Supplementary Information

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

Celina Costas[1,+], Maria de la Paz Sanchez[1,6,+], Hume Stroud[2,+], 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] and Crisanto Gutierrez[1,*]

[1] Centro de Biologia Molecular "Severo Ochoa", CSIC-UAM, Nicolas Cabrera 1, Cantoblanco, 28049 Madrid, Spain

[2] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA

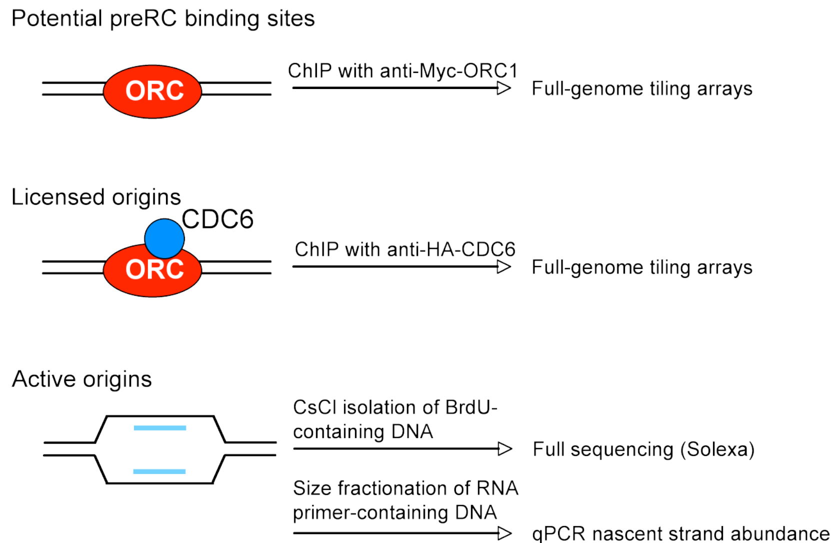
[3] Department of Plant Biology, University of Georgia, Athens, GA30602, USA

[4] Centro Nacional de Biotecnología, CSIC, Cantoblanco, 28049 Madrid, Spain

[5] Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095, USA

[6] Present address, Instituto de Ecología, Universidad Nacional Autónoma de México, 3er Circuito Exterior, Cd. Universitaria, México, DF 04510, Mexico

Supplemental Figures



Supplementary Fig. 1.
Experimental strategy used to identify replication origins in Arabidopsis cells.

The initial event in pre-replication complex (pre-RC) assembly is the binding of origin recognition complex (ORC) at certain genomic locations. To determine all possible sites where ORC can bind in different cell types, we used plants that express constitutively a Myc-tagged ORC1a protein. Chromatin was immunoprecipitated (ChIP) with anti-c-Myc antibodies and the material used to synthesize probes to hybridize full-genome Arabidopsis tiling microarrays, as described in Methods. Total genomic DNA of wild type plants was used as a control to determine the enrichment. Activation of pre-RC initiates when CDC6 protein binds to a subset of the ORC-bound sites. To determine all potential CDC6-binding sites we used plants expressing constitutively an HA-tagged CDC6a protein. Samples were processed as described for ORC1 and the data generated processed as described in Methods.

To determine the locations of putative DNA replication origins we used cultured Arabidopsis cells because:

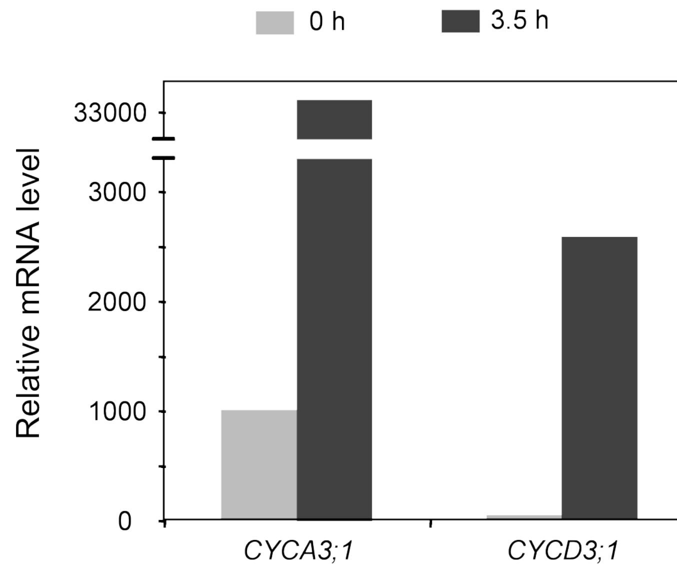
(1) the proportion of actively proliferating cells is relatively very small in whole plants, even in young seedlings,

(2) in the absence of any information on Arabidopsis origins of replication, we wanted to use a synchronized cell population where we can choose the time for BrdU labeling of nascent DNA strands. We are aware that with this strategy we may be loosing origins that are activated at a time not covered by the BrdU pulse,

(3) we wanted to use a treatment with hydroxyurea for two main reasons: (i) this treatment will reduce the dNTP pool, a necessary condition for an efficient BrdU labeling of nascent DNA strands, which otherwise will be below our current detection level. (ii) it will slow-down replication fork movement and will trigger an S-phase checkpoint [1]. In that way, we favor that BrdU incorporation is restricted to sequences around the origins activated during the BrdU pulse.

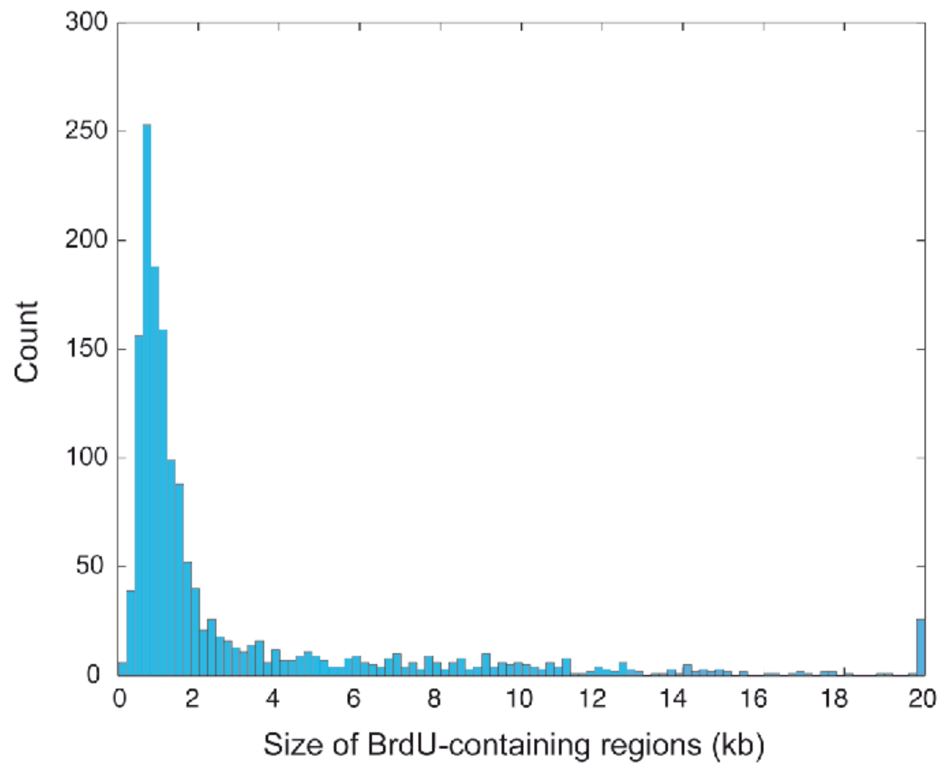
Thus, cultured cells were synchronized in G0 by sucrose deprivation, as described [2], and the purified heavy-light (HL) BrdU-containing DNA subjected to full sequencing using the Solexa/Illumina technology (see Methods). The genomic location of BrdU-sequences provide a genomic map of putative DNA replication origins that were active in cultured cells at the time of BrdU labeling.

Validation of DNA replication origin activity was then carried out by measuring by qPCR the abundance of nascent DNA strands using a purified sample of size-fractionated, RNA primer-containing short DNA molecules [3,4] (see also Methods).



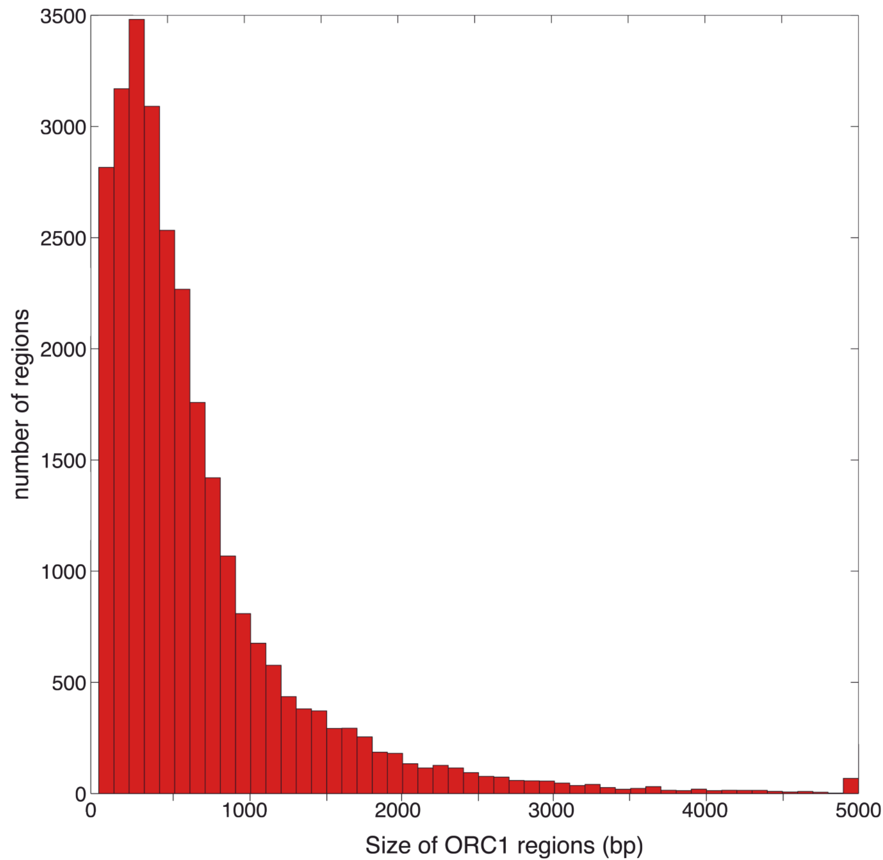
Supplementary Fig. 2.
Assessment of cell cycle progression in synchronized Arabidopsis cells.

Arabidopsis cultured cells were arrested in G0 by sucrose deprivation as described in methods and then released from the block. Samples were taken before the release (time 0 h) and 3.5 h afterwards, coinciding with the time at which parallel cultured were subjected to DNA extractions according to the protocols for isolation of BrdU-labeled DNA and for isolation of short nascent DNA strands, for origin identification. The mRNA level of selected cell cycle marker genes (*CYCA3;1* and *CYCD3;1*, which are upregulated at the G1/S transitions [5]) was determined by real-time PCR.



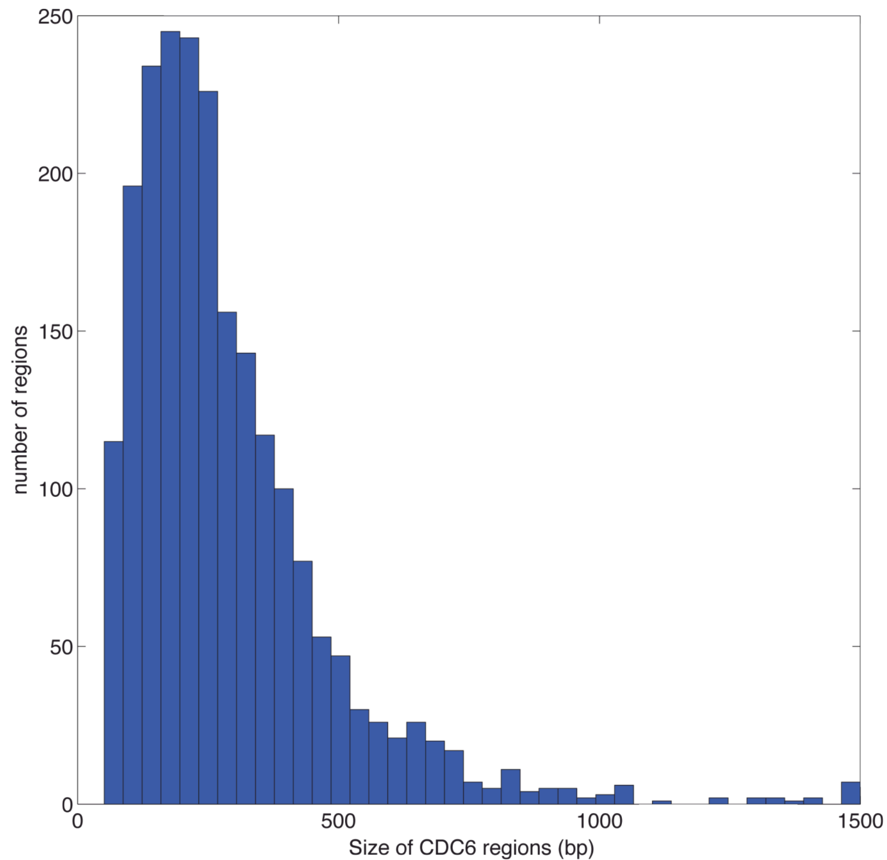
Supplementary Fig. 3.
Size distribution of BrdU-labeled DNA regions.

Origins were defined by the presence of DNA fragments enriched in the BrdU-labeled DNA sample. BrdU positive signals separated within 10kb were combined into a single origin. Here, the size distribution of origin regions is shown (*n* = 1543; *mean* = 3224 nt; *st. d.* = 8274nt; *median* = 1187nt). The data have been capped at 20kb, for simplicity (*min. value* = 28nt; *max. value* = 269861nt).



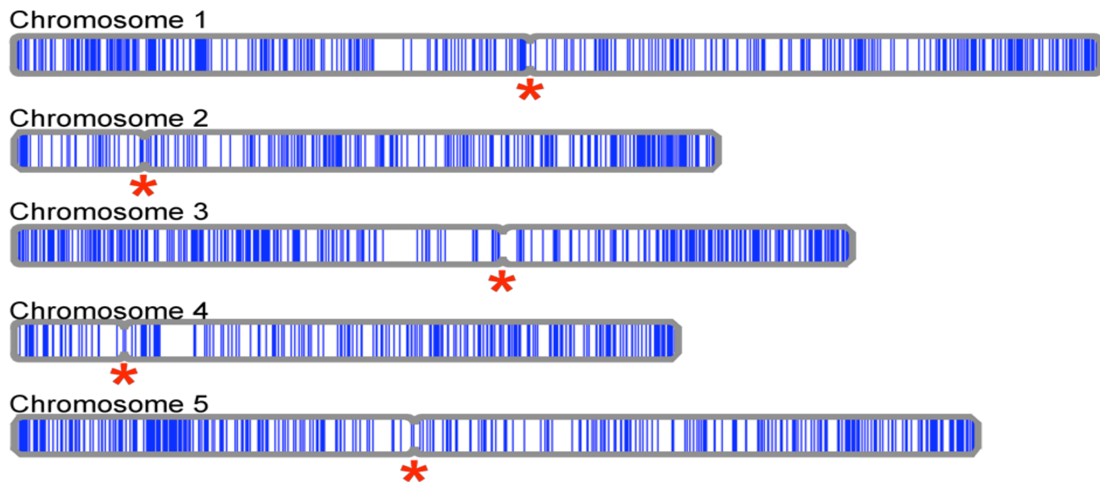
Supplementary Fig. 4.
Size distribution of ORC1-bound DNA regions.

ORC1 positive signals were defined as described in Methods using TileMap with HMM option, with a posterior probability cutoff of 0.5. Here, the size distribution of ORC1-bound regions is shown (*n* = 27405; *mean* = 695nt; *st. d.* = 688nt; *median* = 494nt). The data have been capped at 5kb, for simplicity (*min. value* = 52nt; *max. value* = 10149nt).



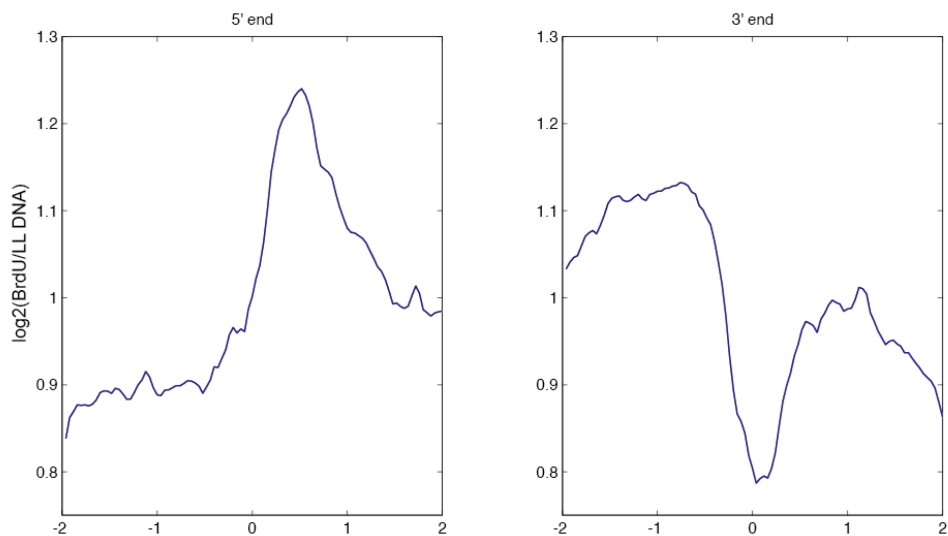
Supplementary Fig. 5.
Size distribution of CDC6-bound DNA regions.

CDC6 positive signals were defined as described in Methods using TileMap with HMM option, with a posterior probability cutoff of 0.5. Here, the size distribution of CDC6-bound regions is shown (*n* = 2157; *mean* = 290nt; *st. d.* = 202nt; *median* = 245nt). The data have been capped at 1.5kb, for simplicity (*min. value* = 51nt; *max. value* = 2272nt).



Supplementary Fig. 6.
Chromosomal view of origin density.

Origins (midpoints of BrdU regions), indicated here as vertical blue bars, were located across each Arabidopsis chromosome. The position of centromeres is indicated by asterisks (red).



Supplementary Fig. 7.
Ends-analysis of BrdU-seq reads relative to the control unlabeled DNA reads.

Reads were plotted over protein-coding genes (TAIR8) that contained origins in the bodies or those that had origins within 2kb of the transcriptional start sites (TSS; 0 in left panel) or the transcription termination sites (TTS; 0 in right panel). Two kb upstream or downstream of genes, and 2 kb into the genes are shown.

Supplementary Table 1.

The code for the different columns is as follows.

N = Origin number (total = 1543)

ORIGIN NAME = Origins are named consecutively according to their location starting with the chromosome number followed by a four-digit number beginning at the right tip of each chromosome. For example, the leftmost origin in chromosome 1 is ori1-0010 and the rightmost origin in the same chromosome is ori1-3760.

CHR = it denotes the chromosomal location of each origin.

START = chromosomal coordinate indicating the first nucleotide of the origin region defined by the BrdU-labeled region.

END = chromosomal coordinate indicating the last nucleotide of the origin region defined by the BrdU-labeled region.

SIZE = Size (in nt) of the corresponding origin region.

MIDPOINT = chromosomal coordinate indicating the midpoint of each origin region.

%G+C = G+C content (%) of a 200bp DNA region (origin midpoint ± 100 bp) at the middle of each origin region.

GENOMIC FEATURE = code of the genomic element overlapped by the origin midpoint (according to TAIR8).

TYPE OF GENOMIC FEATURE = four classes of genomic elements overlapping origins were considered: genes, transposable elements, transposable element genes and none (-) of the above.

DISTANCE TO GENOMIC FEATURE = in the latter case, the distance to the closest genomic element.

CLOSEST GENOMIC FEATURE = code of the closest genomic element.

SEQUENCE = DNA sequence of the 200bp region centered at the origin midpoint.

Supplementary Table 2.

List of oligonucleotides used in the real-time PCR experiments.

OLIGOS USED IN qPCR ORI-VALIDATION

ori1-2300

1F	AACACAATACCACAAACCAAAG
1R	AGTCAATGGAGTATAGATAGAG
2F	TTCCAATCTAAGCCAAAACTC
2R	ATCAGAATCGTCAGCATCAGC
3F	GTATTATCGCTCATGCTTGTG
3R	TGACAACTAAGCAAAGACAAG
4F	TCAATGGATCCAAATACTCGG
4R	TCAACAAGATTACGGAGGAGG
5F	CTTTACTTGTGCCACTTTTAGA
5R	ATTTTAATTTTATGTTTTGCCACG
6F	GTGGGTTTGAATTTCTGGTAG
6R	TGTACCCAATAAAAAGGAAATG

ori2-1340

1F	TAGCCCATGCCATTGATACA
1R	CTGCTGTGGAACACCACTTG
2F	CATCGCTGTATTCCCTTGTGT
2R	AAACCAAATGCACATGCAAA
3F	GTGCGTAGGTCTCTGTCTC
3R	GTCTAGCAGGTGGTGATGG
4F	CGGCACAGGTCTCCAACAC
4R	TGAGTCAGACCTTCCCCGC
5F	AGGAGAGACATCGTAAAAGG
5R	CAAATAACGCTGGACAAGGG

ori2-1430

1F	AGCAAACCTTTGTTGAGCATTC
1R	CTATGAATAGTTATTGCTAGAGC
2F	ACTGAACAAGCAGCATAAAC
2R	AAGACGACGGAAGAAGAAGA
3F	TTCTTCCGTCGTCTTCATCGTC
3R	CCAGCGGCTTGACGAATTTCTT
4F	GTAAGCC CTCTAACCAC TAAG
4R	CTCCACTTGATGATTTTTTCTC
5F	GCTTAATAGAACATTGCTAA C
5R	CTAGAGAGTCAATAAAGATGAC
6F	AGCATATAATAAGAGCCTCCAG
6R	TGTTAGGCTCGAATTGACGG

At4g14700

1F	TCTTCTCTGAGTTCCAAGGC
1R	CTGGAAGATCGATTGGGTC
2F	AGGTTGGAAGGAATCAATGC
2R	ACTTGAATATGAGGTGCCTG
3F	TCACAAGTGGATCTGTAGATG
3R	AGTTCATCCAAGAAAAGACACG

References

1. Santocanale, C. & Diffley, J.F. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature* **395**, 615-618 (1998).

2. Sanchez, M.P. & Gutierrez, C. Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. *Proc. Natl. Acad. Sci. U S A.* **106**, 2065-2070 (2009).

3. 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).

4. Gomez, M. & Antequera, F. Overreplication of short DNA regions during S phase in human cells. *Genes Dev.* **22**, 375-385 (2008).

5. Diaz-Trivino, S. et al. The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differently expressed in proliferating and endoreplicating cells. *Nucleic Acids Res.* **33**, 5404-5414 (2005).