The gastrula transition reorganizes replication-origin selection in Caenorhabditis elegans

Marta Rodríguez-Martínez1, Natalia Pinzón1, Charles Ghommidh2, Emmanuelle Beyne1,3, Hervé Seitz1, Christelle Cayrou1,3 & Marcel Méchali1

Although some features underlying replication-origin activation in metazoan cells have been determined, little is known about their regulation during metazoan development. Using the nascent-strand purification method, we here identified replication origins throughout Caenorhabditis elegans embryonic development and found that the origin repertoire is thoroughly reorganized after gastrulation onset. During the pluripotent embryonic stages (pregastrula), potential cruciform structures and open chromatin are determining factors that establish replication origins. The observed enrichment of replication origins in transcription factor–binding sites and their presence in promoters of highly transcribed genes, particularly operons, suggest that transcriptional activity contributes to replication initiation before gastrulation. After the gastrula transition, when embryonic differentiation programs are set, new origins are selected at enhancers, close to CpG-island-like sequences, and at noncoding genes. Our findings suggest that origin selection coordinates replication initiation with transcriptional programs during metazoan development.

DNA replication initiates from precise sites along the genome that are called replication origins. Prokaryotes and viruses generally replicate from a single origin, whereas, because of its complexity, the eukaryotic genome requires multiple replication origins to complete DNA duplication. In the past few years, several components defining replication origins have been identified1,2. However, the nature and selection of replication origins, particularly their regulation during development, remain elusive in metazoans.

In several organisms, a correlation between transcriptional activity and replication-origin location has been observed, especially the association of a subset of replication origins with promoter regions and gene bodies3–6. These correlations may be favored by the permissive chromatin environment characteristic of actively transcribed units, as revealed by the type of chromatin modifications associated with replication origins. In agreement with this notion, replication origins are also associated with binding sites for remodeling-complex subunits that drive these permissive epigenetic modifications3,7–12.

Sequence specificity for replication origins has been difficult to establish in metazoan cells, in contrast to most yeast species, particularly Saccharomyces cerevisiae, in which replication initiation takes place at an A/T-rich consensus sequence termed ARS13. Interestingly, G/C-rich replication origins have recently been found in the yeast Pichia pastoris14 and have also been found to be associated with transcription start sites (TSSs). In metazoans, the majority of replication origins are associated with G-rich sequences12,15–17 found in the vicinity of actively transcribed genes. A subset of these G-rich sequences form a four-stranded secondary DNA configuration called a G quadruplex (G4)18–20, and point mutations within this DNA element that affect G4 stability impair origin function20–22. It has recently been proposed that asymmetry in the distribution of G/C-rich sequences may also aid in unwinding of DNA at replication origins by altering the stability of the DNA internal structure23,24. Together, these data suggest an important role of DNA secondary structure in replication-origin establishment in metazoans.

Both the nature and the selection of replication origins during metazoan development remain poorly understood. One major developmental change is the maternal-to-zygotic transition, which occurs during either blastulation or gastrulation25. At this stage, a burst of new transcription occurs in the embryo, and maternal RNAs are inactivated, thereby signifying the transition from maternal to embryonic developmental control that is essential for the onset of gastrulation and the initiation of specific differentiation programs. In Drosophila melanogaster and Xenopus laevis, DNA replication origins are closely spaced during the rapid cell cycles of early development, but a more restricted pattern of origin activation is established after the onset of transcription26–29. Here, we performed a genome-wide analysis of replication origins during C. elegans embryo development. C. elegans is a relevant model for replication-origin studies during metazoan development because active gene transcription is already present in pluripotent cells, similarly to transcription during mouse or human development. Transcription starts at the four-cell stage, and gastrulation starts at the 26-cell stage, when a burst of new transcription defining the maternal-to-zygotic transition occurs30.

1Institute of Human Genetics, CNRS, Montpellier, France. 2Agropolymer Engineering and Emerging Technologies, University of Montpellier, Montpellier, France. 3Present addresses: Laboratory Genetics of Rare Diseases, INSERM U827, Montpellier, France (E.B.) and Institute for Research on Cancer and Aging, Nice (IRCAN), CNRS UMR 7284, Nice, France (C.C.). Correspondence should be addressed to M.M. (marcel.mechali@igh.cnrs.fr), C.C. (christelle.cayrou@unice.fr) or H.S. (herve.seitz@igh.cnrs.fr).
Our results identify distinct and specific structural and epigenetic features of origins before and after this crucial embryonic transition.

RESULTS
Mapping active replication origins during *C. elegans* development
We identified replication origins that are established during *C. elegans* early development (pregastrula origins) or that specifically appear after gastrulation onset (26-cell stage)\(^3\) when new transcription programs are set within the embryo (postgastrula-specific origins). For this purpose, we purified nascent strands (NSs) from pregastrula embryos and a mixed embryo population (comprising pre- and postgastrula embryos; Fig. 1a) by using the RNA-primed NS purification method\(^3\) (Supplementary Fig. 1a and Online Methods). This method uses sucrose gradient sedimentation to isolate RNA-primed nascent DNA at replication origins from the bulk of high-molecular-weight DNA and from Okazaki fragments, which are shorter than RNA-primed NSs. Gradient separation is followed by an exhaustive digestion with lambda exonuclease, which removes contaminating DNA pieces but cannot degrade 5’-end RNA-primed DNA fragments. We also developed a new large-scale synchronized culture method in a bioreactor to recover a sufficient number of *C. elegans* embryos for genome-wide studies (Online Methods). Worm growth was synchronized to recover a highly pure population of pregastrula embryos (Supplementary Table 1; 93% of the gravid adult population carried only pregastrula embryos). This synchronization was achieved by seeding embryos in medium without bacteria to arrest worm development at the first larval stage (L1). After all embryos reached this state, food was provided, and development started synchronously (details in Online Methods). Moreover, synchronization was improved by the fine control of parameters, as is possible only in a bioreactor.

We verified embryo quality and viability by flow cytometry analysis (details in Online Methods). Moreover, synchronization was improved by the fine control of parameters, as is possible only in a bioreactor.

To validate the presence of replication origins at IRs, we tested the formation of a replication bubble in its vicinity by using neutral/neutral 2D gel electrophoresis and the bubble-trap method\(^3,7\). This qualitative method, in which DNA restriction fragments are separated according to their size in the first dimension and according to their shape in the second dimension, is very useful to detect replication origins. This method’s relatively low sensitivity makes it technically challenging to use in complex organisms, owing to their large genomes. However, the large number of synchronized embryos that we were able to recover, together with the size of the *C. elegans* genome, allowed us to use this method under our experimental conditions. After isolation of restriction fragments containing a replication bubble from mixed embryos, we detected a bubble arc characteristic of an origin only in the region containing both an IR and one origin (positive region), but not in either the region without a replication origin (negative region) or the region containing only a bubble arc characteristic of an origin only in the region containing both an IR and one origin (positive region), but not in either the region without a replication origin (negative region) or the L1-arrested sample (Fig. 2e). These results obtained with two entirely different methods indicate that an IR structure can facilitate replication-origin establishment. When considering the inverted-repeat (IR) structure with two symmetric stems spaced by a variable loop sequence, we found that pregastrula origins were mostly enriched at the loop border and not at the center, whereas postgastrula-specific origins were somewhat depleted in the same region (Fig. 2b).

To validate the presence of replication origins at IRs, we tested the formation of a replication bubble in its vicinity by using neutral/neutral 2D gel electrophoresis and the bubble-trap method\(^3,7\). This qualitative method, in which DNA restriction fragments are separated according to their size in the first dimension and according to their shape in the second dimension, is very useful to detect replication origins. This method’s relatively low sensitivity makes it technically challenging to use in complex organisms, owing to their large genomes. However, the large number of synchronized embryos that we were able to recover, together with the size of the *C. elegans* genome, allowed us to use this method under our experimental conditions. After isolation of restriction fragments containing a replication bubble from mixed embryos, we detected a bubble arc characteristic of an origin only in the region containing both an IR and one origin (positive region), but not in either the region without a replication origin (negative region) or the L1-arrested sample (Fig. 2e). These results obtained with two entirely different methods indicate that an IR structure can facilitate replication initiation at early developmental stages.

Pregastrula origins, but not postgastrula-specific origins, were also significantly and reciprocally associated with operons (71% of operons; *P* < 10\(^{-20}\)) (Supplementary Fig. 3b). Although operon genes are organized in polycistronic clusters, pregastrula origins were preferentially enriched at operon TSSs (Fig. 2d), at chromosome centers and, to a lesser extent, in the arms (Supplementary Fig. 3c). Operons overlapping a replication origin were significantly more enriched...
We found that open-chromatin marks were highly associated with replication origins and relatively prone to transcription. This association was essentially upstream of the operon TSS (data not shown), a feature that may prevent collisions between DNA and RNA polymerases in these long and active transcription domains.

Highly transcribed regions and open chromatin are prone to initiation of DNA replication during C. elegans early development

Operons represent less than 15% of all coding genes in C. elegans, but they constitute the most abundantly transcribed genes during early embryogenesis. Operons represent less than 15% of all coding genes in C. elegans. As in IRs than were origin-free operons (83% versus 53%; \( P = 9.0756 \times 10^{-28} \); Fig. 2e). This association was essentially upstream of the operon TSS (data not shown), a feature that may prevent collisions between DNA and RNA polymerases in these long and active transcription domains.

Operons represent less than 15% of all coding genes in C. elegans, but they constitute the most abundantly transcribed genes during early development and encode proteins essential for gene expression and energy metabolism. Because replication-origin density is higher at transcriptionally active units in other organisms, we asked whether transcription, and by extension open chromatin areas, were important features in the developmental regulation of replication origins in C. elegans. We found that operon-TSS regions were strongly associated with replication origins, and that transcriptional activity was high at the origins.

Operons represent less than 15% of all coding genes in C. elegans, but they constitute the most abundantly transcribed genes during early development and encode proteins essential for gene expression and energy metabolism. Because replication-origin density is higher at transcriptionally active units in other organisms, we asked whether transcription, and by extension open chromatin areas, were important features in the developmental regulation of replication origins in C. elegans.

We found that open-chromatin marks were highly associated with replication origins. As in mammalian embryonic stem (ES) cells, replication-origin-initiation sites were centered in areas of high nucleosome occupancy and were surrounded by nucleosome-depleted regions.
correlated with the presence of replication origins at promoters, we observed a correlation between replication-origin density and gene expression levels of the associated genes strongly suggested that active transcriptional units are prone to fostering replication-origin establishment, particularly during early development. After gastrulation onset, new replication origins preferentially localize at enhancers and noncoding genes. In contrast to pregastrula origins, postgastrula-specific origins did not correlate with coding genes. We found that they preferentially localize at enhancers and noncoding genes.
associated with noncoding genes (Fig. 1d) and were enriched in the 2-kb upstream regulatory region (Fig. 4a). This enrichment was more pronounced when chromosome arms and center were examined separately (Supplementary Fig. 5a), and the association of origins with noncoding genes was particularly high in the right chromosome arm (Supplementary Fig. 5b). The overrepresentation of noncoding genes in this chromosome domain, owing to the abundance of the 21–U piwi-interacting–RNA class inside the right arm of chromosome IV, and their strong association with postgastrula-specific origins (68%), might explain this result. However, these genes may not be transcribed during embryogenesis (21–U RNAs that can be detected in embryos are most probably parentally deposited in the gametes)41. In agreement with this notion, available embryonic RNA pol II data sets showed a poor overlap between noncoding genes associated with postgastrula-specific origins, RNA pol II and H3K4me3 density (Supplementary Fig. 5c). Our results suggest that promoter transcriptional activity does not explain the observed association between noncoding–RNA genes and postgastrula-specific origins. However, because these data sets were acquired in early embryos, future work will be needed to formally

**Figure 3** Association of replication origins with chromatin marks and transcription features. (a) H3-mark distribution around postgastrula (green) and postgastrula-specific (orange) origin summits (bp). (b,c) Distribution of euchromatin marks (b) and heterochromatin marks (c) around postgastrula origin summits and random centers (gray; mean ± s.d. for 100 sets of random regions; n = 3 independent experiments). (d) mRNA levels in postgastrula (green) and postgastrula-specific (orange) embryos (EB) of genes with promoters that contain (+) or do not contain (−) RNA pol II and/or replication origins. *P < 0.05; **P < 0.0001 by two-sided t test (n = 3 independent experiments). Box plots are defined as follows: lower and upper boundaries of boxes indicate the 25th and 75th percentiles; the horizontal dividing line shows the median, and whiskers extend to minimum and maximum values. Description of the data sets used can be found in Online Methods. (e) Percentage of promoters of three different classes of genes (coding operons, coding nonoperons and noncoding nonoperons) that contain HOT regions and postgastrula origins (green boxes) compared with random sites (gray boxes). Analysis is shown for chromosome arms and centers separately. n = 3 independent experiments (means ± s.d. for 100 sets of random regions). *P < 10−8 by chi-squared test, calculated against 100 sets of random regions, as described in Online Methods. Source data are available online.
Figure 4 Association of replication origins with noncoding genes and enhancers. (a) Distribution of pregastrula (left) and postgastrula-specific (orange) replication origins 3,000 bp around the TSSs of noncoding genes in a 100-bp window. (b) Percentage of enhancers associated with postgastrula-specific (left) or pregastrula origins (right) compared with random sites. n = 3 independent experiments. * P < 10^{-10} (according to the distribution function of the normal distribution), as described in Online Methods. Lower and upper boundaries of boxes indicate the 25th and 75th percentiles; the horizontal dividing line shows the median, and whiskers extend to minimum and maximum values. (c) Chromatin immunoprecipitation (ChIP)–chip signals for histone H3K4me1 ± 3 kb around the summits of postgastrula-specific replication origins (solid orange line) or the centers of random sites (gray lines) (mean ± s.d. for 100 sets of random regions; n = 3 independent experiments). y axis represents average log_{2} value of normalized reads. (d) Heat map of H3K4me1 signal densities in 3-kb regions on each side of postgastrula-specific origin summits. The blue color intensity is proportional to the positive signal. Replication origins were considered to be positive for H3K4me1 signal (n = 4,096 replication origins) when summits were covered by this mark. Source data are available online.

It has previously been shown that in C. elegans, promoter regions of noncoding genes have enhancer-like features. Therefore, we wondered whether enhancers might govern the establishment of postgastrula-specific origins. We found a specific association of intergenic enhancers with postgastrula-specific origins (P < 10^{-15}) (Fig. 4b). Furthermore, the analysis of early established H3K4me1 (the main epigenetic mark known to define enhancers) showed that postgastrula-specific origins are highly associated with enhancers defined before the gastrula transition (Fig. 4c,d). To further characterize this association, we examined the epigenetic profiles of postgastrula-specific origins on the basis of their location with respect to their associations with this set of enhancers. These regions were characterized by the presence of RNA pol II and H3K27ac only at H3K4me1-positive replication origin summits and were surrounded by H3K4me3 (Fig. 5a). We then identified the gene types associated with postgastrula-specific origins linked to early defined enhancers. We examined the early established H3K4me1 profiles of the postgastrula-specific origin summits that were associated with promoters. We found a substantial distribution of early established H3K4me1 at replication origins in noncoding gene promoters in chromosome centers, although the strongest early established H3K4me1 profile was found for replication origins in the promoters of coding genes, especially at chromosome centers (Fig. 5b).

These data revealed that enhancer regulatory regions are associated with the new replication origins established after gastrulation onset. They also suggest the appearance of a new mechanism coordinating DNA replication with cell-type-specific transcription during cell differentiation and involving noncoding genes and enhancers.

Specific sequences are differentially associated with replication origins before and after gastrulation onset

We investigated whether origins contain different sequences depending on the development stage. We considered only motifs showing positional bias distribution around the origin summits compared with the random expectation (P < 10^{-5}), according to the Watson (sense) and Crick (antisense) chromosomal strands. Pregastrula origins were primarily associated with A/T-rich sequences that were asymmetrically distributed around the origin summits (Supplementary Table 3 and Supplementary Fig. 6a). Owing to the preferential association of these origins with promoters and HOT regions, we compared these motifs with TF-binding motifs. More than 60% of these motifs were included in at least one of the 8-mer motifs recognized by each TF (Supplementary Table 3). This observation strengthens the evidence of TF involvement in origin establishment during early development.

In postgastrula-specific origins, we detected G/C-rich motifs around the summits (Supplementary Table 3 and Supplementary Fig. 6b,c). Alignment of CpG-island-like sequences (CGI-like; with a GC content of at least 60% and length >200 bp) with replication origins showed a relatively precise positioning of CGI-like sequences 140 bp from the summit, only in postgastrula-specific origins (Fig. 6a). This enrichment was observed for almost 30% of postgastrula-specific origins found at chromosome centers and nearly 50% of those.
in chromosome arms (Supplementary Fig. 6d,e). We confirmed the presence of replication origins in CGI-like sequences by 2D gel analysis combined with the bubble-trap method by using mixed embryo samples (Supplementary Fig. 6f). We detected bubble arcs only in the region containing both a CGI-like region and a replication origin (positive region), but not in the negative region (no replication origin) or in the L1-arrested sample. We conclude that CGI-like sequences are a crucial element defining postgastrula-specific origins, particularly at chromosome arms.

To further characterize this specific association, we then investigated the genomic locations and epigenetic features of CGI-like-rich origins. Postgastrula-specific origins containing CGI-like regions were significantly associated with coding genes, particularly their promoters (Fig. 6b), in both chromosome arms and centers, and were also enriched in early established H3K4me1. Overall, although enhancers were infrequent in CGI-like sequences (only 23% included at least one CGI-like sequence), postgastrula-specific origins were significantly enriched in CGI-like-containing enhancers (37% versus 19% expected at random) (Fig. 6c). Moreover, the detection of early established H3K4me1 signal enrichment only at the summit of postgastrula-specific origins associated with a CGI-like sequence emphasized these results (Fig. 6d). In contrast, replication origins that did not include CGI-like sequences were depleted in early established H3K4me1. Together, these data support a cooperative effect of CGI-like sequences located in early established cis-regulatory elements, on the new replication initiation profile set in late embryogenesis.

**DISCUSSION**

**DNA secondary structures are common features of metazoan replication origins**

It has been postulated that DNA secondary structures may be features of replication origins. In *C. elegans*, G/C-rich elements are a new feature observed at replication origins only after gastrulation. In metazoans, G-rich elements that can potentially form G4 structures are associated with replication origins15,18–20,24. In *C. elegans*, we did not detect a significant correlation with G4s, in agreement with the weak phenotypes of worms lacking dog-1 (a FANCJ homolog responsible for G4 unfolding) during embryogenesis. However, G/C richness appeared to be conserved in *C. elegans* late embryos. We also found that IRs were significantly associated with replication origins from early stages. IRs can form cruciform structures and are highly conserved in the *Caenorhabditis* family44. The association of IRs and replication origins takes place close to transcribed genes and is maintained throughout development. The association of origins with G4s in mammals and with IRs in *C. elegans* might reflect a similar mechanism whereby secondary structures, instead of strict sequences, are used to control transcription and initiation of replication, possibly through chromatin remodeling.

**Rapid DNA replication in pluripotent embryos through preferential location of replication origins at open chromatin regions**

During early embryogenesis of some metazoans, such as *C. elegans*, *D. melanogaster* and *X. laevis*, genome duplication takes place in an extremely short period. In *X. laevis* early embryos, DNA replication is rapidly completed through stochastic activation of replication origins along the genome27. It has previously been reported that in *C. elegans* pregastrula embryos, replication occurs primarily within 8.5 min (refs. 45, 46), and at least 2,900 active origins might be required to replicate the 100-Mbp genome in each cell cycle (assuming an average fork rate of 2 kb/min and bidirectional forks). This scenario would also suggest a maximum interorigin distance of 34 kb to avoid unreplicated DNA. In our study, we identified 13,581 pregastrula origins, thus suggesting an average origin distance of 6 kb if all embryo cells were using all origins. Our data indicated that at least one-fifth of all identified origins were used in every cell and at every cell cycle during the pregastrula stages, a result in agreement with the excess
Indeed, in all organisms studied to date, including C. elegans, the presence at chromatin regions that coincide with promoters of highly transcribed genes. The presence at these regions statistically favor the location of replication origins at already open chromatin. New replication origins are established. These sites are preferentially localized at regulatory regions that are linked to transcriptional regulation. RNA pol II pausing during development occurs after synthesis of a short RNA primer that may be necessary for initiation of DNA synthesis by DNA pol α. Indeed, RNA polymerase is the primase required for initiation of mitochondrial DNA synthesis. Hence, assembly of a transcriptional complex in X. laevis early embryos favors replication-origin specification.

**Figure 6** Association of replication origins with CGIs. (a) CGI-like-sequence content distribution in 1-kb regions on both sides of the summits of pregastrula and postgastrula-specific origins in a 100-bp window. (b) Pearson’s correlation coefficients (two sided) between CGI-like sequences and genes, promoters, enhancers or H3K4me1 signal in pregastrula (Pre-G) and postgastrula-specific (Post-G-spe) origins and in random sites. Positive correlations are shown in a gradation of orange on the basis of the co-occurrence of CGI-like sequences and all the features described inside replication origins. Negative correlations are shown in a gradation of blue. For postgastrula-specific origins, the analysis on chromosome arms and centers is also shown. (c) Percentage of enhancers with CGI-like regions that overlap postgastrula-specific origins or random sites. Enhancers associated with postgastrula-specific replication origins are significantly more covered by CGI-like sequences than all enhancers or those associated with random regions. n = 3 independent experiments. *P = 1.63 × 10⁻¹⁰ (according to the distribution function of the normal distribution). Data are shown as mean ± s.d. (n = 100 sets of random regions), as described in Online Methods (comparative analysis of replication origins and genome features section). y axis represents average log2 value of normalized reads. (d) H3K4me1 and histone H3 signal around postgastrula-specific origin summits overlapping (solid line) or not (dotted lines) with CGI-like sequences. Source data are available online.

of potential origins and flexibility in origin choice found in the DNA replication origins of eukaryotes including yeast and mammals. However, large parts of the C. elegans genome were not present in the microarray used, owing to the high repeated sequence content. Therefore, the number of replication origins in these regions may have been underestimated.

In contrast to X. laevis, in which transcription is shut off for the first 12 cell cycles, in C. elegans transcription starts at the four-cell stage. Our results suggest that the necessity of rapid cell cycles may opportunistically favor the location of replication origins at already open chromatin regions that coincide with promoters of highly transcribed genes. The presence at C. elegans origins of nucleosome-depleted regions and chromatin marks, such as H3K4me3 or H3K27ac, as has also been observed in ES cells, supports this hypothesis. It has been proposed that the nucleosome at the origin center is labile and that the nucleosome-depleted region upstream would favor the binding of the prereplication complex (preRC), as observed in S. cerevisiae. Indeed, in all organisms studied to date, including Escherichia coli, initiation of DNA synthesis always takes place downstream of the preRC-binding site, simply because of the large DNA sequence coverage and steric hindrance of the preRC complex. However, whether the observed nucleosome organization is the cause or the consequence of the replication-complex regulation remains an open question.

The observed paused RNA pol II at pregastrula origins may aid in regulating this organization. RNA pol II pausing during development occurs after synthesis of a short RNA primer that may be necessary for initiation of DNA synthesis by DNA pol α. Indeed, RNA polymerase is the primase required for initiation of mitochondrial DNA synthesis. Hence, assembly of a transcriptional complex in X. laevis early embryos favors replication-origin specification.

**New transcriptional regulation at gastrulation reorganizes replication-initiation sites**

After gastrula transition, cell differentiation begins, S phase lengthens, and new replication origins are established. These sites are preferentially localized at regulatory regions that are linked to
cell-type-specific transcription, as suggested by their association with intergenic enhancers. Moreover, at least some of these enhancers have been defined before gastrula transition. Even though we cannot analyze the enhancers newly established after gastrula transition, the early established enhancers associated with postgastrula-specific origins have several distinguishing features. In C. elegans, in contrast to mammals, these H3K4me1-rich replication origins do not have nucleosome-depleted regions but contain the open-chromatin mark H3K27ac and are close to promoter regions. As different cell types differentiate after gastrulation, new enhancers may be defined, and therefore whether they would also associate with replication origins remains an open question.

Interestingly, we observed G/C-rich sequences specifically at replication origins in enhancer regions. We have previously suggested that enhancers may serve as selectors of a specific class of replication origins to be activated among the excess of potential replication origins present in each cell. Our results show that specific use of enhancers for differential gene regulation might be an important feature that explains replication-origin selection during late development. CGI-like regions, one of the most abundant features detected at postgastrula-specific origins, are commonly associated with low nucleosome content and can be modulated by methylation. Although the C. elegans genome is considered to be unmethylated, N4 adenine methylation can replace the traditional m3C found in many metazoan genomes. A reciprocal regulation of H3K4 and N4 adenine methylation was observed, in which H3K4me1 increased when m4A decreased. In mouse stem cells, only hemimethylated C was detected at replication origins in enhancer regions, thus potentially favoring a negative regulation of replication initiation by DNA methylation. It should be interesting to determine whether specific DNA-methylation marks contribute to the cell-type-specific enhancer-activity regulation and replication-origin selection after gastrulation onset.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (FP7/2007-2013, grant agreement no. 233339), by the Fondation pour la Recherche Médicale (FRM) and the European Commission Network of Excellence EpiGeNSys (HEALTH-F4-2010-257082 to M.M.). M.-R.-M. was supported by the Fondation ARC. We thank S. Galas (CRBM, Montpellier) for advice regarding C. elegans development and for providing the strain used in this study, as well as J. Bacal for help with the 2D gel technique, D. van Essen for support in bioinformatics, and E. Andermarcher and J. Hutchins for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

M.M. proposed the project and supervised its execution. C.C. supervised the project and its experimental design. C.G. and M.-R.-M. designed the bioreactor conditions for C. elegans synchronous growth. M.-R.-M. and C.C. performed the large-scale embryo preparation. M.-R.-M. performed nascent-strand purification from the embryos, embryo imaging, fluorescence-activated cell-sorting analysis, and bubble-trap and 2D gel experiments. E.B., N.P., H.S., C.C. and M.-R.-M. performed the computational analyses. M.-R.-M. and C.C. performed the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.


© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
**ONLINE METHODS**

**Bacterial cultures.** *E. coli* fermentation was performed at the Recombinant Proteins Facility (ProRec), Montpellier, France. One 1-L flask of preculture of the *E. coli* strain OP50-1 (genetically characterized as a uracil auxotroph and resistant to streptomycin and phage 80) was incubated at 37 °C at 200 r.p.m. overnight and was then used to inoculate 151 of TY medium supplemented with 1 ml Erol 520 SX (Ouvré PMC) as antifoaming agent in a 20-L Biolafitte fermentor. The temperature was maintained at 37 °C; the aeration rate was set at 0.5 air volumes per medium volume per minute (vvm), and the mixing rate was set at 800 r.p.m. *E. coli* cells were harvested approximately 7 h after inoculation by centrifugation in a Beckman JLA8100 rotor at 4,000 r.p.m. for 20 min. Bacterial pellets were washed once with PBS and centrifuged again under the same conditions. Pellets were kept at 4 °C until use.

**C. elegans strain and classical culture.** The wild-type *C. elegans* strain N2 was provided by S. Galas (CRBM, CNRS, Montpellier) and was maintained on agar plates. Gravid adults were collected, and embryos were isolated. Embryos were then inoculated (10,000 embryos/ml) in the absence of bacteria in 500 ml of S medium supplemented with filter-sterilized antibiotics (20 mg/l streptomycin and 5 mg/l nystatin) to prevent contamination. The seed flask was incubated at 16 °C and 220 r.p.m., and after 48 h, when worms had reached the arrested L1 stage, the culture was inoculated with 50 µl/well of *E. coli* suspension and incubated at 20 °C at 220 r.p.m. for 3 d to obtain gravid adults.

**Biorreactor culture.** To obtain large amounts (>4 × 10^8) of synchronized embryos, we maintained embryos in 500-ml flasks (one for 1.5-L and six for 15-L cultures) for 48 h without bacteria, as described above. L1-arrested worms were then inoculated in the bioreactor filled with 151 of S medium and 50 µl (w/v) of bacteria. The culture was maintained at 20 °C and 150 r.p.m. for 60 h. Pressure, airflow, oxygen and pH were adjusted throughout the culture, as described below. Worm development was tracked through microphotography of the samples taken from the bioreactors at regular intervals. *C. elegans* counting allowed us to determine the percentage of synchronized worms at each larval stage (three samples of 500 worms per stage).

Nematodes were cultured in a 20-L Biolafitte fermentor (15-L working volume) fitted with three standard Rushton impellers (10-cm diameter) and operated at 150 r.p.m. and 20 °C. The pH was maintained at 6.5 with 2 M phosphoric acid. The culture was maintained at 20 °C and 220 r.p.m. for 60 h. Pressure, airflow, oxygen and pH were adjusted throughout the culture, as described below. Worm development was tracked through microphotography of the samples taken from the bioreactors at regular intervals. *C. elegans* counting allowed us to determine the percentage of synchronized worms at each larval stage (three samples of 500 worms per stage).

**Embryo purification.** After 60 h of culture in a 15-L bioreactor, worms were harvested in four 5-L flasks on ice and allowed to settle to the bottoms of the flasks for 15 min. Supernatants were discarded by vacuum aspiration. Worms were resuspended in cold 0.1 M NaCl. One volume of ice-cold 60% sucrose solution was added to the suspension and mixed thoroughly. After centrifugation at 1,500 r.p.m., 4 °C for 7 min, viable worms were quickly collected, washed once with cold 0.1 M NaCl and resuspended in cold 0.1 M NaCl. Aliquots of 3 ml were distributed in 50-ml conical tubes, and cold water was added up to 20 ml. Incubation in a mix of 8 ml bleach and 4 ml 5 N NaOH for 5 min resulted in worm disintegration while embryos remained intact. To eliminate dead embryos and worm debris, another cleaning step with 60% sucrose was performed, and then embryos were washed twice with cold 0.1 M NaCl. Embryo purity and developmental stage were assessed through microscopic observation at each step and at the end of the purification. Some embryos were cultured at 20 °C to ensure their viability and ability to progress throughout development.

**Flow cytometry.** After chitinase treatment of embryos (1 U/ml for 10 min), cells were dispersed by gentle pipetting, washed once in egg buffer (118 mM NaCl, 2 mM MgCl₂, 2 mM CaCl₂, 48 mM KCl and 25 mM HEPEs, pH 7.3) and fixed with ethanol. Cells were washed once in PBS before incubation with 50 µg/ml RNase A at 37 °C for 30 min. After addition of PBS, 2% FBS and 50 µg/µl of propidium iodide, cells were incubated at 4 °C for 30 min. Analysis was performed with a FACS Calibur machine (BD Biosciences).

**Immunofluorescence.** After chitinase treatment, embryos were labeled with 50 µM EdU for 20 min and then washed twice in egg buffer. 2,000 embryos were then dropped onto a slide, allowed to dry, fixed with methanol for 10 min at −20 °C, washed with PBS and dried. Slides were kept at 4 °C. EdU detection was carried out with a Click-IT kit protocol (Thermo Fisher Scientific).

**Nascent-strand (NS) purification.** Embryo and L1 pellets were thawed at 56 °C for 2 min and lysed in lysis buffer (0.1 M Tris–HCl, pH 8.5, 0.1 M NaCl, 30 mM EDTA and 0.5% SDS) at 56 °C for 3 h. Samples were digested with 200 µg/ml proteinase K at 37 °C for 2 h. After phenol/chloroform extraction and precipitation with 0.2 M NaCl and 2.5 volumes of 100% ethanol, DNA was transferred into a new tube and washed with 5 ml of 70% ethanol at room temperature (RT) for 5 min and dried at 37 °C. DNA was resuspended in 2 ml TEN20 (10 mM Tris–HCl, pH 7.9, 2 mM EDTA, 20 mM NaCl, 0.1% SDS and 1,000 U RNasin) at 65 °C, boiled for 10–15 min and chilled on ice. For the RNase A negative control, two rounds of RNase A treatment (200 µg/ml) were performed before boiling.

**NS purification was carried out according to the protocol described in ref. 31, except for two main modifications.** The size of the selected fractions ranged from 300 bp to 2 kb, and three rounds of lambda exonuclease digestion were performed to decrease the signal from background DNA and to ensure the specific purification of replication-origin NSs.

**NS amplification and chip data analysis.** Purified NS samples and negative controls were amplified with a WGAII kit (Sigma–Aldrich), with the first step of fragmentation omitted. Amplification products were purified with Qiagen PCR purification columns. Proper unbiased amplification was monitored by RT–qPCR. Hybridization on NimbleGen *C. elegans* 2.1 M microarrays, washing and scanning of microarrays were performed by the NimbleGen Service Laboratory.

**Quantitative real-time PCR analysis.** Real-time PCR analysis of NS samples was performed with SYBR Green PCR master mix (Roche) in a LightCycler 480 real-time PCR thermocycler (Roche). For relative quantification, total genomic DNA dilutions were used to construct the standard curves. All experiments were performed in triplicate.

**Bubble-trap and neutral/neutral 2D gel analyses.** Purification of embryonic DNA was performed as described above. 1 mg of DNA (per 2D gel lane) was digested with Clal (1 U per microgram of DNA) at 37 °C overnight. Digested DNA was subjected to benzoylated naphthoylated DEAE (BND)–cellulose chromatography and then subjected to the bubble-trap method. 100 ng of DNA was analyzed with neutral/neutral 2D agarose gels. The primers used for generating the detection probe of the regions were as follows: for regions containing CGI-like associated replication origins, forward, GTACCATGGCGACCGTAG and reverse, AGACATGCACCTCTGAAGTT (nucleotides 11089575 to 11090412 of chromosome V); for the region containing IRS and CGTAAA-associated replication origins, forward, ACGAGCAGAAACCTCAACCA and reverse, ATCCGGCAATGAACTGGAC (nucleotides 3656327 to 3666062 of chromosome II); for the negative-control forward, ATGGGACATGGCGCTCCTCA and reverse, TTGCTCAAGAAATGGCGCAA (nucleotides 8690511 to 8691255 of chromosome I).

**Bioinformatics analyses.** Description of genomic features. For *C. elegans* genes, the WormBase® genome build WS220 was used. CGI-like sequences were defined as DNA regions of at least 200 bp with a GC content greater than 60%. Inverted repeats were defined with the EMBOSs® package inverted and the parameters used in the WormBase repository with an arm size of up to 10,000 bp. The HOT regions previously identified in ref. 40 and the enhancers from ref. 42 were used. the EMBOSs® package inverted and the parameters used in the WormBase repository with an arm size of up to 10,000 bp. The HOT regions previously identified in ref. 40 and the enhancers from ref. 42 were used.

**Microarray design and correction of cross-hybridization artifacts.** Embryo triplicate samples and negative controls (L1 and RNase A–treated mixed-embryo DNA) were hybridized with NimbleGen *C. elegans* 2.1 M microarrays containing the complete *C. elegans* genome except for repeated sequences.
Despite NimbleGen's efforts to avoid tiling repeated sequences, we noticed that some of the probes matched multiple genome hits. Genomic hits were identified for each probe with BLAST50 and the C. elegans genome WS190 (ce6) with a word size of ten. With a cut-off of 40 bp for the BLAST identity score, the following results were obtained. Among the 2,042,873 array probes, 1,781,108 (87%) had a single genomic hit, 198,437 (10%) had multiple genomic hits (up to 1,767 genomic hits), and 63,328 (3%) did not have any genomic hit. Replication origins were analyzed by using only probes with single genome hits.

**Correlation between biological replicates.** The degree of correlation between biological replicates was evaluated with a scatter plot and subsequent computation of Pearson's correlation coefficient ($R^2$) and Spearman's rank correlation coefficient ($p$).

**Data normalization, smoothening and determination of significant probes.** Experimental (Cy5) and control (Cy3) signal intensities quantified and provided by NimbleGen were converted into log$_2$ ratios (log$_2$(Cy5/Cy3)). The Lowess normalization method was applied to eliminate intensity-dependent variations in dye bias$^61$. According to ref. 15, a sliding median window with a length of 5 oligonucleotide probes was used to smooth the signal. Mode (m) and median absolute deviation (s) of the normalized log$_2$ ratios were computed. Assuming that the normal distribution (specified by m and s) covered the entire background noise (nonsignificant signals) for each probe, one P was computed. Three biologically independent samples for each embryonic stage and one for each control (Li and RNase A–treated embryo DNA) were used. The L1 sample signal (negative control) was subtracted from each replicate independently, the normalized log$_2$ ratios of replicate samples were then combined by averaging the values at the corresponding genomic positions, and the corrected Ps were combined by using a chi-square distribution$^62$. Thus, one probe was denoted as significant if the combined $P$ value was lower than 5% (level of significance).

**Replication-origin definition.** The minimum size of purified NSs was 0.3 kb. Thus, when considering synchronous strand elongation, potential replication origins should be at least 0.6 kb (2 × 0.3 kb for a bidirectional replication origin). We defined replication origins as regions with at least three significant consecutive probes (P < 0.05) in an area containing a minimum of 0.6 kb of positive probes (showing NS enrichment with a log$_2$ ratio >0) on the basis of the scripts used in ref. 15. If two enriched regions were separated by <0.2 kb, they were merged into one. These conditions were used to minimize false-positive events by excluding overly hybridization signals of single probes or small regions, and to score as replication origins peaks that did not overlap with those of a pregastrula origin.

**Sequence-motif analysis.** Detection of oligonucleotides (k-mers) was performed on the basis of their positional biases around peak summits. The occurrences of all k-mers (from 3 to 7 nt) in nonoverlapping windows (100 bp) from −1 kb to +1 kb relative to the origin summits, compared with the random expectation ($P < 10^{-6}$), were scored. All positions were computed according to the Watson (sense) and Crick (antisense) chromosomal strands. K-mer clusters were assembled and used as seeds to extract motifs.

**Comparative analysis of replication origins and genome features.** All data were converted to WS220 (with UCSC Genome Browser's liftOver utility) before analysis. As a negative control, a set of 100 random origins was generated, assembled and used as seeds to extract motifs.

Two different correlation programs were used. The first one counted the overlaps between replication origins and genome features (CGI, inverted repeats, enhancers and HOT regions) at 1 bp minimum; the second one recorded a correlation if the highest peak of the origin coincided with the genome feature (genes, promoters and promotors).

**Analysis of origin and genome feature distribution.** All data were converted to WS220 before analysis. Annotomic genomic inverted repeats were composed of two almost complementary arms and a central loop. For the distribution of epigenetic marks and RNA polymerase II, the signal mean was calculated along the region (0 to 5,000 bp) from the origin highest peak, every 50 or 500 bp, depending on the available data for that mark. These analyses were performed by distinguishing chromosome arms and centers. Data used for distribution around the origin highest peak were obtained from: H3 (modENCODE_2312), RNA polymerase II (GSE50334), H3K4me3 (GSE49739), H3K4me1 (modENCODE_2726), H3K9me3 (GSE49207), H3K27ac (modENCODE_3200), H3K27me3 (modENCODE_3171), H3K36me3 (modENCODE_909 and modENCODE_973), H3K18ac (modENCODE_5152), HTZ-1 (modENCODE_43) and LEM-2 (modENCODE_2729).

The distribution of replication origins around operon TSSs was measured by counting the gene nucleotides that were overlapped by an origin, with a variable distance between the nucleotide of interest and the operon TSS. The result was expressed as the percentage of operons in which a nucleotide was overlapped by an origin.

In Figure 5, the distribution of chromatin marks and RNA polymerase II around postgastrula origins was measured on the origins when their summits were overlapped by H3K4me1 and were within the promoter regions (1 kb upstream from the TSSs) of noncoding or coding nonopener genes, or coding opener genes. These analyses were performed by also distinguishing chromosome arms and centers.

**Analysis of conserved replication origins.** The coverage of the NS log$_2$ ratio from pregastrula and mixed-embryo NS signals was calculated. The ratios between covered regions were divided into three levels according to their absolute values. Within each level, the number of replication origins was identical to the number obtained when the coverage of the NS log$_2$ ratio from mixed embryos was >0 (1,808).

** Transcript-level data, RNA polymerase II and epigenetic marks.** Pregastrula and postgastrula transcript-level data were obtained from L. Hillier and R. Waterston (from modENCODE_6561 and modENCODE_6563). Overlaps of replication origins, gene promoters and RNA polymerase II (GSE50334) were analyzed by aligning 1 kb of sequence upstream of TSS regions with RNA pol II peaks and replication origins. Box plots were generated by taking into account all transcript levels. Transcript-level significance was determined with the t tests.

**Code availability.** The computer code used to generate the results presented in this manuscript is available at: http://www.igh.cnrs.fr/equip/Seitz/Scripts_Rodriguez_Martinez_et_al.tar.

**Data availability.** The data sets generated and analyzed during the current study have been deposited in the NCBI Gene Expression Omnibus database under accession code GSE86651. Source data for Figures 1–6 and Supplementary Figures 2–6 are available online.