which is fully sequenced and annotated, and its small percentage of repetitive sequences (~17%), which are largely confined to the pericentromeric areas[18], make *A. thaliana* a useful system for studying origins. Furthermore, the comparison of replication origin features in organisms with very different developmental and growth strategies could clarify the basic principles governing origin specification and function in eukaryotes. In addition, genome-wide maps of epigenetic marks such as DNA methylation and several histone modifications have already been reported[19–21]. Through massive sequencing of short-pulse 5-bromo-2′-deoxyuridine (BrdU)-labeled DNA, we identified ~1,500 putative replication origins across the *A. thaliana* genome. We also identified ORC1- and CDC6-binding regions using ChIP-chip; notably, these are enriched in BrdU-labeled regions. Furthermore, we validated origin activity independently by measuring nascent DNA strand abundance. Our studies reinforce the idea that some origin features are shared with animal cells whereas others are unique to plants[22,23]. The *A. thaliana* 'originome' reported here provides a basis for identifying the key features of eukaryotic replication origins and delineating their possible regulatory mechanisms.

[1]Centro de Biologia Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas (CSIC)-Universidad autónoma de Madrid (UAM), Madrid, Spain. [2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California, USA. [3]Department of Plant Biology, University of Georgia, Athens, Georgia, USA. [4]Centro Nacional de Biotecnología, CSIC, Madrid, Spain. [5]Howard Hughes Medical Institute, University of California, Los Angeles, California, USA. [6]Present address: Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico DF, Mexico. [7]These authors contributed equally to this work. Correspondence should be addressed to C.G. (cgutierrez@cbm.uam.es) or S.E.J. (jacobsen@ucla.edu).
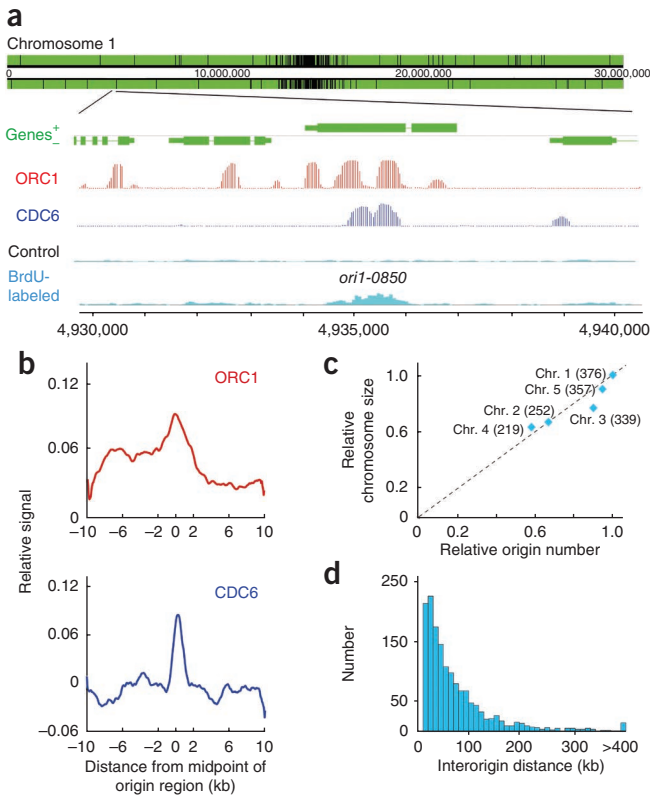
Figure 1 Identification of DNA replication origins in the *A. thaliana* genome. (**a**) Representative genome-browser view of a region in chromosome 1. Genes (green) transcribed from each strand are along the chromosome above and below the position scale. Bottom, an enlarged region containing a replication origin, determined as a region enriched for BrdU-labeled DNA strands (light blue) relative to the unlabeled control DNA (black), together with the ORC1 (red) and CDC6 (dark blue) binding patterns (posterior probabilities for ORC1 and CDC6 data sets). Origins, for example, *ori1-0850*, are named based on their chromosomal location (*ori1* through *ori5*), followed by the four digits indicating the origin number within each chromosome. They are named consecutively starting at the left tip of each chromosome; that is, for chromosome 1, in which we identified 376 origins, the leftmost origin is *ori1-0010* and the rightmost one is *ori1-3760*. (**b**) The pattern of ORC1 and CDC6 binding over origin regions was obtained by plotting their relative binding signal ±10 kb from the origin region midpoint (0) using 50-bp sliding windows (smoothed). The *P* values (two-sided) of the difference in the ChIP-chip signals in origins (midpoint ±300 bp), using a two-tailed Welch test, were 7.25×10^{-6} and 1.29×10^{-10} for ORC1 and CDC6, respectively. (**c**) Number of origins relative to chromosomal size. Chromosome size (relative to chromosome (chr.) 1) versus the number of origins identified in each chromosome (relative to origin number in chromosome 1). Number of origins identified in each chromosome is in parentheses. (**d**) Distribution of interorigin distances, measured as distance between the midpoints of two contiguous origins (median, 51.1 kb; average, 77.2 kb; s.d., 83.4 kb).

RESULTS

### Genome-wide mapping of *A. thaliana* DNA replication origins

Functional origins mark the sites where the synthesis of nascent DNA strands occurs. Thus, our strategy was to sequence purified DNA labeled *in vivo* with a pulse of BrdU and confirm these data with the mapping of pre-replication complex (pre-RC) binding (**Supplementary Fig. 1**). To obtain sufficient amounts of BrdU-labeled DNA, we used *A. thaliana* cultures containing a substantial number of proliferating cells. We synchronized cells in the G0 stage of the cell cycle using sucrose deprivation and labeled them with BrdU a few hours after release from the block when cells were just entering the S-phase[24,25] (**Supplementary Figs. 1** and **2**). DNA was extracted and fractionated by CsCl gradient centrifugation, and the BrdU-labeled material was purified and used to generate genomic libraries for sequencing using Solexa (Illumina) technology. We obtained ~4 million high-quality reads that uniquely mapped to the *A. thaliana* genome. Likewise, we processed a sample of unlabeled DNA as a control (see Online Methods). This BrdU-seq method rendered a comprehensive list of genomic locations with a significant enrichment in BrdU-labeled DNA strands (**Fig. 1a**). To define origin regions using the BrdU-labeled DNA sequencing data, we merged BrdU-positive regions separated by <10 kilobases (kb), as described in Online Methods (see also **Supplementary Fig. 3**). An alignment of DNA sequences of ±100 bp around the midpoint of BrdU-labeled regions did not render any consensus sequence. To corroborate the analysis of BrdU-labeled regions and deal with possible experimental variations, we carried out an independent assay of cell synchronization, BrdU-labeling and CsCl purification followed by massively parallel sequencing. Significantly, 78.2% ($P < 1.0 \times 10^{-6}$) of the BrdU-labeled regions overlapped with the regions defined in the previous experiment, supporting the reproducibility of the two independent experiments.

To identify pre-RC binding sites, in the absence of specific antibodies, we used plants expressing constitutively tagged versions of two pre-RC components, ORC1 (ref. 26) and CDC6 (ref. 27). We purified ORC1- and CDC6-bound DNA fragments by ChIP (**Supplementary Fig. 1**) and hybridized them to whole-genome *A. thaliana* tiling arrays to identify their genome-wide binding sites (**Fig. 1a**). ORC1 binding was spread over many sites (**Supplementary Fig. 4**), whereas CDC6-binding sites were less abundant (**Supplementary Fig. 5**). First, we determined the fraction of the BrdU-labeled regions that contained bound pre-RC components. We found that ~76.7% and 17.0% of BrdU-labeled regions overlapped with ORC1 and CDC6 regions, respectively (midpoint of BrdU region, ±2.1 kb, $P < 0.001$; see colocalization range in **Fig. 1b**). Notably, the midpoints of these regions significantly colocalized with both ORC1- and CDC6-binding sites within ±2 kb regions (**Fig. 1b**). Therefore, we considered the 1,543 regions rendered by our approach to be bona fide replication origins (**Supplementary Table 1**). They seem uniformly distributed across the genome, although it is possible to identify clusters of more closely spaced origins in some genomic locations (**Supplementary Fig. 6**). The number of origins varies for different chromosomes but is roughly correlated with chromosome size (**Fig. 1c**). The distribution of distances between origin region midpoints has a median of 51.1 kb, and a mean of 77.2 kb (**Fig. 1d**).

Assessing origin activity by nascent strand abundance

The BrdU-labeled regions identified in our study and the marked colocalization with ORC1- and CDC6-binding sites support the notion that they represent active DNA replication origins. To assess origin activity directly, we measured the relative abundance of nascent DNA strands of various putative origin regions relative to adjacent regions in a sample of short DNA molecules purified by sucrose-gradient centrifugation and containing a RNA primer at their 5′ ends[28,29]. Thus, we determined origin activity by real-time PCR methods using primer pairs spanning 5–16 kb around putative origin regions. In all cases we analyzed, origin sequences were highly enriched in the short nascent DNA strand sample (**Fig. 2a–c**). Notably, one of the BrdU-labeled regions included in this analysis showed a relatively low CDC6 signal in the ChIP-chip experiment (**Fig. 2a**). Despite this, it showed a
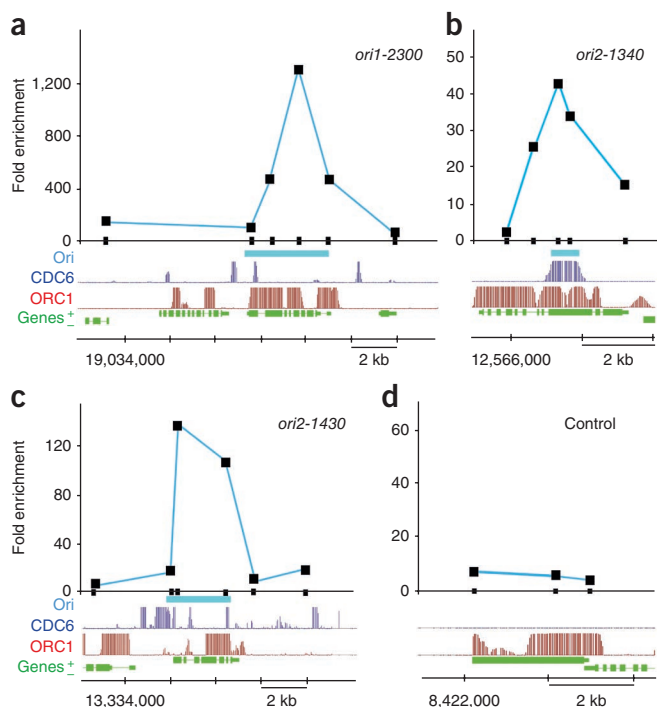
Figure 2 DNA replication origin activity determined by nascent DNA strand abundance. (**a–d**) Several putative origin-containing regions were chosen for detailed measurement by real-time PCR of nascent strand abundance in a sample of short DNA molecules containing an RNA primer at their 5′ end (see Online Methods). Genomic region under study is at bottom of each panel and shows the location of genes (green), ORC1-binding (red) and CDC6-binding (dark blue) signals, and putative origin location (light blue), defined by direct sequencing of the BrdU-labeled DNA sample (see Online Methods). DNA fragments (~200 bp long) amplified by primer pairs scanning each region are small black rectangles on the *x* axis. Coordinates in each chromosome are at the bottom of each panel. Results correspond to PCR amplifications using fraction 5 (see Online Methods). Data for origins *ori1-2300* (**a**), *ori2-1340* (**b**) and *ori2-1430* (**c**). Data for a region used as a negative control around gene *at4g14700* that lacks BrdU-labeled DNA sequences (**d**).

(Wilcoxon rank-sum test, *P* < 0.005) or downstream (Wilcoxon rank-sum test, *P* < 0.01) of genes (**Fig. 3b**).

The body of highly expressed genes in *A. thaliana* is enriched in CG methylation, whereas the three types of C methylation (CG, CHG and CHH, where H is A, T or C) are highly enriched in the repeat-rich pericentromeric regions of the *A. thaliana* genome[19,31]. Notably, we found a slight decrease in CG methylation around origin midpoints as compared with regions flanking them (**Fig. 4a**). Furthermore, we observed that regions ±0.1 kb from the origin midpoints showed higher G+C contents (44.5%), as compared with the whole *A. thaliana* genome (**Fig. 4b**). The histone variant H2A.Z is preferentially deposited near the 5′ end of target genes and is inversely correlated with CG methylation[32]. We found a strong presence of H2A.Z within ±1 kb of the origin midpoints (**Fig. 4c**).

Epigenomic landscape of *A. thaliana* DNA replication origins

To further determine features defining *A. thaliana* replication origins, we next sought to profile the landscape of epigenetic histone marks that seem to associate with replication origins. *A. thaliana*

high abundance of nascent DNA strands measured by quantitative PCR (qPCR), demonstrating the activity of this region as a functional origin as well as the robustness of our approach. A control region, lacking BrdU-labeled DNA sequences, did not show any appreciable enrichment (**Fig. 2d**). These data together led us to conclude that the set of origins identified here provides a starting point for defining their molecular landscape.

**Genomic location of *A. thaliana* DNA replication origins**

To test whether origins are randomly distributed along the genome or show a preferential location, we estimated origin location relative to various genomic elements. We found that 77.7% and 10.2% of origins colocalized with gene units and transposons, respectively. These percentages are significantly different from the proportion of the *A. thaliana* genome represented by these elements (**Fig. 3a**). Next, we analyzed origin density across genes and their 5′ and 3′ upstream regions. We observed that most origins were identified within gene bodies (**Fig. 3b**), but preferentially toward their 5′ ends (**Supplementary Fig. 7**). Origin localization to the bodies of genes did not correlate with gene expression levels (**Fig. 3b**), according to expression data obtained from cell suspensions at the same synchronization time used for BrdU labeling[30]. However, highly expressed genes, compared with weakly expressed genes, tended to have more origins in regions immediately upstream
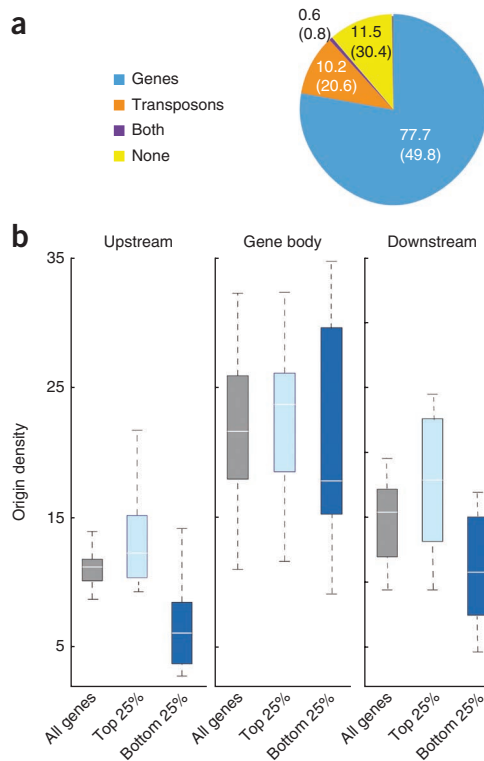
Figure 3 Genomic location of *A. thaliana* replication origins. (**a**) Percentage of origins colocalizing with various genomic elements. Numbers in parentheses, proportion of *A. thaliana* genome represented by each class. (**b**) Origin densities were computed for regions upstream, downstream and within genes of different expression levels (all genes, highest 25%, lowest 25%). Regions 2 kb upstream and downstream of genes, as well as the bodies of genes, were each divided into ten bins, and the origin densities (origins per 10^6 bp) were calculated for each bin and represented as box plots. White lines, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; error bars, minimum and maximum points that fell within 1.5×IQR (interquartile range) below the 25th percentile or above the 75th percentile.

Figure 4 Relationship of *A. thaliana* replication origins to CG methylation and histone H2A.Z. (**a**) Relative levels of CG, CHG and CHH methylation ±10 kb relative to the origin midpoint (0) in 50-bp sliding windows (smoothed). Methylation data have been reported[19]. (**b**) G+C content (%) of replication origins (black) and indicated genomic regions (white). These values were calculated from the sequence data files available at The Arabidopsis Information Resource (TAIR), http://www.arabidopsis.org/. (**c**) Density of the histone variant H2A.Z in a ±10 kb region relative to the origin midpoint (0) in 50-bp sliding windows (smoothed). The genomic distribution of H2A.Z has been reported[32]. The *P* value of the difference in the ChIP-chip signals in origins (calculated as in **Fig. 1b**; see Online Methods), was 9.34×10^{-34}.
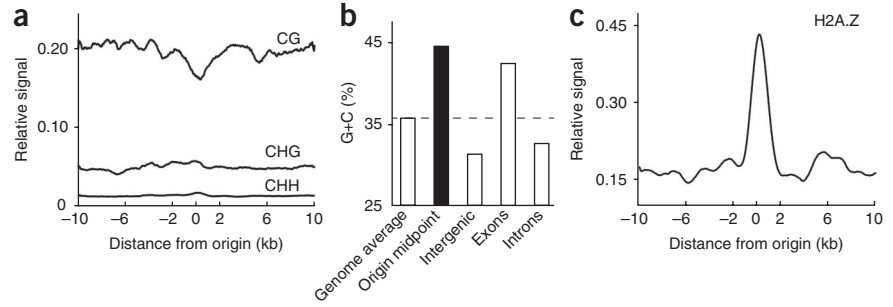
epigenomics data are already available for dimethylation of histone H3 at Lys9 (H3K9me2) and for the three methylated forms of H3 Lys4 (H3K4; refs. 20,21). We found that most origins tend to be depleted in singly methylated H3K4 (H3K4me1; **Fig. 5a**) but are highly enriched in dimethylated and trimethylated H 3K4 (H3K4me2 and H3K4me3; **Fig. 5b,c**). We observed that H3K4me3 and/or H3K4me2, with or without H3K4me1, seems to be a signature of ~80% of origins associated with genes (**Fig. 5**). This is consistent with the preferential localization of origins in 5′ gene body regions observed here and the negative correlation of these marks and CG methylation[21]. Furthermore, H3K9me2 is highly depleted in most of the origins identified in our study (**Fig. 5d**).

Histone hyperacetylation and origin activation are correlated in *Xenopus laevis*[11] and *Drosophila melanogaster* cells[33–35]. Consistent with this, immunofluorescence data obtained in several plant species indicate that increases in histone acetylation occurs during S-phase[22,23]. Recently,

ChIP experiments have shown that H4 Lys 5 (H4K5) and H4K12 (and to a lesser extent H4K8), but not H4K16, need to be acetylated by the HBO1 histone acetylase at origins in human cells to overcome geminin inhibition and facilitate minichromosome maintenance (MCM) loading[36]. Thus, we profiled H4K5ac over the genome by ChIP-chip and found an enrichment of this mark at the origin midpoint (**Fig. 5f**).

DISCUSSION

Initiation of DNA replication in eukaryotes depends on the assembly of pre-RCs in G1 of the cell cycle at certain chromosomal locations and their further activation to initiate DNA replication in S-phase. Both steps must be tightly coordinated to ensure that the genome is duplicated once per cell cycle[2]. We have found that ORC1-binding sites tend to form clusters, similar to the situation in *D. melanogaster* cells[37] but very different from that in *S. cerevisiae*[5]. The presence of ORC1-binding sites across the genome may not only represent broad initiation zones with several potential initiation sites but also reflect the function of ORC1 in other processes, for example, heterochromatin silencing[38], transcriptional control[26,39] or chromatid cohesion[40]. In any case, detection of CDC6 in BrdU regions takes into account the release of CDC6 from the pre-RC once an origin is fired[41].

The distribution of distances between origin region midpoints fell within the range estimated for other eukaryotes[42] and roughly matched estimates of replicon size in *A. thaliana*[13,43]. A fraction of the putative origin regions identified here could correspond to elongating forks rather than to initiation events. However, our direct measurements of origin activity by abundance of RNA primer–containing nascent strands support the idea that the originome reported here is a bona fide list of putative *A. thaliana* DNA replication origins. Future analysis should address this point. The abundance of origin
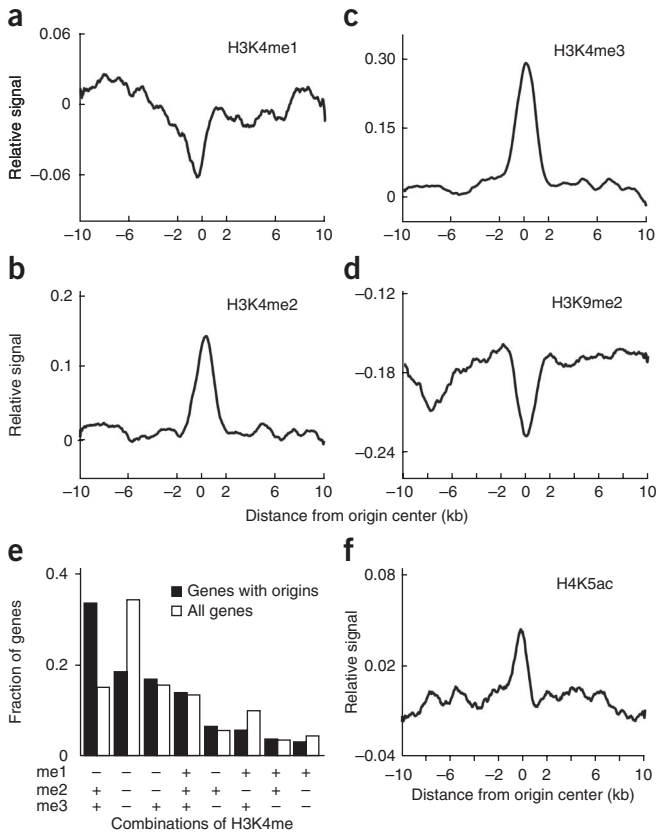
Figure 5 Histone modification landscape around replication origins. (**a–d**) Relative level of indicated histone mark ±10 kb relative to the center of origins (0) in 50-bp sliding windows (smoothed). Data for H3K4me and H3K9me2 have been reported[20,21]. The *P* values of the difference in the ChIP-chip signals in origins (calculated as in **Fig. 1b**; see Online Methods) were 0.86, 3.52×10^{-28}, 1.07×10^{-41} and 7.33×10^{-14} for H3K4me1, H3K4me2, H3K4me3 and H3K9me2, respectively. (**e**) Relationship between H3K4 methylation status and the presence of origins. We calculated the fraction of genes containing origins and different combinations of H3K4 methylation, and compared it with the fraction of all genes containing the same H3K4me combinations[21]. Different classes are ordered with decreasing values of the fraction of genes with origins. (**f**) Relative level of H4K5ac ±10 kb relative to the center of origins (0) in 50-bp sliding windows (smoothed). Calculations are based on the ChIP-chip data set generated in this work. The *P* value of the difference in the ChIP-chip signals in origins (calculated as in **1b**; see Online Methods) was 1.23×10^{-23}.

sequences and the width of the peak of amplified fragments varied for different origins we analyzed, suggesting differences in the efficiency of origin usage or in the usage of initiation sequences within an origin region[10,42].

Notably, the location of most *A. thaliana* origins is different from that in other systems, in which a large proportion of highly efficient origins are associated with gene promoters or transcriptional start sites[16,17,37]. We have found that the ±0.1-kb region around *A. thaliana* DNA replication origins has a higher-than-average G+C content and a slight decrease in CG methylation. Consistent with this observation, early and mid replicons in *A. thaliana* chromosome 4 are also depleted in CG methylation[13]. One possibility is that in *A. thaliana* the relatively high G+C content at origins favors a particular nucleosome organization in these regions. This is reinforced by the colocalization of origins with histone H2A.Z, which affects nucleosome stability[44], and could facilitate pre-RC assembly and/or origin firing. Together, our data show that whereas CG methylation within gene bodies is relevant for gene expression in *A. thaliana*[19], it does not seem to be a requirement for origins. Metazoan origins are highly correlated with unmethylated CpG islands located at the promoters of active genes or in proximity to transcriptional start sites[6,42]. Although CpG islands are not present in the *A. thaliana* genome, our results show a conserved trend of relatively lower CG methylation at origins, and a high correlation among origin activity, a local high G+C content and the presence of H2A.Z.

Post-translational histone modifications can also affect origin specification and function. Most *A. thaliana* origins tend to be enriched in H3K4me2 and H3K4me3, as well as in H4K5ac, similarly to human origins[17,36]. Whether all human origins have the same H4ac pattern, as a consequence of HBO1 activity to overcome geminin inhibition[36], and whether all *A. thaliana* origins require H4ac for activation, remain open questions for the future. However, the H4 acetylation pattern is relevant owing to the presence in *A. thaliana* of (i) an HBO1-related acetyltransferase[45], (ii) increased tetraH4ac residues around ORC1-binding sites[26] and (iii) a CDT1-interacting protein, GEM, structurally unrelated to metazoan geminin[46,47]. Acetylation of other histone residues may be also relevant for origin function, as has been suggested by the presence of H3K56ac in early replicons of chromosome 4 (ref. 13).

How replication origins are specified in large eukaryotic genomes has been a long-standing question. Early-firing origins are associated with transcribed genomic regions[48,49]. The origins that have been studied in mammalian cells, covering only 0.4–1% of their genomes, show a preferential association with active promoters that contain CpG islands[15–17]. We have found that origins located in the upstream regions of genes are preferentially associated with highly expressed genes. However, the differences in the genomic distribution of CG methylation pattern in *A. thaliana* may contribute to the use of different mechanisms to specify origins. In fact, a higher proportion of origins are located in the 5′ half of gene bodies in *A. thaliana* than in mammalian cells.

Our work has defined a landscape of epigenetic marks associated with a genome-wide set of replication origins in *A. thaliana*. The midpoints of most origin regions preferentially colocalize with a higher-than-average G+C content, but lower CG methylation, and are enriched in histone H2A.Z, H3K4me2, H3K4me3 and acetylated H4K5, and depleted in H3K4me1 and H3K9me2. Elucidating how epigenetic mechanisms and gene expression coordinate with DNA replication is important for understanding these processes in a genomic and developmental context. The *A. thaliana* originome reported here provides a foundation for future studies of the mechanisms of origin specification as well as the regulation and function of DNA replication origins in different eukaryotes.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. NCBI GEO: ORC1 and CDC6 ChIP-chip, GSE21928; BrdU-seq and H4K5ac ChIP-chip, GSE21828.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

AUTHOR CONTRIBUTIONS
C.C., M.P.S., Y.Y., S.F., A.B. and I.L.-V. carried out experiments. H.S., J.C.O., C.C., M.P.S., X.Z. and R.S. analyzed data. C.G. and S.E.J. prepared the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Huberman, J.A. & Riggs, A.D. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* **32**, 327–341 (1968).
2. DePamphilis, M.L. *et al.* Regulating the licensing of DNA replication origins in metazoa. *Curr. Opin. Cell Biol.* **18**, 231–239 (2006).
3. Aladjem, M.I. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.* **8**, 588–600 (2007).
4. Bell, S.P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128–134 (1992).
5. 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).
6. Antequera, F. Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J.* **23**, 4365–4370 (2004).
7. Hayashi, M. *et al.* Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J.* **26**, 1327–1339 (2007).
8. Gilbert, D.M. In search of the holy replicator. *Nat. Rev. Mol. Cell Biol.* **5**, 848–855 (2004).
9. Costa, S. & Blow, J.J. The elusive determinants of replication origins. *EMBO Rep.* **8**, 332–334 (2007).
10. Schepers, A. & Papior, P. Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Res.* **18**, 63–77 (2010).
11. Danis, E. *et al.* Specification of a DNA replication origin by a transcription complex. *Nat. Cell Biol.* **6**, 721–730 (2004).
12. MacAlpine, D.M., Rodriguez, H.K. & Bell, S.P. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* **18**, 3094–3105 (2004).
13. Lee, T.J. *et al. Arabidopsis thaliana* chromosome 4 replicates in two phases that correlate with chromatin state. *PLoS Genet.* **6**, e1000982 (2010).
14. Lucas, I. *et al.* High-throughput mapping of origins of replication in human cells. *EMBO Rep.* **8**, 770–777 (2007).
15. Cadoret, J.C. *et al.* Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl. Acad. Sci. USA* **105**, 15837–15842 (2008).
16. Sequeira-Mendes, J. *et al.* Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, e1000446 (2009).
17. Karnani, N., Taylor, C.M., Malhotra, A. & Dutta, A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell* **21**, 393–404 (2010).

18. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
19. Cokus, S.J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
20. Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M. & Jacobsen, S.E. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* **3**, e3156 (2008).
21. Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M. & Jacobsen, S.E. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, R62 (2009).
22. Fuchs, J., Demidov, D., Houben, A. & Schubert, I. Chromosomal histone modification patterns—from conservation to diversity. *Trends Plant Sci.* **11**, 199–208 (2006).
23. Sanchez, M.L., Caro, E., Desvoyes, B., Ramirez-Parra, E. & Gutierrez, C. Chromatin dynamics during the plant cell cycle. *Semin. Cell Dev. Biol.* **19**, 537–546 (2008).
24. Menges, M. & Murray, J.A. Synchronous *Arabidopsis* suspension cultures for analysis of cell-cycle gene activity. *Plant J.* **30**, 203–212 (2002).
25. Menges, M. & Murray, J.A. Synchronization, transformation, and cryopreservation of suspension-cultured cells. *Methods Mol. Biol.* **323**, 45–61 (2006).
26. de la Paz Sanchez, M.P. & Gutierrez, C. *Arabidopsis* ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. *Proc. Natl. Acad. Sci. USA* **106**, 2065–2070 (2009).
27. Castellano, M.M., del Pozo, J.C., Ramirez-Parra, E., Brown, S. & Gutierrez, C. Expression and stability of *Arabidopsis* CDC6 are associated with endoreplication. *Plant Cell* **13**, 2671–2686 (2001).
28. Prioleau, M.N., Gendron, M.C. & Hyrien, O. Replication of the chicken beta-globin locus: early-firing origins at the 5′ HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. *Mol. Cell. Biol.* **23**, 3536–3549 (2003).
29. Gómez, M. & Antequera, F. Overreplication of short DNA regions during S phase in human cells. *Genes Dev.* **22**, 375–385 (2008).
30. Menges, M., Hennig, L., Gruissem, W. & Murray, J.A. Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant Mol. Biol.* **53**, 423–442 (2003).
31. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
32. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).
33. Aggarwal, B.D. & Calvi, B.R. Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature* **430**, 372–376 (2004).
34. Hartl, T., Boswell, C., Orr-Weaver, T.L. & Bosco, G. Developmentally regulated histone modifications in *Drosophila* follicle cells: initiation of gene amplification is associated with histone H3 and H4 hyperacetylation and H1 phosphorylation. *Chromosoma* **116**, 197–214 (2007).
35. Schwaiger, M. *et al.* Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes Dev.* **23**, 589–601 (2009).
36. Miotto, B. & Struhl, K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by geminin. *Mol. Cell* **37**, 57–66 (2010).
37. MacAlpine, H.K., Gordan, R., Powell, S.K., Hartemink, A.J. & Macalpine, D.M. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* **20**, 201–211 (2010).
38. Pak, D.T. *et al.* Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* **91**, 311–323 (1997).
39. Rusche, L.N., Kirchmaier, A.L. & Rine, J. The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* **72**, 481–516 (2003).
40. Takahashi, T.S., Yiu, P., Chou, M.F., Gygi, S. & Walter, J.C. Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nat. Cell Biol.* **6**, 991–996 (2004).
41. Kim, J. & Kipreos, E.T. Control of the Cdc6 replication licensing factor in metazoa: the role of nuclear export and the CUL4 ubiquitin ligase. *Cell Cycle* **7**, 146–150 (2008).
42. Cadoret, J.C. & Prioleau, M.N. Genome-wide approaches to determining origin distribution. *Chromosome Res.* **18**, 79–89 (2010).
43. Van't Hof, J., Kuniyuki, A. & Bjerknes, C.A. The size and number of replicon families of chromosomal DNA of *Arabidopsis thaliana*. *Chromosoma* **68**, 269–285 (1978).
44. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41**, 941–945 (2009).
45. Earley, K.W., Shook, M.S., Brower-Toland, B., Hicks, L. & Pikaard, C.S. *In vitro* specificities of *Arabidopsis* co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. *Plant J.* **52**, 615–626 (2007).
46. Caro, E., Castellano, M.M. & Gutierrez, C. A chromatin link that couples cell division to root epidermis patterning in *Arabidopsis*. *Nature* **447**, 213–217 (2007).
47. Caro, E. & Gutierrez, C. A green GEM: intriguing analogies with animal geminin. *Trends Cell Biol.* **17**, 580–585 (2007).
48. Zhou, J. *et al.* Cell cycle regulation of chromatin at an origin of DNA replication. *EMBO J.* **24**, 1406–1417 (2005).
49. Karnani, N., Taylor, C., Malhotra, A. & Dutta, A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome Res.* **17**, 865–876 (2007).

ONLINE METHODS

Plant material. *A. thaliana* seedlings (Col-0 ecotype) were grown in Murashige and Skoog (MS) salts medium supplemented with 1% (w/v) sucrose and 1% (w/v) agar in a 16 h:8 h light/dark regime at 22 °C. Plants constitutively expressing the Myc-ORC1a–tagged and the hemagglutinin (HA)-tagged *A. thaliana* CDC6a protein have been described[26,27].

Chromatin immunoprecipitation and microarray assays. ChIP assays of plants expressing His-ORC1 and HA-CDC6 were carried out using 10-day-old plants preincubated with 50 μM MG132 before the fixation step. For immunoprecipitation, 10 μl of antibody to Myc (SC-40ac; Santa Cruz Biotechnology) or antibody to HA (A2095; Sigma) were incubated with 1 mg of protein extract, previously sonicated, to obtain DNA fragments of ~500–1,000 bp. After they were washed, immunocomplexes were incubated at 65 °C for 6 h, treated with 20 μg of proteinase K and phenol-chloroform extracted. The DNA obtained (ChIP and input) was purified using Affymetrix cDNA cleanup columns. Three biological replicates for each condition were independently hybridized to GeneChip *A. thaliana* Tiling 1.0R Array (Affymetrix). DNA was amplified using Affymetrix Chromatin Immunoprecipitation Assay protocol. Amplification product (7.5 μg) was fragmented and labeled using GeneChip WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix). Scanning was carried out at a resolution of 0.7 μm using a GeneChip Scanner 3000 7G (Affymetrix). One of the CDC6 replicates was discarded owing to low correlation with the other replicates. The H4K5ac ChIP was carried out using an antibody from Millipore (06-759), and the H3 ChIP control was carried out using an antibody from Abcam (ab1791), as described[21,31,50,51].

Cell culture, BrdU labeling and isolation of nascent strands. *A. thaliana* MM2d cells were grown in MS medium supplemented with 3% (w/v) sucrose (Sigma), 0.5 mg ml^{-1} naphthalene acetic acid (NAA; Sigma) and 0.05 mg ml^{-1} kinetin (Sigma), pH 5.8, and subcultured every 7 d (ref. 52). For cell synchronization, a mid-exponential-phase culture was grown in MS medium lacking sucrose for 24 h, released by changing to medium supplemented with sucrose and cultured for an additional 3.5 h (ref. 25).

Newly synthesized DNA was labeled by adding 200 μM of BrdU, 20 μM 5-fluoro-2′-deoxyuridine and 10 mM hydroxyurea to the medium for the last 60 min of incubation (**Supplementary Fig. 1**). Cells were collected by filtration and DNA was purified[53]. Genomic DNA was digested with BamHI and centrifuged in 5.0 ml of a CsCl solution containing 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and 150 mM NaCl (refractive index adjusted to 1.4000, at 25 °C) in a Beckman Vti 65.2 rotor at 227,300*g* for 21 h at 20 °C. Fractions (100 μl) collected from the bottom were analyzed by immunoblot with antibody to BrdU (Becton-Dickinson) to identify the heavy-light (HL) density fractions. BrdU DNA in the HL fractions and genomic DNA in the unlabeled (LL) fractions, used as control, were pooled separately, dialyzed on TE buffer and analyzed by qPCR or massive sequencing.

To isolate short nascent DNA strands, genomic DNA was isolated from *A. thaliana* MM2d cells treated as described above, under RNase-free conditions. Purified DNA was denatured by heating and size-fractionated in a seven-step sucrose gradient (5–20% (w/v)) by centrifugation at 102,300*g* in a Beckman SW-40Ti rotor for 20 h at 20 °C (ref. 29). Fractions (1 ml) were collected and aliquots were analyzed by electrophoresis in an alkaline agarose gel to monitor size fractionation. Fractions containing replication intermediates (300–2,000 nucleotides in size) were subjected to polynucleotide kinase treatment and λ-exonuclease digestion, which degrades contaminating random sheared DNA and leaves replication intermediates protected by a 5′ RNA primer[28]. The relative abundance of nascent DNA strands around putative origins was monitored by qPCR.

Library preparation for Illumina sequencing of BrdU DNA. DNA libraries for both BrdU DNA and LL DNA were generated and sequenced on Genome Analyzer II according to the manufacturer's instructions (Illumina).

Quantitative real-time PCR. qPCR was carried out with a LightCycler Real-Time PCR System (Roche) using SYBR Green PCR mix and following the manufacturer's instructions. All qPCR reactions were carried out at least in duplicate and analyses were carried out using the commercial software. Oligonucleotides used for fragment amplification are listed in **Supplementary Table 2**.

Data processing and analysis. Sequenced reads were based-called using the standard Illumina software. The reads were trimmed down to 50-mer bases from the 3′ end and then mapped to the *A. thaliana* genome using SeqMap[54]. Uniquely mapping reads with mismatches up to 3 bp were used for the analysis, resulting in 3,849,549 BrdU reads and 9,799,171 unlabeled DNA sequences, after collapsing identical reads. The reads were extended so that the data represent the actual DNA fragments of the libraries (130 bp, determined from distribution of DNA fragments in the library). For all the analyses, identical reads were collapsed into single reads, and each data set was normalized to the total number of uniquely mapping reads. BrdU-positive regions were defined using MACS[55], with the unlabeled DNA sample as a control. A Poisson distribution *P*-value cutoff of 10^{-6}, calculated from the local λ value, was applied. Then, regions separated by <10 kb were combined, a restrictive criterion supported by other origin analysis[42,56].

The CDC6 and ORC1 ChIP data were quantile-normalized to genomic DNA using Affymetrix Tiling Analysis Software (TAS). Intensity peaks were searched using the TileMap algorithm[57] with hidden Markov model for combining neighboring probes (posterior probability cutoffs of 0.5). Histone modification data were quantile normalized to unmodified H3 ChIP data. Each data set was then normalized so that the mean signal across all probes in the genome was zero. The Affymetrix Integrated Genome Browser (IGB) was used by applying thresholds as described[31].

Enrichment of ChIP signals at BrdU midpoints was assessed by first calculating the average scores of probes within ±300 bp of BrdU midpoint of all origin regions and scores in the same number of randomly generated probes in regions flanking the midpoints by 5–10 kb. Then, significance of differences was calculated by Wilcoxon rank-sum test (two-sided *P*-values). The H4K5ac data were analyzed as described[21,31,50,51].

Illustrations were generated using Adobe Illustrator CS2 and Adobe Photoshop CS3.

50. Zhang, X. *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129 (2007).
51. Zhang, X. *et al.* The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* **14**, 869–871 (2007).
52. Menges, M., Hennig, L., Gruissem, W. & Murray, J.A. Cell cycle-regulated gene expression in *Arabidopsis*. *J. Biol. Chem.* **277**, 41987–42002 (2002).
53. Soni, R., Carmichael, J.P., Shah, Z.H. & Murray, J.A. A family of cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. *Plant Cell* **7**, 85–103 (1995).
54. Jiang, H. & Wong, W.H. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24**, 2395–2396 (2008).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Grégoire, D., Brodolin, K. & Mechali, M. HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary. *EMBO Rep.* **7**, 812–816 (2006).
57. Ji, H. & Wong, W.H. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629–3636 (2005).