Specific origin selection and excess functional MCM2-7 loading in ORC-deficient cells.

Yoshiyuki Shibata[1, 4], Mihaela Peycheva [2, 4], Etsuko Shibata[1], Daniel Malzl [2], Rushad Pavri[2, 3] and Anindya Dutta[1, 5]

1: Dept. of Genetics, University of Alabama at Birmingham
2: Research Institute of Molecular Pathology (IMP), Campus-Vienna-Biocenter-1, Vienna Biocenter, 1030 Vienna, Austria
3. Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, King's College London, London SE1 9RT, United Kingdom

4: Co-first authors
5: Corresponding author: duttaa@uab.edu

ABSTRACT

The six subunit Origin Recognition Complex (ORC) loads excess MCM2-7 on chromosomes to promote initiation of DNA replication and is believed to be important for origin specification. Mapping of origins in cancer cell lines engineered to delete three of the subunits, *ORC1*, *ORC2* or *ORC5* shows that specific origins are still used and are mostly at the same sites in the genome as in wild type cells. The few hundred origins that were up-regulated in the absence of ORC suggest that GC/TA skewness and simple repeat sequences facilitate, but are not essential for, origin selection in the absence of the six-subunit ORC. Despite the lack of ORC, excess MCM2-7 is still loaded at comparable rates in G1 phase to license reserve origins and is also repeatedly loaded in the same S phase to permit re-replication. Thus, origin specification and excess MCM2-7 loading on origins do not require the six-subunit ORC in human cancer cell lines.

INTRODUCTION

The six-subunit ORC is essential for DNA replication in eukaryotes because it helps load the double hexameric MCM2-7 on chromatin and this MCM2-7 hexamer forms the core of the replicative helicase that unwinds origins of replication and initiates DNA replication (1-5). The loading of the MCM2-7 double-hexamer around DNA is also called licensing of origins or formation of the pre-Replicative complex (preRC). Because ORC is among the first replication initiation factors to bind DNA and then help load MCM2-7 on DNA, ORC plays an important role in specifying the locations of origins in most eukaryotes. ORC also loads an excess of MCM2-7 on chromatin relative to what is required to fire the normal complement of ~50,000 origins per cell cycle, and this MCM2-7 excess is critical for the firing of reserve origins when replication is interrupted by the premature termination of replication forks emerging from the first set of origins (6, 7). Thus, this reserve origin firing is a measure of whether excess functional MCM2-7 is loaded in the cells. Finally, the licensing and re-licensing of MCM2-7 on origins when certain cell-cycle pathways are de-regulated, leads to re-replication of the genome, and this is another measure of the robustness of functional MCM2-7 loading (8-11).

There is solid evidence that ORC is essential for licensing origins in yeasts, flies and mammalian cells and *in vitro* (2, 3, 12-19). In addition, hypomorphic mutations in ORC subunits lead to microcephalic primordial dwarfism resembling Meier-Gorlin syndrome (20-22). In this syndrome, all tissues are not affected, but specific tissues like the patella, ears and the skeleton are smaller than normal. This phenotype is interpreted as consistent with an essential role of ORC in DNA replication and cell proliferation, though it is difficult to explain why a mutation in a complex essential for DNA replication and cell proliferation globally has such tissue-specific effects.

On the other hand, we have generated cancer cell lines with at least three of the ORC subunits, ORC1, ORC2 or ORC5 mutationally inactivated by CRISPR-Cas9 and the cells are still viable and proliferate with DNA replication (23, 24). The mutational inactivation results in undetectable levels of ORC1 and ORC5 in the *ORC1Δ* and *ORC5Δ* cells, respectively, and <0.1% of a truncated ORC2 (~150 molecules/cell) in the *ORC2Δ* cells (25). We are thus faced with a paradox: while ORC is required for DNA replication in most instances, how do these cancer cell lines replicate and survive without the six-subunit ORC? One possibility is that in the absence of ORC an imprecise loading of MCM2-7 leads to the licensing of diffuse, non-specific origins of replication, so that it would be impossible to map precise origins of replication in a population of cells. Also, the MCM2-7 loading could be so inefficient that reserve origin

licensing or repeated origin licensing in the same S phase are not possible. It is worth noting that some of the cell lines retain different subcomplexes of ORC on chromatin. In the *ORC2Δ* cells ORC1 and ORC6 alone are on the chromatin, in the *ORC1Δ* cells, ORC2-6 load on chromatin at near wild type levels, and the *ORC5Δ* cells allow ORC1, 2, 3 and 6 to associate with chromatin at near wild type levels (23, 24). Thus, another possibility is that different subcomplexes of ORC specify different subsets or origins and have differing ability to load excess MCM2-7.

We therefore decided to stringently test whether, (1) no origins or a different set of origins are selected in the ORC knockout cells or (2) a highly compromised licensing machinery without ORC can load enough MCM2-7 at a reasonable rate in G1 for replication from reserve origins or for re-replication in S phase.

MATERIALS AND METHODS

Cell lines and cell culture.

HCT116 p53−/− cells (26) is a generous gift from Fred Bunz, Johns Hopkins. Cells were maintained in McCoy's 5A-modified medium supplemented with 10 % fetal bovine serum. HCT116 p53-/- *ORC2Δ* (also referred to as ORC2KO), *ORC1Δ* (ORC1KO), or *ORC5Δ* (ORC5KO) cells were generated using CRISPR-Cas9 technologies (23, 24).

Subcellular fractionation and Western blots.

Chromatin fractionation was performed as previously described (27).
In brief, Cells were lysed in buffer A (10mM HEPES [pH7.9], 10 % glycerol, 1mM DTT, Protease inhibitor cocktail) with 0.1% Triton X100, incubated for 8 min on ice and centrifuged at 1300g for 5min. Pellets were dissolved in buffer B (3 mM EDTA, 0.2 mM, EGTA, 1 mM DTT, protease inhibitor cocktail) and incubated for 30 min on ice. Cell lysates were centrifuged at 1700g for 5 min and supernatants were removed. Pellets were lysed in 2 x SDS sample buffer and sonicated (P3). Antibodies used were as follows. ORC1(4731, Cell Signaling Technology), ORC2 (28), ORC5 (29), HSP90 (sc-13119, Santa Cruz Biotechnology), LaminA/C (2032, Cell Signaling Technology), MCM2 BM28 (610700, BD), MCM7 (sc-9966, Santa Cruz Biotechnology), MCM6 (sc-9843, Santa Cruz Biotechnology).

Chromatin flow cytometry

The method has been described in detail in (30). Anti- MCM2 BM28 (610700, BD), anti-Mouse IgG secondary antibody, Alexa Fluor™ 488 (A-11029, Thermo Fisher), DAPI (32670, Millipore Sigma) were used. Fluorescence was measured using a BD LSRFortessa and data was analyzed using Flow Jo. DAPI with secondary antibody alone was used to determine MCM2 positive population.

Ergodic rate analysis

The method has been described in detail in (30) except Flow Jo. was used to export G1-MCM^{DNA} (number of cells with MCM loaded on DNA in G1 phase) raw data to CSV file (Supplementary Table S2). G1-MCM^{DNA} raw data were divided into 10 equal sized bins between the lowest and highest MCM value and used to calculate MCM loading rate. For a mean MCM loading rate, all 10 MCM loading rates were averaged.

The MCM loading rate was calculated using MATLAB (MathWorks).
The rate calculation was

W_n (MCM loading rate in bin N) = $\alpha(2-F)/f_n$

$\alpha = L_n(2)$/doubling time
F = number of G1-MCM^{DNA} cells/total number of cells in sample
f_n = number of cells in bin n/number of G1-MCM^{DNA} cells

Sample MATLAB code was as follows

```
alpha_WT= log(2)/19.5;
F_WT_1 = 0.288143;
%calculate lowest and highest MCM values%
maxMCM_WT_1 = max(WT_1(:,1));
minMCM_WT_1 = min(WT_1(:,1));
%create histogram with 10 bins and specified first and last bin limits%
h10_WT_1 = histogram(WT_1(:,1), 'NumBins', 10, 'BinLimits', [minMCM_WT_1,
maxMCM_WT_1]);
%calculate fn within each bin%
totalf_WT_1 = size(WT_1);
fn10_WT_1=(h10_WT_1.Values)/totalf_WT_1(1,1);
%calculate mean w%
w10mean_WT_1 = mean(alpha_WT.*(2- F_WT_1)./fn10_WT_1)
```

Hydroxyurea sensitivity and molecular combing.
2500 or 10,000 cells were seeded into 6-well plates the day before addition of the indicated amounts of hydroxyurea. After 1 week, cells were fixed and stained in a solution of 20% formaldehyde, 80% methanol, and 0.25% crystal violet, and rinsed with deionized water. Images were quantified with Image J.

For molecular combing assay, cells were treated with 200 μM of hydroxyurea for 4 hrs., then pulse- labeled with 100 μM 5-chlorodeoxyuridine (IdU) followed by 250 μM 5-iododeoxyuridine (CIdU) for 30 min each. Cells were embedded into agarose plugs and DNA was extracted. DNA was combed on silianized coverslips (Genomic Vision), dehydrated at 65 °C for 4 hrs., denatured in 0.5M NaOH and NaCl for 8 min and dehydrated in a series of ethanol bath. CIdU or IdU was immunofluorescently detected using an anti-BrdU antibody that recognizes CIdU (MCA2060, Bio-Rad Laboratories) or an anti-BrdU antibody that recognizes IdU (347583, BD Biosciences).
Images were acquired on a Zeiss AxioObserver Z1, 63 X objective and DNA lengths were measured
using ZEN software. (P value < 0.01, two-sided Wilcoxon rank sum test for two samples; N = number of tracks counted. P: Statistical significance of any difference between HU untreated and HU treated cell.)

Re-replication and FACS.

Cells were treated with 0.3 μM of MLN4924 for 20 hrs. and fixed with 70 % Ethanol. Cells were suspended in 20 μg/ml of propidium iodide (PI) in PBS with 100 μg/ml RNase A. Fluorescence was measured using a BD LSRFortessa and cell cycle phase was estimated from histograms using Flow Jo.

SNS-seq and mapping of origins to genome browser.
These methods have been described in detail in (31)

Statistical analysis

All data analyses were conducted in R 4.1.3 with the `tidyverse`(32), `Biostrings`(33), `BSgenome.Hsapiens.UCSC.hg38`(34), `coin`(35), `effsize`(36), `ggplot2`(37), `rcompanion`(38) packages. If the Shapiro-Wilk test showed that the data were normally distributed, the Welch's t-test was applied to test for mean differences, and the Hedges'g were calculated to measure the effect sizes. If the data were not normally distributed, or if the number of data points exceeded 5000, the Wilcoxon rank sum test was applied, and the Cliff's delta was calculated to measure the effect sizes. In the case of multiple comparisons by repeated testing, the Benjamini-Hochberg procedure was applied to obtain adjusted p-values. All statistical tests are two-sided.

Permutation test

The analysis was performed in R (version 4.1.3), using the `tidyverse`(32), `regioneR` (39) and `plyranges`(40) packages. A permutation test was used to assess the enrichment of origins within the targets. The position of the origins was randomized 1000 times within the genome while retaining canonical chromosomes and avoiding overlaps with masked regions. The number of overlaps between the original origins and the targets was calculated. This calculation was also applied to each randomized set of origins. Statistical significance was assessed by comparing the observed overlap count in the original data to the distribution of overlap counts obtained from the randomized sets. To quantify the enrichment of origins within the target, fold enrichment (ratio of observed overlap to mean randomized overlap) and z-score were calculated.

Venn diagram generation

Venn diagrams and a UpSet plot were generated using R (version 4.1.3), with the `tidyverse`(32), ` VennDiagram`(41) and UpSetR (42) packages. The Venn diagram was used to visualize the overlap between origins identified in various samples. The VennDiagram `findOverlapsOfPeaks` function was used to identify overlapping regions, and the minimum number of involved peaks in each overlapping group included in the overlapping counts (`connectedPeaks="min"`).

Peak overlaps

Overlap of up-regulated or down-regulated peaks, and peak merging was performed using bedtools v2.29.0. The origins up- or down-regulated in the knockout (relative to WT) were intersected with each other. If they overlapped, they were merged.

Data availability

The raw data for SNS-seq have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE278065.

RESULTS

Survival of ORC knockout cells without re-expressing ORC. If a cell survives at a disadvantage because of the absence of ORC subunits, one possibility is that the protein (or a truncated, functional protein) is somehow re-expressed to allow the cells to survive. This was not the case for the ORC knockout cells. The levels of ORC1, ORC2 and ORC5 in the *ORC1Δ*, *ORC2Δ* and *ORC5Δ* cells were still undetectable on Western blots of cell lysates even though these cell lines were derived years ago, and each of them have been cumulatively passaged for several months (**Figure 1A)**. The deleted ORC subunits remained undetectable on the chromatin fraction (**Figure 1B)**.

When we turned to MCM2-7 association with chromatin, there was an interesting difference. MCM2-7 was significantly decreased on the chromatin in the first report of *ORC1Δ* and *ORC2Δ* cells (24), but not the *ORC5Δ* cells (23). We now see that the *ORC1Δ* and *ORC2Δ* cells have adapted during passage because we see more MCM2-7 is associated with the chromatin, with the levels in the *ORC2Δ* cells approaching near normal level, and this has happened without re-expressing the missing ORC subunits (**Figure 1A, B**).

The majority of the origins map to WT sites in the ORC knockout cells. SNS-seq is a well- accepted method to identify origins of replication in a population of cells (43-46). All four cell lines yielded robust SNS-seq peaks (**Figure 2A**).

Origin selection in mammalian cells is notoriously stochastic, so that origins in different cell lines vary significantly, although some overlap is detectable (47). Based on our analysis of the published literature, we believe that Core origins described from 19 human cell samples by Akerman et al. (48, 49) constitute the most reliable set of origins to use as a benchmark. Although the Core origins were not derived from HCT116 colon cancer cells, about 48% of the origins mapped in Wild type HCT116 cells overlapped with the core origins (**Figure 2B**). Nearly 60% of the origins mapped in WT cells were detected in all three ORC knockout cells (**Figure 2C**). It was equally interesting that over 90% of the origins mapped in WT cells were seen in one or more of the ORC knockout cells. This degree of overlap in origins between the four cells suggests that cells are mainly using the same set of origins in the presence or absence of ORC or of some of the subunits of ORC on chromatin.

We have reported that loss of one subunit of ORC decreases the loading of the other subunits in a very specific manner (23, 24). The *ORC2Δ* cells had a great decrease of 4 subunits (ORC2, ORC3, ORC4 and ORC5) and a moderate decrease of 1 subunit (ORC1) on chromatin. The *ORC5Δ* cells had a great decrease of 2 subunits (ORC4 and ORC5) on chromatin. The *ORC1Δ* cells showed a decrease of only 1 subunit (ORC1) on chromatin. The number of origins that were common between WT cells and the *ORC1Δ, ORC2Δ* and *ORC5Δ* cells were 21,527, 23,787 and 24,974, respectively. Thus, the different subcomplexes (varying from 2 to 5 subunits) on the chromatin did not select very different subsets of origins.

Subsets of origins are commonly up-regulated or down-regulated after knockout of individual ORC subunits. We next focused on origins that were selectively upregulated (ORC KO-up) or downregulated (ORC KO-down) in the knockout cells, relative to the WT cells (**Figure 3A**). The Venn diagrams show that 4000 origins were commonly up-regulated in the three different knockout cells, while 2500 origins were commonly down-regulated in all three knockout cells (**Figure 3B**).

Properties of the origin sets. The DNA near origins has been reported to have high nucleotide distribution skew (50). The width (length) of the origin peaks (**Figure 4A**), GC content (**Figure 4B**) and TA skew (**Figure 4C**) were not greatly different between the Core origins (48, 49) and all origins in WT and ORC knockout cells, even though some of the differences were statistically significant. However, the GC skew (**Figure 4D**) was much higher (effect size ≥0.554) in the experimental HCT116 cell lines, both WT and the knockouts, compared to the Core origins, but there was not much of a difference between all the origins in the knockout and WT cells (**Supplemental Table S1**, lines 3-6).

In contrast, when we analyzed the origins that were up-regulated or down-regulated in the knockout cells, an interesting pattern emerged (**Figure 5 and Supplemental Table S1**). The origins up-regulated in all three ORC subunit knockout cells were longer with lower GC content compared to the whole pool of origins in the knockout cells (**Figure 5A, B**). Strikingly, these origins were markedly increased in absolute (mostly positive) GC skew compared to the whole pool of origins in the knockout cells or the origins downregulated in the knockout cells (**Figure 5C, D**).

Overlap of origins with known genomic annotations. It is known that Core origins are enriched relative to random distribution over exons, promoters of genes (2kb upstream of genes) and CpG islands (**Figure 6A**). These properties are preserved in all four experimentally determined origin sets (WT, and the three knockout cells). A surprising exception is seen for the origins that are upregulated when ORC subunits are deleted, where we do not see the enrichment over promoters, exons, 2 kb downstream of genes or CpG islands. The origins that are downregulated in the ORC1 knockout cells are more enriched than even the core origins over promoters, 5'UTR containing exons and CpG islands, but we are not sure of the significance of this, because this enrichment is not seen with the origins downregulated in the ORC2 knockout or ORC5 knockout cells.

As reported before, origins are also enriched over random expectations in areas that form R-loops and G-quadruplexes (**Figure 6B, C**) (51, 52). This was true for all the origins in the WT and knockout cell lines. The enrichment is still seen but is lower for the origins that are up-regulated in any of the ORC knockout cells, and this lower enrichment at G-quadruplexes is likely due to the lack of enrichment of these origins near genes and promoter.

Overlap of origins with (ATCC)n repeat elements. Repeat elements are often enriched near bacterial and eukaryotic origins (51, 53). We wondered whether any of the classes of origins we have identified are significantly different from the general class of origins in terms of their overlap with repetitive elements. The full set of origins we identified in the WT and knockout cells were enriched in low complexity sequences and is slightly enriched in simple repeats (**Figure 7A, B**). There was not much difference between all the origins in the knockout cells compared to the parental WT cells. The origins up-regulated in the ORC knockout cells were atypical in not being enriched in low complexity sequences but twice as much enriched in

simple repeats compared to all core origins or all origins in WT and knockout cells. When we examined what simple repeats were enriched in these up-regulated origins, (ATCC)n motifs (and permutations or complementary sequences) were specifically highly enriched in the KO-up origins relative to random expectation (**Figure 7C**).

Measurement of functional MCM2-7 loading in the ORC knockout cells. A flow cytometry based approach to measure the MCM2 loading on DNA per cell was developed by Jean Cook and collaborators (30). We used this approach to independently confirm that the same maximal level of MCM2 loading on DNA in G1 was seen in the *ORC1Δ*, *ORC2Δ* and *ORC5Δ* cells as in the WT cells (**Figure 8A**).The FACS plots also show that the normal progressive loss of MCM2 from the chromatin during S phase was seen in the mutant cells in a manner similar to the WT cells. Finally, this FACS approach allows us to quantify the relative MCM loading rate in WT and ORCΔ cells using ergodic rate analysis (30, 54) (**Figure 8B**). There was no significant difference in MCM loading rate between WT and ORC1Δ or ORC2Δ cells and ORC5Δ cells were about twice as fast as WT cells.

A 3-10 fold excess of MCM2-7 double hexamers is loaded on chromosomes at the end of G1 than is necessary to fire the ~50,000 origins that are required to completely replicate the genome in each S phase. Normally, the excess loaded MCM2-7 hexamers are passively displaced from chromatin by replication forks originating in adjoining DNA and replicating through the unfired MCM2-7 hexamers. However, the excess loaded MCM2-7 double hexamers provides a reserve of licensed origins that become important under conditions of replication stress when replication forks stall before traversing their normal distance. The reserve origins in the unreplicated parts of the genome now get activated and this is necessary to replicate the entire genome despite the premature stalling of replication forks (6, 7, 55).

To measure this origin reserve, we put the WT and ORC deletion cells under replication stress by exposing them to low doses of hydroxyurea and following their proliferation by clonogenic assays. The knockout cells were not more sensitive to hydroxyurea compared to WT cells, suggesting that there was a normal complement of reserve origins in the knockout cells (**Figure 9A**). Because normally unfired origins are now initiating DNA replication, the firing of reserve origins decreases the inter-origin distance, as measured by molecular combing. Consistent with this, the inter-origin distance decreased upon exposure of WT cells to HU, and a similar decrease was seen in the cells where *ORC2* or *ORC5* were knocked out (**Figure 9B**). Therefore, excess functional MCM2-7 was still loaded (compared to the minimum required to license ~50,000 origins), and the reserve origin function was intact despite the absence of ORC.

We and others have also reported that an artificial increase in the amount of replication initiation factor, Cdt1, relative to an inhibitor of Cdt1, geminin, leads to repeated licensing and firing of origins on already replicated parts of the genome in the same S phase, resulting in re-replication of the DNA (56, 57). This re-replication leads to the production of cells with greater than G2 DNA content and of course, is dependent on the loading of functional MCM2-7 in S phase on parts of the genome that have already replicated once and lost the MCM2-7 that was loaded in G1. Treating cells with MLN4924 inhibits the degradation of Cdt1, increases the amount of Cdt1 relative to geminin and leads to re-replication (58). Re-replication upon addition of MLN4924, as seen by the appearance of cells with >G2 DNA content, was seen in the *ORC* knockout cells (**Figure 9C**), consistent with the suggestion that functional MCM2-7 can be loaded in these cells even in the absence of ORC. However, there may be a limit to the number

of times MCM2-7 can be re-loaded in the knockout cells, accounting for the 30-40% decrease in the number of cells with >G2 DNA content (**Figure 9D**)

DISCUSSION

There are two important questions we set out to answer. Given the obvious importance of ORC in binding DNA and loading MCM2-7 double hexamers in the vicinity, the first question was whether the cells without an ORC hexamer still choose specific sites as origins and if they do, whether those origins have properties different from the origins chosen in the presence of ORC. The second question examined the robustness of the chromatin loading of functional MCM2-7 hexamers in cells without the six-subunit ORC. While MCM2-7 is clearly associated with the chromatin, and sufficient for executing enough DNA replication for cell survival, is there still an excess of functional MCM2-7 hexamers to license the reserve origins that become important for cells to complete replication and survive in the face of replication stress. Furthermore, can functional MCM2-7 hexamers be loaded repeatedly on the same stretch of DNA to facilitate re-replication in the ORC deficient cells.

Origin properties in the knockout cells. Origin selection in mammalian cells is notoriously stochastic, so that it is difficult to find origins in different cell lines that overlap to a great degree. In fact, origin usage varies from cell-cycle to cell-cycle in the same cells, so that different origins are identified in the same cells in different laboratories or at different points in time. Since we intended to compare origins in four cell lines, with different passage histories, we were not sure how much overlap we would see between origins in the four cells. The results were clear. With nearly 50% of the origins mapped by our experiments in HCT116 cells overlapping with the origins reported in the literature as Core origins, we know that our experiments are identifying *bona fide* origins. The surprising discovery is that the vast majority of the origins in the WT cells were seen in one or more of the three ORC subunit deletion cells. This suggests that (a) origin selection still occurs in the absence of all six subunits of ORC and (b) the ORC deletion cells do not use an entirely new set of origins to replicate and survive in the absence of ORC. Note, that for the *ORC1Δ* and *ORC5Δ* cells we do not have any evidence of any detectable molecule related to the subunit that is deleted. For the *ORC2Δ* cells, there is one report that a truncated molecule of ORC can be detected with one particular ORC2 antibody, but even that molecule is present at <150 molecules per cell, unlikely to be enough to individually license the ~50,000 origins that are active in these cells in each cell cycle. Taken together, these results suggest that ORC has a minimal or no role in the selection of origins in these cancer cells.

None of the subcomplexes of ORC subunits that associated with DNA in the three different knockout cells were very different from the full six-subunit ORC at selecting specific origins. No matter whether five subunits or two subunits of ORC persist on chromatin, the overlap with origins in WT cells is similar. This is consistent with the idea that ORC or subcomplexes of ORC have minimal or no role in selection of origins in these cancer cells.

The origins seen in the ORC knockout cells were generally similar to the origins in WT cells regarding GC content and preference for promoters, CpG islands and G quadruplexes. Thus, the general features that promote origin selection, GC content, promoter, CpG islands and G quadruplexes continue to select origins in the absence of ORC. The higher GC skew in the origins of the ORC knockout cells will be discussed below.

The origins that are activated preferentially in the absence of ORC. Because there were a few hundred origins that were upregulated when ORC subunits were deleted (compared to WT cells), we examined them closer to see whether they had any sequence properties that set them apart from the vast majority of origins. These origins had a higher GC skew and TA skew and a lower overlap with promoters, CpG islands, R loops and G quadruplexes than the origins that overlapped with those in WT cells. As mentioned above, promoters, CpG islands and G quadruplexes have been reported in multiple laboratories to favor origin selection, so the decrease in their enrichment in the origins upregulated in ORC-KO cells suggest that the knockout cells may use slightly different properties of DNA or chromatin to select origins. Indeed, the increase in GC skew, TA skew and unique enrichment of (ATCC)n simple repeats suggest that these features are particularly helpful to license or fire origins in the absence of ORC and are not as helpful when ORC is present to load MCM2-7 double hexamers.

In bacteria and in eukaryotes, it has been reported that the count of G exceeds that of C (GC skew positive) and T over A (TA skew positive) on the leading strand (59-62) and this trend peaks near origins (where the leading strand-lagging strand switch occurs). This increased GC and TA skew near origins has been suggested to be an evolutionary footprint arising from deamination of cytosines on the lagging strand template that is more often left single-stranded during DNA replication. The higher GC skew in the few origins that are up-regulated in the ORC knockout cells either suggests that these origins have a longer evolutionary history or that increased GC skew favors MCM2-7 loading in the absence of ORC.

(ATCC)n repeats can on the opposite strand be represented as (TGGA)n repeats. It is thus very interesting that (TGGA)n repeats have been reported to be reluctant to assemble nucleosomes in *in vitro* selection experiments with a nearly 50% decrease in affinity for histone octamers compared to average DNA (63). Thus, one reason why these sequences promote origin function in the ORC knockout cells could be because they are less chromatinized and so more susceptible to MCM2-7 loading.

None of the ORC subcomplexes on chromatin are better than the others at origin specification. There was a gradient of loading of ORC subunits in the knockout cells. The *ORC2Δ* cells had a great decrease of 4 subunits (ORC2, ORC3, ORC4 and ORC5) and a moderate decrease of 1 subunit (ORC1) on chromatin. The *ORC5Δ* cells had a great decrease of 2 subunits (ORC4 and ORC5) on chromatin. The *ORC1Δ* cells showed a decrease of only 1 subunit (ORC1) on chromatin. Our results show that if the remaining subunits formed subcomplexes that can assist in MCM2-7 loading, they are not much worse than the holocomplex for origin selection. One formal possibility is that ORC1 on its own and ORC2-5 on the other side can independently load MCM2-7 in unusual circumstances when the entire six-subunit ORC is absent. However, our current understanding of ORC structure and function makes this unlikely (2, 3, 18, 19), and we have experimentally determined that hepatocytes without both ORC1 and ORC2 are still capable of loading functional MCM2-7 leading to normal liver development, liver regeneration, endo-reduplication and DNA synthesis (64).

Functional and excess MCM2-7 loading. The ability of the ORC knockout cells to load functional MCM2-7 in excess (for reserve origin function), repeatedly (for re-replication), and at nearly the same kinetics in G1 as in WT cells highlights that these cells have a robust mechanism to load the MCM2-7 double hexamers in the virtual absence of ORC. Till now it could be argued that equal levels of MCM2-7 *association* with chromatin in the knockout cells compared to WT cells does not necessarily equate with comparable levels of *functional loading*

of MCM2-7 double hexamers around the DNA. It could also be argued that an MCM2-7 loading system inactivated by the loss of ORC could barely load enough of the hexamers to permit the one round of replication required for cell cycle and cell proliferation. However, the results with hydroxyurea stressed cells or with stabilized Cdt1, suggest that the MCM2-7 loading machinery is still capable of loading excess functional MCM2-7 and/or repeatedly loading functional MCM2-7 in the same cell-cycle despite the absence of ORC.

ORC-independent MCM2-7 loading. The near normal behavior of the cancer cells in the absence of the six-subunit ORC is a puzzle. In *S. cerevisiae*, ORC binding to the DNA causes it to bend, but it is not clear whether this bending is essential for the subsequent MCM2-7 loading (65). Structural studies on yeast MCM2-7 hexamer complexed with CDT1 and on human MCM2-7 hexamer show that the MCM ring has an open gap between MCM2 and MCM5 (66-68). The size of the gap for human MCM2-7 has been estimated to be 10-15 Angstroms, which is too narrow to admit the 20 Angstrom diameter DNA double helix, but ORC is not needed to separate the MCM2 and 5 subunits to create the gap. Thus, there may be other ways, besides ORC, to either widen the gap to load a double-stranded DNA or to take advantage of naturally occurring denatured parts of DNA to let the MCM2-7 ring load on to a single strand of DNA. In addition, while the current understanding of yeast MCM2-7 loading is that ORC is required to load two MCM2-7 hexamers sequentially on DNA to produce the functional head-to-head double hexamer, human MCM2-7 spontaneously forms head-to-head double hexamers even without DNA or any loading proteins (68). Also, new pathways of ORC-dependent loading of human MCM2-7 on DNA *in vitro* suggest that doubling of the hexamers can happen after independent loading of single hexamers (69). We have now ruled out that the origins used in these cells are very non-specific or very different compared to WT cells. We have also ruled out that the MCM2-7 loading in these cells is not robust and barely sufficient to carry out one round of DNA replication. The results suggest that while ORC may be essential to load MCM2-7 in many physiological conditions, during development, and even in current *in vitro* MCM2-7 loading reactions, the mammalian cancer cell is capable of loading functional and excess MCM2-7 in the absence of the six-subunit ORC.

Cellular adaptation that facilitates ORC-independent MCM2-7 loading. The increased MCM2-7 association with chromatin over months of passage of the *ORC2Δ* and *ORC1Δ* cells is consistent with the presence of an adaptive mechanism that allows better ORC-independent MCM2-7 loading. Such adaptive behavior could indicate that the mechanism involves some form of genetic or epigenetic change. A strong role of epigenetics in origin specification in eukaryotic cells has been reported multiple times (70-74), particularly in view of the enrichment of origins on euchromatin and near promoters, but it has always been assumed that the chromatin decompaction at those sites help recruit ORC to those sites, which then leads to MCM2-7 loading. The fact that those epigenetically open sites are still the preferred sites of origins even in the absence of ORC, makes it likely that the open chromatin is also important for directly enabling MCM2-7 loading and even perhaps origin firing in the absence of ORC.

ACKNOWLEDGEMENT

The work was supported by R01 CA060499 from the NIH to AD and by P29163-B26 from the Austrian Science Fund to RP and core-funding to the IMP by Boehringer Ingelheim. We thank all members of the Dutta and Pavri labs for helpful discussions. MP did the SNS-seq experiments and DM did the initial mapping, YS did all the data analysis, ES and YS did the

experiments in Figures 1, 8 and 9. YS and AD designed the project and wrote the paper. AD and RP supervised the experiments.

FIGURE LEGENDS

Figure 1 Chromatin loading of MCM2-7 in ORC KO cell lines. A**)** Immunoblot of total protein in the ORC KO cell line. B**)** Immunoblot of chromatin-associated proteins in the ORC KO cell lines.

Figure 2 A) An example of a typical origin peak seen in WT and ORC subunit knockout cells as seen with the UCSC browser. Each black colored region indicates an ~1 kb origin from two replicates of each cell line. B) C) Overlap of origin peaks in WT and ORC subunit knockout cells. Numbers in parentheses indicate the total number of peaks in each cell lines. Because multiple peaks from one group may overlap with a single peak from the other, the overlapping peak number is the minimum number of overlapping peaks from either of the two overlapping groups. (B) Overlap between CORE origins and origins from WT cells. (C) Overlap of origins derived from WT and ORC subunit knockout cells.

Figure 3 A) Examples of a typical up-regulated or down-regulated origins seen in ORC subunit knockout cells as seen with the UCSC browser. Each black colored region indicates origins from two replicates of each cell line. The red arrows indicate differential origins. B) Venn Diagram of origins that were up-regulated in ORC subunit knockout cells (compared to WT cells) to show overlap between the knockout cells (left). A similar analysis was done for origins that were down-regulated in the knockout cells (right). Rest as in Fig. 2.

Figure 4 Distribution of (A) length, (B) GC content, (C) TA skew and (D) GC skew of origins in the four cell lines compared to Core origins. Box plot shows median and interquartile range. For the TA skew and GC skew, the values were converted to absolute values. P values determined by two-sided Wilcoxon rank sum test with the Benjamini–Hochberg procedure (FDR correction). * adjusted p-value < 0.05, ** adjusted p-value < 0.01, *** adjusted p-value < 0.001, ns means not significant.

Figure 5 Distribution of (A) length, (B) GC content, (C) TA skew and (D) GC skew of origins in ORC subunit knockout cells to test whether the merged up- or down-regulated origins (KOup_Ovlp or KOdown_Ovlp) in the knockout (relative to WT) are significantly different from the origins in WT cells. The rest as in Figure 4.

Figure 6 Permutation test to determine whether the colocalizations indicated are enriched relative to random expectation. Colocalization of sets of origins and (A) indicated genomic sites, (B) Level 5 R-loops from (75) and (C) G quadruplexes (76) (G4H1.52). 1,000 randomized region sets were obtained for each origin set and intersected with each feature. The mean of the overlap seen with the 1000 random sets is the expected overlap. The X-axis indicates fold enrichment of the experimentally determined overlap with this expected overlap

Figure 7 Permutation test results of colocalization between various origin peaks and (A) low complexity regions, (B) simple repeat regions and (C) simple repeat regions with ATCC motif. The rest as in Figure 6.

Figure 8 A) Top: Flow cytometry plots of MCM2 loading on DNA (Y axis) against DNA content as measured by DAPI (X-axis). The yellow window marked as G1, MCM positive, contains G1 cells that have loaded MCM2 on DNA. Bottom: Distribution of MCM2 staining intensity and cell count from the G1 populations of indicated cell lines. The maximum X-axis value for each curve is a measure of the maximum amount of MCM2 loaded on chromatin at the end of G1 in

that cell line. B) Left: Parameters used to calculate Right: Calculated mean MCM loading rate per hour by ergodic rate analysis (n = 3 replicate), unpaired two tailed t-test

Figure 9 DNA replication in ORC KO cell lines. A) Clonogenic assay of Hydroxyurea treated ORC KO cell lines. Cells were treated with indicated amount of Hydroxyurea for 7 days and colonies detected by Crystal Violet staining (Top). Crystal violet-stained colony density was measured. Data was normalized to the colony density of Day 1 (Bottom). B) DNA molecular combing of cells treated with or without Hydroxyurea for 4 hrs. Box and whiskers plot for inter-origin distance of indicated clones of cells. C) FACS profile of propidium iodide-stained cell treated with MLN4924 for 20 hrs. D) The percentage of cells in G1, G2, >G2 in (C) is plotted.

Supplementary Table S1 Adjusted p-values and effect sizes of comparison tests (Fig. 4 and 5). Bold: Effect size > 0.15 or <-0.15

Supplementary Table S2 MCM2 Fluorescence signals in G1-MCM^{DNA} positive cells used for Ergodic Rate Analysis calculation. Each value is fluorescence intensity for MCM2 in one cell. There are 3 biological replicates for each cell line.

REFERENCES

1. Bell, S. P., andDutta, A. (2002) DNA replication in eukaryotic cells Annu Rev Biochem **71**, 333-374 10.1146/annurev.biochem.71.110601.135425
2. Bleichert, F. (2019) Mechanisms of replication origin licensing: a structural perspective Curr Opin Struct Biol **59**, 195-204 10.1016/j.sbi.2019.08.007
3. Costa, A., andDiffley, J. F. X. (2022) The Initiation of Eukaryotic DNA Replication Annu Rev Biochem **91**, 107-131 10.1146/annurev-biochem-072321-110228
4. Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., andDiffley, J. F. (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing Cell **139**, 719-730 10.1016/j.cell.2009.10.015
5. Yeeles, J. T., Deegan, T. D., Janska, A., Early, A., andDiffley, J. F. (2015) Regulated eukaryotic DNA replication origin firing with purified proteins Nature **519**, 431-435 10.1038/nature14285
6. Alver, R. C., Chadha, G. S., andBlow, J. J. (2014) The contribution of dormant origins to genome stability: from cell biology to human genetics DNA Repair (Amst) **19**, 182-189 10.1016/j.dnarep.2014.03.012
7. Woodward, A. M., Gohler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A. *et al.* (2006) Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress J Cell Biol **173**, 673-683 10.1083/jcb.200602108
8. Hook, S. S., Lin, J. J., andDutta, A. (2007) Mechanisms to control rereplication and implications for cancer Curr Opin Cell Biol **19**, 663-671 10.1016/j.ceb.2007.10.007
9. Blow, J. J., andGillespie, P. J. (2008) Replication licensing and cancer--a fatal entanglement? Nat Rev Cancer **8**, 799-806 10.1038/nrc2500
10. Truong, L. N., andWu, X. (2011) Prevention of DNA re-replication in eukaryotic cells J Mol Cell Biol **3**, 13-22 10.1093/jmcb/mjq052
11. Munoz, S., Bua, S., Rodriguez-Acebes, S., Megias, D., Ortega, S., de Martino, A. *et al.* (2017) In Vivo DNA Re-replication Elicits Lethal Tissue Dysplasias Cell Rep **19**, 928-938 10.1016/j.celrep.2017.04.032
12. Bell, S. P., Kobayashi, R., andStillman, B. (1993) Yeast origin recognition complex functions in transcription silencing and DNA replication Science **262**, 1844-1849 10.1126/science.8266072
13. Foss, M., McNally, F. J., Laurenson, P., andRine, J. (1993) Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae Science **262**, 1838-1844 10.1126/science.8266071
14. Micklem, G., Rowley, A., Harwood, J., Nasmyth, K., andDiffley, J. F. (1993) Yeast origin recognition complex is involved in DNA replication and transcriptional silencing Nature **366**, 87-89 10.1038/366087a0
15. Loo, S., Fox, C. A., Rine, J., Kobayashi, R., Stillman, B., andBell, S. (1995) The origin recognition complex in silencing, cell cycle progression, and DNA replication Mol Biol Cell **6**, 741-756 10.1091/mbc.6.6.741
16. Semple, J. W., Da-Silva, L. F., Jervis, E. J., Ah-Kee, J., Al-Attar, H., Kummer, L. *et al.* (2006) An essential role for Orc6 in DNA replication through maintenance of pre-replicative complexes EMBO J **25**, 5150-5158 10.1038/sj.emboj.7601391
17. Prasanth, S. G., Prasanth, K. V., Siddiqui, K., Spector, D. L., andStillman, B. (2004) Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance EMBO J **23**, 2651-2663 10.1038/sj.emboj.7600255
18. Stillman, B. (2022) The remarkable gymnastics of ORC Elife **11**, 10.7554/eLife.76475

19. Hu, Y., andStillman, B. (2023) Origins of DNA replication in eukaryotes Mol Cell **83**, 352-372 10.1016/j.molcel.2022.12.024
20. Bicknell, L. S., Walker, S., Klingseisen, A., Stiff, T., Leitch, A., Kerzendorfer, C. *et al.* (2011) Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome Nat Genet **43**, 350-355 10.1038/ng.776
21. Bicknell, L. S., Bongers, E. M., Leitch, A., Brown, S., Schoots, J., Harley, M. E. *et al.* (2011) Mutations in the pre-replication complex cause Meier-Gorlin syndrome Nat Genet **43**, 356-359 10.1038/ng.775
22. Guernsey, D. L., Matsuoka, M., Jiang, H., Evans, S., Macgillivray, C., Nightingale, M. *et al.* (2011) Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome Nat Genet **43**, 360-364 10.1038/ng.777
23. Shibata, E., andDutta, A. (2020) A human cancer cell line initiates DNA replication normally in the absence of ORC5 and ORC2 proteins J Biol Chem **295**, 16949-16959 10.1074/jbc.RA120.015450
24. Shibata, E., Kiran, M., Shibata, Y., Singh, S., Kiran, S., andDutta, A. (2016) Two subunits of human ORC are dispensable for DNA replication and proliferation Elife **5**, 10.7554/eLife.19084
25. Chou, H. C., Bhalla, K., Demerdesh, O. E., Klingbeil, O., Hanington, K., Aganezov, S. *et al.* (2021) The human origin recognition complex is essential for pre-RC assembly, mitosis, and maintenance of nuclear structure Elife **10**, 10.7554/eLife.61797
26. Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J. P. *et al.* (1998) Requirement for p53 and p21 to sustain G2 arrest after DNA damage Science **282**, 1497-1501 10.1126/science.282.5393.1497
27. Mendez, J., andStillman, B. (2000) Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis Mol Cell Biol **20**, 8602-8612 10.1128/MCB.20.22.8602-8612.2000
28. Dhar, S. K., Yoshida, K., Machida, Y., Khaira, P., Chaudhuri, B., Wohlschlegel, J. A. *et al.* (2001) Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin Cell **106**, 287-296 10.1016/s0092-8674(01)00458-5
29. Machida, Y. J., Teer, J. K., andDutta, A. (2005) Acute reduction of an origin recognition complex (ORC) subunit in human cells reveals a requirement of ORC for Cdk2 activation J Biol Chem **280**, 27624-27630 10.1074/jbc.M502615200
30. Matson, J. P., Dumitru, R., Coryell, P., Baxley, R. M., Chen, W., Twaroski, K. *et al.* (2017) Rapid DNA replication origin licensing protects stem cell pluripotency Elife **6**, 10.7554/eLife.30473
31. Peycheva, M., Neumann, T., Malzl, D., Nazarova, M., Schoeberl, U. E., andPavri, R. (2022) DNA replication timing directly regulates the frequency of oncogenic chromosomal translocations Science **377**, eabj5502 10.1126/science.abj5502
32. al, W. e. (2019) Welcome to the Tidyverse Journal of Open Source Software **4**, 1686, https://doi.org/10.21105/joss.01686
33. Pagès H, A. P., Gentleman R, DebRoy S (2024) Biostrings: Efficient manipulation of biological strings. R package version 2.72.1 Ed. Biostrings
34. Team, T. B. D. (2020) BSgenome.Hsapiens.UCSC.hg38: Full genomic sequences for Homo sapiens (UCSC genome hg38). R package version 1.4.5. Ed. BSgenome.Hsapiens.UCSC.hg38
35. Hothorn, T., Hornik, K., van de Wiel, M. A., andZeileis, A. (2006) A Lego System for Conditional Inference The American Statistician **60**, 257-263 10.1198/000313006X118430
36. M, T. (2020) effsize: Efficient Effect Size Computation R package version 0.8.1 Ed.

37. Kassambara, A. (2023) ggpubr: 'ggplot2' Based Publication Ready Plots R package version 0.4.0 Ed.
38. Mangiafico, S. S. (2024) Functions to Support Extension Education Program Evaluation version 2.4.36 Ed.
39. Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A., andMalinverni, R. (2016) regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests Bioinformatics **32**, 289-291 10.1093/bioinformatics/btv562
40. Lee, S., Cook, D., andLawrence, M. (2019) plyranges: a grammar of genomic data transformation Genome Biol **20**, 4 10.1186/s13059-018-1597-8
41. Chen, H. (2018) VennDiagram: Generate High-Resolution Venn and Euler Plots R package version 1.6.20. Ed.
42. Conway, J. R., Lex, A., andGehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties Bioinformatics **33**, 2938-2940 10.1093/bioinformatics/btx364
43. Cadoret, J. C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L. *et al.* (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation Proc Natl Acad Sci U S A **105**, 15837-15842 10.1073/pnas.0805208105
44. Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I. *et al.* (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features Genome Res **21**, 1438-1449 10.1101/gr.121830.111
45. Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C. *et al.* (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs Nat Struct Mol Biol **19**, 837-844 10.1038/nsmb.2339
46. Picard, F., Cadoret, J. C., Audit, B., Arneodo, A., Alberti, A., Battail, C. *et al.* (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells PLoS Genet **10**, e1004282 10.1371/journal.pgen.1004282
47. Tian, M., Wang, Z., Su, Z., Shibata, E., Shibata, Y., Dutta, A. *et al.* (2024) Integrative analysis of DNA replication origins and ORC-/MCM-binding sites in human cells reveals a lack of overlap Elife **12**, 10.7554/eLife.89548
48. Guilbaud, G., Murat, P., Wilkes, H. S., Lerner, L. K., Sale, J. E., andKrude, T. (2022) Determination of human DNA replication origin position and efficiency reveals principles of initiation zone organisation Nucleic Acids Res **50**, 7436-7450 10.1093/nar/gkac555
49. Akerman, I., Kasaai, B., Bazarova, A., Sang, P. B., Peiffer, I., Artufel, M. *et al.* (2020) A predictable conserved DNA base composition signature defines human core DNA replication origins Nat Commun **11**, 4826 10.1038/s41467-020-18527-0
50. Langley, A. R., Graf, S., Smith, J. C., andKrude, T. (2016) Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq) Nucleic Acids Res **44**, 10230-10247 10.1093/nar/gkw760
51. Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L. *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity Nat Commun **10**, 3274 10.1038/s41467-019-11104-0
52. Lombrana, R., Almeida, R., Alvarez, A., andGomez, M. (2015) R-loops and initiation of DNA replication in human cells: a missing link? Front Genet **6**, 158 10.3389/fgene.2015.00158

53. Rajewska, M., Wegrzyn, K., andKonieczny, I. (2012) AT-rich region and repeated sequences - the essential elements of replication origins of bacterial replicons FEMS Microbiol Rev **36**, 408-434 10.1111/j.1574-6976.2011.00300.x
54. Kafri, R., Levy, J., Ginzberg, M. B., Oh, S., Lahav, G., andKirschner, M. W. (2013) Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle Nature **494**, 480-483 10.1038/nature11897
55. Karnani, N., andDutta, A. (2011) The effect of the intra-S-phase checkpoint on origins of replication in human cells Genes Dev **25**, 621-633 10.1101/gad.2029711
56. Nishitani, H., Lygerou, Z., Nishimoto, T., andNurse, P. (2000) The Cdt1 protein is required to license DNA for replication in fission yeast Nature **404**, 625-628 10.1038/35007110
57. Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C., andDutta, A. (2000) Inhibition of eukaryotic DNA replication by geminin binding to Cdt1 Science **290**, 2309-2312 10.1126/science.290.5500.2309
58. Lin, J. J., Milhollen, M. A., Smith, P. G., Narayanan, U., andDutta, A. (2010) NEDD8-targeting drug MLN4924 elicits DNA rereplication by stabilizing Cdt1 in S phase, triggering checkpoint activation, apoptosis, and senescence in cancer cells Cancer Res **70**, 10310-10320 10.1158/0008-5472.CAN-10-2062
59. Brodie Of Brodie, E. B., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Thermes, C. *et al.* (2005) From DNA sequence analysis to modeling replication in the human genome Phys Rev Lett **94**, 248103 10.1103/PhysRevLett.94.248103
60. Touchon, M., Nicolay, S., Audit, B., Brodie of Brodie, E. B., d'Aubenton-Carafa, Y., Arneodo, A. *et al.* (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins Proc Natl Acad Sci U S A **102**, 9836-9841 10.1073/pnas.0500577102
61. Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A. *et al.* (2007) Human gene organization driven by the coordination of replication and transcription Genome Res **17**, 1278-1285 10.1101/gr.6533407
62. Foss, E. J., Lichauco, C., Gatbonton-Schwager, T., Gonske, S. J., Lofts, B., Lao, U. *et al.* (2024) Identification of 1600 replication origins in S. cerevisiae Elife **12**, 10.7554/eLife.88087
63. Cao, H., Widlund, H. R., Simonsson, T., andKubista, M. (1998) TGGA repeats impair nucleosome formation J Mol Biol **281**, 253-260 10.1006/jmbi.1998.1925
64. Przanowska, R. K., Chen, Y., Uchida, T.-O., Shibata, E., Hao, X., Rueda, I. S. *et al.* (2024) DNA replication in primary hepatocytes without the six-subunit ORC bioRxiv 2024.2004.2004.588006 10.1101/2024.04.04.588006
65. Li, N., Lam, W. H., Zhai, Y., Cheng, J., Cheng, E., Zhao, Y. *et al.* (2018) Structure of the origin recognition complex bound to DNA replication origin Nature **559**, 217-222 10.1038/s41586-018-0293-x
66. Boskovic, J., Bragado-Nilsson, E., Saligram Prabhakar, B., Yefimenko, I., Martinez-Gago, J., Munoz, S. *et al.* (2016) Molecular architecture of the recombinant human MCM2-7 helicase in complex with nucleotides and DNA Cell Cycle **15**, 2431-2440 10.1080/15384101.2016.1191712
67. Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P. Y., Gao, N. *et al.* (2017) Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer Nat Struct Mol Biol **24**, 300-308 10.1038/nsmb.3374
68. Xu, N., Lin, Q., Tian, H., Liu, C., Wang, P., Suen, C. M. *et al.* (2022) Cryo-EM structure of human hexameric MCM2-7 complex iScience **25**, 104976 10.1016/j.isci.2022.104976
69. Yang, R., Hunker, O., Wise, M., andBleichert, F. (2024) Multiple pathways for licensing human replication origins bioRxiv 10.1101/2024.04.10.588796

70. Ferguson, B. M., andFangman, W. L. (1992) A position effect on the time of replication origin activation in yeast Cell **68**, 333-339 10.1016/0092-8674(92)90474-q
71. Ding, Q., andMacAlpine, D. M. (2011) Defining the replication program through the chromatin landscape Crit Rev Biochem Mol Biol **46**, 165-179 10.3109/10409238.2011.560139
72. Shen, Z., Chakraborty, A., Jain, A., Giri, S., Ha, T., Prasanth, K. V. *et al.* (2012) Dynamic association of ORCA with prereplicative complex components regulates DNA replication initiation Mol Cell Biol **32**, 3107-3120 10.1128/MCB.00362-12
73. Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J. C. *et al.* (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes Genome Res **25**, 1873-1885 10.1101/gr.192799.115
74. Mei, L., Kedziora, K. M., Song, E. A., Purvis, J. E., andCook, J. G. (2022) The consequences of differential origin licensing dynamics in distinct chromatin environments Nucleic Acids Res **50**, 9601-9620 10.1093/nar/gkac003
75. Lin, R., Zhong, X., Zhou, Y., Geng, H., Hu, Q., Huang, Z. *et al.* (2022) R-loopBase: a knowledgebase for genome-wide R-loop formation and regulation Nucleic Acids Res **50**, D303-D315 10.1093/nar/gkab1103
76. Bedrat, A., Lacroix, L., andMergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter Nucleic Acids Res **44**, 1746-1759 10.1093/nar/gkw006

Fig. 1

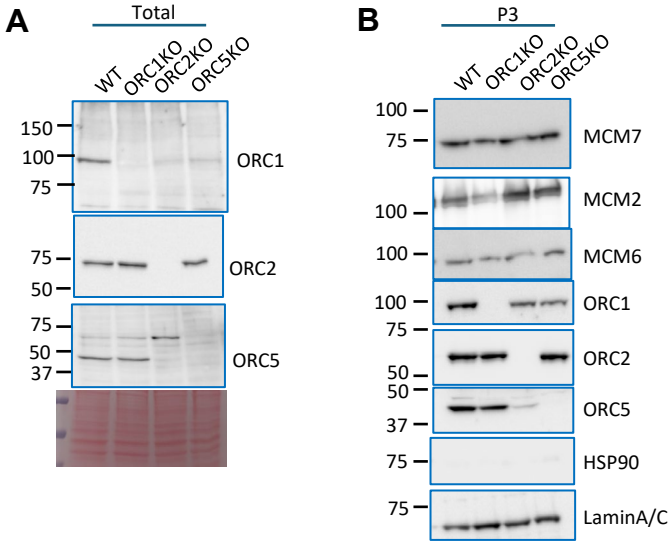
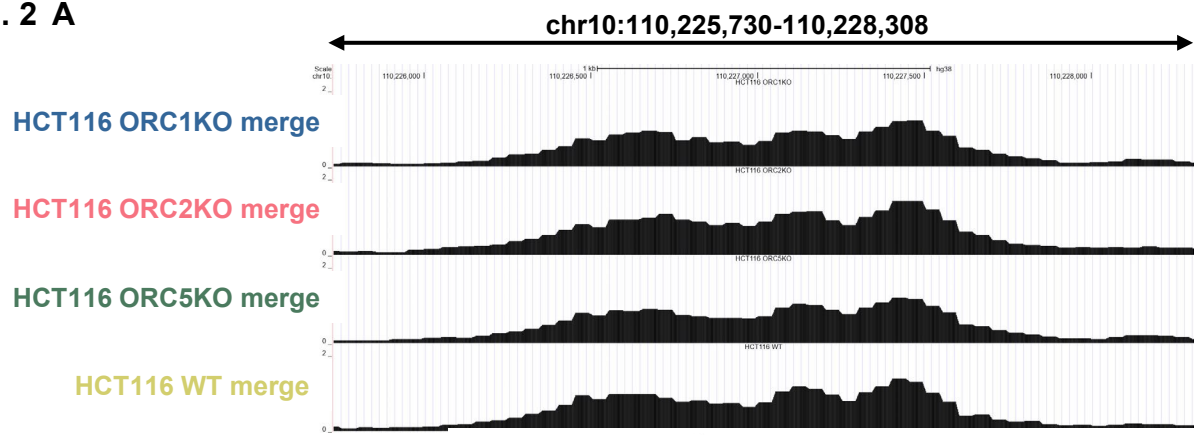
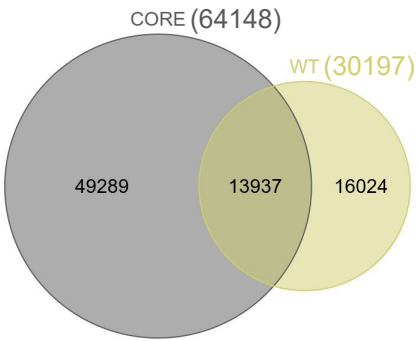

A
Total
WT
ORC1KO
ORC2KO
ORC5KO
150
100
75
ORC1
75
50
ORC2
75
50
37
ORC5
B
P3
WT
ORC1KO
ORC2KO
ORC5KO
100
75
MCM7
100
MCM2
100
MCM6
100
ORC1
75
50
ORC2
50
37
ORC5
75
HSP90
75
LaminA/C

Fig. 2 A


chr10:110,225,730-110,228,308
HCT116 ORC1KO merge
HCT116 ORC2KO merge
HCT116 ORC5KO merge
HCT116 WT merge

B


CORE (64148)
WT (30197)
49289
13937
16024

C

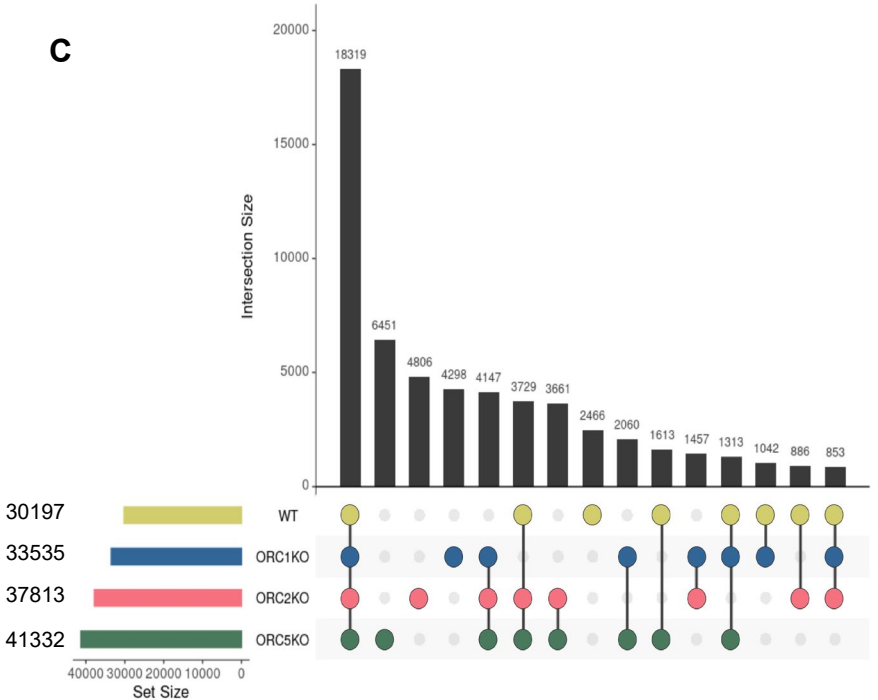

Intersection Size
20000
15000
10000
5000
0
18319
6451
4806
4298
4147
3729
3661
2466
2060
1613
1457
1313
1042
886
853
30197
33535
37813
41332
WT
ORC1KO
ORC2KO
ORC5KO
40000 30000 20000 10000 0
Set Size

Fig. 3

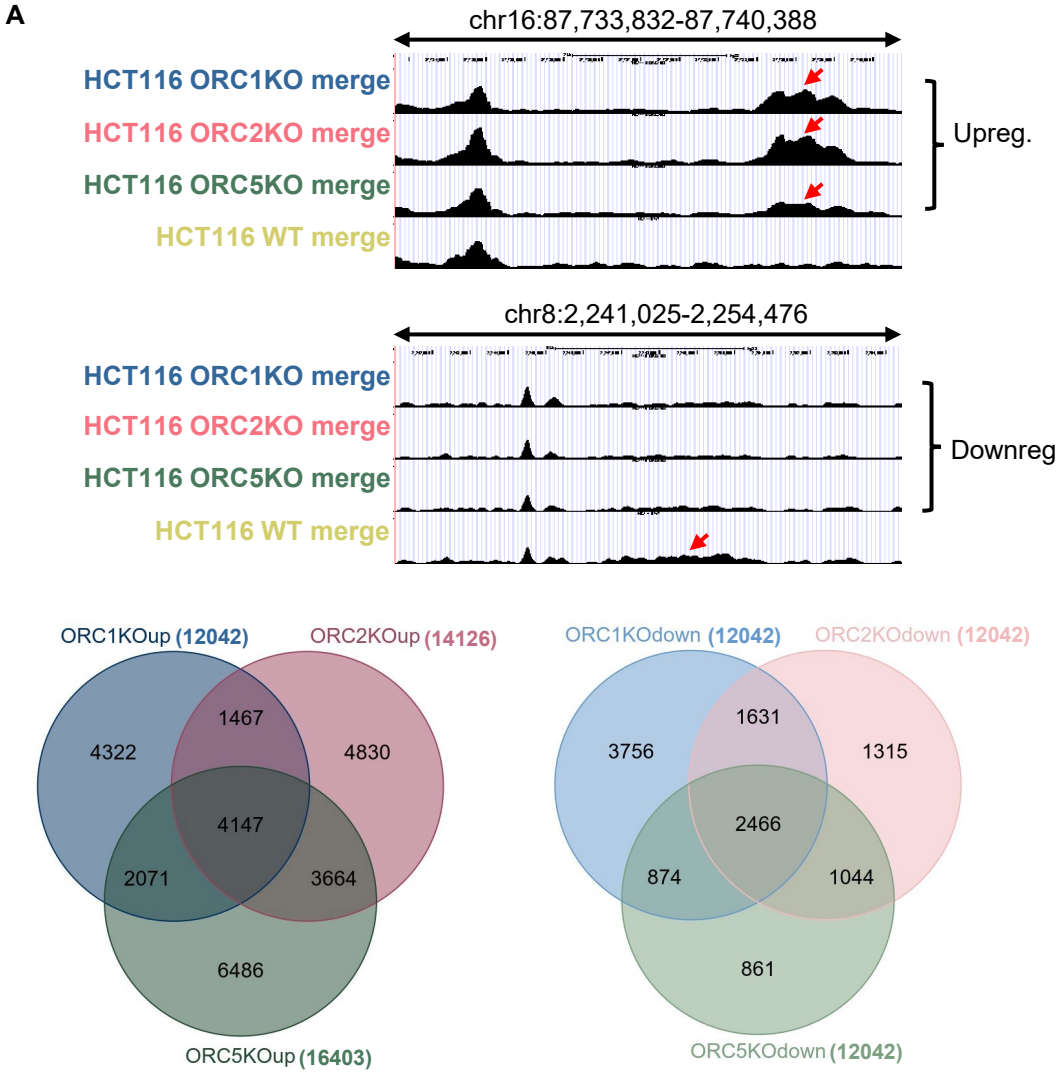

A
chr16:87,733,832-87,740,388
HCT116 ORC1KO merge
HCT116 ORC2KO merge
HCT116 ORC5KO merge
HCT116 WT merge
Upreg.
chr8:2,241,025-2,254,476
HCT116 ORC1KO merge
HCT116 ORC2KO merge
HCT116 ORC5KO merge
HCT116 WT merge
Downreg.
B
ORC1KOup (12042)
ORC2KOup (14126)
ORC5KOup (16403)
4322
1467
4830
4147
2071
3664
6486
ORC1KOdown (12042)
ORC2KOdown (12042)
ORC5KOdown (12042)
3756
1631
1315
2466
874
1044
861

Fig. 4

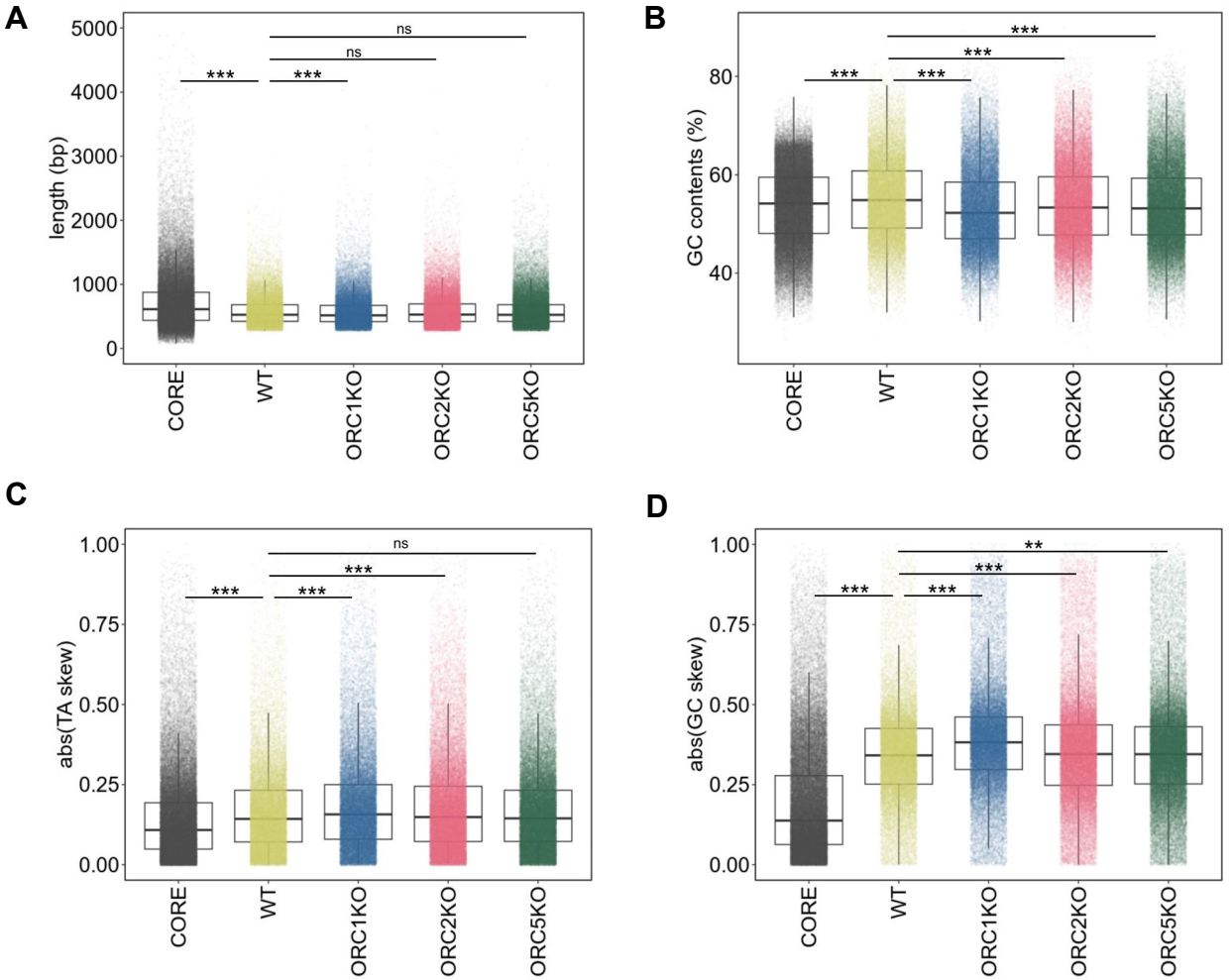

A
length (bp)
5000
4000
3000
2000
1000
0
***
***
ns
ns
CORE
WT
ORC1KO
ORC2KO
ORC5KO
B
GC contents (%)
80
60
40
***
***
***
***
CORE
WT
ORC1KO
ORC2KO
ORC5KO
C
abs(TA skew)
1.00
0.75
0.50
0.25
0.00
***
***
***
ns
CORE
WT
ORC1KO
ORC2KO
ORC5KO
D
abs(GC skew)
1.00
0.75
0.50
0.25
0.00
***
***
***
**
CORE
WT
ORC1KO
ORC2KO
ORC5KO

Fig. 5

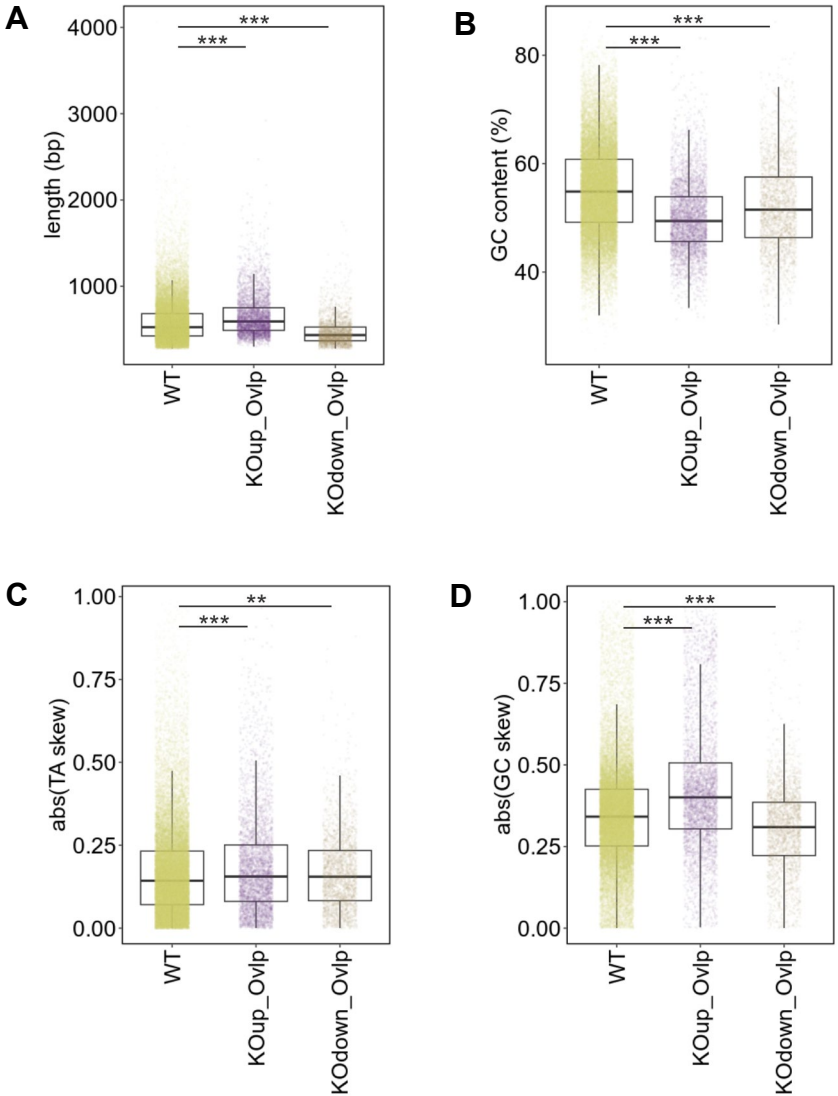
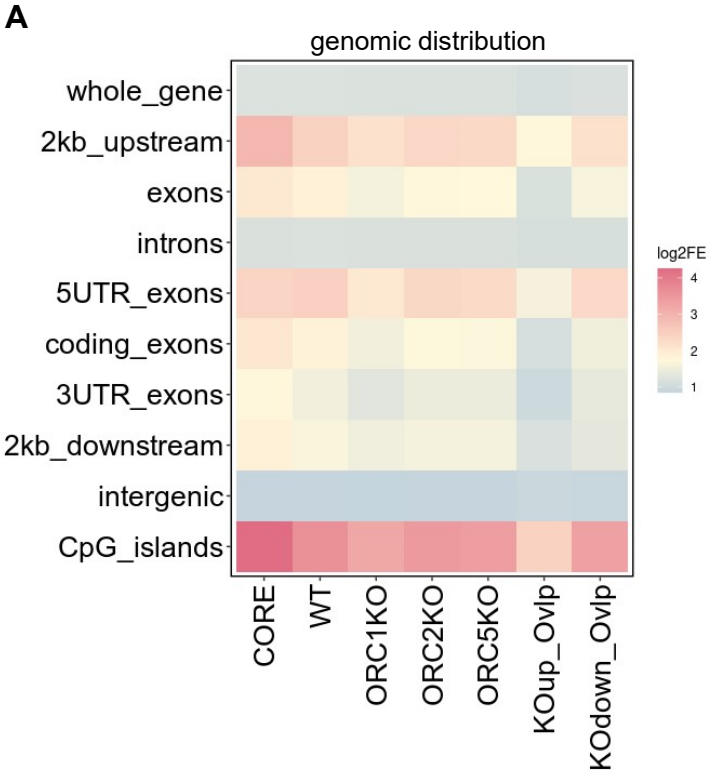
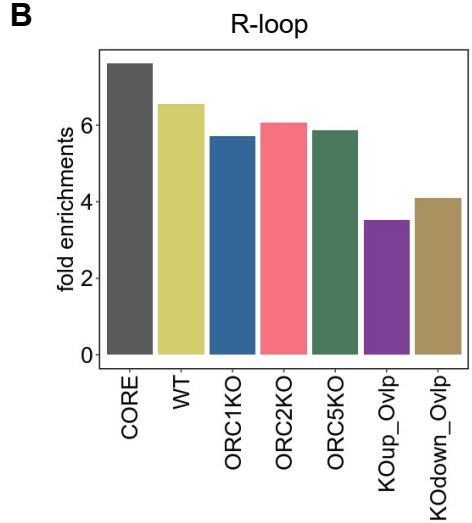

A
length (bp)
4000
3000
2000
1000
***
***
WT
KOup_Ovlp
KOdown_Ovlp
B
GC content (%)
80
60
40
***
***
WT
KOup_Ovlp
KOdown_Ovlp
C
abs(TA skew)
1.00
0.75
0.50
0.25
0.00
**
***
WT
KOup_Ovlp
KOdown_Ovlp
D
abs(GC skew)
1.00
0.75
0.50
0.25
0.00
***
***
WT
KOup_Ovlp
KOdown_Ovlp

Fig. 6


A
genomic distribution
whole_gene
2kb_upstream
exons
introns
5UTR_exons
coding_exons
3UTR_exons
2kb_downstream
intergenic
CpG_islands
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp
log2FE
4
3
2
1


B
R-loop
fold enrichments
0
2
4
6
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp

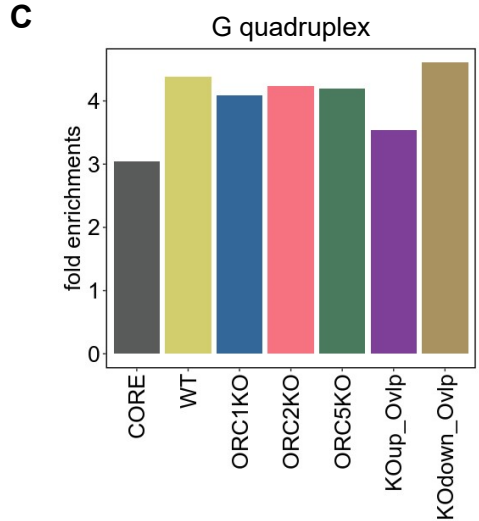

C
G quadruplex
fold enrichments
0
1
2
3
4
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp

Fig. 7

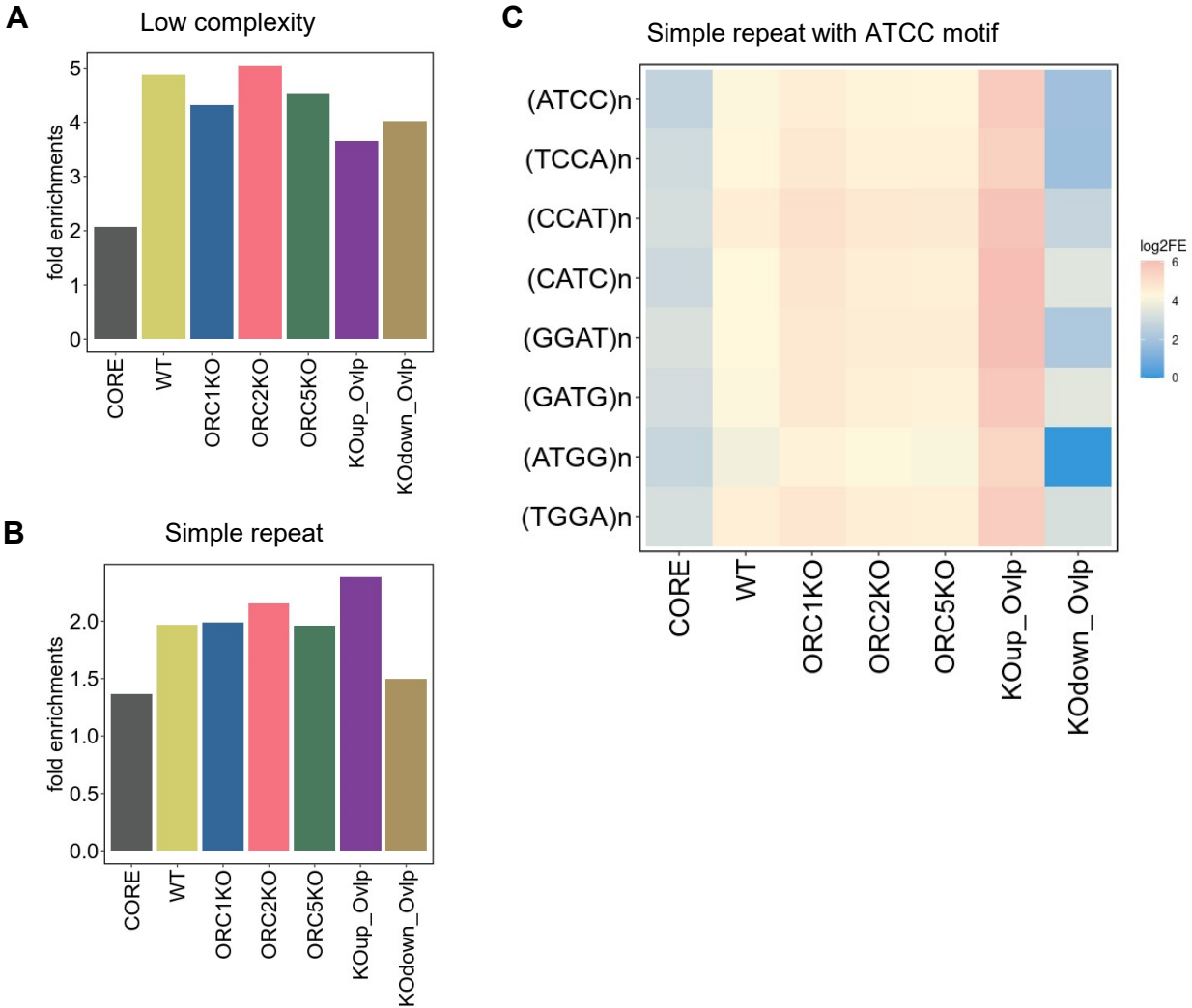
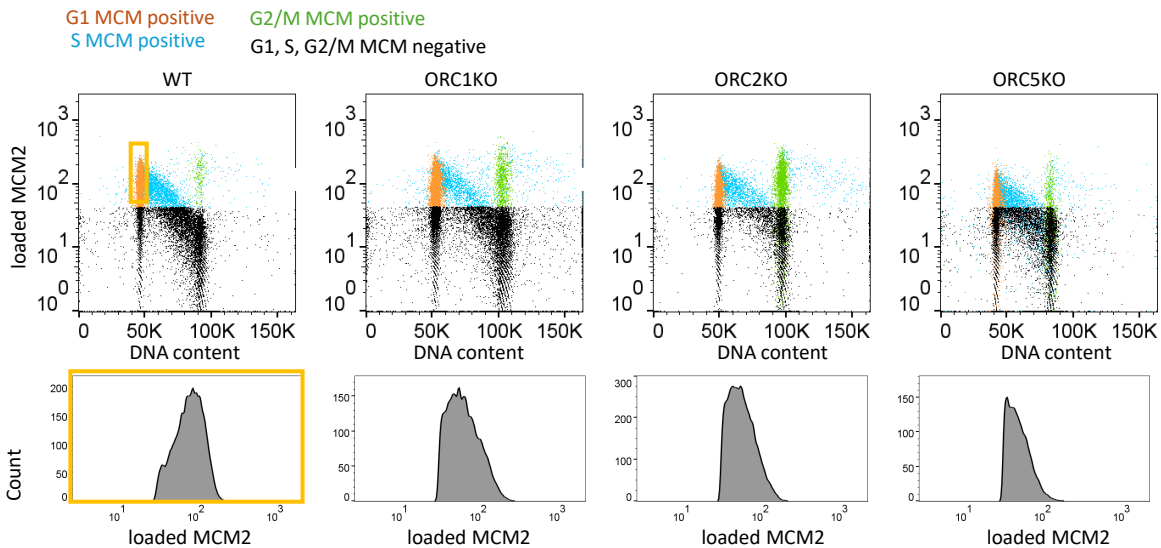

A
Low complexity
fold enrichments
0
1
2
3
4
5
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp
B
Simple repeat
fold enrichments
0.0
0.5
1.0
1.5
2.0
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp
C
Simple repeat with ATCC motif
(ATCC)n
(TCCA)n
(CCAT)n
(CATC)n
(GGAT)n
(GATG)n
(ATGG)n
(TGGA)n
CORE
WT
ORC1KO
ORC2KO
ORC5KO
KOup_Ovlp
KOdown_Ovlp
log2FE
6
4
2
0

Fig. 8

A



B

	Doubling time	F		
		replicate1	replicate2	replicate3
WT	19.5	0.29	0.24	0.27
ORC1KO	27.7	0.30	0.18	0.19
ORC2KO	30.3	0.40	0.22	0.21
ORC5KO	21.9	0.24	0.10	0.12

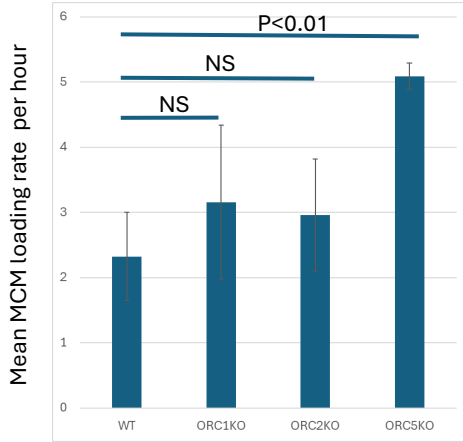
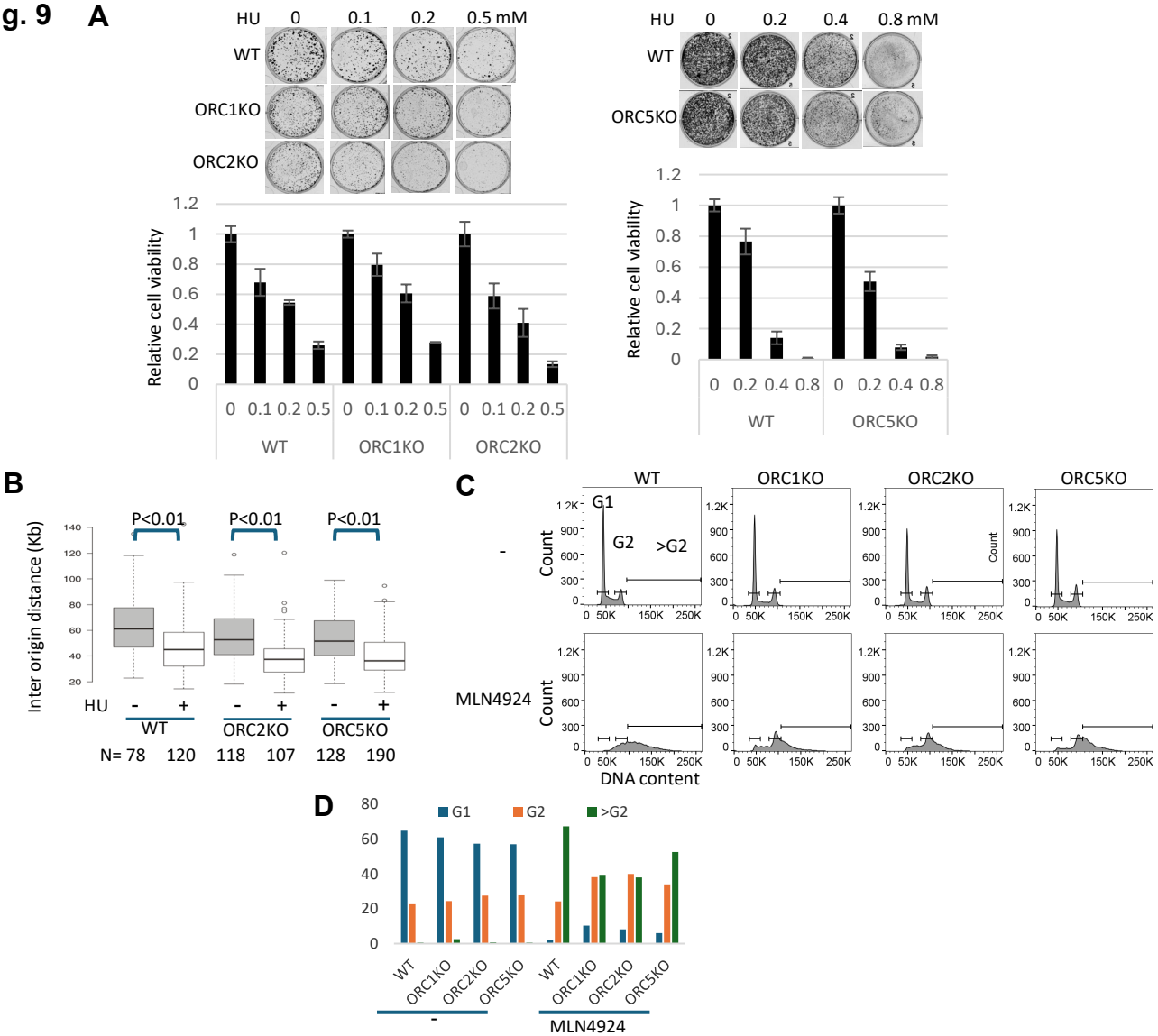


Fig. 9


A
HU 0 0.1 0.2 0.5 mM
WT
ORC1KO
ORC2KO
Relative cell viability
1.2
1
0.8
0.6
0.4
0.2
0
0 0.1 0.2 0.5 0 0.1 0.2 0.5 0 0.1 0.2 0.5
WT ORC1KO ORC2KO
HU 0 0.2 0.4 0.8 mM
WT
ORC5KO
Relative cell viability
0 0.2 0.4 0.8 0 0.2 0.4 0.8
WT ORC5KO
B
Inter origin distance (Kb)
140
120
100
80
60
40
20
P<0.01
P<0.01
P<0.01
HU - + - + - +
WT ORC2KO ORC5KO
N= 78 120 118 107 128 190
C
WT ORC1KO ORC2KO ORC5KO
-
MLN4924
Count
G1
G2
>G2
1.2K
900
600
300
0
0 50K 150K 250K
DNA content
D
G1 G2 >G2
80
60
40
20
0
WT ORC1KO ORC2KO ORC5KO WT ORC1KO ORC2KO ORC5KO
-
MLN4924