

# ORC1 binds to *cis*-transcribed RNAs for efficient activation of replication origins

Received: 21 April 2023

Accepted: 11 July 2023

Published online: 24 July 2023

Check for updates

Aina Maria Mas <sup>1,2</sup>, Enrique Goñi <sup>1,2,13</sup>, Igor Ruiz de los Mozos <sup>1,2,13</sup>, Aida Arcas <sup>1,2,13</sup>, Luisa Statello <sup>1,2</sup>, Jovanna González<sup>1,2</sup>, Lorea Blázquez<sup>3,11,12</sup>, Wei Ting Chelsea Lee<sup>4</sup>, Dipika Gupta <sup>4</sup>, Álvaro Sejas<sup>1,2</sup>, Shoko Hoshina<sup>5</sup>, Alexandros Armaos<sup>6</sup>, Gian Gaetano Tartaglia<sup>6,7,8,9</sup>, Shou Waga <sup>5</sup>, Jernej Ule <sup>3</sup>, Eli Rothenberg<sup>4</sup>, María Gómez <sup>10</sup> & Maite Huarte <sup>1,2</sup> ✉

Cells must coordinate the activation of thousands of replication origins dispersed throughout their genome. Active transcription is known to favor the formation of mammalian origins, although the role that RNA plays in this process remains unclear. We show that the ORC1 subunit of the human Origin Recognition Complex interacts with RNAs transcribed from genes with origins in their transcription start sites (TSSs), displaying a positive correlation between RNA binding and origin activity. RNA depletion, or the use of ORC1 RNA-binding mutant, result in inefficient activation of proximal origins, linked to impaired ORC1 chromatin release. ORC1 RNA binding activity resides in its intrinsically disordered region, involved in intra- and inter-molecular interactions, regulation by phosphorylation, and phase-separation. We show that RNA binding favors ORC1 chromatin release, by regulating its phosphorylation and subsequent degradation. Our results unveil a non-coding function of RNA as a dynamic component of the chromatin, orchestrating the activation of replication origins.

The initiation of DNA replication involves a sequential assembly and disassembly of protein complexes on genomic DNA, which is tightly controlled along the cell cycle. Initiation occurs at specific sites throughout the genome where the Origin Recognition Complex (ORC) associates in M/G1. ORC is composed of six subunits (ORC1-6), of which ORC1 is the pioneering subunit in the binding to the chromatin. This binding recruits other members of the complex, followed by additional initiation factors (CDC6, CDT1,

MCM helicases, CDC45, and CDC7) in a sequential manner, leading to, (i) licensing and (ii) firing of replication origins<sup>1</sup>. Linked to the firing in the S phase, some components of the initiation complex are disassembled and targeted for degradation, which is followed by the activation of DNA helicases and loading of replication factors, avoiding DNA re-duplication events in the same cycle<sup>2,3</sup>. ORC1 stands out among ORC components in its distinct regulation throughout the cell cycle, which is consistent with its crucial

<sup>1</sup>Center for Applied Medical Research, University of Navarra, Pio XII 55 Ave, 31008 Pamplona, Spain. <sup>2</sup>Institute of Health Research of Navarra (IdiSNA), Cancer Center Clínica Universidad de Navarra (CCUN), Pamplona, Spain. <sup>3</sup>RNA Networks Lab, The Francis Crick Institute, NW11BF London, UK. <sup>4</sup>Department of Biochemistry and Molecular Pharmacology, Perlmutter Cancer Center, New York University School of Medicine, New York, NY 10016, USA. <sup>5</sup>Department of Chemical and Biological Sciences, Japan Women's University, Tokyo 112-8681, Japan. <sup>6</sup>Center for Human Technologies, Istituto Italiano di Tecnologia, Genova, Italy. <sup>7</sup>Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>8</sup>Department of Biology 'Charles Darwin', Sapienza University of Rome, Rome, Italy. <sup>9</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. <sup>10</sup>Centro de Biología Molecular Severo Ochoa (CBMSO), Consejo Superior de Investigaciones Científicas/Universidad Autónoma de Madrid (CSIC/UAM), Nicolás Cabrera 1, 28049 Madrid, Spain. <sup>11</sup>Present address: Neurosciences Area, Biodonostia Health Research Institute, 20014 San Sebastian, Spain. <sup>12</sup>Present address: Ikerbasque, Basque Foundation for Science, 48009 Bilbao, Spain. <sup>13</sup>These authors contributed equally: Enrique Goñi, Igor Ruiz de los Mozos, Aida Arcas. ✉e-mail: [maitehuarte@unav.es](mailto:maitehuarte@unav.es)

function in the initiation of replication and the maintenance of undamaged cell propagation<sup>2,4</sup>.

In *S. cerevisiae*, the position of replication origins is defined by ORC recognition of DNA sequence-dependent elements<sup>5</sup>. In contrast, how origins are positioned in mammalian genomes is still an outstanding question, since ORC does not recognize a known consensus DNA sequence<sup>6</sup>. Furthermore, only ~20% of licensed origins in a given cell fire in the S phase<sup>7</sup>, but how this activation is regulated is not fully understood. Multiple chromatin features are known to influence the flexible selection and activation of origins, which, in combination, dictate the probability of stochastic origin activation. These include chromatin accessibility, specific histone marks such as H4K20me<sup>28,9</sup> and H2AZ<sup>10</sup>, and the presence of DNA sequences prone to form G-quadruplexes (G4)<sup>11–13</sup>. While the simultaneous replication and transcription of a precise DNA position are strictly incompatible, the most active origins are localized at transcription start sites (TSSs), with origin activity correlating with the level of gene expression<sup>14–20</sup>. Thus, in the entry of the S phase, when early replication origins are fired, RNA is produced in close proximity. This raises the possibility that RNAs could influence origin selection or activation. While previous reports have pointed to specific roles for RNA in replication initiation in *X. laevis* shortly after fertilization<sup>21</sup> and at Epstein Barr virus OriP<sup>22</sup>, whether transcribed RNA plays a general role in the activation of mammalian origins yet remains unknown.

Here, we show that ORC1 interacts with RNAs transcribed at active origins, which is linked to ORC1 dynamic association with the chromatin and has an impact on replication origin activity.

## Results

### ORC1 interacts with RNA in vivo

We hypothesized that ORC1, the first subunit of the initiation complex associated with the chromatin, binds to RNA in cells. In vitro RNA binding had been described for human ORC1, mapping to a region between amino acids 413 and 511<sup>23,24</sup>, part of an Intrinsically Disordered Region (IDR) and separated from other domains mediating, among other functions, its binding to nucleosomes<sup>4,25</sup> (Fig. 1a and Supplementary Fig. 1a). In line with this hypothesis, stochastic optical reconstruction microscopy (STORM) on the chromatin fraction of G1-synchronized cells detected a significant cross-correlation between ORC1 and EU-labeled RNA, when compared to randomized images, and not detected in the experimental negative control (Fig. 1b and Supplementary Fig. 1b). These results indicate that ORC1 is in very close proximity to RNA in vivo.

We then set out to identify RNAs bound by ORC1 by applying complementary approaches (Fig. 1c). First, we performed native ORC1 RNA immunoprecipitation coupled to sequencing (RIP-seq) of total RNA (ribo-depleted polyA+ and polyA- RNA) from nuclei of dividing HCT116 cells (Supplementary Fig. 2a, b), using an anti-ORC1 antibody. These experiments identified 2203 RNAs enriched in ORC1 RIP relative to the input samples of nuclear RNA (Log<sub>2</sub> fold change >1, *p* value <0.05), while the control IgG did not recover detectable amounts of RNA. To confirm the capacity of immunoprecipitated ORC1 to retrieve RNA, we also performed anti-Flag RIP-seq from nuclear extracts of HCT116 cells transiently expressing ORC1-3xFlag. Most of the identified transcripts (68%) overlapped with the ones interacting with the endogenous ORC1 (Hypergeometric test, *p* value 1.89e-191) (Supplementary Data 1). Next, theoretical binding predictions were calculated to examine the plausibility of direct interaction between ORC1 and the co-immunoprecipitated RNAs. *catRAPID* algorithm, which computes an interaction probability based on the biophysical characteristics of proteins and RNAs<sup>26</sup>, showed that enriched RNAs by RIP also present higher predicted binding compared to the depleted ones (*p* value <2.2 e<sup>-16</sup>). Moreover, it showed a correlation between the theoretical binding score and experimental fold changes in RIP-seq (Supplementary Fig. 2c, d), thus suggesting that

direct binding occurs between the experimentally-defined set of RNAs and ORC1.

We therefore set to detect the direct RNA-protein interactions with nucleotide resolution by applying the iCLIP protocol<sup>27</sup>. HCT116 cells expressing ORC1-3xFlag were UV-crosslinked prior to nuclear isolation and anti-Flag immunoprecipitation. Affinity-purified complexes were partially RNase digested, and protected RNAs at and above ORC1-expected molecular weight were extracted, sequenced, and compared with the negative control (i.e., anti-Flag IP in cells that do not express ORC1-3xFlag), which only recovered neglectable RNA amounts (Fig. 1c, Supplementary Fig. 2e, and Supplementary Table 1). Read alignment to the human genome identified the vast majority of ORC1 crosslink RNA-binding sites (>95%) covering genic regions on the same direction of transcription. While ORC1-RNA crosslinks peaked at the 5' of genes, they mapped to both exonic and intronic regions, indicating binding to nascent transcripts (Fig. 1d and Supplementary Fig. 2f).

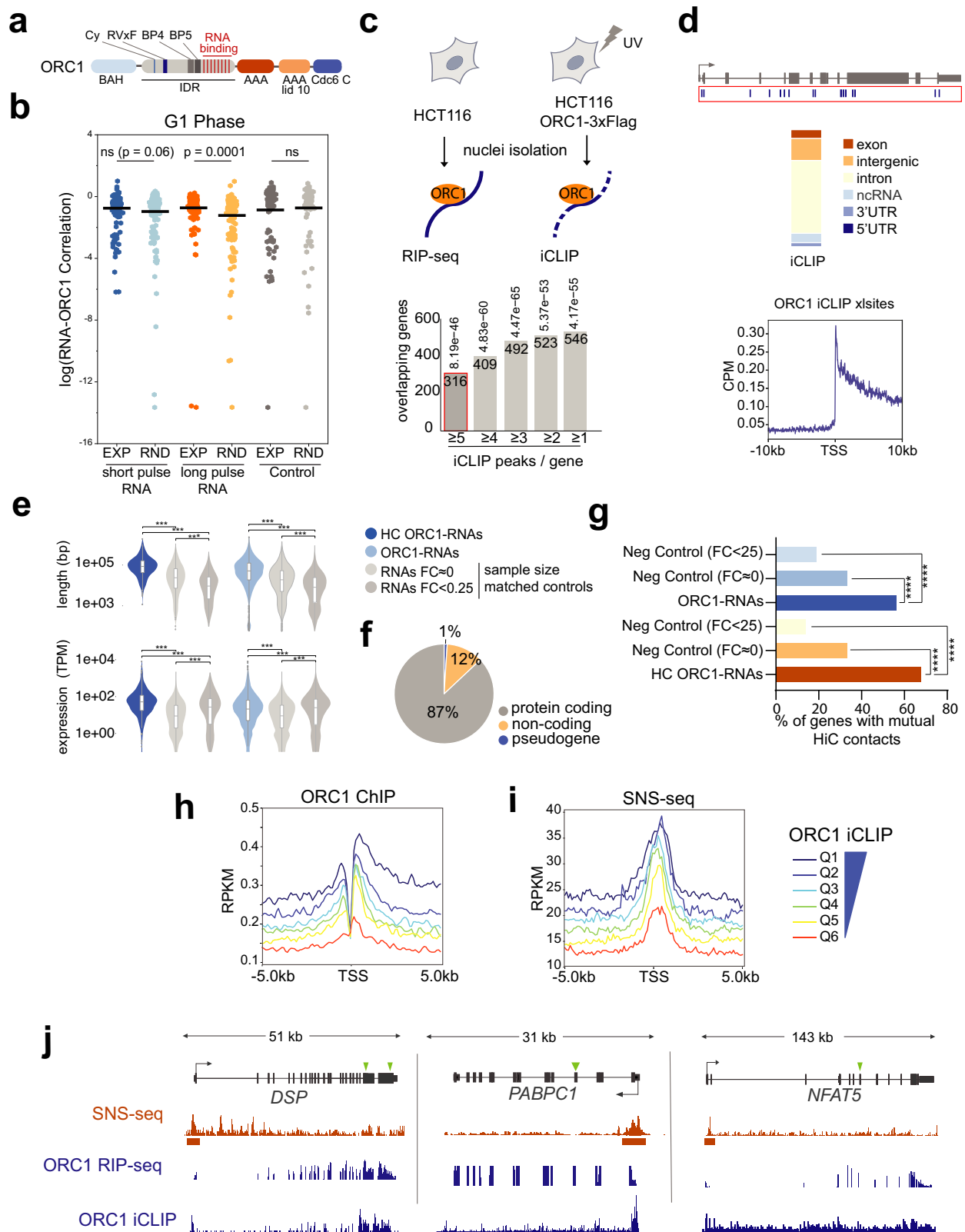
Together, these results demonstrate that ORC1 interacts with RNA in the nucleus of living cells, which could shape protein functionality along the cell cycle.

### ORC1 binds GAA-rich and highly transcribed RNAs, sequence-independently

Once established that ORC1 interacts with RNA in the nucleus of cells, we next explored whether it prefers to associate through specific RNA motifs. To do that, we searched for motifs in 200-nt windows centered around the iCLIP peaks defined with iCOUNT peak caller. This analysis identified the snoRNA C box UGAUGA motif (*e*-value 6.0e-246) (Supplementary Fig. 2g), in line with the presence of a high number of crosslinks (35%) corresponding to highly expressed snoRNAs, although only mapping to 6% of the genes of ORC1-bound RNAs (Supplementary Data 2). Consistently, when snoRNAs were filtered out, no motif was found enriched, indicating that the binding of ORC1 to the majority of RNAs (94% of genes) is not mediated through a well-defined sequence. Since ORC1 had been reported to preferentially bind to G-quadruplex RNA structures<sup>24,28</sup>, we also performed G4 predictions around iCLIP peaks, which found a mild enrichment of RNA sequences prone to form this type of secondary structures (Supplementary Fig. 2h). These analyses suggest that ORC1 does not bind to RNA through a specific sequence, although it shows some preference toward structured RNA elements.

To understand the nature of ORC1 RNA interactome, we compared the results of ORC1 RIP-seq and iCLIP experiments. Interestingly, RIP-seq and iCLIP showed good agreement, since ORC1 RNA interactors identified by iCLIP also showed higher fold enrichments by RIP (Supplementary Fig. 2i), even when RIP interactors are defined relative to their input, and therefore not biased by expression level. Moreover, by astringently selecting RNAs containing 5 or more iCLIP peaks (1887 genes), a significant overlap (Hypergeometric test, *p* value 8.19e-46) was found with the RNAs identified by RIP-seq (Fig. 1c), representing high confidence RNAs directly interacting with ORC1 (HC ORC1-RNAs) (Supplementary Data 2). When the selection was extended to genes with more than 4, 3, or 2 iCLIP peaks, the overlap was increased and also highly significant (Fig. 1c). In addition, several RNA interactors detected by RIP-seq and/or iCLIP were validated in independent RIP and CLIP experiments, with no enrichment detected for the unspecific IgG (Supplementary Fig. 2j, k), supporting the validity of the used complementary analyses (Supplementary Fig. 2l).

We then looked at the characteristics of both the overlapping and the broader set of identified RNAs. HC ORC1-RNAs, as well as the larger set of ORC1-RNAs (Supplementary Data 2), are mostly mRNAs (95% and 87%), with higher expression levels and length than negative control genes (Fig. 1e, f). Similar features were found when assessing RIP-RNAs or iCLIP-RNAs separately (Supplementary Fig. 3a, b), confirming that ORC1 preferentially binds to RNAs of these characteristics in physiological conditions. According to Repli-seq and Hi-C analyses, the

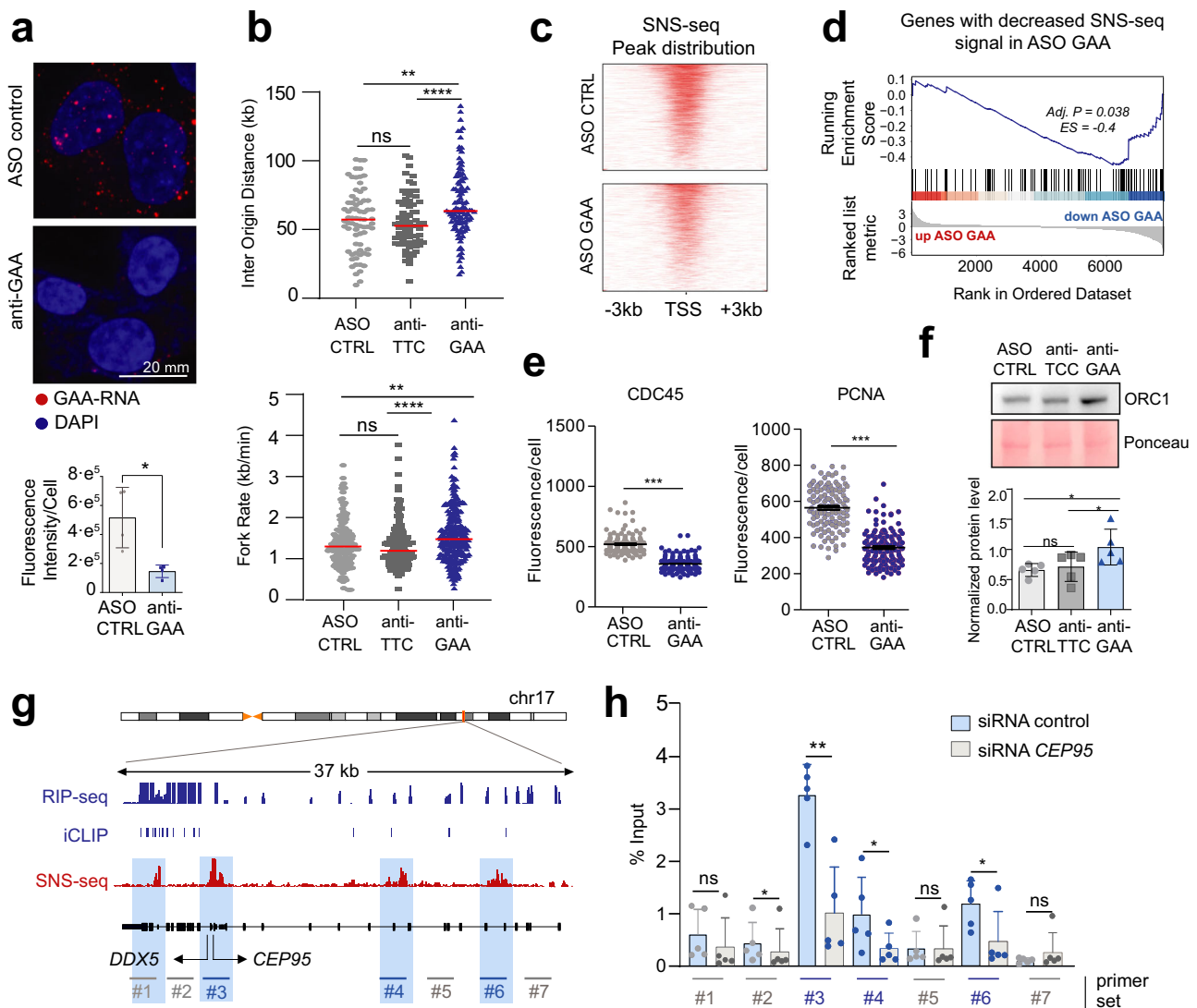


majority of genes encoding for HC ORC1-RNAs and ORC1-RNAs are localized within the early-replicating regions of the genome (80%,  $p$  value  $4e^{-253}$  and 66.47%,  $p$  value  $2e^{-164}$ ), and interact with each other in the 3D nucleus with higher frequency ( $p$  value  $<2.2e^{-16}$  for both sets, Fig. 1g). Interestingly, sequence analyses performed on full-length HC ORC1-RNAs showed the enrichment of tandem GAA repeats ( $e$ -value  $7.5e^{-11}$ , Supplementary Fig. 3c), known to be linked to nuclear retention

of certain mRNAs<sup>29</sup>. The sequences were also highly enriched in RNAs detected by RIP-seq of endogenous ORC1 and ORC1-3xFlag, with 82 and 94% of the RNAs containing this type of sequence, respectively (Supplementary Fig. 3d, e). However, the position of iCLIP peaks indicates that ORC1 does not bind RNA through it, as confirmed by the similar EMSA behavior of ORC1-RNAs regardless of the presence of this sequence (Supplementary Fig. 3f–h). We then concluded that the







**Fig. 2 | RNAs bound by ORC1 are necessary for optimal origin activation.** **a** RNA-FISH representative images, and signal quantification (below), of RNAs containing GAA repeats, in ASO-transfected HCT116 cells. Dots represent mean values ( $n = 4$  biologically independent samples)  $\pm$ SEM. \* denotes  $p$  value  $< 0.05$ , derived from unpaired two-tailed Student's  $t$ -test. **b** DNA fiber quantification of inter-origin distances and fork rates of ASO-transfected HCT116 cells. Red lines indicate the median. ns denotes  $p$  value  $> 0.05$ , \*\* denotes  $p$  value  $< 0.01$ , \*\*\*\* denotes  $p$  value  $< 0.0001$ , derived from unpaired two-tailed Mann-Whitney  $t$ -test. **c** SNS-seq peak count frequency and distribution relative to TSS positions ( $\pm 3$  kb) in ASO control and anti-GAA treated HCT116 cells. **d** GSEA showing the reduction of SNS-seq signal (peaks enriched in the control condition) in anti-GAA downregulated genes. Statistical significance (adjusted  $p$  value 0.038) of the enrichment score (ES) derives from a permutation test. **e** CDC45 and PCNA chromatin immunofluorescences per

cell (HCT116) upon ASO knockdown, after soluble protein washout. Data were presented as mean values ( $n > 100$  cells)  $\pm$ SEM. \*\*\* denotes  $p$  value  $< 0.001$ , derived from unpaired two-tailed Mann-Whitney  $t$ -test. **f** ORC1 western blot and protein quantification, in chromatin extracts of ASO-transfected HCT116 cells. Data were presented as mean values ( $n = 5$  biologically independent experiments)  $\pm$ SEM. ns denotes  $p$  value  $> 0.05$ , \* denotes  $p$  value  $< 0.05$ , derived from paired two-tailed Student's  $t$ -test. **g** Browser snapshot at DDX5-CEP95 locus showing ORC1 RIP-seq enrichment, ORC1 iCLIP peaks, and SNS-seq reads in HCT116 cells, with the position of qPCR primers (#) indicated, and origins highlighted in blue. **h** Enrichment of nascent strands determined by SNS-qPCR at genomic positions indicated in Fig. 2g, in siRNA-treated HCT116 cells. Data were presented as mean values ( $n = 5$  biologically independent experiments)  $\pm$ SEM. ns denotes  $p$  value  $> 0.05$ , \* denotes  $p$  value  $< 0.05$ , \*\* denotes  $p$  value  $< 0.01$ , derived from paired two-tailed Student's  $t$ -test.

origin activation was also observed by CHIP of CDC45 and PCNA, showing decreased chromatin enrichments at local origins while distant origins were unaffected (Supplementary Fig. 6d–g). In contrast, *CEP95* overexpression from a plasmid (i.e. in trans) did not cause changes in origin activity (Supplementary Fig. 6d–g), in agreement with a model where transcribing RNAs would be regulating the activity of replication origins found at their site of production. Analogous initiation defects *in cis* were observed when the HC ORC1 RNA *HSP90AA1* was individually depleted (Supplementary Fig. 6a–c). The decrease of local origin activity, visualized in the enrichment of nascent strands, was evident, while no consistent effect was appreciated on distant origins at non-related loci (Supplementary Fig. 6h–j).

### ORC1 RNA binding mutant is impaired in origin activation

To determine how the capacity to bind to RNA is relevant to ORC1 function, we next studied an RNA binding mutant (Fig. 3a). Based on a previous work<sup>24</sup>, and as evidenced by RNA electrophoretic mobility shift assays (Fig. 3b and Supplementary Fig. 7a, b), the substitution to Alanines of three Arginines inside the ORC1 RNA binding region (R441A, R444A and R465A) (Fig. 3a) results in the loss of *in vitro* RNA binding. The impaired recognition of ORC1-RNAs in cells by the mutant protein was also predicted by *Clever Suite* and *catRAPID* omics *v2* (Supplementary Fig. 7c, d), while it did not predict affection of ORC1 ability to bind to DNA (Supplementary Fig. 7c). Importantly, the absence of UV-crosslinked RNA in iCLIP experiments (Supplementary

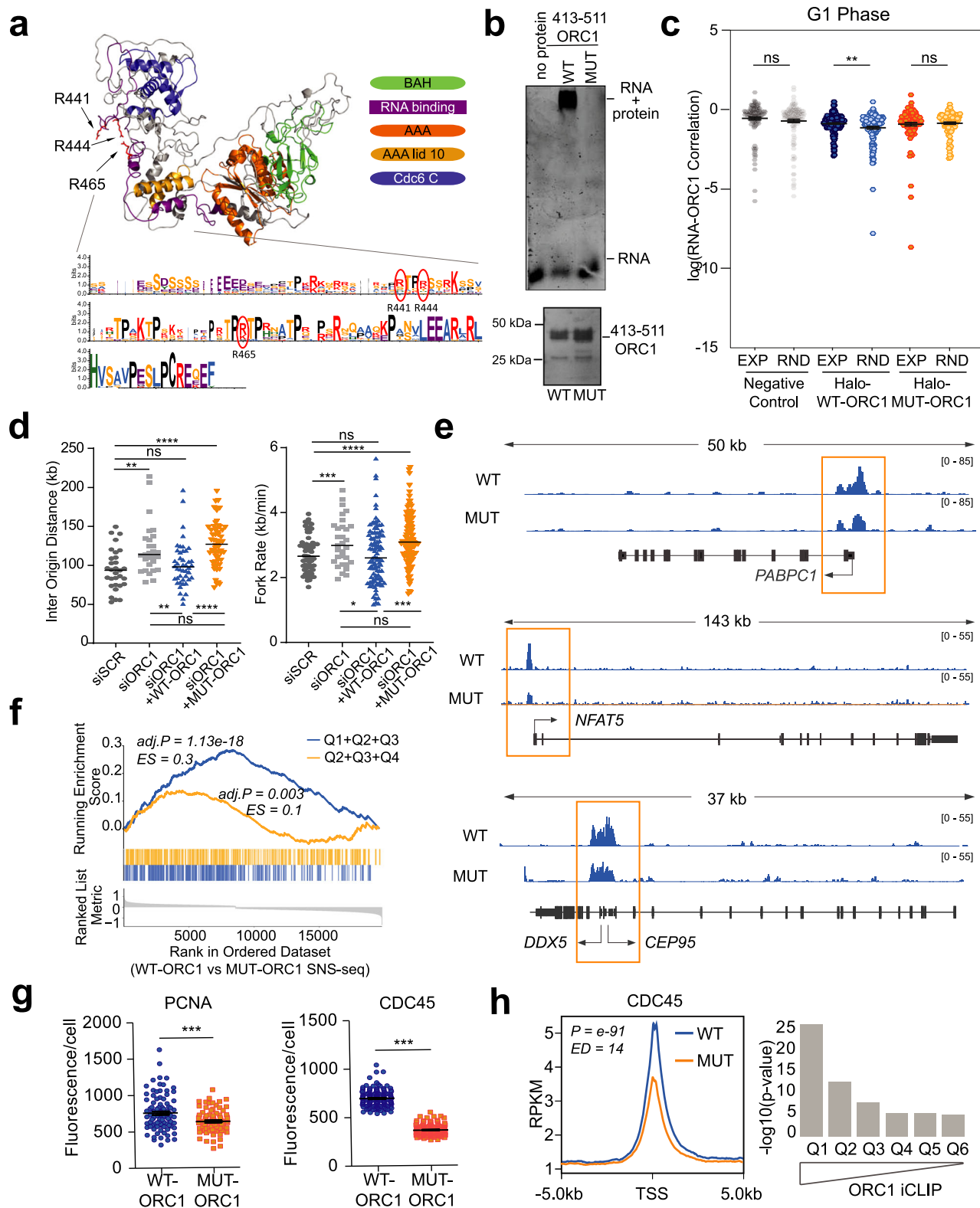


Fig. 7e) and the decreased colocalization with RNA by STORM (Fig. 3c) showed experimentally the lack of RNA binding by the mutant ORC1 in cells.

We next investigated whether the usage of ORC1 RNA binding mutant could affect origin firing. Cells depleted of the endogenous ORC1 and transfected with plasmids expressing WT or MUT-ORC1 were subjected to fiber assays. As expected, depletion of the endogenous ORC1 resulted in the decreased number of fired origins,

reflected by increased distances between origins<sup>35</sup> and compensatory increased fork speed (Fig. 3d and Supplementary Fig. 7f, g). While WT-ORC1 rescued the normal firing and fork rates, MUT-ORC1 did not (Fig. 3d), suggesting that the RNA-binding activity of ORC1 is needed for optimal DNA replication initiation. Notably, SNS-seq showed decreased origin activity in cells stably expressing MUT-ORC1, and preferentially affecting origins at genes that produce the ORC1-RNAs with a stronger level of binding to ORC1 as determined by iCLIP (GSEA

**Fig. 3 | ORC1 RNA-binding mutant is impaired in origin activation.** **a** 3D model of human ORC1 showing domains in colors, and residues R441, R444, and R465 (involved in RNA-binding) in red. Below, the vertebrate consensus of ORC1 RNA-binding region, circles indicating mutated residues in MUT-ORC1. **b** RNA staining of EMSA assays, with GST-tagged purified WT and MUT-ORC1 (amino acids 413–511) (2.5  $\mu$ M) incubated with fragmented cellular RNA (2.5  $\mu$ M). Below, the silver staining of proteins used in the assay. **c** Cross-correlation between ORC1 and EU-labeled RNA (long pulse) in G1-synchronized U2OS cells, untransfected or transfected with Halo-tagged WT and MUT-ORC1, comparing STORM experimental (EXP) and randomized (RND) samples. Data were presented as mean values ( $n > 50$  cells)  $\pm$  SEM. ns denotes  $p$  value  $> 0.05$ , \*\* denotes  $p$  value  $< 0.01$ , derived from unpaired two-sample  $t$ -test. **d** DNA fiber quantification of inter-origin distances and fork rates in HCT116 cells transfected with the indicated siRNAs,  $\pm$  plasmids expressing Flag-tagged WT or MUT-ORC1. Black lines indicate the median. ns denotes  $p$  value  $> 0.05$ ,

\* denotes  $p$  value  $< 0.05$ , \*\* denotes  $p$  value  $< 0.01$ , \*\*\*\* denotes  $p$  value  $< 0.0001$ , derived from unpaired two-tailed Mann–Whitney  $t$ -test. **e** Browser snapshot at ORC1-RNA *PABPC1*, *NFATS*, and *DDX5-CEP95* loci, showing SNS-seq normalized signal of HCT116 cells stably expressing WT or MUT-ORC1. **f** GSEA showing enrichment of ORC1-RNAs in merged iCLIP-defined quantiles (Q), toward ranked genes according to their WT vs MUT (log<sub>2</sub> fold change) SNS-seq coverage at TSSs. Statistical significance (adjusted  $p$  value  $1.13\text{e-}18$  or  $0.003$ ) of the enrichment scores (ES) were calculated by permutation tests. **g** CDC45 and PCNA chromatin immunofluorescences per cell, in HCT116 cells stably expressing WT or MUT-ORC1, after soluble protein washout. Data were presented as mean values ( $n > 100$  cells)  $\pm$  SEM. \*\*\* denotes  $p$  value  $< 0.001$ , derived from unpaired two-tailed Mann–Whitney  $t$ -test. **h** Coverage plot of CDC45 ChIP-seq data at TSSs, in WT or MUT-ORC1 HCT116 stably expressing cells, and two-tailed  $t$ -test statistical results between their coverage at TSSs of ORC1 iCLIP-defined gene quantiles (Q).

Adj.  $p$  value =  $1.13\text{e-}18$  for genes within Q1–Q3 iCLIP quantiles vs  $0.003$  for Q4–Q6, and Enrichment scores of  $0.3$  and  $0.1$  respectively, Fig. 4e, f and Supplementary Fig. 7h). These results support the idea that ORC1 binding to RNA has a *cis*-regulatory effect on replication origins, as observed upon global (Fig. 2d and Supplementary Fig. 5a–c) or individual (Fig. 2h and Supplementary Fig. 6) RNA knockdown. Also in line with fewer origin activation events observed in GAA-RNA-depleted cells (Fig. 2e and Supplementary Fig. 5d), PCNA and CDC45 were significantly reduced at the chromatin in cells expressing ORC1 RNA-binding mutant (Fig. 3g and Supplementary Fig. 7i), also revealed by ChIP-seq signals, primarily decreased at TSSs of transcribing RNAs bound by ORC1 (Fig. 3h).

We then concluded that the use of ORC1 RNA-binding mutant recapitulates the origin activation defects observed upon knockdown of RNA interactors of ORC1, suggesting that the RNA-binding activity of ORC1 is important for normal initiation of DNA replication.

### RNA binding facilitates ORC1 phosphorylation and chromatin release

To better understand how RNA-binding could be important for ORC1 function, we further investigated the behavior of the RNA-binding mutant by analyzing its chromatin association. Notably, MUT-ORC1 was more associated with the chromatin than its wild-type counterpart (Supplementary Fig. 8a), as was the endogenous ORC1 when ORC1 RNA interactors were depleted (Fig. 2f and Supplementary Fig. 5e). Indeed, MUT-ORC1 presented longer half-life than WT-ORC1, as observed upon translational inhibition with cycloheximide treatment, while both proteins responded to proteasomal inhibition at a similar rate (Fig. 4a), suggesting that the RNA binding activity of ORC1 decreases its stability. Importantly, while the knockdown of GAA-RNAs also led to increased levels of WT-ORC1 on chromatin, it had no effect on the levels of MUT-ORC1 (Supplementary Fig. 8b), indicating that this phenotype is RNA-dependent.

Since origin licensing and firing is governed by the temporal chromatin association and release of ORC along the cell cycle, we decided to dissect this phenotype in synchronized cells. Even though both WT- and MUT-ORC1-expressing cells progressed through S phase at similar rates (Supplementary Fig. 8c), they showed striking differences in ORC1 chromatin association dynamics (Fig. 4b). WT-ORC1, but not MUT-ORC1, was released from the chromatin at the entry of S phase, suggesting that the observed persistence of MUT-ORC1 on chromatin was due to an inefficient protein release. Importantly, endogenous ORC1 also showed delayed chromatin release when GAA-RNAs were depleted (Fig. 4c). Similar effects were observed for WT-ORC1 when synchronized cells were permeabilized and treated with RNase A, resulting in an enhanced chromatin association along the S phase that is progressively lost with no RNase A treatment (Fig. 4d and Supplementary Fig. 8d). In contrast, increased MUT-ORC1 chromatin persistence was observed regardless of RNA degradation (Fig. 4d and Supplementary Fig. 8d). Of note, the delay in MUT-ORC1 release was

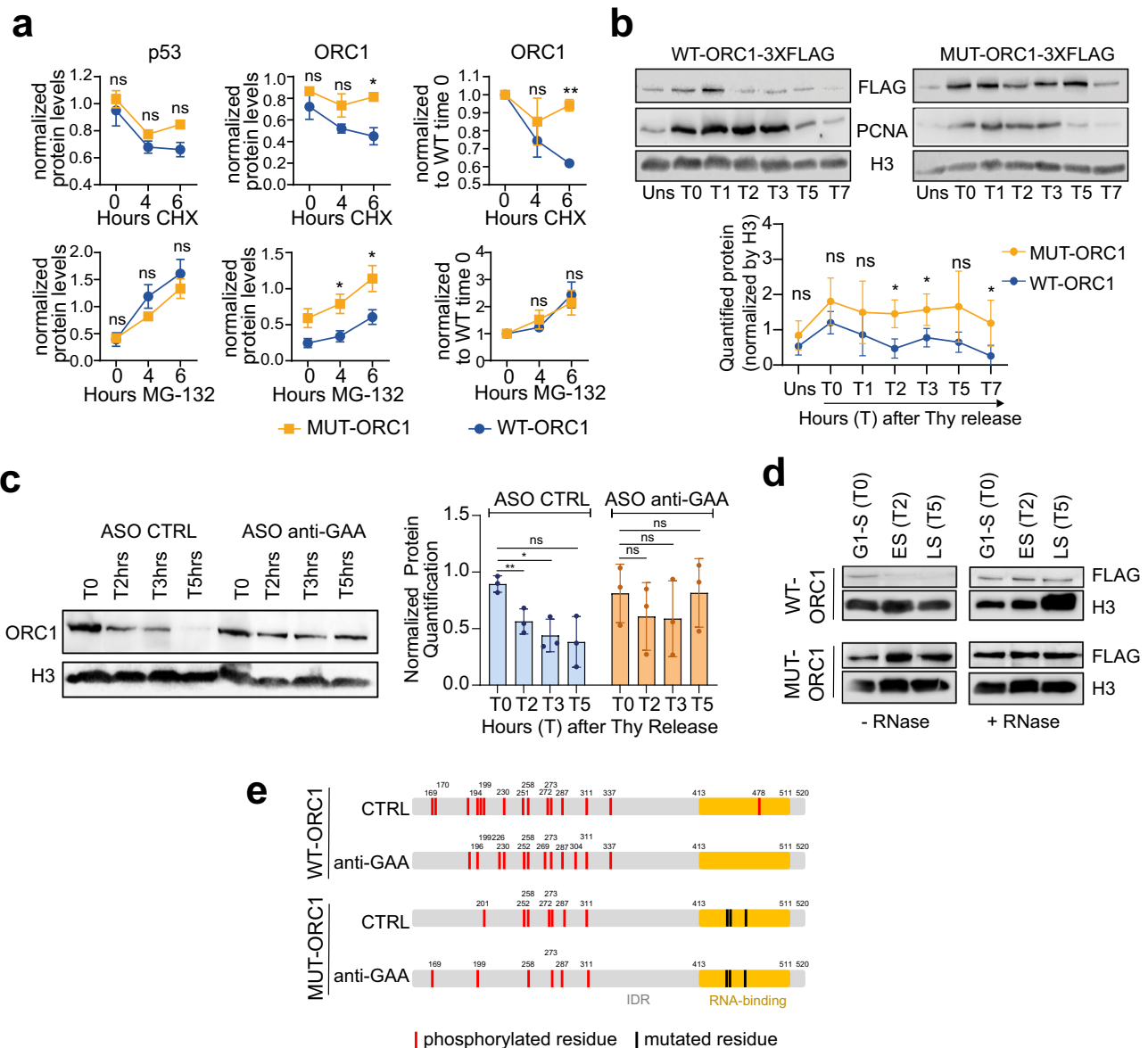
associated with reduced levels of PCNA loaded on chromatin (Fig. 4b), indicating that the RNA-dependent release of ORC1 could be linked to origin activation.

Given that both origin firing and ORC1 stability are governed by phosphorylation<sup>1</sup>, we hypothesized that the phenotypes observed may account for an RNA-dependent modulation of ORC1 phosphorylation. In line with this hypothesis, MUT-ORC1 showed lower mobility in gels, as did the endogenous ORC1 upon the knockdown of GAA-RNA (Supplementary Fig. 8e). Phosphoproteomic analysis confirmed the hypo-phosphorylation of MUT-ORC1 (Supplementary Data 3), particularly affecting its IDR, where phosphorylated sites are known to regulate ORC1 chromatin release and proteasomal degradation<sup>2</sup> (Fig. 4e and Supplementary Fig. 8f, g). Decreased phosphorylation of WT-ORC1 was also observed upon GAA-RNA knockdown, while total phosphorylation levels of ORC1 RNA-binding mutant, as well as the number of phosphorylated residues, were less affected by the transient downregulation of ORC1 RNA interactors (Fig. 4e and Supplementary Fig. 8g), indicating that the interaction with RNA plays a role in regulating ORC1 phosphorylation, and subsequent chromatin release and degradation in S phase. Our results suggest a reciprocal regulation of RNA binding and phosphorylation at the IDR of ORC1, which is linked to its chromatin release, and altogether demonstrate that RNA binding dynamically modulates ORC1 function.

### Discussion

Here, through orthogonal approaches, we show that RNA, the product of transcription, is needed for optimal origin activation mediated through the intrinsic RNA binding capacity of ORC1. We demonstrate that RNA transcribed in the proximity of the origins, preferentially found at TSS regions, influences the efficiency at which they are activated. In line with this model, we find an enrichment of ORC1 iCLIP crosslinks at the 5' region of the RNAs, although not exclusively restricted to it. This is not unexpected given the described mechanism, and the fact that RNA transcripts remain close to the chromatin while co-transcriptionally processed. Moreover, actively transcribed genes are known to form loops, maintaining 3D proximity between the 5' and 3' ends<sup>36</sup>. Specifically, we found that the binding of RNA by ORC1 favors the efficient release of the protein at the entry of the S phase. As suggested in previous studies<sup>3</sup>, our data indicates that ORC1 release could be linked to origin firing, although the precise sequence of molecular events at this critical step will require further studies to be fully understood.

Of note, we failed to detect ORC1 RNA binding at a fraction of active origins. While this could be due to limited sensitivity, the RNA-dependent regulation of origins may be restricted to the genes with active transcription, known to replicate in the early S phase, where we find the genes producing ORC1-RNAs. Origins found at transcriptional deserts, associated with late replication, might be activated through a mechanism not involving RNA interaction with ORC1. On the other hand, ORC1 might not be the only sensor of RNA at origins from



**Fig. 4 | RNA regulates ORC1 chromatin release.** **a** p53 and ORC1-3xFlag protein quantification from western blots with total extracts of HCT116 cells, transfected with WT or MUT-ORC1, and treated with cycloheximide (CHX) or MG-132. Dots represent mean values ( $n = 3$  biologically independent experiments)  $\pm$  SEM. ns denotes  $p$  value  $>0.05$ , \* denotes  $p$  value  $<0.05$ , \*\* denotes  $p$  value  $<0.01$ , derived from paired two-tailed Student's  $t$ -test. **b** Western blot on chromatin extracts of HCT116 cells, transfected with Flag-tagged WT-ORC1 and MUT-ORC1, unsynchronized (Uns) or synchronized in G1/S and released at different times (T as in Supplementary Fig. 8c). Below, normalized protein quantification. Dots represent mean values ( $n = 4$  biologically independent experiments)  $\pm$  SEM. ns denotes  $p$  value  $>0.05$ , \* denotes  $p$  value  $<0.05$ , derived from paired two-tailed Student's  $t$ -test. **c** Western blot and quantification of endogenous ORC1 on chromatin in

different stages of the cell cycle (T as in Supplementary Fig. 8c), upon depletion of GAA-RNAs (ASO anti-GAA) or control conditions (ASO CTRL). Bars represent mean values ( $n = 3$  biologically independent experiments)  $\pm$  SEM. ns denotes  $p$  value  $>0.05$ , \* denotes  $p$  value  $<0.05$ , \*\* denotes  $p$  value  $<0.01$ , derived from paired two-tailed Student's  $t$ -test. **d** Western blot showing the effect of RNase A treatment on WT and MUT-ORC1 chromatin association, along the cell cycle of synchronized cells (T as in Supplementary Fig. 8c). Quantification of independent biological replicates ( $n = 4$ ) is shown in Supplementary Fig. 8d. **e** Representation of the IDR in WT and MUT-ORC1, showing RNA-binding regions (orange), and the discrete positions of RNA-binding mutations (black) and phosphorylated residues (red) detected by mass spectrometry, in control or GAA-knockdown conditions. Source data are provided as a Source Data file.

transcriptionally active regions. RNA may interact with additional proteins in the initiation complex. Among them, CDC6 and CDT1 possess positively charged IDRs<sup>37</sup> with repetitive sites similar to the observed in ORC1 IDR, although their interaction with RNA has not been studied.

Short linear protein motifs (SLiMs) inside ORC1 IDR have been implicated in ORC1 self-interaction as well as interaction with Protein Phosphatase 1 $\alpha$  (PP1 $\alpha$ ) and CDC6<sup>4</sup>, and, regulated by phosphorylation and dephosphorylation, they contribute to the control of ORC1 levels during the cell cycle. Interestingly, the described SLiMs do not overlap

with ORC1 RNA binding domain albeit are part of the same highly flexible region<sup>4</sup> (Fig. 1a). Thus, the interaction of ORC1 with RNA may influence interactions with these factors and/or other properties linked to the IDR. Among those, ORC1 IDR has been implicated in *in vitro* liquid-liquid phase separation enhanced by DNA<sup>4,37</sup>, being CDK/Cyclin phosphorylation a key inhibitor of liquid phase condensation by replication initiation factors<sup>37</sup>. While the effect of RNA was not tested in those studies, RNA biophysical properties are consistent with an analogous role. Supporting this notion, we observed that RNA induces the formation of droplets of WT- but not of MUT-ORC1 in a concentration-



dependent manner, reaching a threshold where RNA leads to droplet dissolution (Supplementary Fig. 9a–c).

Based on our data, and by analogy to the RNA feedback mechanism described in transcription<sup>38</sup>, where the balance between positive and negative charges determines whether RNA promotes the formation or dissolution of IDR-containing protein condensates, RNA could be playing a sequential role in the initiation of replication (Supplementary Fig. 9d). We speculate that in late M phase, incipient transcription may favor the formation of coacervates that nucleate ORC1 around TSSs, explaining the presence of ORC1 bound to RNAs transcribed from active origins at early-replicating genomic regions, as well as the enrichment of iCLIP binding sites at 5' ends of genes. Nevertheless, the need of other factors for ORC1 chromatin recruitment should not be excluded. Once preinitiation complexes are assembled on DNA, the property of forming condensates would not be further required. On the contrary, higher local RNA concentration resulting from transcription elongation, particularly at longer genes, would help ORC1 disassembly from chromatin, coupled to an RNA-dependent control of ORC1 phosphorylation balance, and followed by its targeting for degradation and origin firing. This would also explain the enrichment of RNAs containing GAA repeats among ORC1 partners, since these purine-rich sequences are thought to retain RNAs in the nucleus through a saturable nuclear retention factor<sup>29</sup>.

In addition to the described mechanisms, ORC1 is known to have cellular functions beyond the origins of replication<sup>39,40</sup>. It is likely that some of the ORC1-RNA interactions are occurring at those locations. For instance, the interaction between ORC1 and RNA may also be involved in ORC1 transport in and out of the nucleus<sup>41</sup>, which may account for the presence of fully spliced RNAs among ORC1 binders; all questions deserve future investigation.

Overall, our results unveil a novel non-coding function for RNA as a dynamic component of the chromatin, which helps to coordinate transcription and replication in the nucleus.

## Methods

### Cell lines, growth conditions, and culture treatments

HCT116 cells (CCL-247) were cultured in RPMI-1640 (GIBCO), and U2OS cells (HTB-96) in DMEM (GIBCO), all mediums supplemented with 10% fetal bovine serum (GIBCO) and 1x penicillin/streptomycin (Lonza), and maintained at 37 °C and 5% CO<sub>2</sub>. To generate HCT116 cells stably expressing WT-ORC1-3xFlag or MUT-ORC1-3xFlag, cells were transfected with pcDNA3.1 vectors containing codon-optimized wild type or mutated cDNA sequences of ORC1 (Supplementary Table 2), and treated with Neomycin-G418 (Sigma) 500 µg/mL for 10 days.

Proteasome inhibition was achieved with short pulses (4 and 6 h) of MG-132 (MilliporeSigma) 50 µM. Cycloheximide (Sigma) was incubated in a culture medium at 100 µg/mL for 4 or 6 h for translation inhibition.

For RNA in vivo labeling, synchronized U2OS cells were incubated with 5-ethynyl uridine (EU; Thermo Fisher) at 0.2 mM final concentration in complete culture medium, in short (20 minutes) or long (16 and 3 h release) pulses.

For fiber assays, exponentially growing HCT116 cells were first pulsed with 50 µM CldU (Sigma) for 20 minutes, washed, and subjected to a second 20 min pulse with 250 µM IdU (Sigma).

In experiments with RNase A treatment, synchronized cells at different points of the S phase (see cellular synchronization section) were trypsinized, permeabilized with 0.05% Tween-20 in PBS for 10 min on ice, and mock-treated or treated with 1 mg/mL RNase A (Sigma) for 30 min at RT, as previously described<sup>42</sup>.

### Cellular transfection

Cellular transfections were performed with Lipofectamine 2000 (Invitrogen) in Serum-free Opti-MEM (GIBCO), following manufacturer instructions.

For RNA knockdown, siRNAs or Antisense Oligonucleotides were transfected for 24 h at final concentrations of 40 or 80 nM, respectively, except for the 48 h transfection to knockdown ORC1. siRNAs were designed using the i-Score designer tool and purchased from Sigma (Supplementary Table 3). Control ASO was designed and synthesized by Ionis Pharmaceuticals. anti-TTC and anti-GAA ASOs were self-designed and synthesized by iDT, with six 2'-O-methoxyethyl nucleotides on the 5' and 3' ends, and nine consecutive oligodeoxynucleotides to support RNaseH activity, as previously described<sup>31</sup> (Supplementary Table 3).

For exogenous ORC1 expression, pcDNA3.1 or pBABE vectors containing codon-optimized wild type or mutated cDNA sequences of ORC1, tagged with 3xFlag or Halo (Supplementary Table 2), were transfected for 48 hours, while, in rescue experiments, plasmid transfection was preceded by 24 h endogenous ORC1 siRNA-mediated depletion (Supplementary Table 3). Exogenous *CEP95* was over-expressed by 48 h transfection of a pcDNA3.1 plasmid purchased from GenScript (Supplementary Table 2).

### Cellular synchronization and cell cycle analysis

For cellular synchronization at G1/S, U2OS cells were subjected to serum starvation for 48 h, while HCT116 cells were blocked by double thymidine shock as previously described in ref. 43. Synchronization of HCT116 cells was released by PBS washing and incubation in a complete culture medium, after which cells were harvested at different time points (1, 2, 3, 5, 6, 7, or 9 h) covering the entire S phase.

For cell cycle analysis, cells were fixed in PBS 2% PFA, and incubated in 2 N HCl 0.5% Triton for 30 min. Followed by 0.1 M Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> incubation, cells were treated with RNase A (Promega), resuspended in PBS, and the DNA stained with propidium iodide 1 mg/mL (Sigma). In flow cytometer FACSCalibur, DNA staining was recorded by the BD CellQuest program. Cell cycle profiles were determined by considering the amount of labeled DNA (FL2-H) per cell.

### RNA extraction, processing, and RT-qPCR

Cell preparations were fixed with TRIzol (Sigma), and RNA precipitated with isopropanol.

In bulk RNA-seq experiments of HCT116 cells treated with control or anti-GAA ASOs, total RNA extraction was followed by Turbo DNase (Invitrogen) digestion and library preparation with TruSeq Stranded mRNA kit from Illumina. Duplicate experiments were sequenced with Illumina NextSeq 500.

For RT-qPCR, up to 1 µg RNA was treated with DNase I (Invitrogen) and reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystem) with random hexamer primers, following manufacturer instructions. The obtained cDNA was analyzed by quantitative PCR (qPCR) using iTaq Universal SYBR Green supermix (Bio-Rad) in a ViiA™ 7 Real-Time PCR System machine (Thermo Fisher), all reactions performed in quadruplicate. *HPRT1* or *GAPDH* RNA levels were used for the normalization of total and cytoplasmic cellular extracts, while *MALAT1* was used to normalize RNA levels in nuclear extracts. For RIP and UV-RIP validations, RNA levels were normalized by their levels in a 10% experimental input of nuclear RNA. Statistical differences between relative RNA levels or relative enrichments were calculated by unpaired two-tailed Student's *t*-test. RT-qPCR primers were self-designed or designed with the Universal Probe Library Assay Design Center (Roche), and purchased from Metabion (Supplementary Data 4).

### Protein extraction and western blot

Soluble-chromatin cellular fractionation was performed as previously described in ref. 44, while subcellular fractionation in cytoplasm and nucleus, and nucleoplasm if indicated, was performed as described elsewhere<sup>45</sup>.

Proteins from cell preparations were quantified with Pierce BCA Protein Assay Kit (Thermo Fisher). Samples were run in denaturing polyacrylamide gels by electrophoresis, and then transferred to nitrocellulose membranes (Bio-Rad). Membranes were blocked with 5% milk in PBS-Tween, and incubated overnight with primary antibodies (Supplementary Table 4). HRP-conjugated secondary antibodies (Cell Signaling Technology, dilution 1:10,000) on the membrane were detected with the use of enhanced chemiluminescence (ECL) reagent (PerkinElmer) in Odyssey CLx (LI-COR), and recorded with Image Studio Lite software. Full scan blots are provided in the Source Data file. Relative protein levels were obtained based on the intensity of the western blot bands using Fiji software. Quantified intensities were normalized to those of the loading reference and, if indicated, fold changes relative to other conditions were calculated. Statistical differences between normalized intensity values were calculated by paired two-tailed Student's *t*-test.

### ORC1 RIP-seq on nuclear extracts

Native RNA immunoprecipitation was performed as previously described<sup>46</sup> with minor modifications. Briefly,  $40 \times 10^6$  asynchronous HCT116 cells, untreated or transiently expressing WT-ORC1-3xFlag, were harvested and lysed. Nuclear lysates were dounced and sonicated (Bioruptor Diagenode) for ten cycles, pre-cleared with protein A/G Dynabeads, and incubated with 5  $\mu$ g of control IgG (sc-2025) or antibody of interest (anti-FLAG [M2, F1804, Sigma] or anti-ORC1 78-1-172 [Bruce Stillman laboratory<sup>3</sup>]). Protein A/G Dynabeads were added to sequester the antibody, and washed five times. For protein analysis, 10% of beads and inputs were resuspended in 2X Laemmli sample loading buffer, and run in acrylamide gels for western blot. RNA from beads and inputs was obtained following the RNA extraction protocol, to perform RT-qPCR (methods section RNA extraction, processing, and RT-qPCR) or library preparation. For sequencing, samples were first treated with Turbo DNase (Invitrogen), and libraries were prepared with the TruSeq Stranded Total RNA kit from Illumina. Sequencing of triplicate (anti-ORC1) or duplicate (anti-Flag) experiments was done with Illumina NextSeq 500.

### ORC1 IP and phosphoproteomics

Transfected WT-ORC1-3xFlag and MUT-ORC1-3xFlag proteins in HCT116 cells (Supplementary Table 2), in control or anti-GAA-knockdown conditions (Supplementary Table 3), were immunoprecipitated from nuclear extracts of  $160 \times 10^6$  cells, as previously described in ref. 3, with minor modifications. In brief, after cellular lysis (260 mM Sucrose, 8 mM Tris-HCl pH 7.4, 4 mM MgCl<sub>2</sub>, 0.8% Triton X-100), nuclei were resuspended and incubated for 30 min in high salt buffer (20 mM Tris-HCl pH 7.5, 400 mM NaCl, 0.4% Igepal, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 1 mM CaCl<sub>2</sub>, 10% Glycerol, and 0.1 mM DTT), supplemented with phosphatase and protease inhibitors 1x (Roche), as well as benzonase to digest DNA. After sonication (Bioruptor Diagenode) for 30 cycles, NaCl and Igepal concentrations were brought down to 200 mM and 0.2%, respectively, by adding an equal volume of dilution buffer. Clear nuclear lysates were then pre-cleared with protein G Dynabeads, and incubated with 5  $\mu$ g of anti-FLAG antibody (M2, F1804, Sigma), untransfected cells being the negative control. Protein G Dynabeads were added to sequester the antibody, and washed five times with complete washing buffer (20 mM Tris, 100 mM NaCl, 0.15% Igepal, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, and 10% Glycerol). Beads and inputs were resuspended in 2X Laemmli sample loading buffer, and run in acrylamide gels for Coomassie blue staining.

Bands at the expected molecular weight were cut, and containing proteins were subjected to mass spectrometry. The samples were reduced with 1 mM DTT for 30 min at 60 °C and then alkylated with 5 mM iodoacetamide for 15 min in the dark at room temperature. Gel pieces were then subjected to a modified in-gel trypsin digestion procedure<sup>47</sup>. Gel pieces were washed and dehydrated with acetonitrile

for 10 min, followed by removal of acetonitrile, and then completely dried in a speed-vac. Rehydration was done with 50 mM ammonium bicarbonate solution containing 12.5 ng/ $\mu$ L modified sequencing-grade trypsin (Promega, Madison, WI) at 4 °C. Samples were then placed in a 37 °C room overnight. Peptides were later extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 1% formic acid. The extracts were dried in a speed-vac (-1 h). For the analysis, samples were reconstituted in 5–10  $\mu$ L of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 2.6  $\mu$ m C18 spherical silica beads into a fused silica capillary (100  $\mu$ m inner diameter  $\times$  -30 cm length) with a flame-drawn tip<sup>48</sup>. After equilibrating the column, each sample was loaded via a Famos autosampler (LC Packings, San Francisco, CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile and 0.1% formic acid). As each peptide was eluted, they were subjected to electrospray ionization and then they entered into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Eluting peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein or translated nucleotide databases with the acquired fragmentation pattern by the software program, Sequest (ThermoFinnigan, San Jose, CA)<sup>49</sup>. The modification of 79.9663 mass units to serine, threonine, and tyrosine was included in the database searches to determine phosphopeptides. Phosphorylation assignments were determined by the Ascore algorithm<sup>50</sup>. All databases include a reversed version of all the sequences and the data were filtered to between a one and two percent peptide false discovery rate. The position and amount of detected phosphopeptides is presented in Supplementary Data 3. Raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD043232.

### ORC1 UV-RIP and iCLIP on nuclear extracts

About  $40 \times 10^6$  asynchronous HCT116 cells were rinsed in cold PBS and irradiated with 150 mJ/cm<sup>2</sup> in a Stratelinker 2400 at 254 nm. To isolate cellular nuclei, the first steps of fractionation iCLIP protocol were performed<sup>51</sup>, followed by UV-RIP or iCLIP protocols.

For UV-RIP, fixed nuclear pellets of HCT116 cells transiently expressing Flag-tagged WT-ORC1 were resuspended in RIPA buffer and sonicated (Bioruptor diagenode) for 15 cycles. Solubilized nuclear extracts were pre-cleared with protein G Dynabeads, and 5  $\mu$ g of control IgG (sc-2025) or anti-FLAG (M2, F1804, Sigma) antibodies incubated overnight at 4 °C. Protein G Dynabeads were then added to sequester the antibody, and washed six times. Immunoprecipitates and inputs were eluted, and Proteinase K (NEB) was incubated for 45 min at 45 °C for protein digestion, followed by RNA extraction and RT-qPCR (see methods section RNA extraction, processing, and RT-qPCR).

iCLIP data of ORC1 was generated by following the iCLIP method described elsewhere<sup>52</sup>, using  $10^6$  HCT116 UV-fixed nuclei previously transfected with ORC1-WT-3xFlag ( $n=5$ ), ORC1-MUT-3xFlag ( $n=2$ ) or untransfected cells as control ( $n=1$ ) (Supplementary Table 2). First, nuclei were lysed in 1 mL of lysis buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.4, 1% Igepal CA-630, 0.1% SDS, and 0.5% Na-Deoxycholate) supplemented with protease inhibitors 1x (Roche). Nuclear lysates were sonicated (Bioruptor Diagenode) for ten cycles at low intensity. Afterward, RNase digestion was performed for 3 min at 37 °C with 0.4 U ( $n=3$ ) or 1 U ( $n=2$ ) of RNase (Thermo Scientific, EN0602) in the presence of 4U of Turbo DNase (Invitrogen), to avoid DNA contamination. Cleared supernatant was immunoprecipitated overnight at 4 °C with 5  $\mu$ g of anti-Flag antibody (M2, F1804, Sigma). Protein-RNA complexes were visualized using pre-adenylated, infrared dye-labeled L3 adapter with the following sequence: /5rApp/AG

ATCGGAAGAGCGGTTTCAGAAAAAAAAAAAAA /iAzideN/AAAAAAAAAAA  
AAA/3Bio/. Reverse transcription was performed using RNA-  
dependent retrotranscriptase Superscript IV (Invitrogen) and bar-  
coded primers (XXXXX) containing UMIs (NNNN): /5Phos/ WWW  
XXXXX NNNN AGATCGGAAGAGCGTCTGAT /iSp18/ GGATCC /iSp18/  
TACTGAACCGC. Purification of cDNAs following reverse transcription  
and circularization was performed using AMPure XP beads (Beckman-  
Coulter, USA) and isopropanol. Libraries were sequenced as single-end  
100 bp reads on Illumina HiSeq 4000.

### Chromatin immunoprecipitation (ChIP)

About  $40 \times 10^6$  asynchronous HCT116 cells, stably expressing Flag-  
tagged wild type and mutated ORC1 (RNA-binding mutant) (Supple-  
mentary Table 2), or transfected to knockdown or overexpress *CEP95*  
mRNA (Supplementary Tables 2, 3) and synchronized in S phase (see  
cellular synchronization section), were fixed for 30 min with 2 mM  
DSG, and then crosslinked with 1% formaldehyde for 10 min. Pelleted  
cells were lysed (5 mM Tris-HCl pH 8, 85 mM KCl, and 0.5% Igepal),  
nuclei resuspended in RIPA buffer, supplemented with protease inhi-  
bitors 1x, and sonicated (Bioruptor Diagenode) for 30 cycles. Solubi-  
lized nuclear extracts were pre-cleared with protein A/G Dynabeads,  
and 5  $\mu$ g of control IgG (2729 Cell Signaling or sc-2025) and anti-CDC45  
(11882, Cell Signalling Technology) or anti-PCNA (ab29, Abcam) anti-  
bodies incubated overnight at 4 °C. Protein A/G Dynabeads were then  
added to sequester the antibody, and washed with low salt (0.1% SDS,  
1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl), high  
salt (Low Salt buffer with 500 mM NaCl) and LiCl (0.25 M LiCl, 1% Ige-  
pal, 1% Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8) buffers,  
supplemented with protease inhibitors 1x (Roche). Immunoprecipitates  
and inputs were eluted, RNA digested with RNase A (Promega) for  
30 minutes at 37 °C, and Proteinase K (NEB) incubated for 45 min at  
45 °C for protein digestion. Samples were de-crosslinked overnight at  
65 °C, followed by phenol:chloroform DNA extraction and ethanol  
precipitation.

qPCR of precipitated DNA was done as cDNA samples (see RNA  
processing section), with self-designed primers at genomic DNA  
replication origins or control regions (Supplementary Data 4), having  
SNS-seq data in wild-type HCT116 cells as a reference and purchased  
from Metabion. ChIP-qPCR enrichment was done compared to a 10%  
input, and relativized to transfection controls. Statistical differences  
between relative enrichments were calculated by unpaired two-tailed  
Student's *t*-test.

For sequencing of CDC45 ChIP samples, libraries of duplicate  
experiments were generated as previously described in ref. 53, and  
sequenced with Illumina NextSeq 2000.

### Nascent strand preparation by $\lambda$ -exonuclease method and RT-qPCR analysis

Nascent strand preparation was performed as previously described<sup>54</sup>,  
in HCT116 cells, wild type or stably expressing WT-ORC1-3xFlag or  
MUT-ORC1-3xFlag (Supplementary Table 2), untreated or transfected  
with control or specific siRNAs or ASOs for RNA knockdowns (Sup-  
plementary Table 3). SNS-seq from wild-type HCT116 cells was per-  
formed in triplicate experiments. To assess origin activity in HCT116  
cells knocked down for GAA-RNAs, duplicate SNS-seq experiments  
were analyzed in cells transfected with a control ASO, or the anti-GAA  
ASO (Supplementary Table 3). SNS-seq from HCT116 cells stably  
expressing WT or MUT-ORC1 were performed in duplicates.

Briefly, genomic DNA of  $1-2 \times 10^8$  exponentially growing cells was  
digested with Proteinase K (NEB), precipitated with ethanol, and  
solubilized in TE buffer pH 8 supplemented with RNase OUT (Invitro-  
gen) for 48 h at 4 °C. Denatured DNA was ultracentrifuged in a seven-  
step 5 to 20% discontinuous sucrose gradient, with SW40Ti rotor at  
24,000 rpm for 20 h at 20 °C. About 1 mL fractions were collected, and  
the DNA precipitated with ethanol. 10% volume of DNA in each fraction

was denatured with 0.2 M NaOH and run onto a 1% alkaline agarose gel  
(50 mM NaOH and 1 mM EDTA). Followed by neutralization with 1x  
TAE, Syber Gold (Thermo Fisher) staining was used to visualize the  
fractionation profile with a UV-Biorad camera. Fractions of interest  
(0.5–2 kb) were treated twice with PNK (Thermo Fisher) and  $\lambda$ -  
exonuclease (Thermo Fisher) to remove short-cut DNA and keep  
newly synthesized DNA, which has a 5' RNA segment. The efficiency of  
 $\lambda$ -exonuclease digestion was evaluated by incubation of 10% reaction  
volume with a digested control plasmid.

qPCR of purified nascent strands was done as cDNA samples (see  
RNA processing section), with self-designed primers at genomic DNA  
replication origins or control regions having SNS-seq data in wild-type  
HCT116 cells as reference (Supplementary Data 4), and purchased  
from Metabion. SNS-qPCR enrichment was done compared to a 10%  
input of genomic DNA, and statistical differences of relative enrich-  
ments, between control or RNA knocked-down cells, were calculated  
by paired two-tailed Student's *t*-test. The associated *p* values were used  
for heatmap representations.

For nascent strand sequencing (SNS-seq), ssDNA fragments were  
converted to dsDNA as previously described in ref. 54. First, short  
nascent strands were digested with RNase A/TI mixture (Thermo  
Fisher) to eliminate both mRNAs and RNA primers, and facilitate  
adapters ligation for sequencing. Samples were mixed with random  
hexamer primer phosphate (Roche), and a reaction with the Klenow  
exo- polymerase (NEB) was used to extend the primers and synthesize  
the complementary DNA strand. Taq DNA ligase (NEB) was used to  
ligate the synthesized fragments, and the dsDNA was extracted and  
precipitated with ethanol. For library preparation, a protocol primarily  
designed for High-throughput chromatin immunoprecipitation (HT-  
ChIP) was followed<sup>53</sup>. Duplicate or triplicate experiments were  
sequenced with Illumina NextSeq 2000.

### RNA-FISH

Cultured HCT116 cells on coverslips, and transfected with control or  
anti-GAA ASOs (Supplementary Table 3), were fixed in PBS 3% PFA, and  
washed in 2x SSC 50% formamide. Meanwhile, anti-GAA FISH probes,  
which were self-designed and purchased from iDT, were denatured  
at 92 °C for 4 min, cooled down, and diluted in hybridization buffer  
(50% Deionized Formamide, 2x SSC, 10% Dextran Sulfate) to 25 nM  
final concentration. FISH probes were incubated on the cells overnight  
at 37 °C in humidity.

/56-FAM/\*T\*T\*C\*T\*T\*C\*T\*T\*C\*T\*T\*C\*T\*T\*C\*T\*T\*C\*T\*T\*C\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*  
T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*T\*T\*C\*\*/36-FAM/ Cells  
were washed with 2x SCC 50% Formamide at 55 °C, followed by a wash  
with 2x SCC at 55 °C, and a wash with 2x SCC. Preparations were  
blocked (PBS 0.5% Tween, 10% Heat-inactivated Goat Serum, 0.5%  
Blocking Reagent [Roche]), and incubated with  $\alpha$ -FAM-POD antibody  
(Roche). After washing with 4x SSC, preparations were incubated with  
TSA-Cy3 (PerkinElmer) diluted in Amplification Diluent (PerkinElmer).  
Secondary antibodies were washed with 4x SSC, followed by a wash  
with 4x SSC 0.1% Triton, and a last wash with 4x SSC. Cells were  
mounted on microscope glass slides with the mounting solution with  
DAPI (Vectashield) and imaged with confocal fluorescence microscope  
Zeiss LSM 880 NLO 63x objective, and images were captured with the  
ZEN microscopy software (Zeiss). Fiji software was employed for  
stacks deconvolution and signal quantifications. Probe fluorescence  
per cell (15 cells each) in 4 independent biological replicates were  
compared between control and anti-GAA treated cells by unpaired  
two-tailed Student's *t*-test.

### Protein imaging (chromatin immunofluorescence, STORM)

Protein imaging at chromatin of HCT116 cells, after removing soluble  
cellular fractions, was performed as previously described in ref. 55,  
with anti-CDC45 (Cell Signalling 11881) and anti-PCNA (sc-56) anti-  
bodies. After immunofluorescence staining with secondary antibodies,



coverslips were mounted on microscope glass slides with a mounting solution with DAPI (Vectashield). Preparations were imaged using the 63x objective of the automated microscope Zeiss Axio Imager M1, and images were captured with the ZEN microscopy software (Zeiss). Fiji software was employed for stacks deconvolution and signal quantifications. After delimiting cellular nuclei, fluorescence intensities were measured and normalized by nuclei area. Comparisons between cells (>100) in different experimental conditions were done by unpaired two-tailed Mann–Whitney *t*-test. Results were validated in two independent biological replicates.

In STORM experiments, synchronized U2OS cells (see cellular synchronization section), untreated, mock-transfected, or transiently expressing Halo-tagged wild type and mutated ORC1 (Supplementary Table 2), were permeabilized before fixation, to capture protein and RNA imaging at chromatin, as previously described in refs. 55,56. Incorporated EU (see culture treatments section) was detected with the Click-it Plus Kit (Thermo Fisher) AF647 (Thermo Fisher), and ORC1 with anti-ORC1 antibody (F-10) (sc-398734) or Janelia Fluor® 549 HaloTag® Ligand (Promega, GA1110). After immunofluorescence staining with secondary antibodies, coverslips were mounted on microscope glass slides with freshly prepared super-resolution imaging buffer (PBS 1 mg/mL Glucose Oxidase [Sigma], 0.02 mg/mL Catalase [Sigma], 10% Glucose [Sigma], 100 mM Mercaptoethylamine [Thermo Fisher]) flowed through. All raw images were acquired using a custom-built inverse microscope platform (Applied Scientific Instrumentation). Briefly, 639 nm (UltraLaser, MRL-FN-639-1000) and 561 nm (Coherent, Sapphire 561 LPX – 500) laser lines were adjusted to 1.5 and 0.8 kW/cm<sup>2</sup>, respectively. Fluorescence emission was expanded with a 1.67× achromatic lens tube and was collected on a sCMOS camera (Photometrics, Prime 95B). Fluorescence signals were collected sequentially using the AF647 (Semrock, FF01-676/37) and AF568 (Semrock, FF01-607/36) single-band pass filters in a filter wheel (ThorLabs, FW102C). A 405 nm laser line (UltraLaser, MDL-III-405-500) was introduced to enhance recovery of dark state fluorophores when required. About 2000 Frames at 33 Hz were acquired for each color.

Localization of single molecules and the mapping of the two different channels were carried out in algorithms written in MATLAB as previously described in ref. 57. To quantify the degree of colocalization between RNA and ORC1, cross-correlation (Eq. (1)) between the two species was calculated. Similar to the radial distribution function, for two images  $Im_1$  and  $Im_2$ , the correlation magnitude at displacement  $\mathbf{r} = (r, \theta)$  is defined as

$$c(\mathbf{r}) = \frac{\langle \delta\rho_1(\mathbf{R})\delta\rho_2(\mathbf{R} + \mathbf{r}) \rangle_{\mathbf{R}}}{\langle \rho_1(\mathbf{R}) \rangle_{\mathbf{R}} \langle \rho_2(\mathbf{R}) \rangle_{\mathbf{R}}} \quad (1)$$

where  $\rho_i(\mathbf{R})$  denotes the local density of  $Im_i$  at location  $\mathbf{R}$  and  $\langle \rho_i(\mathbf{R}) \rangle_{\mathbf{R}}$  denotes the average density over the entire image, where  $\langle \bullet \rangle_{\mathbf{R}}$  denotes the average operator over all the location  $\mathbf{R}$ ;  $\delta\rho_i(\mathbf{R}) = \rho_i(\mathbf{R}) - \langle \rho_i(\mathbf{R}) \rangle_{\mathbf{R}}$  denotes the fluctuation of the local density at location  $\mathbf{R}$ . As the correlation is not orientation-specific, the 2D  $c(\mathbf{r}) = c(r, \theta)$  was further averaged over  $\theta$  and plotted as the correlation profile as the function of the radial distance  $r$ .

In brief, each nucleus was first manually outlined to generate a ROI for independent analysis. EU and ORC1 signals from the same ROI were submitted for cross-correlation analysis to obtain their association magnitudes, whilst the cross-correlation between the two species from different ROIs serve as a control describing random distributions<sup>56</sup>. An unpaired two-sample *t*-test between experimental and randomized data was done to determine the significance of the correlation. The same analyses on cells with no EU incubation were used as the experimental negative control. Results were validated in two independent biological replicates.

## Fiber stretching and staining

500 HCT116 cells transfected with siRNAs (Supplementary Table 3), plasmids to express exogenous ORC1 (Supplementary Table 2), or ASOs (Supplementary Table 3), and pulsed with CldU and IdU (see culture treatments section), were dropped on Superforst Thermo Scientific microscope slides and lysed with spreading buffer (0.5% SDS, 200 mM Tris-HCl pH 7.5, and 50 mM EDTA) in humidity. Slides were tilted at a 10–15° angle to allow the DNA suspension to run slowly down the slide and air dried. DNA fibers were fixed in –20° cold 3:1 methanol:acetic acid and air dried. For fiber staining, DNA was denatured in 2.5 M HCl for 30 min at room temperature, washed with PBS, and blocked (PBS 1% BSA and 0.1% Triton). Slides were then incubated with primary antibodies detecting CldU (ab6326, Abcam) and IdU (347580, BD), in a humidity chamber overnight at 4 °C. After PBS washing, slides were incubated with fluorescent secondary antibodies. After washing, primary anti-ssDNA clone 16–19 (MAB3034) and secondary fluorescent antibodies to label DNA fibers were incubated for 30 min. Preparations were then air dried and mounted with Prolong diamond (Invitrogen). Preparations were imaged using the 40x objective of the automated microscope Zeiss Axio Imager M1, and images were captured with the ZEN microscopy software (Zeiss).

DNA fiber images were analyzed with Fiji software, considering a conversion factor of 1 μm = 2.59 kb. Two parameters were analyzed: fork rate, measuring the length (in kb) of the IdU track and dividing it by the 20 min of the duration of the pulse; inter-origin distance, measuring the distance between adjacent origins (recognized as IdU-CldU-IdU tracks). Comparisons of fork rate and inter-origin distances between experimental conditions were analyzed by unpaired two-tailed Mann–Whitney *t*-test. Results were validated in two independent biological replicates.

## Protein purification

Purified wildtype and mutated (R441A, R444A, and R465A) ORC1 RNA-binding regions (amino acids 413–511), fused to GST, were produced in Shou Waga laboratory<sup>24</sup>. Protein concentrations were estimated by Coomassie blue or silver staining, compared to BSA known concentrations.

Wildtype and mutated full-length MBP-PP-GFP-ORC1-6xHis were expressed and submitted to Ni-NTA beads purification, followed by a second purification with amylose beads, as previously described in ref. 4. Briefly, *E.coli* BL21 cells were transformed with fusion plasmids (tagged full-length ORC1) (Supplementary Table 2), grown in LB media at 37 °C until 0.7–0.9 O.D., and induced overnight with 0.3 mM IPTG at 16 °C. Bacterial cells were then pelleted, washed, and lysed with 100 mg/mL lysozyme in buffer A (25 mM Tris pH 7.5, 150 mM NaCl, 0.02% Igepal, 5 mM MgCl<sub>2</sub>, 5 mM Benzamidine-HCl, 1 mM PMSF, 1x protease inhibitors [Roche], and 10% Glycerol). After centrifugation, the clarified supernatant was incubated with pre-washed Ni-NTA beads for 3 h at 4 °C. Bead-bound proteins were washed with lysis buffer, and eluted with 300 mM imidazole. WT and MUT MBP-PP-GFP-ORC1-6xHis proteins were further purified with amylose beads and eluted with 20 mM maltose. Protein concentrations were estimated by Coomassie blue or silver staining, compared to BSA known concentrations.

## In vitro assays (EMSA, LLPS)

In electrophoretic mobility shift assays (EMSA), wildtype and mutated GST-ORC1 (413–511) (concentrations indicated in figure captions) or control buffer (25 mM HEPES pH 8, 300 mM NaCl, 10% Glycerol, 1 mM DTT), were incubated with RNA (concentration indicated in figure captions) in 20 μL binding buffer (25 mM HEPES pH 8, 10 mM Mg(C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>)<sub>2</sub>, 0.1 mM EDTA, 5% Glycerol), supplemented with BSA 2 mg/mL, 3 mM ATP (NEB), and RNAs in ribonuclease inhibitors (Promega). The binding reaction was incubated at 30 °C for 30 min, and samples were immediately loaded into a pre-run non-denaturing gel for electrophoresis at 4 °C. RNA was stained in Sybr Gold (Thermo



Fisher), visualized with a UV-Biorad camera, and quantified with Fiji software. In experiments with total RNA, RNA fragments were obtained by sonicating (Bioruptor diagenode) total extracts of HCT116 cells, and purifying short RNA fragments with miRNA columns (PureLink). Electrophoresis was done in 7.5% polyacrylamide gels. In EMSAs with CEP95 RNA, RNA fragments were in vitro synthesized with T7 RNA polymerase (Promega), using as a template the PCR products amplified from the pCDNA3-CEP95 vector (Supplementary Table 2 and Supplementary Data 4). Electrophoresis was done in 0.7% agarose gel.

In liquid-liquid phase separation (LLPS) assays, full-length wild type and mutated MBP-PP-GFP-ORC1-6xHis (4  $\mu$ M) were resuspended in LLPS reaction buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, and 1 mM DTT) supplemented with Precission protease, in presence or absence of 4  $\mu$ M of RNA probes previously used (same sequence) in DNA phase separation experiments<sup>4,37</sup> (GAAGCTAGACTTAGGTGCA-TATTGAACCTACTATGCCGAAGTAGTTACGAGCTATAAC), and incubated at 4 °C for 16 h. Following incubation, the reactions were spotted on microscope slides with coverslips and observed immediately with 63x oil immersion objective of the automated microscope Zeiss Axio Imager M1. Images were captured with the ZEN microscopy software (Zeiss). GFP-droplet size was measured with Fiji software.

### CatRAPID analysis

The *catRAPID* algorithm estimates the binding potential of a protein–RNA pair through van der Waals, hydrogen bonding, and secondary structure propensities, allowing the identification of binding partners with an accuracy of 0.78 or higher<sup>26,58</sup>.

The *catRAPID* analysis to predict ORC1 direct interactions with ORC1 RIP-RNAs was performed following standard pipelines<sup>59</sup>. Briefly, we used the major RNA isoform for each gene reported in ORC1 RIP-seq experiments, retrieved the ORC1 interaction scores from RNAc<sup>60</sup>, and all transcripts with *p* value >0.01 were filtered out. Two classes were analyzed: depleted RNA, when log<sub>2</sub> fold change <0, and enriched RNA, if log<sub>2</sub> fold change >0, and the difference in the predicted *catRAPID* scores of enriched and depleted was represented (*z*-score) and computed with unpaired two-tailed Student's *t*-test. We also used the experimental log<sub>2</sub> fold change to rank the two groups of transcripts, enriched and depleted. Equal fractions of enriched (i.e., highest fold changes) and depleted (i.e., lowest fold changes) RNAs were compared. The discriminative power, measured as the Area under the ROC curve (AUC), increases proportionally to the experimental signal and reaches a value of 0.80, indicating strong enrichment of predicted physical interactions.

*catRAPID omics v2* was used to predict RNA interactions of the wild type and mutated ORC1 protein sequences (R441A, R444A, and R465A)<sup>61</sup>. We divided the Interaction Propensity score in bins of width 20 (a.u.). For each of the Interaction Propensity Bins, we calculated the fraction of RNAs that obtained decreased Interaction Propensity score against the mutated ORC1 sequence. We found that the RNAs exhibiting ‘High Interaction Propensity’ scores were enriched targets of WT-ORC1, while those exhibiting ‘Low Interaction Propensity’ scores were enriched targets of MUT-ORC1. This finding reinforces the notion that the identified RNAs by RIP-seq are bona-fide ORC1 direct interactors.

### cleverSuite analysis

To investigate the effect of amino acid mutations on ORC1, we used the *CleverSuite* approach<sup>62</sup>, which uses two protein sequence sets (Positive and Control/Negative) to build a model able to separate them based on physicochemical features (hydrophobicity, secondary structure, charge, etc.). The model can be reused to predict the classification of other sets. We trained 2 models: RNA and DNA binding ability. For the RNA model we used RNA binding proteins<sup>63</sup> and, as control, a set of proteins that were found in the lysate from the same study. For the DNA model, we used Proteins annotated as ‘DNA binding’ from

UniProt and, as control, a sample set of similar size as the DNA binding proteins that were not annotated as ‘DNA binding’ nor ‘RNA binding’. The two models can be found at the links (<http://crg-webservice.s3.amazonaws.com/submissions/2021-09/393018/output/index.html?unlock=8bc230ac56> and <http://crg-webservice.s3.amazonaws.com/submissions/2013-12/17868/output/index.html?unlock=f3a7ffa08f>).

We used the predictive ability of the *CleverSuite* to assess if the ORC1 mutant (R441A, R444A, and R465A) belongs to the positive or control set. Specifically, we tested the hypothesis of whether the ORC1 region, including the mutations, affected its ability to interact with RNA (positive) and/or DNA (control). The results unequivocally showed that mutant ORC1 has decreased RNA binding activity, while its DNA binding activity remains unaffected. In the analysis, we considered different ORC1 fragments centered around the mutations to control the signal-to-noise ratio of *CleverSuite* scores. Specifically, in the RNA-Binding model, all ORC1 WT fragments are predicted as ‘Positive’ (RNA binding) and all the Mutated ORC1 fragments were predicted as ‘Control’. Importantly, the same regions were predicted as ‘Positive’ for both WT and mutant (DNA binding) when using the second (Control) model. Thus, the DNA binding ability is predicted to be unaffected by RNA-binding mutations.

### Statistical analysis

Experimental data were plotted and analyzed using the GraphPad statistical software, following the statistical analysis for each type of data, specified in each method section and/or figure captions. Most experimental data are represented as the average of at least three biological replicates, indicated at figure captions. Imaging data is presented as a representative experiment with multiple measurements, which was validated in additional biological replicates. The number of replicates in sequencing experiments is specified in each method section.

R software was used for bioinformatic analysis, using the R package ggplot2 to generate different types of plots (<https://cran.r-project.org/web/packages/ggplot2/index.html>). Significance was obtained using the statistical test corresponding to each type of data analyzed, as explained in each analysis section. In all cases, *p* values were given using the following thresholds: ns for *p* value > 0.05; \* for *p* value ≤ 0.05; \*\* for *p* value ≤ 0.01; \*\*\* for *p* value ≤ 0.001.

### RNA-seq, RIP-seq, and ChIP-seq Pipelines

In ASO control and anti-GAA (Supplementary Table 3) RNA-seq of HCT116 cells, and ORC1 and ORC1-3xFlag RIP-seq experiments, QC of sequencing files was performed with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Fastq files were trimmed with Trimmomatic (v 0.38)<sup>64</sup>, aligned with STAR<sup>65</sup>, reads aligning to GL contigs were removed, and FeatureCounts<sup>66</sup> was used to quantify the number of reads falling in annotated genes in hg19 human reference genome, downloaded from Ensembl<sup>67</sup>. DESeq2<sup>68</sup> was used to measure differential expression, being ASO control, and inputs the reference conditions for RNA-seq and RIP-seq analyses, respectively.

RNA-seq public sequences from untreated HCT116 cells<sup>69</sup> (GSE118051) were trimmed with Trimmomatic (v 0.38)<sup>64</sup>, aligned to the hg38 reference genome, downloaded from GENCODE<sup>70</sup>, using STAR<sup>65</sup> with parameters: ‘winAnchorMultimapNmax20 -outFilterMultimapNmax 20 -twopassMode Basic’.

Public ChIP-seq sequencing reads from duplicate experiments (1,2) were analyzed using a Nextflow pipeline<sup>71</sup> with the golden standard of the NF-Core consortia (<https://nf-co.re/>). From ENCODE<sup>72</sup>: H3K27me3 (ENCFF457PEW), H3K9me3 (ENCFF020CHJ), H3K4me1 (ENCFF531IUP), H3K27ac (ENCFF227RRY), H3K4me3 (ENCFF213WKK), and H3K36me3 (ENCFF059WYR). From Sequence Read Archive (SRA) the following published data<sup>10</sup>: H2A.Z (SRR9850576, SRR9850577), H4K20me1 (SRR9850580, SRR9850581), and H4K20me2 (SRR9850582 and SRR9850583), ORC1 ChIP input (SRR9850586) and IP

(SRR9850594 and SRR9850595). Raw sequencing reads were trimmed and quality control was performed to remove poor quality sequences. Adapter-trimmed reads were then aligned to hg38 human reference genome, downloaded from GENCODE<sup>70</sup>, with BWA<sup>73</sup> under default parameters. Replicates were merged and deduplicated to remove optical reads<sup>74</sup>. To obtain robust estimates of the results, we respected the best practices as suggested by ENCODE consortia<sup>75</sup>. Then, aligned reads passing all ChIP-seq QC metrics were submitted to MACS2 peak calling<sup>76</sup> comparing no antibody input and corresponding samples using the function “callpeak -g hs -B -q 0.05 -fe-cutoff 1.5 -broad”.

Self-generated CDC45-treated ChIP-seq samples were trimmed to remove adapted and low-quality sequenced reads with Trimmomatic (v 0.38)<sup>64</sup>. Bowtie2<sup>77</sup> was used to align the reads to the hg19 human reference genome downloaded from ENSEMBL<sup>67</sup>. Coverage tracks were generated with bamCoverage<sup>78</sup>, and the average signal at TSS positions ( $\pm 5$  kb) was divided in bins of 10 bp length, with computeMatrix from deeptools, to be analyzed by paired *t*-test (Euclidean distance of 14.0163511182744 and 13.8052506781907 for replicates 1 and 2, respectively; *p* value 3.08e-16 and 7.73e-65 for replicates 1 and 2, respectively). Coverage differences (and associated *p* values) between WT and MUT-ORC1 cells were also measured at TSSs of genes in individual iCLIP-defined quantiles.

### iCLIP analysis

ORC1 iCLIP sequencing reads were analyzed on the iMaps server (Genialis Workspace) using the iCount software<sup>79</sup> (<https://github.com/tomazc/iCount>). Briefly, experimental barcodes were removed and sequencing reads aligned with STAR<sup>65</sup> to hg38 human reference genome downloaded from GENCODE<sup>70</sup>, allowing two mismatches and ten secondary alignments. DNA or chromatin contamination was excluded by aligned data interrogation with infer\_experiment.py package from RSeQC software<sup>80</sup>. Unique Molecular Identifiers (UMIs), were used to distinguish and remove PCR duplicates. To determine protein–RNA contact sites, the sequencing read preceding nucleotide was allocated as the crosslink site event.

The presented data refers to analysis from the three experimental replicates of ORC1-3xFlag transfected cells and low (0.4 U) RNase treatment, while non-transfected control was used to corroborate data specificity (Supplementary Table 1). Replicates were merged and a summary of cDNA counts within genes and genic regions were generated with iCount summary function. Assignment of crosslink sites to coding transcripts, non-coding or biotype features, was done by following segmentation hierarchy rules (<https://github.com/tomazc/iCount/blob/master/iCount/genomes/segment.py>).

For data representation, the iCLIP signal was normalized by sequencing deep and millions of tags (CPM) and binned per nucleotide. Coverage tracks were generated using deepTools<sup>78</sup>, and meta-gene plots were drawn using normalized coverages between the transcriptional start site (TSS) and the transcriptional termination site (TTS) of genes, defined by the GENCODE<sup>70</sup> annotation from hg38 human reference genome, using 100 nucleotide bins. Normalized iCLIP data was also plotted in metagenes, in  $\pm 10$  kb windows around genomic TSS positions.

The significant crosslink signal was normalized by sequencing deep and millions of tags (CPM). Significant contact sites were identified as iCLIP peaks, using the iCount peak function, based on a false discovery rate (FDR)  $< 0.05$  comparing specific sites within a window of three nucleotides with randomized data (100 permutations) and within co-transcribed regions (<https://github.com/tomazc/iCount/blob/master/iCount/analysis/peaks.py>).

To identify RNA motifs mediating ORC1 binding, iCLIP peaks with more than five crosslinks per nucleotide were slop 100 nucleotides both sides and submitted to MEME motif finding algorithm<sup>81</sup> with parameters “-rna -maxw 6 -maxsize 1000000000 -neg”, comparing positive sites with negative randomized data from SNS-seq

experiments and from homologous genomic regions. Although several motif sizes were tested, 6-mers appeared to be the most reliable. G4 predictions were also obtained, by using TetraplexFinder from the QuadBase2 web server ([quadbase.igib.res.in](http://quadbase.igib.res.in))<sup>82</sup>, using pre-set motif configuration (medium stringency G3 L1-7, Greedy search algorithm, Bulge size = 0) and search “+” strand only. Statistical significance was analyzed by two-proportions *z*-test.

ORC1 binding preferences for small nuclear RNAs (snoRNAs) annotated hg38 reference genome, downloaded from GENCODE<sup>70</sup>, were separately and further assessed in experimental snoDB database v.1.2.1 (<http://scottgroup.med.usherbrooke.ca/snoDB/> downloaded in December 2020).

### ORC1 interactome analysis (iCLIP-RIP comparison, genomic data, and MEME motifs)

SAMtools<sup>83</sup> and Bedops<sup>84</sup> were used to do different types of operations with genomic data, and the UCSC liftOver tool was used to convert coordinates between different genome versions.

Effective ORC1 protein–RNA contact sites from iCLIP were pooled overlapping Ensembl IDs from iCLIP peaks ( $> 5$  crosslink sites and  $< 0.05$  FDR) and RIP-seq technique ( $\log_2$  fold change  $> 1$  and *p* value  $< 0.05$ ) based on annotation of hg38 human reference genome, downloaded from GENCODE<sup>70</sup>, defining ORC1 iCLIP-RNAs, RIP-RNAs, and ORC1-RNAs (union iCLIP and RIP) or HC ORC1-RNAs (overlap iCLIP and RIP) (Supplementary Data 2). The hypergeometric test confirmed the significance of the overlap and the union of both techniques, also when applying different iCLIP cut-offs. Furthermore, to study general agreement between ORC1 RIP and iCLIP in identifying the same groups of transcripts, ORC1 bound and not bound groups of transcripts were defined by iCLIP CTPM parting from general transcriptomic data in HCT116 cells<sup>69</sup>. Statistical differences between these two groups of transcripts in terms of CTPM and RIP enrichment ( $\log_2$  fold change) were analyzed by unpaired two-tailed Student’s *t*-test.

Biotypes and gene length of different groups of genes (ORC1-RNAs, HC ORC1-RNAs, RIP-RNAs, and iCLIP-RNAs) were obtained from Ensembl BioMart<sup>67</sup>, and HCT116 expression data were obtained from Array Express<sup>85</sup> study E-MTAB-2770 (RNA-seq of 934 human cancer cell lines from the Broad-Novartis Cancer Cell Line Encyclopedia<sup>86</sup>). Negative controls of ORC1-RNAs and HC ORC1-RNAs consisted in an equal number of RNAs with  $\log_2$  fold change  $\leq -0.25$  and *p* adj  $< 0.5$ , or with  $-0.16 < \log_2$  fold change  $< 0.16$  in ORC1 RIP-seq experiment, to consider genes with no ORC1 RNA-binding, with expression in HCT116 cells. Comparisons of gene length and expression between the different groups of genes were statistically interrogated by unpaired two-tailed Student’s *t*-test. The same controls were used to determine whether ORC1-RNAs and HC ORC1-RNAs interact with each other in the 3D nucleus with different frequencies than controls, and Hi-C data of HCT116 cells were obtained from GEO series accession number GSE104333<sup>87</sup> (untreated synchronized combined MAPQ  $\geq 30$ ). Juicer<sup>88</sup> tools command dump was used to extract data from the.hic file and obtain KR-normalized intra-chromosomal contact information at 100 kb resolution. Statistical significance in terms of Hi-C contacts between control or genes of interest was analyzed by a two-proportions *z*-test.

To establish ORC1-RNAs localization within early or late replicating regions of the genome, Repli-seq data for HCT116 cells were obtained from ReplicationDomain (<https://www2.replicationdomain.com/index.php>)<sup>89</sup> database, hg19 human reference genome (Int90617792 and Int97243322). Significant enrichment of ORC1-RNA and HC ORC1-RNA genes for early-replicating regions was assessed by hypergeometric test, considering the universe of early and late replicating regions genome-wide.

To detect enriched sequence motifs through entire ORC1-bound RNAs, sequences of the largest transcripts were extracted

from the list of candidate genes. Input datasets of RNA sequences were run to find significantly enriched motifs on any length with MEME<sup>91</sup>, with respect to a control dataset of the same number of transcript sequences (option `-neg`) by using the differential enrichment objective function. Negative controls consisted of the same number of randomly selected transcripts from ORC1 RIP-seq data, thus expressed in HCT116 cells.

### SNS-seq analysis

The quality of the sequencing files was assessed with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Bowtie2<sup>77</sup> was used to align the reads to the human hg19 reference genome, downloaded from Ensembl<sup>67</sup>, and Picard (<http://broadinstitute.github.io/picard>) was used to remove duplicate reads.

To compare SNS-seq in control vs anti-GAA cells, statistical analyses were done for each of the experimental replicates, by considering the normalized raw sequencing signal (bamCoverage, RPKM normalization) at TSS ( $\pm 7.5$  kb) genomic positions. The average signal was divided in bins of 10 bp length, with computeMatrix from deeptools, and analyzed by paired *t*-test (Euclidean distances of 1.471843 and 2.139965, for replicates 1 and 2, respectively; Fold change ASO anti-GAA vs ASO control of 0.9872622 and 0.9704698, for replicates 1 and 2, respectively; *p* value 8.85552e-36 and 6.840894e-167 for replicates 1 and 2, respectively).

To compare SNS-seq experiments in WT vs MUT-ORC1 expressing HCT116 cells, we merged raw sequencing signal (bamCoverage, RPKM normalization) at TSS ( $\pm 5$  kb) genomic positions, from the two experimental replicates. The average signal was divided in bins of 10 bp length, with computeMatrix from deeptools, and compared by paired *t*-test (Euclidean distance of 2.96711709492902; Fold change WT vs MUT 1.059799; *p* value 2.2e-16).

Origin peaks were determined following the analysis pipeline in<sup>77</sup>, that uses two peak callers: MACS2<sup>76</sup> with a threshold of  $q = 0.1$  to identify narrow peaks, and EPIC2<sup>90</sup> with FDR = 0.1, to detect diffuse peaks. To obtain the common peaks detected with both methods, we used intersectBed from BedTools<sup>91</sup>, with parameters `-wa -u` (non-reciprocal, report any overlapping features). The final set of peaks considered for subsequent analyses includes those common MACS2 + EPIC2 peaks that are present in at least two out of the three SNS-seq replicates in the wild-type HCT116 samples, and in both replicates anti-GAA and control ASOs.

SNS-seq identified 37725 origins that were consistent with previously published origin mapping in other cell types<sup>13</sup>, since 63% of the HCT116 origins overlapped with those in quantiles Q1 and Q2 of the ten defined by the mentioned study, which represent the most robust origins with the highest conservation among cell types (called core origins)<sup>13</sup>. Significant enrichment of SNS-seq peaks at TSSs of ORC1-RNA genes, also visualized in metagenes of normalized reads, was assessed by a hypergeometric test, considering the presence of SNS-seq peaks at TSS of genes in the entire genome. To compare anti-GAA samples against control samples, peaks were divided into different groups: present only in control replicates, present only in anti-GAA replicates, common peaks (those overlapping anti-GAA and control peaks), and peaks differentially bound in anti-GAA or in control. Differentially bound peaks were determined with DiffBind<sup>92</sup> (v.2.10.0) and DESeq2<sup>68</sup> (v.1.22.1) R packages. ChiPseeker<sup>93</sup> was used for SNS-seq peak annotation, comparison, and visualization in heatmaps around TSS positions (13108 peaks in ASO control; 12198 peaks in ASO anti-GAA).

Differences in SNS-seq (ASO control vs ASO anti-GAA and WT-ORC1 vs MUT-ORC1) in combination with other data (RNA-seq ASO control vs ASO anti-GAA or ORC1 iCLIP) were assessed by Gene Set Enrichment analyses (GSEA, see Combined data and correlation analyses section).

### Combined data and correlation analyses (ORC1 iCLIP, RIP-seq, SNS-seq, RNA-seq)

ChIP-seq, SNS-seq, and RNA-seq data were normalized by RPKM, in non-overlapping bins of 10 nucleotides. Genome-wide tracks of normalized ORC1 iCLIP (crosslinks and iCLIP peaks), SNS-seq, ORC1 RIP-seq, and public ChIP-seq data<sup>10</sup>, were visualized and plotted in the IGV (integrative genomics viewer) browser<sup>94</sup>. Normalized sequencing data was also presented in coverage tracks and metagenes, using deepTools<sup>78</sup>, heatmaps, and violin or bar plots.

To study the correlation between replication origins and chromatin marks at TSSs of genes encoding for ORC1-RNAs, normalized ChIP-seq, SNS-seq, and RNA-seq data were represented in a correlation heatmap, with associated Spearman's Correlation values.

ORC1 RNA-binding iCLIP data, normalized by sequencing depth and millions of tags and binned per nucleotide (CTPM), was used to define the division of ORC1-RNA genes (union RIP-seq and iCLIP—see Genomic data analysis section) in 6 quantiles (Q1 to Q6), defining levels of direct ORC1-RNA binding. Having determined iCLIP-defined groups of genes, normalized sequencing data was represented and subjected to statistical analyses, to study the correlation between replication origins, chromatin marks, and ORC1 RNA-binding. Statistical differences between quantiles of genes in terms of RNA levels (RNA-seq), SNS-seq, and ChIP-seq, represented in profile, heatmap, violin, or bar plots, were analyzed by unpaired two-tailed Student's *t*-test.

Gene set enrichment analysis (GSEA) analyses were performed using fgsea<sup>95</sup> (v.1.22.0) R package with 10,000 permutations to calculate statistical significance. Genes were filtered ranked according to their log<sub>2</sub> fold change between anti-GAA vs Control RNA-seq samples, or between the signal across the TSS ( $\pm 5$  Kb) in WT-ORC1 against MUT-ORC1 SNS-seq replicates. In both cases, genes with significant (adjusted *p* value < 0.05) differences between conditions were considered for generating the ranking lists.

Ranked genes according to their upregulation or downregulation in RNA-seq experiments (ASO control vs ASO anti-GAA) were crossed with a list of genes with reduced SNS-seq signal in the anti-GAA condition. This second list of genes was defined by DiffBind<sup>92</sup> (v.2.10.0), with a cut-off of *p* value < 0.05, to select genes with TSS SNS-seq peaks (see SNS-seq analysis section) enriched in the control condition.

Ranked genes according to their enrichment of SNS-seq signal at TSS (WT vs MUT SNS-seq) were crossed with the list of genes in iCLIP-defined quantiles, individually (Q1, Q2, Q3, Q4, Q5, and Q6) or in combination (Q1 + Q2 + Q3 and Q4 + Q5 + Q6).

### In silico analysis of ORC1 sequence, structure, and conservation

To trace orthologues of human ORC1, sequence datasets for 132 proteomes (Supplementary Data 5) were downloaded from the available databases comprising 53 prokaryotes and 79 eukaryotes.

Homologous sequences of human proteins were identified using Inparanoid<sup>96</sup>, an automatic method that uses pair-wise similarity scores between two proteomes for constructing orthology clusters, calculated using NCBI-Blast. The program was run using default parameters except for the in-paralog confidence cut-off, which we made more stringent (from 0.05 to 0.25). All Inparanoid blasts were run using a threshold *e*-value of 0.01 and different matrices were used in pair-wise comparisons to account for different evolutionary distances: Blossum45 to compare prokaryotes, Blossum62 for eukaryotes, and Blossum80 for comparisons between metazoans.

The L-INS-i model in Mafft<sup>97</sup> was used to build a multiple sequence alignment (MSA) with the ORC1 orthologous proteins from vertebrates and metazoans (Supplementary Table 5). The alignment was visualized using Jalview<sup>98</sup> and its quality was manually checked. Consensus sequences logos were generated with WebLogo<sup>99</sup>. We used MEME Suite web platform<sup>100</sup> to find motifs within the consensus sequence of the vertebrate RNA-binding sequence of ORC1 orthologues and MEME



FIMO<sup>101</sup> to search for the TPR/K motif in the *H. sapiens* genome, Ensembl<sup>57</sup>.

To analyze the domain repertoire of ORC1 orthologues, we ran the Hmmscan program from HMMER 3.2 (hmmer.org)<sup>102</sup> against the Pfam database (version 32, September 2018)<sup>103</sup>. Non-overlapping hits with scores above the conditional e-value threshold of 0.05 were considered significant.

ORC1 protein secondary structure was predicted with PsiPred<sup>104</sup>, and the MetaDisorder server<sup>105</sup> was used to predict intrinsic protein disorder using iPDA<sup>106</sup>, PrDOS<sup>107</sup>, Pdisorder (<http://www.softberry.com/>) and IUPred long<sup>108</sup>.

Phyre2<sup>109</sup> was used to predict the 3D protein structure of human ORC1 and the model was visualized and colored using the PyMOL Molecular Graphics System (Schrödinger, LLC.).

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The raw data for RIP-seq, iCLIP, RNA-seq, and SNS-seq have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number [GSE173452](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173452). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD043232](https://www.ebi.ac.uk/pride/archive/study/SDA1232). Raw data for all figures and Supplementary figures is provided in Supplementary Data files, Supplementary Tables, and Source Data file.

RNA-seq public sequences from untreated HCT116 cells were obtained from GEO series [GSE118051](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118051).

ChIP-seq public sequences were downloaded from the ENCODE portal (<https://www.encodeproject.org/>):

[ENCF457PEW](https://www.encodeproject.org/track/ENCF457PEW) for H3K27me3, [ENCF020CHJ](https://www.encodeproject.org/track/ENCF020CHJ) for H3K9me3, [ENCF531IUP](https://www.encodeproject.org/track/ENCF531IUP) for H3K4me1, [ENCF227RRY](https://www.encodeproject.org/track/ENCF227RRY) for H3K27ac, [ENCF213WKK](https://www.encodeproject.org/track/ENCF213WKK) for H3K4me3, [ENCF059WYR](https://www.encodeproject.org/track/ENCF059WYR) for H3K36me3.

Public Repli-seq data for HCT116 cells were obtained from ReplicationDomain (<https://www2.replicationdomain.com/index.php>) database, Homo sapiens build hg19, files Int90617792 and Int97243322. snoRNAs were confirmed from public dataset snoDB database v.1.2.1 (<http://scottgroup.med.usherbrooke.ca/snoDB/>). Public HCT116 RNA-seq expression data were obtained from Array Express study E-[E-MTAB-2770](https://www.ebi.ac.uk/arrayexpress/study/E-MTAB-2770). Public Hi-C data of HCT116 cells were obtained from GEO series [GSE104333](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104333) (untreated synchronized combined MAPQ  $\geq 30$ ). Human OCR1 protein post-translational modification sites were obtained with PhosphositePlus (v.6.5.9.3) (<https://www.phosphosite.org/>). Pfam database (version 32, September 2018) was used for protein domain identification. Source data are provided with this paper.

## References

- Fragkos, M., Ganier, O., Coulombe, P. & Méchali, M. DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015).
- Méndez, J. et al. Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol. Cell* **9**, 481–491 (2002).
- Kara, N., Hossain, M., Prasanth, S. G. & Stillman, B. Orc1 binding to mitotic chromosomes precedes spatial patterning during G1 phase and assembly of the origin recognition complex in human cells. *J. Biol. Chem.* **290**, 12355–12369 (2015).
- Hossain, M., Bhalla, K. & Stillman, B. Multiple, short protein binding motifs in ORC1 and CDC6 control the initiation of DNA replication. *Mol. Cell* **6**, 1951–1969 (2021).
- Marahrens, Y. & Stillman, B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**, 817–823 (1992).
- Vashee, S. et al. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* **17**, 1894–1908 (2003).
- Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell* **17**, 5337–5345 (2006).
- Tardat, M. et al. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat. Cell Biol.* **12**, 1086–1093 (2010).
- Kuo, A. J. et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115–119 (2012).
- Long, H. et al. H2A.Z facilitates licensing and activation of early replication origins. *Nature* **577**, 576–581 (2020).
- Valton, A.-L. et al. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **33**, 732–746 (2014).
- Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
- Akerman, I. et al. A predictable conserved DNA base composition signature defines human core DNA replication origins. *Nat. Commun.* **11**, 4826 (2020).
- Sequeira-Mendes, J. et al. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, e1000446 (2009).
- 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).
- Valenzuela, M. S. et al. Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. *PLoS ONE* **6**, e17308 (2011).
- Mesner, L. D. et al. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Res.* **21**, 377–389 (2011).
- Dellino, G. I. et al. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res.* **23**, 1–11 (2013).
- Langley, A. R., Gräf, S., Smith, J. C. & Krude, T. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic Acids Res.* **44**, 10230–10247 (2016).
- Sugimoto, N., Maehara, K., Yoshida, K., Ohkawa, Y. & Fujita, M. Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. *Nucleic Acids Res.* **46**, 6683–6696 (2018).
- Aze, A., Fragkos, M., Bocquet, S., Cau, J. & Méchali, M. RNAs coordinate nuclear envelope assembly and DNA replication through ELYS recruitment to chromatin. *Nat. Commun.* **8**, 2130 (2017).
- Norseen, J. et al. RNA-dependent recruitment of the origin recognition complex. *EMBO J.* **27**, 3024–3035 (2008).
- Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).



24. Hoshina, S. et al. Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* **288**, 30161–30171 (2013).
25. Duncker, B. P., Chesnokov, I. N. & McConkey, B. J. The origin recognition complex protein family. *Genome Biol.* **10**, 214 (2009).
26. Bellucci, M., Agostini, F., Masin, M. & Tartaglia, G. G. Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445 (2011).
27. Huppertz, I. et al. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
28. Norseen, J., Johnson, F. B. & Lieberman, P. M. Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* **83**, 10336–10346 (2009).
29. Taniguchi, I., Masuyama, K. & Ohno, M. Role of purine-rich exonic splicing enhancers in nuclear retention of pre-mRNAs. *Proc. Natl Acad. Sci. USA* **104**, 13684–13689 (2007).
30. Cayrou, C., Grégoire, D., Coulombe, P., Danis, E. & Méchali, M. Genome-scale identification of active DNA replication origins. *Methods* **57**, 158–164 (2012).
31. Zheng, R. et al. Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. *J. Cell Sci.* **123**, 3734–3744 (2010).
32. Conti, C. et al. Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol. Biol. Cell* **18**, 3059–3067 (2007).
33. Simon, A. C., Sannino, V., Costanzo, V. & Pellegrini, L. Structure of human Cdc45 and implications for CMG helicase function. *Nat. Commun.* **7**, 11638 (2016).
34. Filipowicz, W., Jaskiewicz, L., Kolb, F. A. & Pillai, R. S. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* **15**, 331–341 (2005).
35. Coulombe, P. et al. The ORC ubiquitin ligase OBI1 promotes DNA replication origin firing. *Nat. Commun.* **10**, 2426 (2019).
36. Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* **12**, 283–294 (2011).
37. Parker, M. W. et al. A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife* **8**, e48562 (2019).
38. Henninger, J. E. et al. RNA-mediated feedback control of transcriptional condensates. *Cell* **184**, 207–225.e24 (2021).
39. Prasanth, S. G., Shen, Z., Prasanth, K. V. & Stillman, B. Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. *Proc. Natl Acad. Sci. USA* **107**, 15093–15098 (2010).
40. Popova, V. V., Brechalov, A. V., Georgieva, S. G. & Kopytova, D. V. Nonreplicative functions of the origin recognition complex. *Nucleus* **9**, 460–473 (2018).
41. Kopytova, D. et al. ORC interacts with THSC/TREX-2 and its subunits promote Nxf1 association with mRNP and mRNA export in *Drosophila*. *Nucleic Acids Res.* **44**, 4920–4933 (2016).
42. Beltran, M. et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
43. Chen, G. & Deng, X. Cell synchronization by double thymidine block. *Bio Protoc.* **8**, e2994 (2018).
44. Shoab, M. et al. Histone H4K20 methylation mediated chromatin compaction threshold ensures genome integrity by limiting DNA replication licensing. *Nat. Commun.* **9**, 3704 (2018).
45. Mayer, A. & Churchman, L. S. A detailed protocol for subcellular RNA sequencing (subRNA-seq). *Curr. Protoc. Mol. Biol.* **120**, 4291–42918 (2017).
46. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
47. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
48. Peng, J. & Gygi, S. P. Proteomics: the move to mixtures. *J. Mass Spectrom.* **36**, 1083–1091 (2001).
49. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
50. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292 (2006).
51. Brugiolo, M., Botti, V., Liu, N., Müller-McNicoll, M. & Neugebauer, K. M. Fractionation iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm and cytoplasm. *Nucleic Acids Res.* **45**, 10452–10465 (2017).
52. Blazquez, L. et al. Exon junction complex shapes the transcriptome by repressing recursive splicing. *Mol. Cell* **72**, 496–509 (2018).
53. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).
54. Almeida, R. et al. Chromatin conformation regulates the coordination between DNA replication and transcription. *Nat. Commun.* **9**, 1590 (2018).
55. Scovassi, A. I. & Prosperi, E. Analysis of proliferating cell nuclear antigen (PCNA) associated with DNA. *Methods Mol. Biol.* **314**, 457–475 (2006).
56. Chen, Y. H. et al. ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol. Cell* **58**, 323–338 (2015).
57. Yin, Y., Lee, W. T. C. & Rothenberg, E. Ultrafast data mining of molecular assemblies in multiplexed high-density super-resolution images. *Nat. Commun.* **10**, 1–7 (2019).
58. Cirillo, D. et al. Quantitative predictions of protein interactions with long noncoding RNAs. *Nat. Methods* **14**, 5–6 (2016).
59. Agostini, F. et al. CatRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* **29**, 2928–2930 (2013).
60. Lang, B., Armaos, A. & Tartaglia, G. G. RNAct: protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res.* **47**, D601–D606 (2019).
61. Armaos, A., Colantoni, A., Prioletti, G., Rupert, J. & Tartaglia, G. G. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. *Nucleic Acids Res.* **49**, W72–W79 (2021).
62. Klus, P. et al. The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinformatics* **30**, 1601–1608 (2014).
63. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
64. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
65. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).

68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
69. Gajdušková, P. et al. CDK11 is required for transcription of replication-dependent histone genes. *Nat. Struct. Mol. Biol.* **27**, 500–510 (2020).
70. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
71. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
72. Consortium, E. P. The ENCODE (ENCyclopedia of DNA Elements) project. *Science* **306**, 636–640 (2004).
73. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
74. Picard Tools - By Broad Institute. <https://broadinstitute.github.io/picard/> (Accessed: 31st March 2021).
75. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
76. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
79. Ule, J. et al. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* **8**, e1000530 (2010).
80. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
81. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
82. Dhapola, P. & Chowdhury, S. QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res.* **44**, W277–W283 (2016).
83. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
84. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
85. Athar, A. et al. ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.* **47**, D711–D715 (2019).
86. Barretina, J. et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
87. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
88. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
89. Weddington, N. et al. ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* **9**, 530 (2008).
90. Stovner, E. B. & Sætrom, P. Epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392–4393 (2019).
91. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
92. Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP-Seq peak data. <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>.
93. Yu, G., Wang, L. G. & He, Q. Y. ChIP seeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
94. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
95. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/060012> (2021).
96. Remm, M., Storm, C. E. V. & Sonnhammer, E. L. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
97. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
98. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
99. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
100. Bailey, T. L. et al. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
101. Grant, C. E., Bailey, T. L. & Noble, W. S. Sequence analysis FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
102. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
103. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
104. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
105. Kozłowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111 (2012).
106. Su, C. T., Chen, C. Y. & Hsu, C. M. IPDA: integrated protein disorder analyzer. *Nucleic Acids Res.* **35**, W465–W472 (2007).
107. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464 (2007).
108. Dosztányi, Z., Csizsmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
109. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

## Acknowledgements

We thank Dr. Bruce Stillman for the kind gift of anti-ORC1 antibody and ORC1 constructs; Dr. Juan José Lasarte for helping with protein purification; Amaya Abad for NGS library preparation and Dr. Francesco Marchese and Tomás Aragón for critical reading of the manuscript. This work was supported by grants to M.H.: PID2020-113683GB-I00/funded by MCIN/AEI/10.13039/501100011033, La Caixa Foundation [LCF/PR/HR21/00176], European Research Council Consolidator 771425 and Worldwide Cancer Research grants 20-0204; and to G.G.T.: European Research Council [RIBOMYLOME\_309545 and ASTRA\_855923]. A.M.M. was supported by Ph.D. fellowship from Ministerio de Economía y Competitividad REF: BES2015-074569 cofounded by FSE.

## Author contributions

A.M.M. designed, performed, and analyzed most of the experiments; E.G. performed computational and statistical analyses; I.R.d.l.M. performed iCLIP and RIP-seq data analyses and metanalyses; A.Arc performed data analyses and protein modeling and predictions; L.S. performed experimental work; J.G. assisted with experimental work; L.B. performed iCLIP experiments supervised by J.U.; W.T.C.L. and D.G.

performed STORM experiments in collaboration with A.M.M. and supervised by E.R.; A.S. assisted with experimental work; S.H. purified recombinant ORC1 protein supervised by S.W.; A. Arm and G.G.T. performed catRapid analyses; M.G. supervised A.M.M. in SNS-seq and DNA fiber assays; M.H. conceived and designed the study, supervised the work, obtained funding, and wrote the manuscript with the input from all authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40105-3>.

**Correspondence** and requests for materials should be addressed to Maite Huarte.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023