SAF-A promotes origin licensing and replication fork progression to ensure robust DNA replication

Caitlin Connolly[1], Saori Takahashi[3], Hisashi Miura[3], Ichiro Hiratani[3], Nick Gilbert[2], Anne D. Donaldson[1], and Shin-ichiro Hiraga[1*]

[1]Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK
[2]MRC Human Genetics Unit, The University of Edinburgh, Crewe Rd, Edinburgh, EH4 2XU, UK
[3]RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo 650-0047, Japan
* Corresponding author (s.hiraga@abdn.ac.uk)

Keywords:
DNA replication, Chromatin, Replication stress

Running title:
SAF-A promotes robust replication

Abstract

The organisation of chromatin is closely intertwined with biological activities of chromosome domains, including transcription and DNA replication status. Scaffold attachment factor A (SAF-A), also known as Heteronuclear Ribonucleoprotein Protein U (HNRNPU), contributes to the formation of open chromatin structure. Here we demonstrate that SAF-A promotes the normal progression of DNA replication, and enables resumption of replication after inhibition. We report that cells depleted for SAF-A show reduced origin licensing in G1 phase, and consequently reduced origin activation frequency in S phase. Replication forks also progress less consistently in cells depleted for SAF-A, contributing to reduced DNA synthesis rate. Single-cell replication timing analysis revealed two distinct effects of SAF-A depletion: first, the boundaries between early- and late-replicating domains become more blurred; and second, SAF-A depletion causes replication timing changes that tend to bring regions of discordant domain compartmentalisation and replication

timing into concordance. Associated with these defects, SAF-A-depleted cells show elevated γ-H2AX formation and tend to enter quiescence. Overall we find that SAF-A protein promotes robust DNA replication to ensure continuing cell proliferation.

Summary Statement

Scaffold attachment factor A (SAF-A/HNRNPU) protein contributes to the formation of open chromatin structure. We have discovered that SAF-A supports DNA replication at multiple steps.

Introduction

DNA replication in eukaryotic genomes initiates from discrete sites termed DNA replication origins. Potential replication origin sites are defined by stepwise assembly of a protein complex, the pre-replication complex (pre-RC), during G1 phase of the cell cycle (Fragkos et al., 2015). During pre-RC formation the Origin Recognition Complex (ORC) and CDT1 cooperate to load the heterohexameric MCM complex, leading to 'origin licensing' (McIntosh and Blow, 2012). MCM plays a critical role when DNA replication initiates at each origin, forming the central component of the replicative helicase (Fragkos et al., 2015). Cells monitor the level of replication licensing and prevent cell cycle progression if an insufficient number of sites are licensed (Feng et al., 2003; Lau et al., 2009; Nevis et al., 2009; Shreeram et al., 2002; Zimmerman et al., 2013). This "licensing checkpoint" mechanism appears to be compromised in cancer cells (Feng et al., 2003; Lau et al., 2009; Nevis et al., 2009; Shreeram et al., 2002; Zimmerman et al., 2013).

A recent study demonstrates that replication licensing is impacted by the state of chromatin packaging. The histone methyltransferase SET8 (also called PR-SET7 or KMT5A) can stimulate origin licensing at specific sites (Tardat et al., 2010), but also prevents over-licensing by enhancing chromatin compaction as cells exit mitosis (Shoaib et al., 2018). SET8 is responsible for the methylation of histone H4 at lysine 20, and for maintaining chromatin compaction at the M/G1 boundary (Shoaib et al., 2018). Replication licensing is therefore impacted both by local chromatin changes and broader changes occurring at chromosome domain level. How chromatin packaging status affects origin licensing and the subsequent steps in DNA replication is still not fully understood, but there is a well-established connection between chromatin packaging and the temporal programme of replication (Fu et al., 2018; Gilbert, 2010; Gilbert et al., 2010), in which euchromatin domains containing active genes generally replicate early in S phase, while heterochromatic, highly packaged domains containing mainly inactive genes replicate late. Replication timing of some domains is modulated during development, often reflecting changes in gene activity (Hiratani et al., 2008). The replication timing programme is established at early G1 phase, which coincides with chromatin decompaction and chromatin remodelling as cells exit M phase (Shoaib et al., 2018), suggesting that the dynamic controls over chromatin structure imposed as cells exit mitosis determine the

replication potential and subsequent replication timing of local chromatin domains (Dimitrova and Gilbert, 1999; Dimitrova et al., 2002).

Chromatin packaging also impacts on replication progression, with processive replication of heterochromatin regions requiring local decompaction (Chagin et al., 2019). Recent studies also highlight numerous “difficult-to-replicate” regions (Cortez, 2015; Gadaleta and Noguchi, 2017), including DNA-protein complexes, repetitive DNA such as centromeres and telomeres, and secondary DNA structures. Replicating such regions requires support by specific proteins, without which replication tends to fail, leading to genome instability (Cortez, 2015; Gadaleta and Noguchi, 2017) and the formation of fragile sites (Boteva et al., 2020). These observations highlight the importance of modulating chromatin structure during the replication process.

Scaffold Attachment Factor A (SAF-A; also known as Heterogeneous Nuclear Ribonucleoprotein U) is an RNA- and DNA-binding protein which modulates chromatin structure by tethering chromatin-associated RNA (caRNA) to chromatin (Creamer et al., 2021; Fackelmayer et al., 1994; Kiledjian and Dreyfuss, 1992; Nozawa et al., 2017; Sharp et al., 2020). SAF-A oligomerisation contributes to de-compacted chromatin (Creamer et al., 2021; Nozawa et al., 2017), and depletion of SAF-A causes global chromatin condensation (Fan et al., 2018). A super-resolution microscopy study also implicates SAF-A in the establishment of correct chromatin structure, and SAF-A has been shown to regulate both active chromatin and also X-chromosome inactivation (Hasegawa et al., 2010; Lu et al., 2020; Smeets et al., 2014). SAF-A interacts and colocalises with proteins that define chromatin domain boundaries, namely CTCF and cohesin, and plays a role in defining boundaries of Topologically-Associated Domains that form smaller units of chromosome organisation (Fan et al., 2018; Zhang et al., 2019). Although SAF-A promotes open chromatin, depletion of SAF-A has fairly minor effects on transcription (Nozawa et al., 2017). SAF-A is however reported to be implicated in alternative splicing (Xiao et al., 2012; Ye et al., 2015), mRNA stability (Yugami et al., 2007) and nuclear retention of mRNA (Huang et al., 2021).

Association of SAF-A with chromatin is cell cycle-regulated: SAF-A is chromatin-associated throughout interphase but is removed from chromatin in M phase (Sharp et al., 2020). Regulated dissociation of SAF-A from the mitotic chromosome, triggered by phosphorylation of SAF-A by Aurora B protein kinase, is

essential for proper progression of mitosis (Douglas et al., 2015; Sharp et al., 2020). SAF-A re-associates with chromatin as cells exit from mitosis, implicating SAF-A in chromatin decompaction at this cell cycle stage.

SAF-A also localises to DNA damage sites swiftly after γ-ray irradiation (perhaps to modulate chromatin structure), and then at a later stage appears to be excluded from damage sites (Hegde et al., 2016). Interestingly, the expression of the SAF-A gene tends to increase in a wide range of cancers, particularly breast invasive carcinoma (The Cancer Genome Atlas). This increased expression suggests that SAF-A contributes to the formation or survival of cancer cells in a dose-dependent manner. Conversely, SAF-A loss-of-function alleles are linked to developmental disorders including microcephaly (Durkin et al., 2020; Leduc et al., 2017; Yates et al., 2017). Overall, these observations suggest a positive role for SAF-A in promoting cell proliferation. While roles of SAF-A in mitosis have been investigated (Sharp et al., 2020), its contribution to cell proliferation during interphase, particularly to DNA replication, has not been studied.

Here we investigate the effects of SAF-A on DNA replication. We show that SAF-A protein is required for full replication licensing in the G1 phase of the cell cycle, and depleting SAF-A leads to increased spacing between replication origins. We find moreover that replication fork progression is compromised in cells depleted for SAF-A, and that SAF-A protein plays a role in defining the boundaries of early/late replication domains in the genome-wide replication programme. Loss of these functions leads to spontaneous replication stress and increases cellular entry to quiescence, explaining the need for SAF-A for normal cell proliferation.

Results

SAF-A is required for robust DNA replication

To assess the general impact of SAF-A on subnuclear organisation of chromatin, we examined the distribution of DNA within nuclei of hTERT-RPE1 cells treated with siRNA targeting SAF-A (siSAF-A) (Fig 1A). hTERT-RPE1 is a non-cancer cell line derived from retinal pigment epithelial cell immortalised by expression of human telomerase (hTERT) (Bodnar et al., 1998). Using super-resolution microscopy to examine very thin sections, we found that control nuclei show relatively

142 homogeneous DNA density distribution (i.e. mostly green), with smaller areas of
143 higher (yellow/red) or lower (cyan/blue) DNA density. In siSAF-A cells the DNA
144 density distribution shows larger areas with low DNA density (blue) interspersed with
145 high DNA density areas (red), indicative of a more polarised distribution with sections
146 of genomic DNA densely packed in abnormally compact domains. Unbiased
147 classification of “DAPI high” and “DAPI low” areas in each nucleus (Fig 1B)
148 confirmed the formation of larger “DAPI low” areas in siSAF-A nuclei, with chromatin
149 confined into smaller areas. These data suggest that SAF-A promotes proper
150 dispersal of chromatin distribution within nuclei, and prevents the formation of over-
151 compacted chromatin. This microscopic observation is consistent with SAF-A
152 function in maintaining correct chromatin architecture as revealed by microscopy
153 (Creamer et al., 2021; Nozawa et al., 2017) and Hi-C methods (Fan et al., 2018).
154 Cells depleted for SAF-A were reported to show proliferation defects (Nozawa
155 et al., 2017), but the exact nature of the defect was not studied in detail. We
156 examined the cell proliferation and DNA replication profiles of cells depleted for SAF-
157 A. siSAF-A cells showed a significant and reproducible reduction in cell proliferation
158 rates, compared with siControl cells (Fig 1C), consistent with a previous report
159 (Nozawa et al., 2017). Flow cytometry analysis of DNA content in asynchronous
160 cultures (Fig S1A) however showed no specific cell cycle arrest point, but did reveal
161 a slight reduction of S phase population, suggesting that loss of SAF-A may cause
162 problems with DNA replication.
163 Cells depleted for SAF-A were reported to be defective in recovery from
164 replication inhibition by the DNA polymerase inhibitor aphidicolin (Nozawa et al.,
165 2017). We therefore tested whether SAF-A-depleted cells also fail to recover from
166 the DNA replication inhibitor hydroxyurea (HU), an inhibitor of ribonucleotide
167 reductases that causes stalled replication forks. After treating siControl and siSAF-A
168 cells with 4 mM HU for 24 hr to cause early S phase arrest (Fig 1B, 0 hr), we
169 examined recovery by monitoring DNA content and incorporation of a thymidine
170 analogue ethynyl deoxyuridine (EdU). Control cells recovered from arrest efficiently
171 and reached mid-S phase by 4 hr after release from HU (Fig 1D, siControl). In
172 contrast, very few siSAF-A cells recovered to reach a similar stage by 6 hr (Fig 1D,
173 siSAF-A). Assessment of EdU-positive cells further indicated that a reduced number
174 of siSAF-A cells were able to resume DNA synthesis compared with siControl (Fig
175 1E), and that the rate of DNA synthesised in EdU-positive siSAF-A cells was lower

than that in siControl cells until 6 hr after release (Fig S1B & C). By 8 hr after the release, the majority of siControl cells had finished DNA replication, whereas a notable fraction of siSAF-A cells were still synthesising DNA (Fig S1B & C). These observations indicate that cells depleted for SAF-A have difficulty in recovering from HU. Taken together, deficiency of SAF-A causes cells to be severely impaired in recovery from replication stress.

We tested whether depletion of SAF-A impacts DNA replication in the absence of exogenous stress, by measuring DNA synthesis rate based on pulse-labelling nascent DNA with EdU followed by flow cytometry analysis. Cells depleted for SAF-A showed a significantly reduced percentage of EdU-positive cells (Fig 1F; 37.2% in siControl and 24.7% in siSAF-A). Moreover, the EdU-positive population of siSAF-A cells showed reduced DNA synthesis rate compared to siControl (Fig 1G). This DNA synthesis defect of siSAF-A cells is not confined to a specific stage of S phase (Fig S1D), suggesting that SAF-A function is required throughout DNA replication.

Together, these results indicate that SAF-A is required for robust DNA replication without exogenous replication stress, and also supports the recovery of cells after replication stress.

SAF-A is important for replication licensing

Changes in chromatin due to SAF-A depletion could potentially affect multiple steps of DNA replication, including origin licensing, replication fork progression, and fork restart. We decided to assess the requirement for SAF-A for each of these steps in a series of experiments.

Since SAF-A plays a positive role in open chromatin structure, and prevents over-compaction (Fig 1A & B) we hypothesised that SAF-A may play a positive role in stimulating origin licensing by promoting open chromatin. To test this hypothesis, we used a flow cytometry ‘3D licensing assay’ (Moreno et al., 2016) (Fig 2A & B), which simultaneously measures MCM loading on chromatin and EdU incorporation EdU (to assess cell cycle stage). In this assay (Fig 2A) G1 (red box), S (cyan), and G2/M (orange) phase cells can be clearly distinguished, and the amount of chromatin-associated MCM3 assessed in each cell cycle population (Fig 2B). As clearly seen in Fig 2B, siSAF-A-treated hTERT-RPE1 cells show reduced levels of

chromatin-associated MCM3 in individual cells both in G1 phase (red) and in cells entering S phase (left hand part of cyan population). This observation indicates that siSAF-A cells show compromised levels of MCM loading in G1 phase cells that persist into S phase, suggestive of a defect in origin licensing.

We next tested whether SAF-A affects other licensing proteins. CDT1 interacts with the MCM complex and assists its loading onto chromatin in G1 phase (Frigola et al., 2017; Zhai et al., 2017a): outside of G1 phase CDT1 is negatively regulated by Geminin (Blow and Tanaka, 2005). *In vitro* studies show that CDT1 dissociates from MCM after the assembly of the double MCM hexamer on DNA (Zhai et al., 2017b). Consistently, flow cytometry analysis showed that CDT1 associates with chromatin predominantly in G1 phase (vertical spikes in Fig 2C). We found that depletion of SAF-A in hTERT-RPE1 cells caused reduced chromatin association of CDT1 (Fig 2C, siSAF-A), consistent with the reduced MCM licensing in G1 phase (Fig 2B).

We next tested the chromatin association of ORC1 protein. ORC1 protein is a subunit of the ORC protein complex that initially defines MCM loading sites. ORC1 protein expression and stability is cell cycle-regulated so that it is present predominantly in G1 phase, helping to confine MCM loading to G1 phase of the cell cycle (Mendez et al., 2002; Ohta et al., 2003; Tatsumi et al., 2003). In the absence of an ORC1 antibody suitable for flow cytometry analysis, we made use of a HEK293-based cell line expressing FLAG-tagged ORC1 protein (Tatsumi et al., 2003) to analyse chromatin association of ORC1 (Fig 2D). As expected and previously reported (Hiraga et al., 2017), ORC1 chromatin association is detected predominantly in the G1 phase of the cell cycle (Fig 2D, top panel). Depletion of SAF-A results in a reduction of chromatin-associated ORC1 (Fig 2D, top panel, right). We also observed that CDT1 and MCM licensing are reduced when SAF-A is depleted in the HEK293 FLAG-ORC1 cell line (Fig 2D, middle and bottom panels, respectively), similar to the effects in hTERT-RPE1 cells (Fig 2B). In contrast, chromatin association of ORC2, which does not fluctuate during cell cycle (Mendez and Stillman, 2000; Mendez et al., 2002), was not affected by SAF-A depletion (Fig S2A). We confirmed the reproducibility of these effects in multiple independent experiments in both HEK293-derived cells and hTERT-RPE1 cells (Fig S2B&C).

Reduced association of licensing factors in SAF-A-depleted hTERT-RPE1 cells was confirmed by Western analysis of chromatin-associated proteins. Western

blotting of chromatin-enriched fractions confirmed the reduced association of CDT1 protein (Fig S2E left panel, lanes 3 & 4) and ORC1 protein (Fig S2E right panel, lanes 7 & 8 top panel) after SAF-A depletion. CDC6, another protein required for replication licensing (Blow and Tanaka, 2005), however did not show such a reduction in chromatin association (Fig S2E, lanes 7 & 8 middle panel). Chromatin association of ORC2 was not affected by SAF-A depletion (Fig S2E, lanes 11 & 12), consistent with the flow cytometry result (Fig S2A).

The effects of SAF-A depletion on chromatin association of FLAG-ORC1 and CDT1 appeared very similar in HEK293 FLAG-ORC1 cells extracted with CSK buffer (which contains 100 mM NaCl, as compared to standard conditions for such flow cytometry involving extraction with only 10 mM NaCl) (Fig. S3). CSK buffer is widely used for biochemical preparation of chromatin-associated protein fractions in HEK293-derived cells. In hTERT-RPE1 cells we noticed that CDT1 chromatin association appeared more salt sensitive (data not shown), suggesting that chromatin association may differ between cell types.

Since the western analysis suggests some reduction in total CDT1 and ORC1 levels (Fig S2E lanes 1-2 and 5-6, respectively), we investigated expression of these proteins per cell using flow cytometry. CDT1 expression was not affected by SAF-A depletion in hTERT-RPE1 cells. In FLAG-ORC1 cells depleted for SAF-A we found that levels of FLAG-ORC1 and CDT1 were somewhat reduced (Fig S4), although results were variable between experimental repeats (compare first and second panels from top). Overall we conclude that reduced ORC1 and CDT1 expression may contribute to phenotypes of SAF-A depletion, but cannot fully account for the reduced licensing levels particularly in hTERT-RPE1 cells.

Overall, the data presented show that SAF-A promotes the G1 phase chromatin association of several origin licensing components, including loading of the MCM complex itself.

SAF-A depletion results in reduced origin activation

Impaired replication licensing in cells depleted for SAF-A suggests there will be a reduced number of potential replication origins available for activation. Therefore, we next tested whether a reduced origin frequency is observed on chromosomes in

SAF-A-depleted cells, by measuring inter-origin distances using single-molecule DNA fibre analysis.

To detect origin activation on single DNA molecules, nascent DNA was sequentially labelled with thymidine analogues 5-chloro-2'-deoxyuridine (CldU) and 5-Iodo-2'-deoxyuridine (IdU), as illustrated in Fig 3A. Analogue incorporation was analysed by immunostaining of DNA fibres stretched by molecular DNA combing (Bianco et al., 2012). Replication origins can be identified as illustrated in the top panel of Fig 3B, with the mid-point between divergent replication forks assigned as a replication origin. In these experiments, increased distance between replication origins (inter-origin distances; IOD) is indicative of fewer active origins. We found that depletion of SAF-A caused an increase in IOD compared with the control (Fig 3B), suggesting that the number of active origins is indeed reduced by SAF-A depletion. This reduction in origin activation frequency probably reflects inefficient origin licensing.

Stalling of DNA replication forks due to replication stress causes activation of nearby dormant origins (Ge et al., 2007), believed to protect cells from replication stress by guarding against the formation of unreplicated stretches between two stalled or collapsed replication forks (Blow and Ge, 2009; Kawabata et al., 2011). Since the licensing defect of SAF-A-depleted cells might affect the number of available dormant origins, we assessed whether cells depleted for SAF-A activate dormant origins normally (Fig 3B and Fig 3C). siRNA-treated cells were incubated for 4 hr with 0.1 mM HU to slow replication forks. At the end of the HU treatment, nascent DNA was labelled with CldU and IdU as in Fig 3A. Under this condition, replication forks continue to progress but with significantly reduced speed (Fig 4A). As expected, HU treatment induced the activation of dormant origins near stalled forks, evidenced by a reduction in IOD and appearance of very short IODs below 20 kb (below blue dotted line), both in siControl and siSAF-A cells (Fig 3C). In cells depleted for SAF-A, however, short IODs (in the range 0-30 kb) occurred at reduced frequency compared to siControl, suggesting impaired dormant origin activation.

Overall these data confirm that the number of active DNA replication origins is reduced in cells depleted for SAF-A, and that SAF-A is required for activation of dormant origins at normal frequency under replication stress.

SAF-A supports DNA replication fork progression

SAF-A depletion leads to decreased cellular DNA synthesis rate in unperturbed S phase (Fig 1F & G), as well as reduced origin licensing (Fig 2) and activation (Fig 3). However, it was unclear whether reduced origin activation fully accounts for the decreased cellular DNA synthesis. To explore whether altered DNA replication fork speed also affects DNA synthesis when SAF-A is deficient, we investigated replication fork speed using the same DNA combing technique as in Fig 3A. The lengths of IdU tracts in stretched DNA molecules were taken as a proxy for DNA synthesis rate. The average replication fork speed was not affected by SAF-A depletion (Fig 4A, compare siControl and siSAF-A). However, we repeatedly detected wider variance of replication fork speed in siSAF-A cells treated with HU, compared with siControl cells with HU (Fig 4A, compare siControl +HU and siSAF-A +HU), suggesting that in SAF-A-depleted cells replication fork speed is less tightly regulated during replication stress.

We next tested whether SAF-A depletion affects the processivity of DNA replication forks. If processivity is high, the rate of DNA synthesis will stay consistent through the CldU and IdU labelling periods, and the log value of (IdU tract length / CldU tract length) is expected to be close to 0 (for example, Fig 4B siControl). Frequent pause or collapse of forks will lead to a wider spread in $\log_2$(IdU /CldU) values. We found cells depleted for SAF-A show wider spreading of the $\log_2$(IdU /CldU) values under HU-treated condition, indicating increased probability of fork slowing, pause, or collapse in cells depleted for SAF-A (Fig 4B) (Note that forks pausing or collapsing in the CldU labelling period will not be counted, since as they produce only CldU labelling they are indistinguishable from termination sites). This result suggests that SAF-A is required to support processive DNA synthesis under replication stress.

The nascent DNA labelling experiments demonstrate that SAF-A is required for robust replication fork progression, as well as to support origin licensing.

SAF-A affects replication timing at domain boundaries, and at regions of discordance between compartmentalisation and replication timing

Given its effect on replication origin activation and fork progression, we examined whether SAF-A is important for DNA replication timing. To enable detection of

changes that might not be evident in a population analysis (such as increased variability that does not affect the average replication time of a locus), we examined the replication timing programme in single cells (Miura et al., 2020; Takahashi et al., 2019). Briefly, we used a recently described method in which single mid-S phase cells are collected by cell sorting based on their DNA content, then NGS library preparation and copy number sequencing carried out for each individual single cell. As a control, we carried out similar analysis using a pool of 100 mid-S cells. The relative copy number of 200 kb segments was calculated based on the number of sequencing reads, normalised against reads obtained from G1 phase cells.

We compared the replication timing profiles of 33 single mid-S hTERT-RPE1 cells for siControl, and 25 single mid-S cells for siSAF-A (Fig 5A). We found that, as previously proposed, replication timing of single siControl cells generally reflected A/B compartment distribution as determined by hTERT-RPE1 cell Hi-C analysis (Darrow et al., 2016; Miura et al., 2018). While in siSAF-A cells the overall replication timing profiles are largely similar to siControl, we found that the boundaries of the replication timing domains are less uniform. For example, in the regions shown magnified at the bottom of Fig 5A, siControl cells show clear, fairly uniform boundaries between unreplicated (blue) and replicated (red) domains, across the 33 analysed cells. In siSAF-A cells in contrast, the boundary position shows more variation between single cells, resulting in a lack of clear boundaries when viewed across the population. The 'RT changes' plot shows the difference between single siControl and siSAF-A cells in average replication timing. Statistical comparison of single-cell replication timing between siControl and siSAF-A cells confirms the impression that boundaries are blurred (see '$-\log_{10}P$' plot in Fig 5A), with peaks indicating regions showing significant difference between the siControl and siSAF-A profiles, generally coinciding with timing domain boundaries. We identified 420 "$-\log_{10}P$" peaks genome-wide, of which 173 overlap with replication timing domain boundaries (RT boundaries). One-tailed Fisher's exact test find the co-occurrence of "$-\log_{10}P$" peaks and RT boundaries is statistically significant (p-value = 2.22×10^{-18}), whereas no statistical significance was found if genomic locations of "$-\log_{10}P$" peaks were randomly shuffled (p-values = 0.99; 76 co-occurrences out of 420).

Depletion of SAF-A does not lead to any clear trend in replication timing changes genome-wide (with similar numbers of regions becoming earlier or later, Fig

5B (i)), nor does it cause a consistent shift in timing at all RT domain boundaries (Fig 5B (ii)). If however we consider all regions showing significant difference in replication time between siControl and siSAF-A cells (defined by $-\log_{10}$P value > 3), then SAF-A depletion results in some tendency towards earlier replication timing (Fig 5B (iv)), although the changes vary substantially in direction and magnitude with some sequences replicating later than normal. Furthermore, at RT boundaries where siControl and siSAF-A cells show substantial differences, the effect of SAF-A depletion is noticeably bimodal, with the majority of such boundary regions replicating earlier but some replicating later (Fig 5B (iii)). Our finding that many RT boundaries are sensitive to SAF-A depletion is consistent with the proposed function of SAF-A in defining chromatin domain boundaries (Fan et al., 2018).

We also noticed that chromosomal locations that show replication timing shift in siSAF-A cells tend to coincide with genomic locations where A/B compartment and replication timing patterns are discordant or 'disagree' (see Fig S5A for specimen loci). For statistical comparison, we picked the top 10% of genomic segments showing later replication timing in siSAF-A cells compared to siControl ('EtoL' sites), and the top 10% of segments showing earlier replication in siSAF-A cells ('LtoE' sites). These EtoL and LtoE sites show a significantly higher proportion of discordance between compartment and replication timing than the genomic average (Table 1). Moreover, at these loci, the observed replication timing shift in siSAF-A cells tends to bring replication timing into alignment with the A/B compartment pattern (Fig. S5B and Table 1).

To quantitatively confirm the variability in replication timing between individual siSAF-A cells, we compared the number of NGS reads per 200 kb sliding window (i.e. tag density) at 40-kb intervals (Miura et al., 2020). In the pool of 100 mid-S cells from the siControl, distribution of the tag density forms two overlapped peaks (Fig 5C left), representing unreplicated (left peak) and replicated (right peak) portions of the genome. The separation of these peaks in the 100 pooled cells means that (1) unreplicated and replicated domains are distinct in each cell and (2) this distinct pattern is essentially conserved in the 100 cells. In other words, the replication timing programme is well-conserved in these 100 siControl cells. In contrast, tag density from the 100 mid-S siSAF-A cells does not show clear peak separation (Fig 5C right). Note however that we do see clear separation of two peaks when analysing single siSAF-A cells at mid-S (Fig S5B), similar to single siControl cells, indicating

409 that our single cell analysis does effectively distinguish unreplicated and replicated
410 domains. Therefore, the poor peak separation of tag density in the 100 mid-S siSAF-
411 A cell pool is due to compromised conservation of the replication timing programme
412 between single cells.
413 Although the overall replication timing profiles appear fairly similar (Fig 5A)
414 and only a limited number of loci showed altered replication timing (Fig 5B, S5A, and
415 Table 1), t-SNE clustering analysis (van der Maaten, 2014; van der Maaten and
416 Hinton, 2008) of the distribution of early and late domains in single cells showed a
417 clear separation of the siControl and siSAF-A populations (Fig 5D), indicating that
418 genome-wide replication timing program is indeed altered in siSAF-A cells.
419 Taken together, we conclude that in cells depleted for SAF-A, the genome-
420 wide DNA replication timing programme is less well-defined, becoming more 'blurred'
421 and unstable particularly at replication timing domain boundaries, and in regions
422 where A/B compartment and replication timing are normally discordant.
423
424 **SAF-A prevents spontaneous quiescence**
425 Our data suggest that DNA replication is aberrant at various stages in cells depleted
426 for SAF-A, even without exogenous replication stress (Fig 1F, 1G, and S1D). Recent
427 studies suggest that cells with incomplete DNA replication and/or DNA damage can
428 progress through mitosis but may activate the p53-mediated G1 checkpoint in the
429 subsequent cell cycle, leading to a transient quiescence of daughter cells (Arora et
430 al., 2017; Barr et al., 2017). Such delayed progression can be monitored by
431 examining expression of the CDK inhibitor p21 (also called $p21^{WAF1}$). We tested the
432 possibility that replication problems in siSAF-A cells leads to spontaneous
433 quiescence, by looking at the expression of p21. Cells depleted for SAF-A show
434 clear expression of p21 without any exogenous damage (Fig 6A), whereas the
435 expression of p21 is barely detectable in siControl cells. Flow cytometry (Fig 6B)
436 reveals that a significant proportion of siSAF-A cells with a 'G1 phase' DNA content
437 show p21 expression, suggesting these cells are in quiescence (= G0 phase).
438 Interestingly, a fraction of G2 phase siSAF-A cells already express p21. Expression
439 of p21 in G2 cells has been reported for cells that have undergone DNA damage and
440 are destined to enter quiescence (Arora et al., 2017; Barr et al., 2017). The
441 tendency of SAF-A depleted cells to enter quiescence was also confirmed by

measuring the phosphorylation of retinoblastoma (Rb) protein. Rb is a negative regulator of the cell cycle, and is phosphorylated in proliferating cells but remains unphosphorylated in quiescent cells (Giacinti and Giordano, 2006). Measurements of the cellular levels of Rb phosphorylation at Ser-807/811 demonstrate that a significantly higher proportion of siSAF-A cells are in quiescence, as evidenced by dominance of cells with unphosphorylated Rb (Fig 6C and 6D).

Depletion of SAF-A leads to replication stress

A previous study demonstrated that depletion of SAF-A increased the proportion of cells showing diffuse localisation of the histone variant H2AX phosphorylated at its C-terminus (called γ-H2AX) (Nozawa et al., 2017). γ-H2AX has been commonly used as a DNA damage marker, but recent studies suggest that diffuse localisation of γ-H2AX within the nucleus is indicative of replication stress rather than DNA damage, whereas a more focal γ-H2AX localisation pattern represents DNA damage (Dhuppar et al., 2020; Moeglin et al., 2019). We assessed the impact of depletion of SAF-A with or without replication stress based on γ-H2AX localisation pattern. Fig 7A shows a specimen image with 'diffuse' and 'focal' γ-H2AX localisation patterns. Without replication stress, few Control cells exhibit either γ-H2AX pattern, with most cells showing no apparent γ-H2AX signal (Fig 7B, siCont). Replication stress (induced by 3 hr HU treatment) significantly increased the proportion of cells with ‘diffuse’ γ-H2AX (29%; Fig S7A), consistent with the suggestion that diffuse γ-H2AX signal is indicative of DNA replication stress. In contrast, 29% of cells depleted for SAF-A have diffuse γ-H2AX even without HU treatment (Fig 7B, siSAF-A), suggesting that depletion of SAF-A imposes replication stress on cells. In a separate experiment, we confirmed that virtually all cells with diffuse γ-H2AX signal are in S phase (Fig S7B), and that a large fraction of S phase cells have diffuse γ-H2AX signal when SAF-A is depleted (Fig S7C).

Multiple protein kinases, including ATM, ATR and DNA-PK, have been implicated in γ-H2AX formation upon DNA damage (Rogakou et al., 1998; Wang et al., 2005). However during replication stress, ATR is generally believed to be the kinase responsible for γ-H2AX (Ward and Chen, 2001), although some reports implicate other protein kinases (Buisson et al., 2015; Chanoux et al., 2009; Serrano et al., 2013). To investigate which protein kinase is responsible for increased γ-

H2AX in siSAF-A cells, we tested the impact of inhibiting ATM (using KU-60019), ATR (VE-821), and DNA-PK (NU-7441) on nuclear γ-H2AX signal in S-phase cells (Fig 7C and S7C). Inhibition of ATR almost completely suppressed γ-H2AX induction upon SAF-A depletion (Fig 7C, compare right halves of Untreated and ATRi), consistent with the increased γ-H2AX signal in siSAF-A cells being caused by replication stress. Unexpectedly, ATM inhibition caused an increase in basal γ-H2AX signal even in control cells (Fig 7C, compare left halves of Untreated and ATMi), but did not notably impact the γ-H2AX levels in siSAF-A cells. DNA-PKi has a slight effect on γ-H2AX levels. In conclusion, the increased γ-H2AX in SAF-depleted cells appears to be mediated mainly by ATR.

These data demonstrate that cells depleted for SAF-A suffer from constant replication stress, leading to more frequent (or more extended) quiescence than in control cells, which can at least partly explain the slower cell proliferation (Fig 1C) and the reduced fraction of S-phase cells (Fig 1F).

Taken all these observations together, our results demonstrate that SAF-A supports DNA replication by promoting origin licensing, fork progression speed, and fork processivity, probably by modulating chromatin compaction to ensure the optimal structure for robust DNA replication.

Discussion

Our investigation of effects of SAF-A on DNA replication establishes that SAF-A promotes replication licensing (Fig 2). Consistent with this effect, cells depleted for SAF-A showed increased origin spacing when compared to control cells, as well as reduced ability to activate dormant origins under replication stress (Fig 3B&C). It was recently demonstrated that the histone methyltransferase SET8 in contrast limits replication licensing (Shoaib et al., 2018), presumably through its activity in histone H4K20 methylation. Given that SAF-A mediates the establishment of open chromatin structure (Creamer et al., 2021; Nozawa et al., 2017; Sharp et al., 2020), it therefore appears that the correct level of origin licensing requires an appropriate balance between oppositely acting cellular mechanisms that specify chromatin compaction.

In addition to its implication in chromatin structure, SAF-A is known to affect gene expression by modulating mRNA splicing, mRNA stability and mRNA localisation (Huang et al., 2021; Xiao et al., 2012; Ye et al., 2015; Yugami et al.,

2007). We found a slight reduction in protein levels of FLAG-ORC1 and CDT1 (Fig S4) that may contribute to the reduced licensing in cells depleted for SAF-A. SAF-A depletion appears to compromise replication licensing more severely than it does licensing factor expression, so we suspect that altered chromatin structure is the major determinant of reduced licensing in siSAF-A cells, with reduced protein expression an additional contributing factor.

Once origins have initiated, replication fork progression is also affected by SAF-A depletion, with fork speed less tightly regulated (Fig 4A) and higher variance in fork processivity under replication stress (Fig 4B). These changes may reflect an increased incidence of 'chromatin obstacles' in the absence of SAF-A, corresponding to hard-to-replicate sites that challenge the replication machinery (Gadaleta and Noguchi, 2017). It has been demonstrated that processive replication through heterochromatin regions is coupled with local chromatin decompaction (Chagin et al., 2019), so that chromatin over-compaction in the absence of SAF-A may increase the number of replication fork impediments, causing the observed inconsistent fork processivity (Fig 4B). Interestingly, it is proposed that unactivated ('non-CMG-assembled') MCM proteins reduce replication fork speed to prevent DNA damage during DNA replication (Sedlackova et al., 2020). Reduced chromatin association of MCM in siSAF-A cells could be envisaged to increase the variance in fork processivity in this manner as well.

Despite having a moderate effect on origin licensing, depletion of SAF-A has a fairly mild effect on EdU incorporation levels without additional replication stress (Fig 1F, 1G, and S1D). This probably reflects the fact that under normal circumstances, MCM complex is loaded at a larger number of sites than will be utilised, so that a modest reduction in origin licensing has only slight impact on cellular DNA replication dynamics under unperturbed conditions (Ge et al., 2007; Woodward et al., 2006). We find however that there is a stronger requirement for SAF-A in enabling cells to recover from replication stress (Fig 1D&E, S1B&C). One possible explanation, based on the origin licensing defect of SAF-A-depleted cells, is that insufficient 'dormant' origins are available for activation to enable proper recovery from stress (Fig 3B&C). Inadequate licensing that fails to provide enough dormant origins may lead to chromosome segments remaining unreplicated, if incoming replication forks from both directions collapse under replication stress conditions (McIntosh and Blow, 2012).

Depletion of SAF-A leads to increased γ-H2AX signal without exogenous damage (Fig 7A&B). Aiming to identify the protein kinase (ATM, ATR, and DNA-PK) responsible for γ-H2AX, we treated cells with inhibitors against these kinases (Fig 7C). The result confirmed that the increased phosphorylation of H2AX is mainly mediated by ATR, consistent with our inference based on the diffuse γ-H2AX localisation pattern that depletion of SAF-A causes replication stress.

Although depletion of SAF-A leads to reduced overall licensing and increased origin spacing, we find that the genome-wide replication timing programme is not severely affected (Fig 5A). This is perhaps consistent with the fact that each replication timing domain contains multiple replication origins whose activation is concomitantly regulated. Therefore, even if some licenced origins are lost from a replication timing domain, the domain can still retain its programmed replication timing, enabled by correctly regulated initiation at the remaining origins. We did however observe that the sharp boundaries that normally delineate replication timing domains tend to be blurred, showing considerably increased cell-to-cell heterogeneity (Fig 5A&B). Such increased heterogeneity in timing domain boundaries will contribute to the poor separation of 'early' and 'late' replication peaks in the analysis of genome-wide tag density distribution (Fig 5C), and is likely to be an important parameter in the separation of siControl cells and siSAF-A cells by t-SNE analysis (Fig 5D). Interestingly, SAF-A protein has been reported to interact with chromatin domain boundary proteins including CTCF and cohesion subunit RAD21 (Fan et al., 2018; Zhang et al., 2019) and to be involved in defining chromatin domain boundaries (Fan et al., 2018). Depletion of SAF-A also tends to cause replication timing shift at loci where A/B compartment and Early/Late replication timing disagree (Table 1 and Fig S5A). Although in general the A/B compartment pattern corresponds well to the Early/Late replication pattern, there are some exceptional loci are where replication timing and A/B compartment patterns disagree. At some such loci, SAF-A is required for the establishment or maintenance of the replication timing pattern discordance with chromatin compartmentalisation.

SAF-A has been implicated in inactivation of X chromosomes (Hasegawa et al., 2010; Lu et al., 2020; Smeets et al., 2014). The hTERT-RPE1 cells used for our timing analysis are female, but we did not observe any obvious impact of depleting SAF-A on X chromosome replication timing (not shown). However, subtle changes in

the inactive X chromosome replication timing might not be detected, given our methodology did not separately analyse the two X chromosomes.

RNA appears to be a functional component of chromatin (Brockdorff, 2019; Michieletto and Gilbert, 2019; Rodriguez-Campos and Azorin, 2007). One recent study proposes that chromatin-associated RNA promotes open chromatin structures by neutralising the positive charges on histone tails (Dueva et al., 2019). The 'polarised' chromatin distribution we observed (Fig 1A&B) may reflect a role for SAF-A in tethering RNA to decompact chromatin (Creamer et al., 2021).

In summary, we have demonstrated that SAF-A is required for robust DNA replication, both in unperturbed conditions and in the recovery from replication stress. Moreover, we show that depletion of SAF-A leads to spontaneous replication stress and increased quiescence. Elevated expression of SAF-A in a wide range of cancers (The Cancer Genome Atlas) suggests that SAF-A is important for cancer cell survival, possibly through the management of replication stress in cancer cells (Gaillard et al., 2015; Macheret and Halazonetis, 2015). Overall, our findings reported here show that the promotion of robust DNA replication by SAF-A is crucial for its role in supporting cellular capacity for proliferation.

Materials and methods

Cell lines

Cell lines hTERT-RPE1 (Bodnar et al., 1998) and HEK293 FLAG-ORC1 were as previously described (Tatsumi et al., 2003). Cells were checked for mycoplasma contamination at regular intervals.

Cell culture

All human cell lines were cultivated in synthetic defined media (described below) supplemented with 10% foetal bovine serum (tetracycline-free), 100 U/ml penicillin, and 100 µg/ml streptomycin in 5% CO_2, ambient O_2 and at 37ºC. hTERT-RPE1 cells were generally cultivated in DMEM/F12 (Gibco), except for experiments involving dNTP analogue (e.g., EdU, CldU, and IdU) labelling where DMEM (Gibco) was used. Other cell lines were cultivated in DMEM.

siRNA

siRNA used are:

SAF-A siRNA - Human HNRNPU (3192) ON-TARGETplus SMARTpool (Dharmacon Cat#L-013501-00-0005, Horizon Discovery)

Control siRNA - Luciferase (GL2) (Dharmacon Cat#D-001100–01, Horizon Discovery)

Cells were transfected with 10 nM siRNA using Lipofectamine RNAiMAX reagent (Invitrogen, ThermoFisher).

Antibodies

Primary antibodies used were:

SAF-A - Mouse monoclonal [3G6] (Abcam, ab10297), 0.5 µg / 10^6 cells for flow cytometry and 1/10000 for western

FLAG - Mouse monoclonal [M2] (Sigma-Aldrich, F-1804), 1/200 for flow cytometry

MCM3 - Goat polyclonal (N-19) IgG (Santa Cruz Biotechnology, sc-9850), 1/200 for flow cytometry

CDT1 - Rabbit monoclonal [EPR17891] (Abcam, ab202067), 1/200 for flow cytometry

ORC1 - M monoclonal [F-10] IgG1 (Santa Cruz Biotechnology Biotechnology, sc-398734)

ORC2 - Rabbit polyclonal (Bethyl, A302-734A), 1/100 for flow cytometry

p21 - Rabbit polyclonal (C-19) (Santa Cruz Biotechnology, sc-397)

Histone H3 - Rabbit polyclonal (Abcam, ab1791)

CldU - Rat monoclonal anti-BrdU [BU1/75 (ICR1)], (Abcam, ab6326), 1/100 for DNA combing

IdU - Mouse monoclonal anti-BrdU (BD Biosciences, Cat# 347580), 1/100 for DNA combing

ssDNA - Mouse monoclonal IgG3 [16-19] (Millipore MAB3868), 1/100 for DNA comging

γ-H2AX - Rabbit monoclonal [20E3]. (Cell Signaling Technology, #9718) used at 1/400 for immunofluorescence; Alexa Fluor 647 Mouse Monoclonal [N1-431] (BD Pharmingen, 560447).

Phospho-Rb (Ser807/811) - Rabbit monoclonal [D20B12] IgG (Cell Signaling Technology, #8516), 1/800 for flow cytometry

Secondary antibodies used were:
AlexaFluor647 Donkey anti-rabbit IgG (H+L) (Abcam, ab150063)
AlexaFluor 488 Donkay anti-rabbit IgG (H+L) (Abcam, ab150065)
AlexaFluor 488 Donkay anti-mouse IgG (H+L) (Abcam, ab150109)
AlexaFluor647 Donkeyanti-goat IgG (H+L) (Abcam, ab150135)
AlexaFluor488 Goat anti-mouse IgG (H+L) (Abcam, ab150117)
Alexa Fluor 594 Goat anti-rat IgG (H+L) (Molecular Probes A-11007, ThermoFisher)
Alexa Fluor 350 Goat anti-Mouse IgG (H+L) (Molecular Probes A-11045, ThermoFisher)
Alexa Fluor 488 Goat anti-mouse IgG1 (Molecular Probes A-21121, ThermoFisher)
They were all used at 1/2000 dilution.

Chromatin fractionation

To prepare chromatin-enriched fractions (Fig S2E) cells were lysed in cytoskeleton (CSK) buffer (10 mM HEPES-KOH [pH 7.4], 100 mM NaCl, 3 mM MgCl2, 300 mM sucrose), containing 0.2% Triton X-100, 1X cOmplete protease inhibitor cocktail EDTA-free (Roche, 04693159001) and 1X HALT protease and phosphatase inhibitor (Thermo Scientific, 78446) for 10 min on ice. Lysed cells were then centrifuged for 3 min at 2000xg. The pellet was washed once with CSK buffer, centrifuged for 4 min at 3200 rpm, and resuspended in CSK buffer containing 10 µl/ml Benzonase for 30 min on ice. Samples were boiled in 1X Laemmli sample buffer for 10 min and 5% β-mercaptoethanol was added.

To prepare chromatin-enriched fractions for analysis of p21 (Fig 7a), CDT1 (Fig S2C left), CDC6 and ORC1 (Fig S2C middle), cells were lysed in Low Salt Extraction (LSE) buffer (10mM K-phosphate [pH 7.4], 10 mM NaCl, 5 mM MgCl2) containing 0.1% Igepal CA-630 and 1mM PMSF for 5 min on ice. Lysed cells were then centrifuged and the pellet was washed once with LSE buffer. The pellet was resuspended and boiled in 1X Laemmli sample buffer for 10 min and 5% β-mercaptoethanol was added.

Protein concentrations in whole cell extracts were determined using the Bio-Rad RC DC Protein assay kit. For Western blots, an equal amount of total protein was loaded on each WCE (whole cell extract) lane, and loading for the corresponding chromatin fractions was calculated based on cell-equivalency. Equal loading was further confirmed by examining total protein using Mini-PROTEAN stain-free gels (Bio-Rad).

DNA combing

For analysis of nascent DNA on DNA fibres, cells were pulse-labelled sequentially with CldU and IdU for 20 min each. Cells were then collected and DNA combing carried using FiberComb instrument (Genomic Vision, Bagneux, France) according to the manufacturer's instructions. Detection of CldU and IdU was as previously described (Garzon et al., 2019). Images were acquired on Zeiss Axio Imager M2 microscope and 63x/NA1.4 objective equipped with ORCA-Flash 4.0LT CMOS camera (Hamamatsu Photonics). Images were analysed as previously described (Garzon et al., 2019). For inter-origin distance measurements, 1 μm was converted to 2 kb based on a predetermined value (Bensimon et al., 1994; Bensimon et al., 1995).

Flow cytometry

Cell cycle analysis of cells stained with DAPI was performed as described (Hiraga et al., 2017; Watts et al., 2020). EdU labelling and its detection by flow cytometry have been previously described (Hiraga et al., 2017). Cells were extracted before fixation with low salt extraction buffer (0.1% Igepal CA-630, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 mM PMSF, 10 mM Potassium phosphate buffer (pH 7.4) unless otherwise noted. Detection and analysis of chromatin-bound proteins by flow cytometry were performed as previously described (Hiraga et al., 2017) with multiplexing as described below. For analysis of total proteins, cells were fixed with formaldehyde prior to permeabilisation. Data were acquired on Becton Dickinson LSRII or Fortessa flow cytometers with FACSDiva software (Beckton Dickinson), and analysed using FlowJo software Ver. 10.4.2 (FlowJo LLC., Ashland, OR, USA).

We found that apparent MCM levels per cell are very sensitive to the number of cells analysed (i.e. the ratio of cells to antibody during immunostaining), causing

tube-to-tube variations. To avoid this issue, we adopted a "multiplexing" strategy. In brief, before immunostaining, samples were differentially labelled with CellTrace Yellow (Molecular Probes, ThermoFisher) at a concentration unique to each sample (between 0 μM and 0.5 μM final concentration). Differentially stained samples were then mixed, and immunostained in a single tube, to eliminate tube-to-tube variations. After data acquisition by flow cytometry, cell populations were separated based on their CellTrace Yellow signal levels. We confirmed that the CellTrace Yellow signal does not affect the quantification of AlexaFluor 488 and AlexaFluor 647 signals.

Microscopy

For visualisation of chromatin DNA within the nucleus, cells were grown on ibidi chambered slides (ibi-treated) (ibidi, obtained from Thistle Scientific Ltd., Glasgow, UK). Cells were washed with PBS, and fixed with neutral buffered 4% formaldehyde (Sigma) for 15 min at RT, then permeabilised with 0.1% Triton X-100 in PBS for 15 min at RT. After washing cells three times with PBS + 0.1% Igepal CA-630, DNA was stained with 0.25 μg/ml DAPI for 30 min. Cells were finally washed and mounted in ibidi mounting medium. Eleven Z-section images were acquired at 170 nm intervals on Zeiss LSM-880/AiryScan microscope with 63x/NA 1.3 objective (with Zen Black software (Zeiss)). The middle section of the Z-stacks was assigned as the plane where each nucleus has the largest XY projection. After AiryScan processing (with the automatic 3-D AiryScan processing condition), the middle section was used for analysis. The areas of DAPI high and DAPI low regions were determined in an unbiased manner by a custom pipeline utilising Minimal Cross-Entropy on CellProfiler 3.19 (McQuin et al., 2018).

For visualisation of EdU incorporation and immunofluorescence detection of γ-H2AX, cells were grown and fixed as above, and kept in 70% ethanol. Cells were then washed with PBS, permeabilised with 0.5% Triton X-100 in PBS, and incorporated EdU was visualised using Alexa Fluor 488 EdU imaging kit (Molecular Probes C10337) according to the manufacturer's instruction, followed by indirect immunofluorescence staining of γ-H2AX. Antibodies used were p-Histone H2A.X S139 (20E3) Rabbit mAb (Cell Signalling Technology, #9718) and AlexaFluor 647 anti-rabbit IgG (Abcam, ab150063). Z-stack images were acquired at 250 nm intervals to cover entire nuclei in the field. After AiryScan processing, maximum

intensity z-projection images were created for downstream analysis using ImageJ. Detection of cells with diffuse γ-H2AX or γ-H2AX foci were carried out by using a custom CellProfiler pipeline. For the data presented in Fig 7C, Zeiss Axio Observer Z1 with Plan-Apochromat 63x/1.40 objective were used. Image were taken with ORCA-Flash4.0 V3 CMOS camera (Hamamatsu Photonics).

Single-cell replication timing analysis and bioinformatics

Homo sapiens (human) genome assembly GRCh38 (hg38) from Genome Reference Consortium was used throughout the analysis. Single-cell replication timing of siControl and siSAF-A hTERT-RPE1 cells were analysed as described (Miura et al., 2020; Takahashi et al., 2019). Replication timing boundaries were defined as corresponding to the transition point for binarized replication values (from -1 to 1 or 1 to -1) of 100 siControl mid-S phase cells. Replication timing changes (RT changes) between siControl and siSAF-A cells were calculated by comparing the average replication timing of single siControl and siSAF-A cells. The "$-\log_{10}P$" values were calculated by comparing the distribution of single-cell replication timing of 100-kb segments between siControl and siSAF-A cells using t-test. The "$-\log_{10}P$" peaks were defined as those with "$-\log_{10}P$" values above 3, which corresponds to p-values below 0.0001. Bedtools (version 2.30.0) was used for additional data analysis (Quinlan and Hall, 2010). Bedtools intersect was used to identify intersections between different data tracks, such (e.g., $-\log_{10}P$ peaks and RT boundaries).

Bedtools fisher was used to perform Fisher's exact test to evaluate the co-occurrence between $-\log_{10}P$ peaks and RT boundaries (Quinlan and Hall, 2010). One-tailed test was performed, with the null hypothesis that occurrences of $-\log_{10}P$ peaks have no correlation with RT boundaries, and the alternative hypothesis that occurrences of $-\log_{10}P$ peaks have positive correlation with RT boundaries.

t-SNE clustering analysis of replication timing profile data was performed using Rtsne R library (Krijthe, 2015), with R version 3.6.1. For t-SNE analysis, all the 40-kb chromosome segments with measurements from all the singles cells (i.e., no missing values) were used. The number of 40-kb chromosome segments used were 71979, which covers 93% of the genome. Following parameters were supplied to Rtsne; Perplexity = 19, PCA-scaling = T.

Chromatin A/B compartments in hTERT-RPE1 were determined as described (Miura et al., 2018) using a previously published Hi-C dataset in 100-kb bins (Darrow et al., 2016). The original genome coordinates in genome assembly GRCh37 (hg19) to GRch38 (hg38) were converted using UCSC liftOver tool (Hinrichs et al., 2006) and remapped using BEDOPS bedmap (with a weighted average option) (Neph et al., 2012), so that both datasets can be compared in a common genomic segmentation system. For the analysis presented in Table 1, A/B compartment and replication timing in siControl cells was compared at each 100-kb segment of the genome.

Other software and statistical analysis

GraphPad Prism (version 7; Graphpad Software, San Diego, CA, USA) and R Studio (version 1.4.1717 with R 4.1.0) were used for statistical analysis of experimental data and creating graphs. Student's t-test was used for data with normal distributions, and Mann-Whitney-Wilcoxn test was used for non-normal data (in Fig 1G, 3B&C, and 7C). Two-tailed Fisher's exact test was used in Fig 6D and Table 1. Specific conditions used were stated either in the main text or in the figure legends when necessary. Beanplot R package (version 1.2) was used for creating two-sided bean plots (Kampstra, 2008).

Acknowledgements

Information for SAF-A expression was obtained at The Cancer Genome Atlas (TCGA) Research Network (https://www.cancer.gov/tcga). We thank Dr Ryu-suke Nozawa for help in the early stage of the project, and Professor Julian Blow for advice on the 3D licensing assay. Thanks to the staff of the Iain Fraser Cytometry Centre, and Microscopy and Histology facility at the University of Aberdeen.

Footnotes

Author contributions

Conceptualisation: S.H., A.D.D.; Methodology: S.T., H.M., I.H.; Formal analysis: C.C., S.H., H.M.; Investigation: C.C., S.H., S.T.; Resources: N.G.; Writing – original

draft: S.H.; Writing – review & editing: S.H., C.C., A.D.D., I.H., N.G.; Supervision: S.H., A.D.D., N.G.; Project administration: S.H., A.D.D.; Funding acquisition, A.D.D., S.H., N. G.

Competing interests

The authors declare no competing or financial interests.

Funding

CC was supported by a BBSRC EASTBIO Doctoral Training programme PhD studentship. SH was supported by Daiwa Anglo-Japanese Foundation (12928/13746). Work in the Hiraga-Donaldson lab supported by Cancer Research UK awards C1445/A19059 and DRCPGM\100013. NG is supported by Medical Research Council (MC_UU_00007/13).

Data availability

The Illumina sequence dataset has been deposited to ArrayExpress (accession number E-MTAB-10234).

References

Arora, M., Moser, J., Phadke, H., Basha, A. A. and Spencer, S. L. (2017). Endogenous Replication Stress in Mother Cells Leads to Quiescence of Daughter Cells. *Cell Rep* **19**, 1351-1364.

Barr, A. R., Cooper, S., Heldt, F. S., Butera, F., Stoy, H., Mansfeld, J., Novak, B. and Bakal, C. (2017). DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nat Commun* **8**, 14728.

Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F. and Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science* **265**, 2096-8.

Bensimon, D., Simon, A. J., Croquette, V. V. and Bensimon, A. (1995). Stretching DNA with a receding meniscus: Experiments and models. *Phys Rev Lett* **74**, 4754-4757.

Bianco, J. N., Poli, J., Saksouk, J., Bacal, J., Silva, M. J., Yoshida, K., Lin, Y. L., Tourriere, H., Lengronne, A. and Pasero, P. (2012). Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods* **57**, 149-57.

Blow, J. J. and Ge, X. Q. (2009). A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep* **10**, 406-12.

Blow, J. J. and Tanaka, T. U. (2005). The chromosome cycle: coordinating replication and segregation. Second in the cycles review series. *EMBO Rep* **6**, 1028-34.

Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S. and Wright, W. E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349-52.

Boteva, L., Nozawa, R. S., Naughton, C., Samejima, K., Earnshaw, W. C. and Gilbert, N. (2020). Common Fragile Sites Are Characterized by Faulty Condensin Loading after Replication Stress. *Cell Rep* **32**, 108177.

Brockdorff, N. (2019). Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biol* **9**, 190213.

Buisson, R., Boisvert, J. L., Benes, C. H. and Zou, L. (2015). Distinct but Concerted Roles of ATR, DNA-PK, and Chk1 in Countering Replication Stress during S Phase. *Mol Cell* **59**, 1011-24.

Chagin, V. O., Reinhart, B., Becker, A., Mortusewicz, O., Jost, K. L., Rapp, A., Leonhardt, H. and Cardoso, M. C. (2019). Processive DNA synthesis is associated with localized decompaction of constitutive heterochromatin at the sites of DNA replication and repair. *Nucleus* **10**, 231-253.

Chanoux, R. A., Yin, B., Urtishak, K. A., Asare, A., Bassing, C. H. and Brown, E. J. (2009). ATR and H2AX cooperate in maintaining genome stability under replication stress. *J Biol Chem* **284**, 5994-6003.

Cortez, D. (2015). Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst)* **32**, 149-157.

Creamer, K. M., Kolpa, H. J. and Lawrence, J. B. (2021). Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction. *Mol Cell* **81**, 3509-3525 e5.

Darrow, E. M., Huntley, M. H., Dudchenko, O., Stamenova, E. K., Durand, N. C., Sun, Z., Huang, S. C., Sanborn, A. L., Machol, I., Shamim, M. et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**, E4504-12.

Dhuppar, S., Roy, S. and Mazumder, A. (2020). gammaH2AX in the S Phase after UV Irradiation Corresponds to DNA Replication and Does Not Report on the Extent of DNA Damage. *Mol Cell Biol* **40**.

Dimitrova, D. S. and Gilbert, D. M. (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell* **4**, 983-93.

Dimitrova, D. S., Prokhorova, T. A., Blow, J. J., Todorov, I. T. and Gilbert, D. M. (2002). Mammalian nuclei become licensed for DNA replication during late telophase. *J Cell Sci* **115**, 51-9.

Douglas, P., Ye, R., Morrice, N., Britton, S., Trinkle-Mulcahy, L. and Lees-Miller, S. P. (2015). Phosphorylation of SAF-A/hnRNP-U Serine 59 by Polo-Like Kinase 1 Is Required for Mitosis. *Mol Cell Biol* **35**, 2699-713.

Dueva, R., Akopyan, K., Pederiva, C., Trevisan, D., Dhanjal, S., Lindqvist, A. and Farnebo, M. (2019). Neutralization of the Positive Charges on Histone Tails by RNA Promotes an Open Chromatin Structure. *Cell Chem Biol* **26**, 1436-1449 e5.

Durkin, A., Albaba, S., Fry, A. E., Morton, J. E., Douglas, A., Beleza, A., Williams, D., Volker-Touw, C. M. L., Lynch, S. A., Canham, N. et al. (2020). Clinical findings of 21 previously unreported probands with HNRNPU-related syndrome and comprehensive literature review. *Am J Med Genet A* **182**, 1637-1654.

Fackelmayer, F. O., Dahm, K., Renz, A., Ramsperger, U. and Richter, A. (1994). Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur J Biochem* **221**, 749-57.

Fan, H., Lv, P., Huo, X., Wu, J., Wang, Q., Cheng, L., Liu, Y., Tang, Q. Q., Zhang, L., Zhang, F. et al. (2018). The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes. *Genome Res* **28**, 192-202.

Feng, D., Tu, Z., Wu, W. and Liang, C. (2003). Inhibiting the expression of DNA replication-initiation proteins induces apoptosis in human cancer cells. *Cancer Res* **63**, 7356-64.

Fragkos, M., Ganier, O., Coulombe, P. and Mechali, M. (2015). DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* **16**, 360-74.

Frigola, J., He, J., Kinkelin, K., Pye, V. E., Renault, L., Douglas, M. E., Remus, D., Cherepanov, P., Costa, A. and Diffley, J. F. X. (2017). Cdt1 stabilizes an open MCM ring for helicase loading. *Nat Commun* **8**, 15720.

Fu, H., Baris, A. and Aladjem, M. I. (2018). Replication timing and nuclear structure. *Curr Opin Cell Biol* **52**, 43-50.

Gadaleta, M. C. and Noguchi, E. (2017). Regulation of DNA Replication through Natural Impediments in the Eukaryotic Genome. *Genes (Basel)* **8**.

Gaillard, H., Garcia-Muse, T. and Aguilera, A. (2015). Replication stress and cancer. *Nat Rev Cancer* **15**, 276-89.

Garzon, J., Ursich, S., Lopes, M., Hiraga, S. I. and Donaldson, A. D. (2019). Human RIF1-Protein Phosphatase 1 Prevents Degradation and Breakage of Nascent DNA on Replication Stalling. *Cell Rep* **27**, 2558-2566 e4.

Ge, X. Q., Jackson, D. A. and Blow, J. J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* **21**, 3331-41.

Giacinti, C. and Giordano, A. (2006). RB and cell cycle progression. *Oncogene* **25**, 5220-7.

Gilbert, D. M. (2010). Cell fate transitions and the replication timing decision point. *J Cell Biol* **191**, 899-903.

Gilbert, D. M., Takebayashi, S. I., Ryba, T., Lu, J., Pope, B. D., Wilson, K. A. and Hiratani, I. (2010). Space and time in the nucleus: developmental control of replication timing and chromosome architecture. *Cold Spring Harb Symp Quant Biol* **75**, 143-53.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K. and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469-76.

Hegde, M. L., Dutta, A., Yang, C., Mantha, A. K., Hegde, P. M., Pandey, A., Sengupta, S., Yu, Y., Calsou, P., Chen, D. et al. (2016). Scaffold attachment factor A (SAF-A) and Ku temporally regulate repair of radiation-induced clustered genome lesions. *Oncotarget* **7**, 54430-54444.

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F. et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-8.

Hiraga, S. I., Ly, T., Garzon, J., Horejsi, Z., Ohkubo, Y. N., Endo, A., Obuse, C., Boulton, S. J., Lamond, A. I. and Donaldson, A. D. (2017). Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation. *EMBO Rep* **18**, 403-419.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C. W., Lyou, Y., Townes, T. M., Schubeler, D. and Gilbert, D. M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6**, e245.

Huang, Y., Qiao, Y., Zhao, Y., Li, Y., Yuan, J., Zhou, J., Sun, H. and Wang, H. (2021). Large scale RNA-binding proteins/LncRNAs interaction analysis to uncover lncRNA nuclear localization mechanisms. *Brief Bioinform*.

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software* **28**.

Kawabata, T., Luebben, S. W., Yamaguchi, S., Ilves, I., Matise, I., Buske, T., Botchan, M. R. and Shima, N. (2011). Stalled fork rescue via dormant replication origins in unchallenged S phase promotes proper chromosome segregation and tumor suppression. *Mol Cell* **41**, 543-53.

Kiledjian, M. and Dreyfuss, G. (1992). Primary Structure and Binding-Activity of the Hnrnp U-Protein - Binding Rna through Rgg Box. *Embo Journal* **11**, 2655-2664.

Krijthe, J. H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation.

Lau, E., Chiang, G. G., Abraham, R. T. and Jiang, W. (2009). Divergent S phase checkpoint activation arising from prereplicative complex deficiency controls cell survival. *Mol Biol Cell* **20**, 3953-64.

Leduc, M. S., Chao, H. T., Qu, C., Walkiewicz, M., Xiao, R., Magoulas, P., Pan, S., Beuten, J., He, W., Bernstein, J. A. et al. (2017). Clinical and molecular characterization of de novo loss of function variants in HNRNPU. *Am J Med Genet A* **173**, 2680-2689.

Lu, Z., Guo, J. K., Wei, Y., Dou, D. R., Zarnegar, B., Ma, Q., Li, R., Zhao, Y., Liu, F., Choudhry, H. et al. (2020). Structural modularity of the XIST ribonucleoprotein complex. *Nat Commun* **11**, 6163.

Macheret, M. and Halazonetis, T. D. (2015). DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* **10**, 425-48.

McIntosh, D. and Blow, J. J. (2012). Dormant origins, the licensing checkpoint, and the response to replicative stresses. *Cold Spring Harb Perspect Biol* **4**.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A., Karhohs, K. W., Doan, M., Ding, L., Rafelski, S. M., Thirstrup, D. et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol* **16**, e2005970.

Mendez, J. and Stillman, 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**, 8602-12.

Mendez, J., Zou-Yang, X. H., Kim, S. Y., Hidaka, M., Tansey, W. P. and Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol Cell* **9**, 481-91.

Michieletto, D. and Gilbert, N. (2019). Role of nuclear RNA in regulating chromatin structure and transcription. *Curr Opin Cell Biol* **58**, 120-125.

Miura, H., Poonperm, R., Takahashi, S. and Hiratani, I. (2018). Practical Analysis of Hi-C Data: Generating A/B Compartment Profiles. *Methods Mol Biol* **1861**, 221-245.

Miura, H., Takahashi, S., Shibata, T., Nagao, K., Obuse, C., Okumura, K., Ogata, M., Hiratani, I. and Takebayashi, S. I. (2020). Mapping replication timing domains genome wide in single mammalian cells with single-cell DNA replication sequencing. *Nat Protoc* **15**, 4058-4100.

Moeglin, E., Desplancq, D., Conic, S., Oulad-Abdelghani, M., Stoessel, A., Chiper, M., Vigneron, M., Didier, P., Tora, L. and Weiss, E. (2019). Uniform Widespread Nuclear Phosphorylation of Histone H2AX Is an Indicator of Lethal DNA Replication Stress. *Cancers (Basel)* **11**.

Moreno, A., Carrington, J. T., Albergante, L., Al Mamun, M., Haagensen, E. J., Komseli, E. S., Gorgoulis, V. G., Newman, T. J. and Blow, J. J. (2016). Unreplicated DNA remaining from unperturbed S phases passes through mitosis for resolution in daughter cells. *Proc Natl Acad Sci U S A* **113**, E5757-64.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S. et al. (2012).

BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-20.

Nevis, K. R., Cordeiro-Stone, M. and Cook, J. G. (2009). Origin licensing and p53 status regulate Cdk2 activity during G(1). *Cell Cycle* **8**, 1952-63.

Nozawa, R. S., Boteva, L., Soares, D. C., Naughton, C., Dun, A. R., Buckle, A., Ramsahoye, B., Bruton, P. C., Saleeb, R. S., Arnedo, M. et al. (2017). SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* **169**, 1214-1227 e18.

Ohta, S., Tatsumi, Y., Fujita, M., Tsurimoto, T. and Obuse, C. (2003). The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle. *J Biol Chem* **278**, 41535-40.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2.

Rodriguez-Campos, A. and Azorin, F. (2007). RNA is an integral component of chromatin that contributes to its structural organization. *PLoS One* **2**, e1182.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858-68.

Sedlackova, H., Rask, M. B., Gupta, R., Choudhary, C., Somyajit, K. and Lukas, J. (2020). Equilibrium between nascent and parental MCM proteins protects replicating genomes. *Nature* **587**, 297-302.

Serrano, M. A., Li, Z., Dangeti, M., Musich, P. R., Patrick, S., Roginskaya, M., Cartwright, B. and Zou, Y. (2013). DNA-PK, ATM and ATR collaboratively regulate p53-RPA interaction to facilitate homologous recombination DNA repair. *Oncogene* **32**, 2452-62.

Sharp, J. A., Perea-Resa, C., Wang, W. and Blower, M. D. (2020). Cell division requires RNA eviction from condensing chromosomes. *J Cell Biol* **219**.

Shoaib, M., Walter, D., Gillespie, P. J., Izard, F., Fahrenkrog, B., Lleres, D., Lerdrup, M., Johansen, J. V., Hansen, K., Julien, E. et al. (2018). Histone H4K20 methylation mediated chromatin compaction threshold ensures genome integrity by limiting DNA replication licensing. *Nat Commun* **9**, 3704.

Shreeram, S., Sparks, A., Lane, D. P. and Blow, J. J. (2002). Cell type-specific responses of human cells to inhibition of replication licensing. *Oncogene* **21**, 6624-32.

Smeets, D., Markaki, Y., Schmid, V. J., Kraus, F., Tattermusch, A., Cerase, A., Sterr, M., Fiedler, S., Demmerle, J., Popken, J. et al. (2014). Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin* **7**, 8.

Takahashi, S., Miura, H., Shibata, T., Nagao, K., Okumura, K., Ogata, M., Obuse, C., Takebayashi, S. I. and Hiratani, I. (2019). Genome-wide stability of the DNA replication program in single mammalian cells. *Nat Genet* **51**, 529-540.

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C. and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat Cell Biol* **12**, 1086-93.

Tatsumi, Y., Ohta, S., Kimura, H., Tsurimoto, T. and Obuse, C. (2003). The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. *J Biol Chem* **278**, 41528-34.

van der Maaten, L. J. P. (2014). Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 3221-3245.

van der Maaten, L. J. P. and Hinton, G. E. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605.

Wang, H., Wang, M., Wang, H., Bocker, W. and Iliakis, G. (2005). Complex H2AX phosphorylation patterns by multiple kinases including ATM and DNA-PK in human cells exposed to ionizing radiation and treated with kinase inhibitors. *J Cell Physiol* **202**, 492-502.

Ward, I. M. and Chen, J. (2001). Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. *J Biol Chem* **276**, 47759-62.

Watts, L. P., Natsume, T., Saito, Y., Garzon, J., Dong, Q., Boteva, L., Gilbert, N., Kanemaki, M. T., Hiraga, S. I. and Donaldson, A. D. (2020). The RIF1-long splice variant promotes G1 phase 53BP1 nuclear bodies to protect against replication stress. *Elife* **9**.

Woodward, A. M., Gohler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D. A. and Blow, J. J. (2006). Excess Mcm2-7 license

dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol* **173**, 673-83.

Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., Li, H., Sun, H., Zhang, Y. and Fu, X. D. (2012). Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell* **45**, 656-68.

Yates, T. M., Vasudevan, P. C., Chandler, K. E., Donnelly, D. E., Stark, Z., Sadedin, S., Willoughby, J., Broad Center for Mendelian, G., study, D. D. D. and Balasubramanian, M. (2017). De novo mutations in HNRNPU result in a neurodevelopmental syndrome. *Am J Med Genet A* **173**, 3003-3012.

Ye, J., Beetz, N., O'Keeffe, S., Tapia, J. C., Macpherson, L., Chen, W. V., Bassel-Duby, R., Olson, E. N. and Maniatis, T. (2015). hnRNP U protein is required for normal pre-mRNA splicing and postnatal heart development and function. *Proc Natl Acad Sci U S A* **112**, E3020-9.

Yugami, M., Kabe, Y., Yamaguchi, Y., Wada, T. and Handa, H. (2007). hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett* **581**, 1-7.

Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P. Y., Gao, N. and Tye, B. K. (2017a). Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer. *Nat Struct Mol Biol* **24**, 300-308.

Zhai, Y., Li, N., Jiang, H., Huang, X., Gao, N. and Tye, B. K. (2017b). Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing. *Mol Cell* **67**, 168-179.

Zhang, L., Song, D., Zhu, B. and Wang, X. (2019). The role of nuclear matrix protein HNRNPU in maintaining the architecture of 3D genome. *Semin Cell Dev Biol* **90**, 161-167.

Zimmerman, K. M., Jones, R. M., Petermann, E. and Jeggo, P. A. (2013). Diminished origin-licensing capacity specifically sensitizes tumor cells to replication stress. *Mol Cancer Res* **11**, 370-80.

Figure legends

Figure 1: SAF-A is required for robust DNA replication

(A) Specimen images showing the distribution of DNA within nuclei of cells treated with control siRNA (siControl) and SAF-A siRNA (siSAF-A). Super-resolution images of DAPI-stained DNA were identically processed and are displayed in pseudo-colour scale as shown below the images. Bar is 5 μm. (B) Quantification of "DAPI high" areas. Images at top show DAPI intensity in greyscale with identified “DAPI high” areas overlaid in red. Plots below show the proportion of nuclear area that is “DAPI high” in two independent experiments (shown in red and blue) with average values and standard deviations. Statistical significance was calculated for each biological replicate by t-test. (C) Growth of hTERT-RPE1 cells treated with control siRNA (siControl) or SAF-A siRNA (siSAF-A) in DMEM media, measured by counting cell number at each passage. Average values and standard deviations from three biological replicates are shown. Y-axis shown in $\log_2$ scale. (D) Cells depleted for SAF-A are defective in recovery from replication stress. siControl and siSAF-A cells were arrested with 4 mM HU for 24 hr, then released into fresh media. Cells were sampled at indicated time points and DNA content analysed by flow cytometry. Results from one of 5 biological replicates are shown. (E) Percentages of EdU-positive cells after removal of HU. After removal of HU, cells were pulse-labelled for 20 min with EdU at the indicated time points and EdU-positive cells were identified by flow cytometry. See Fig S1C for gating strategy. (F) Cells depleted for SAF-A show reduced DNA synthesis rate. Asynchronously growing cells were pulse-labelled with 20 μM EdU for 1 hr and collected. DNA content and the amount of EdU was measured by flow cytometry. Contour intervals are set with 5% of cells falling between successive contour lines. Results from one of 4 biological replicates are shown. (G) Incorporation of EdU in S phase cells. EdU incorporation per cell was measured in EdU-positive (S phase) cells. Violin plots show the median (solid line) and quartiles (dotted lines). Results from one of two biological replicates are shown. ** p<0.01; *** p<0.001; **** p<0.0001.

Figure 2: SAF-A is important for replication licensing.

(A) Analysis of hTERT-RPE1 cell cycle phases by DNA content and EdU incorporation. Gates used in (B) are indicated by coloured dotted parallelograms.

Contour intervals are set with 5% of cells falling between successive contour lines. (B) 3-D licensing assay in hTERT-RPE1 cells. G1 (red), S (cyan), and G2/M (orange) cell populations were distinguished as in (A). Chromatin-associated MCM3 was measured as previously described (Hiraga et al., 2017). (C) SAF-A promotes chromatin association of CDT1 protein in G1 phase. Chromatin association of CDT1 protein in control and SAF-A depleted cells was tested in hTERT-RPE1 cells. Contour line intervals are set to 5%. (D) SAF-A is required for full chromatin association of ORC1 and CDT1 proteins in G1 phase. Chromatin association of FLAG-tagged ORC1 protein in a HEK293-derived cell line was tested using anti-FLAG antibody (top panels). Chromatin association of CDT1 (middle panels) and MCM3 (bottom panels) proteins was tested in the same batch of cells. Contour line intervals are set to 5%.

Figure 3: Cells depleted for SAF-A has a reduced origin activation potential and is defective in the activation of dormant origins.

(A) Scheme of experiment. Cells were treated with either control siRNA (siControl) or SAF-A siRNA (siSAF-A) for 72 hr then pulse-labelled sequentially with CldU then IdU for 20 min each. Cells were collected, and genomic DNA subjected to DNA combing. Specimen image shows visualised CldU and IdU. (B) Inter-origin distance was measured as illustrated in cells treated with control siRNA or SAF-A siRNA, and with or without hydroxyurea (HU) treatment. Superplots below show results from 3 independent experiments (blue, orange, and grey). For HU-treated samples, CldU and IdU labellings were done at the end of 4 hr HU treatment at 0.1 mM. 0.1 mM HU does not stop DNA synthesis completely (Fig 4A). Averages from each experiment were statistically tested by pairwise t-test. (C) Inter-origin distance was measured as in Fig 3A: results from one representative experiment shown in two-sided bean plots. Note that the Y-axis is in log scale.

ns, not significant; * $p<0.05$; ** $p<0.01$.

Figure 4: SAF-A supports replication fork processivity

(A) SAF-A depletion does not affect replication fork speed. Nascent DNA was labelled as in Fig 3A, and replication fork speed measured based on IdU tract length. SuperPlots showing replication fork speed from biological replicates (4 experiments

in red, blue, orange and grey for untreated, and 2 experiments in red and blue for HU-treated conditions). Averages from each experiment were statistically tested by pairwise t-test. Forks speeds under HU-treated conditions in each experiment were also tested by F-test to compare the variances. (B) SAF-A is required for fork processivity. IdU:CldU tract length ratios were measured where the two tracts were consecutive, and $\log_2$ values are plotted. SuperPlots from multiple biological replicates (4 experiments for untreated, and 2 experiments for HU-treated conditions) are shown. Variances under HU-treated conditions were tested by F-test for each experiment.

ns, not significant; * $p<0.05$; ** $p<0.01$; **** $p<0.0001$.

Figure 5. Replication timing is affected by SAF-A

(A) 60 Mbase region of Chromosome 8 illustrating the impact of depleting SAF-A on single-cell replication timing profiles. Heat maps show replication in single mid-S phase cells (red: early replicating, blue: late replicating). Each horizontal line represents the replication profile of a single cell (33 siControl cells and 25 siSAF-A cells), in 200-kb windows. The "$-\log_{10}P$" plot (green) shows statistical significance of the difference between single-cell replication timing of siControl and siSAF-A cells. Two specimen regions showing differences between siControl and siSAF-A are magnified at the bottom. (B) Replication timing changes caused by SAF-A depletion. Violin plots show changes in (i) the whole genome, (ii) all replication timing domain boundaries, (iii) replication timing domains boundaries overlapping with $-\log_{10}P$ peaks, (iv) $-\log_{10}P$ peak genomic loci, and (v) the same set of $-\log_{10}P$ peaks scrambled to random genomic loci. Note that (iii) is an intersection of (ii) and (iv). (C) Distribution of NGS tag density in 100 cells of siControl and siSAF-A. One hundred mid-S cells were collected by a cell sorter, and NGS libraries were prepared. Tag densities were calculated for 200 kb sliding windows at 40 kb intervals across the genome. (D) t-SNE clustering analysis of replication timing in siControl and siSAF-A cells. Each dot represents a single cell.

Figure 6: Loss of SAF-A leads to spontaneous replication stress and quiescence

(A) Depletion of SAF-A leads to p21 expression. Whole-cell extracts were prepared from cells treated with control siRNA (-) and SAF-A siRNA (+), then abundance of SAF-A and p21 examined by the western blotting. Stain-free gel image represents loading control. Results from one of two biological replicates are shown. (B) Cell cycle analysis of p21 expression. Cells treated with control siRNA (siControl) or SAF-A siRNA (siSAF-A) were analysed for DNA content and p21 expression by flow cytometry. Gates used to designate p21-positive cells are shown. Results from one of two biological replicates shown. (C) Analysis of phosphorylated Rb protein in SAF-A depleted cells. Cells were treated with siRNA as in (B), fixed, and abundance of Rb protein phosphorylated at Ser-807/811 was analysed by flow cytometry. (D) Quantification of phosphorylated Rb protein. Cells with 2N DNA content were classified either into “High-P” or “Low-P” based on their levels for Phospho-Rb signals as in Fig S6B. Results from one of two biological replicates are shown. **** $p<0.0001$

Figure 7: SAF-A depletion causes replication stress

(A) Specimen image showing different γ-H2AX localisation patterns. White arrowheads indicate cell with ‘diffuse’ γ-H2AX localisation, and amber arrows indicate cells with γ-H2AX foci. Scale bar is 10 μm. (B) Depletion of SAF-A leads to spontaneous replication stress. Cells treated with control siRNA (siControl) and SAF-A siRNA (siSAF-A) were analysed for the localisation of γ-H2AX by immunofluorescence. Percentage of cells with either γ-H2AX foci or diffuse γ-H2AX localisations were scored. Averages and standard error of the means from 4 independent experiments are shown. At least 50 cells were analysed for each condition. The p-values were calculated by t-test. (C) Effects of inhibiting checkpoint kinases on γ-H2AX signals. Cells were treated with siCont or siSAF-A for 3 days. Protein kinase inhibitors and EdU were added 24 hr and 20 min before the cell fixation, respectively. Inhibitors were ATR inhibitor (1 μM VE-821), ATM inhibitor (2 μM KU-60019), and DNA-PK inhibitor (1 μM NU-7441). Integrated intensities of nuclear γ-H2AX signals in EdU-positive cells are shown in two-sided bean plots, with average indicated as black line. At least 50 EdU-positive cells were analysed for

each condition. Results from one of two independent experiments are shown. Note that the Y-axis is log_{10} scale.

ns: not significant; * p<0.05; **** <p<0.0001

Table 1. A/B compartment and Replication timing discord

Loci	Sites with Compartment & Replication timing Discord			Total
	A & Late	B & Early	Total	
All Genome	1852	2413	4265	30299
EtoL[2]	238	604	842[3]	3048
LtoE[4]	500	313	813[3]	3042

[1]Top 10% of 100-kb genomic segments showing earlier replication timing in siSAF-A cells compared to siControl cells

[2]Enrichment statistically significant ($p<0.0001$; Fisher's exact test)

[3]Top 10% of 100-kb genomic segments showing later replication timing in siSAF-A cells compared with siControl cells

Figure 1: SAF-A is required for robust DNA replication

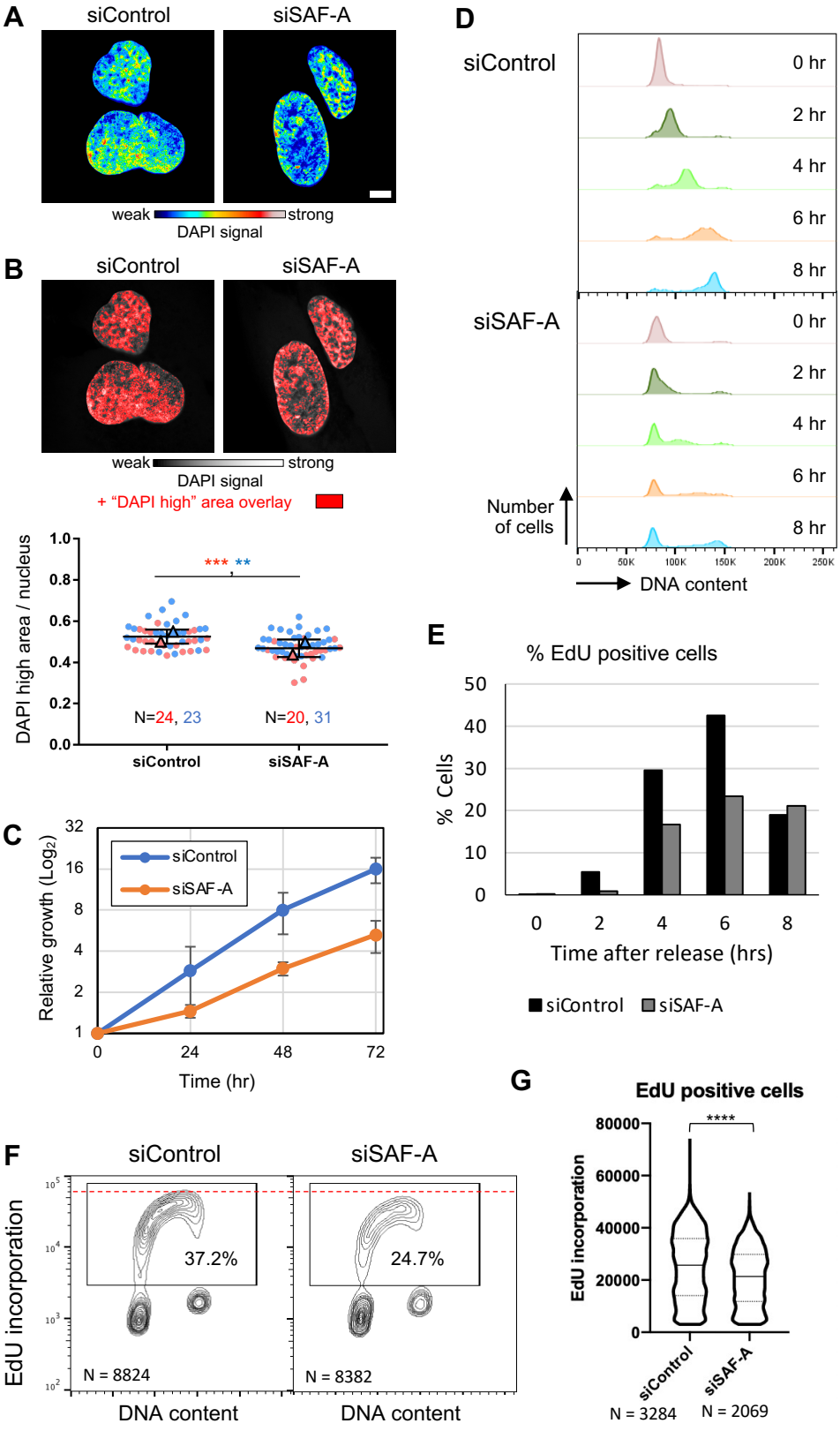


Figure 2: SAF-A is important for replication licensing

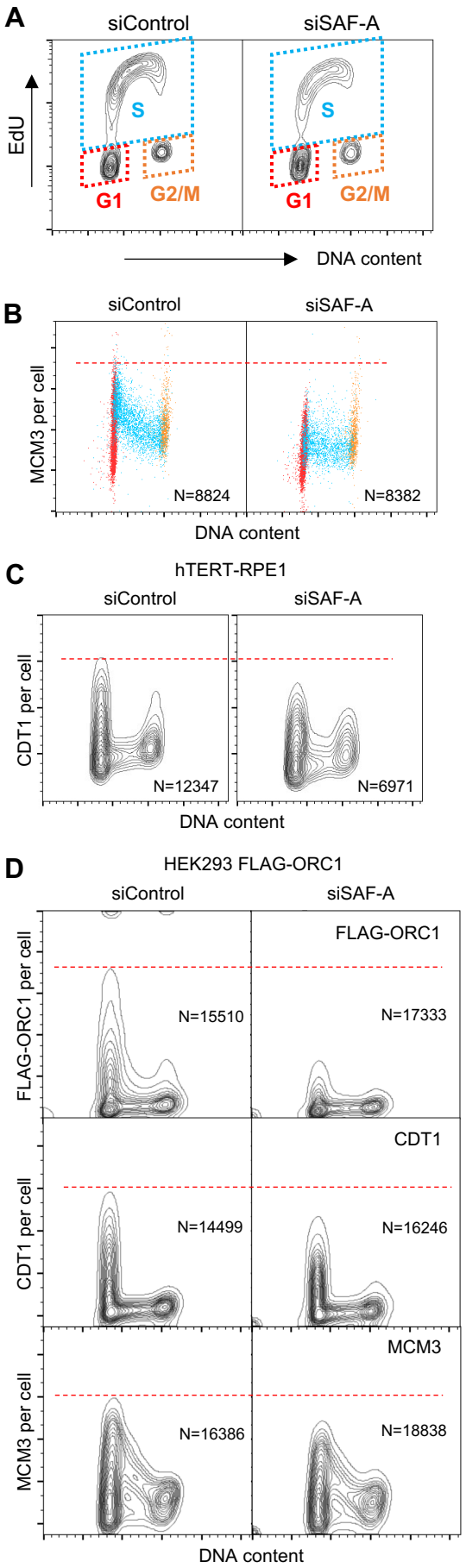


Figure 3: SAF-A depletion results in reduced origin activation

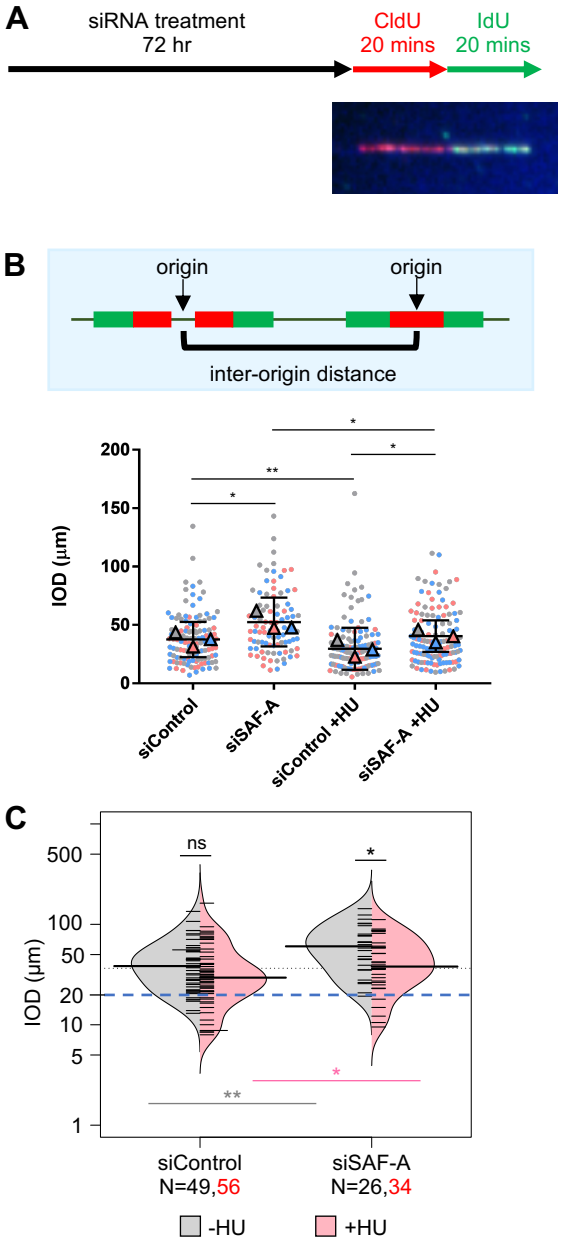


Figure 4: SAF-A supports replication fork processivity

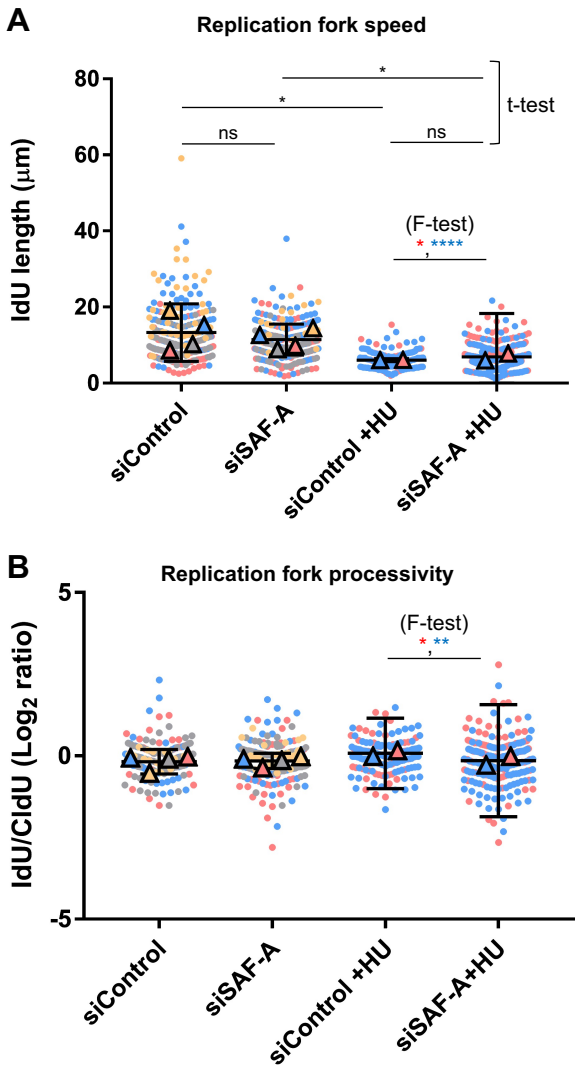
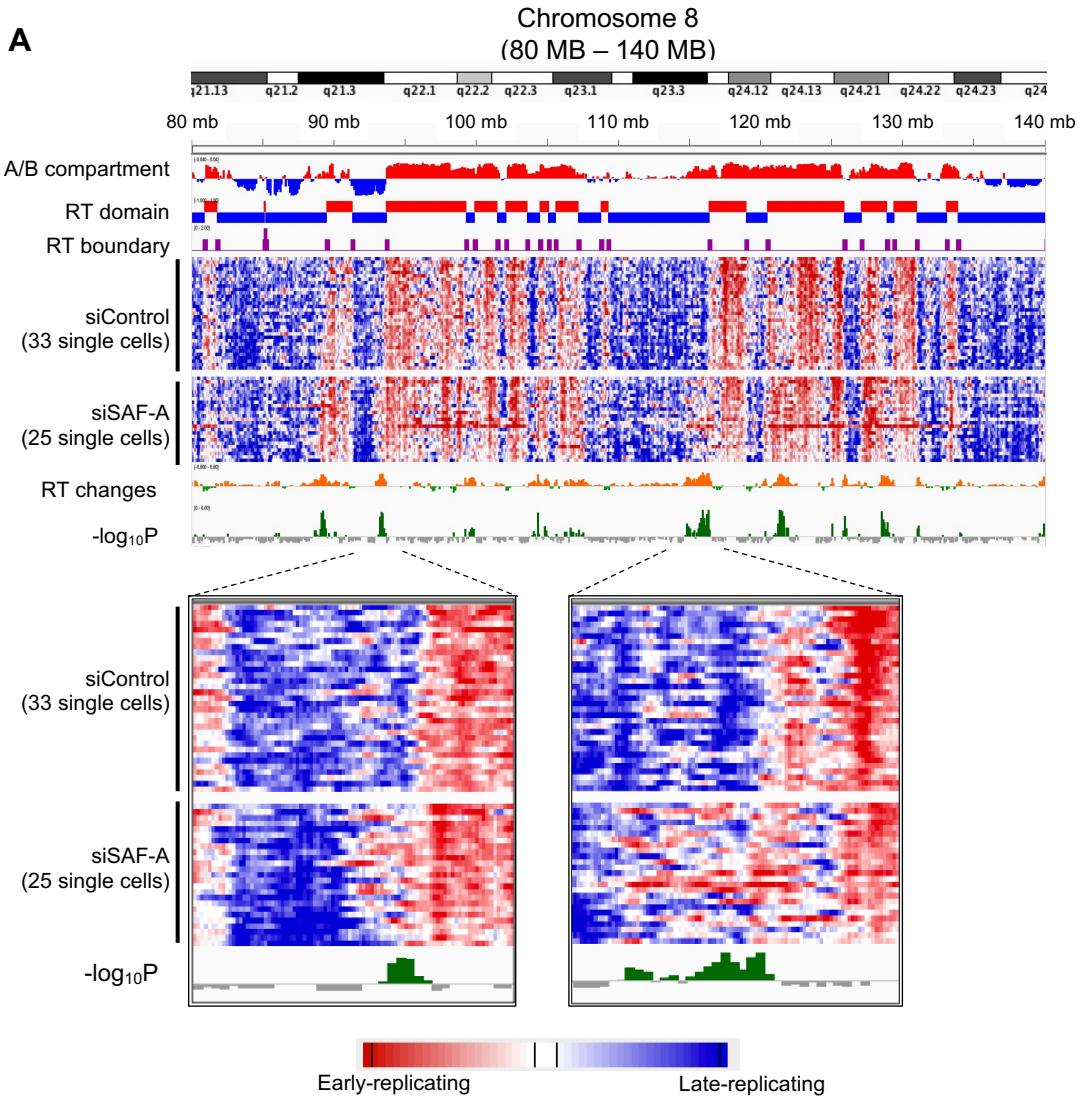
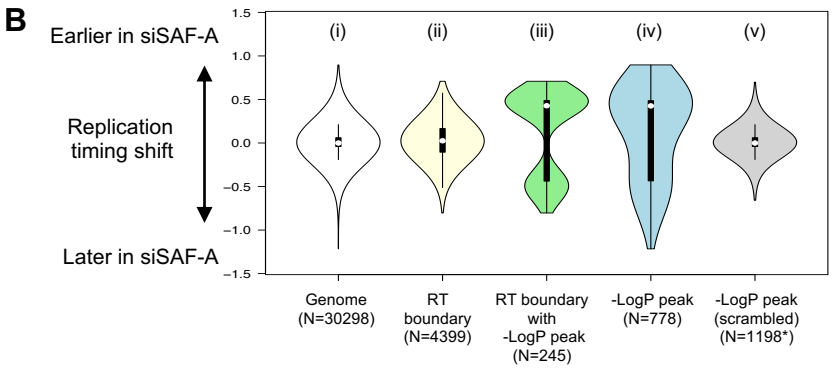
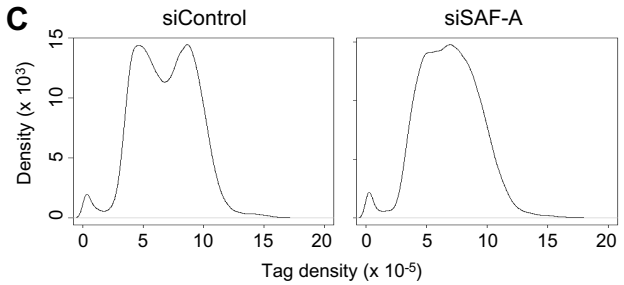


Figure 5: Replication timing is affected by SAF-A







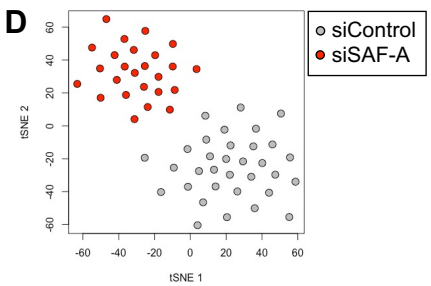


Figure 6: Loss of SAF-A leads to quiescence

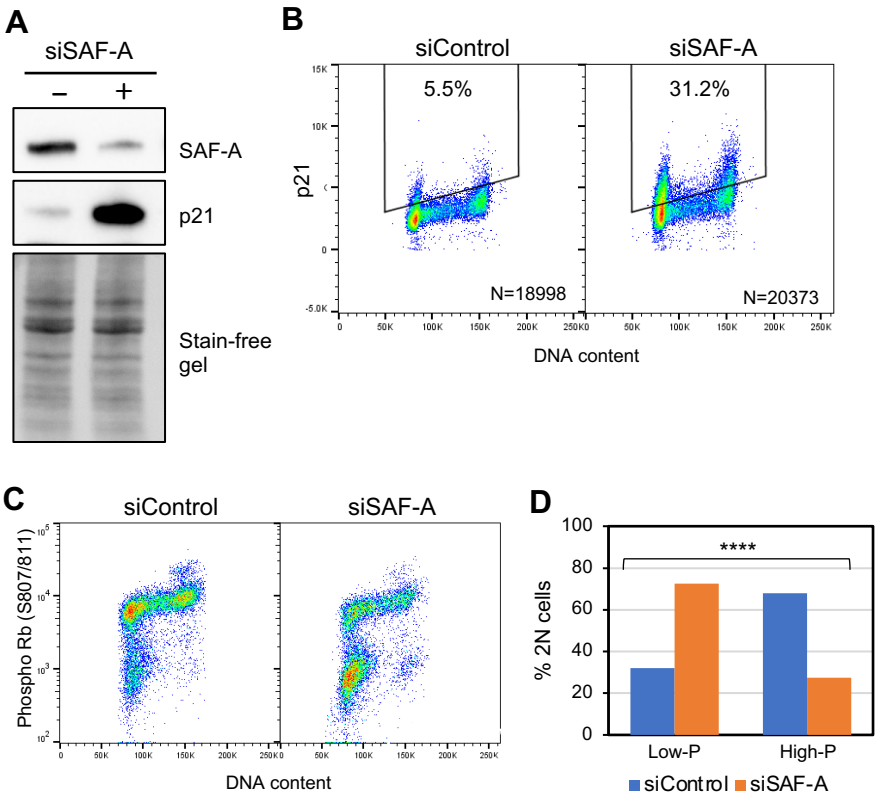
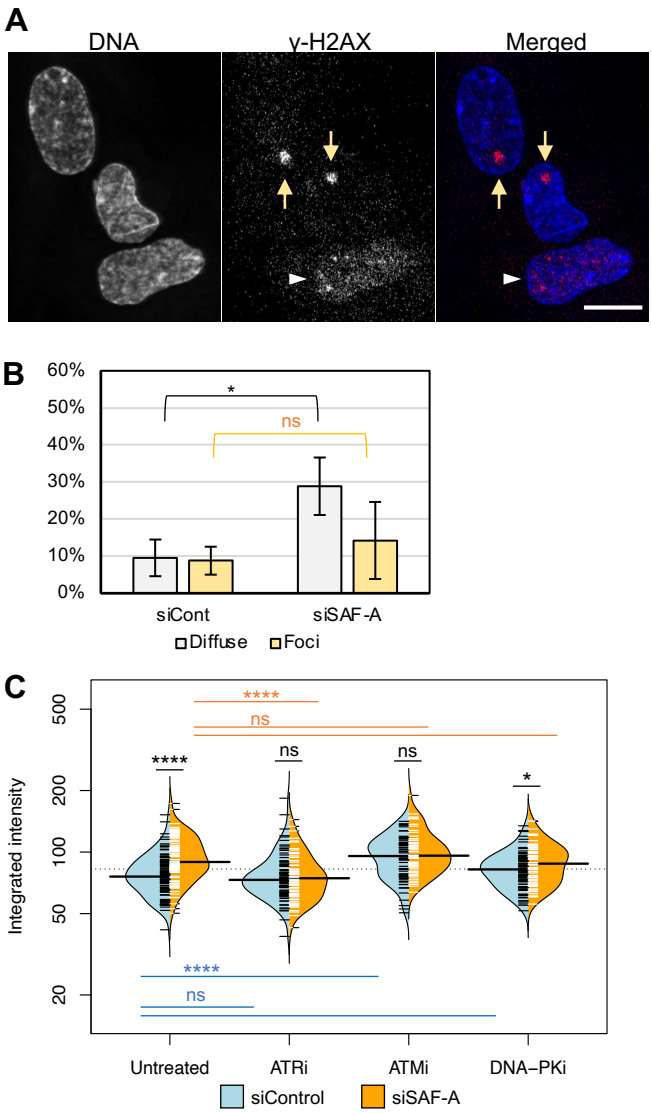


Figure 7: SAF-A depletion causes replication stress



Supplementary figures for Connolly et al. **“SAF-A promotes origin licensing and replication fork progression to ensure robust DNA replication.”**

Table of contents

Figure S1: SAF-A is required for robust DNA replication
Figure S2: Impact of SAF-A depletion on replication licensing
Figure S3: Retention of licensing proteins in CSK buffer.
Figure S4: Expression levels of licensing proteins in SAF-A depleted cells
Figure S5: Further replication timing analysis of SAF-A depleted cells
Figure S6: SAF-A depleted cells tend to enter quiescence
Figure S6: SAF-A depleted cells tend to enter quiescence

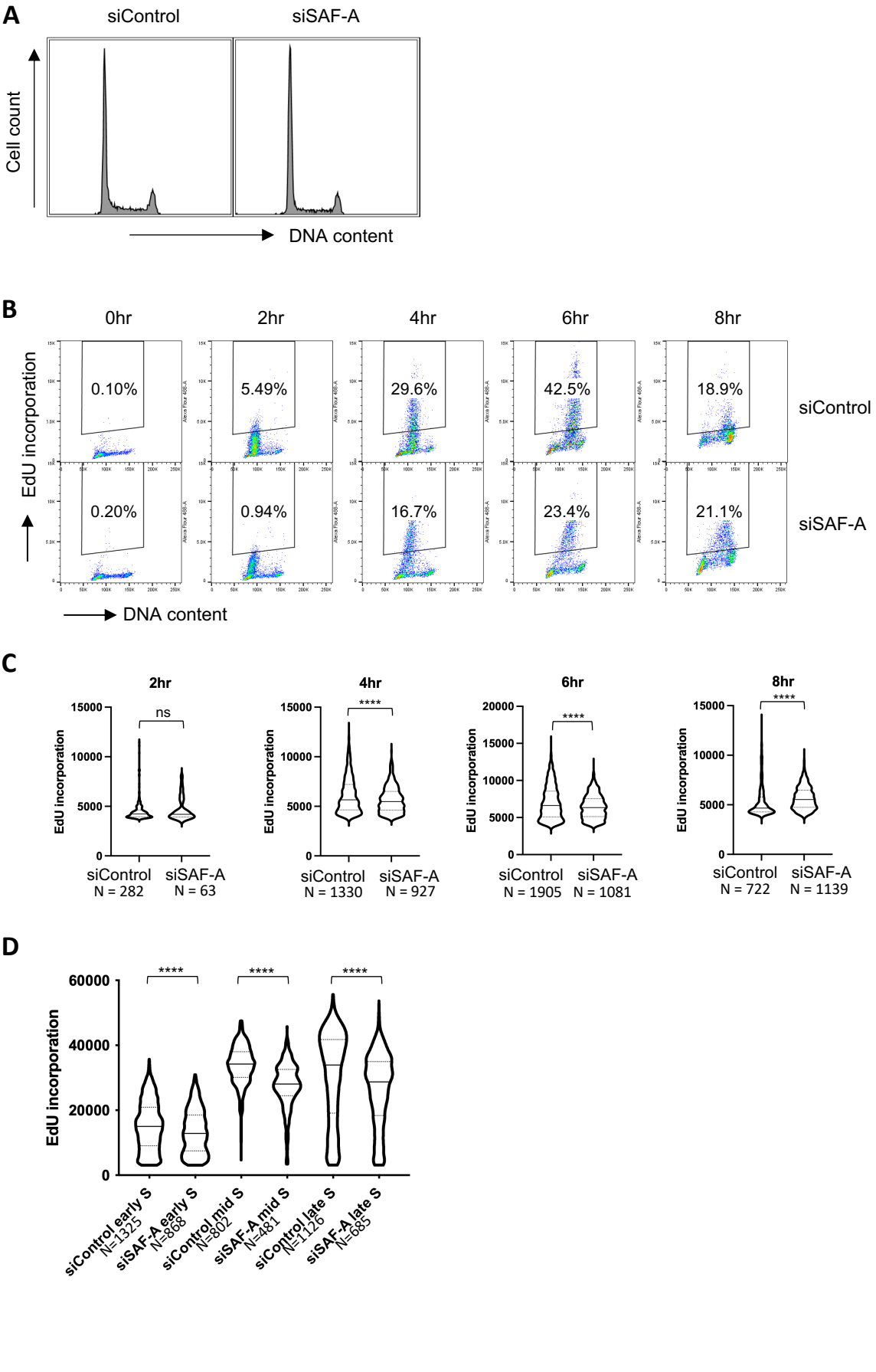

A
siControl
siSAF-A
Cell count
DNA content
B
0hr
2hr
4hr
6hr
8hr
EdU incorporation
0.10%
5.49%
29.6%
42.5%
18.9%
siControl
0.20%
0.94%
16.7%
23.4%
21.1%
siSAF-A
DNA content
C
2hr
4hr
6hr
8hr
EdU incorporation
ns
****
siControl N = 282
siSAF-A N = 63
siControl N = 1330
siSAF-A N = 927
siControl N = 1905
siSAF-A N = 1081
siControl N = 722
siSAF-A N = 1139
D
EdU incorporation
****
siControl early S N=1325
siSAF-A early S N=868
siControl mid S N=802
siSAF-A mid S N=481
siControl late S N=1126
siSAF-A late S N=685

Fig S1

Figure S1: SAF-A is required for robust DNA replication

(A) DNA content analysis of asynchronously growing control siControl- and siSAF-A-treated cells analysed by flow cytometry. (B) Impact of SAF-A depletion on recovery from replication inhibition. Cells were arrested by HU and released as in Fig 1D & E, and changes in DNA content and EdU incorporation analysed at times indicated. Gates were set as indicated, to account for shift in basal EdU signal over time. (C) EdU incorporation in EdU-positive populations in control and siSAF-A cells. Cells were treated as described in Fig 1D & E, and the amount of EdU incorporated in each EdU-positive cell was analysed by flow cytometry at indicated time points. Note that this measurement is based on only EdU-positive cells identified as in (B), and is unaffected by the EdU-negative populations. Violin plots show the median (solid grey line) and quartiles (finer dotted lines). As distribution was not normal, the p-value was calculated by Mann-Whitney-Wilcoxon test. (D) EdU incorporation in cells at various stages within S phase. The S phase cells were gated into early-, mid-, and late-S phase populations based on DNA content and EdU incorporation measured in each population. The p-value was calculated by Mann-Whitney-Wilcoxon test. Results from one of two biological replicates shown.

ns, not significant; **** $p<0.0001$.

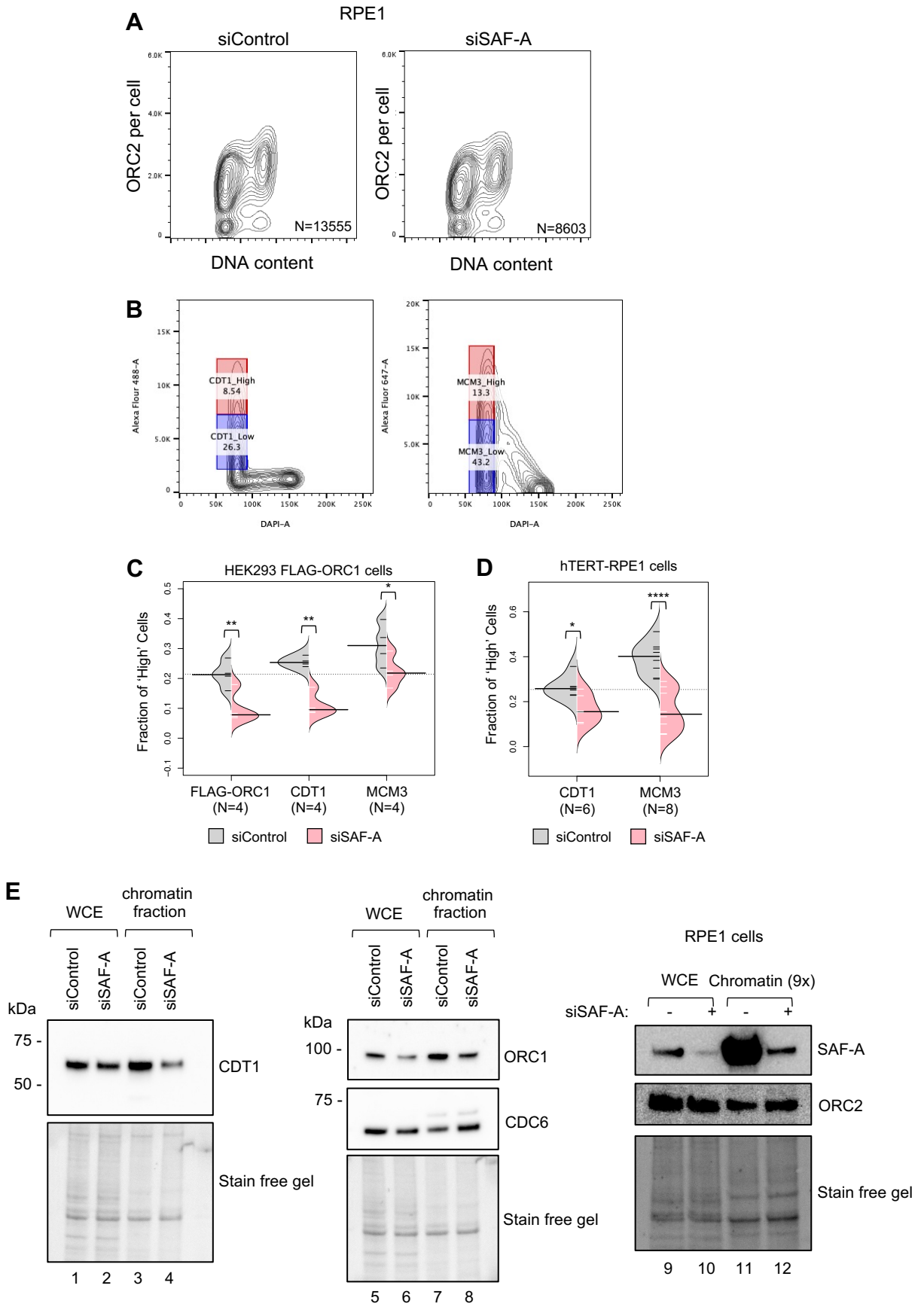

A
RPE1
siControl
siSAF-A
ORC2 per cell
DNA content
N=13555
N=8603
B
CDT1_High
8.54
CDT1_Low
26.3
MCM3_High
13.3
MCM3_Low
43.2
Alexa Fluor 488-A
Alexa Fluor 647-A
DAPI-A
C
HEK293 FLAG-ORC1 cells
Fraction of 'High' Cells
FLAG-ORC1
(N=4)
CDT1
(N=4)
MCM3
(N=4)
siControl
siSAF-A
D
hTERT-RPE1 cells
CDT1
(N=6)
MCM3
(N=8)
E
WCE
chromatin
fraction
kDa
75 -
50 -
CDT1
Stain free gel
100 -
ORC1
CDC6
RPE1 cells
WCE
Chromatin (9x)
siSAF-A:
SAF-A
ORC2
1
2
3
4
5
6
7
8
9
10
11
12

Fig S2

Figure S2: Impact of SAF-A depletion on replication licensing

(A) SAF-A depletion does not affect chromatin association of ORC2. Chromatin association of ORC2 was assessed by flow cytometry in hTERT-RPE1 cells as in Fig 3A. Contour line intervals are set to 5%. (B) Gating strategies for quantification of licensing defects in SAF-A depleted cells. G1 cells are gated as shown in left panel for quantification of chromatin-associated CDT1 or FLAG-ORC1. and as in right panel for MCM3. (C) Quantification of licensing from multiple biological experiments in HEK293 FLAG-ORC1 cells. Cells populations with 'High' and 'Low" chromatin association were assessed for each experiment with gates as illustrated in Fig S2B. 'N' refers to number of experiments shown in each bean plot. In each of the experiments, at least 3000 gated cells were analysed. Statistical significance was tested using pairwise t-test. (D) Quantification of licensing from multiple biological experiments in hTERT-RPE1 cells. Chromatin association of CDT1 and MCM were quantified as in Fig S2C. (E) Impact of SAF-A depletion on replication licensing was confirmed by western analysis of chromatin-enriched fractions of hTERT-RPE1 cells. Stain-free gel was used as a loading control. We did not assess CDC6 chromatin association by flow cytometry, because commercially available antibodies tested were unsuitable for flow cytometry (data not shown)

$*$ $p<0.05$; ** $p<0.01$; **** $p<0.0001$

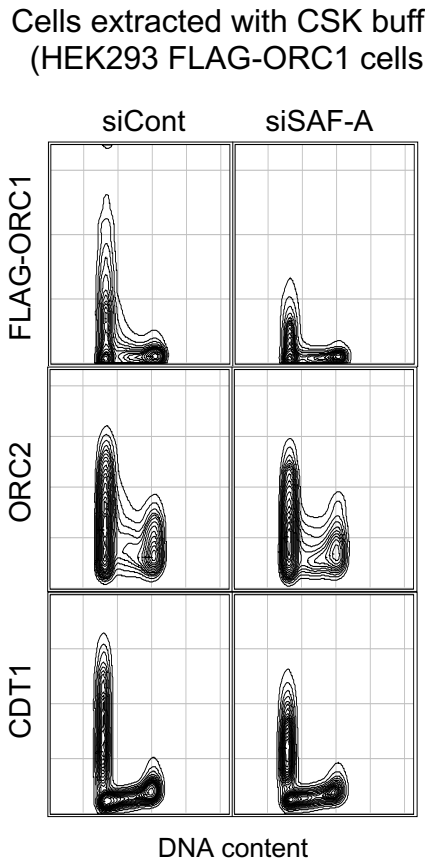


Fig S3

Figure S3: Retention of licensing proteins in CSK buffer.
HEK293 FLAG-ORC1 cells were treated with siRNA as in Fig 2C&D, and extracted with CSK buffer (10 mM HEPES-KOH (pH7.4), 300 mM sucrose, 100 mM NaCl, 1 mM $MgCl_2$). Chromatin-associated FLAG-ORC1 and CDT1 were analysed as in Fig 3B. Contour line intervals are set to 5%. Results from one of two biological replicates are shown.

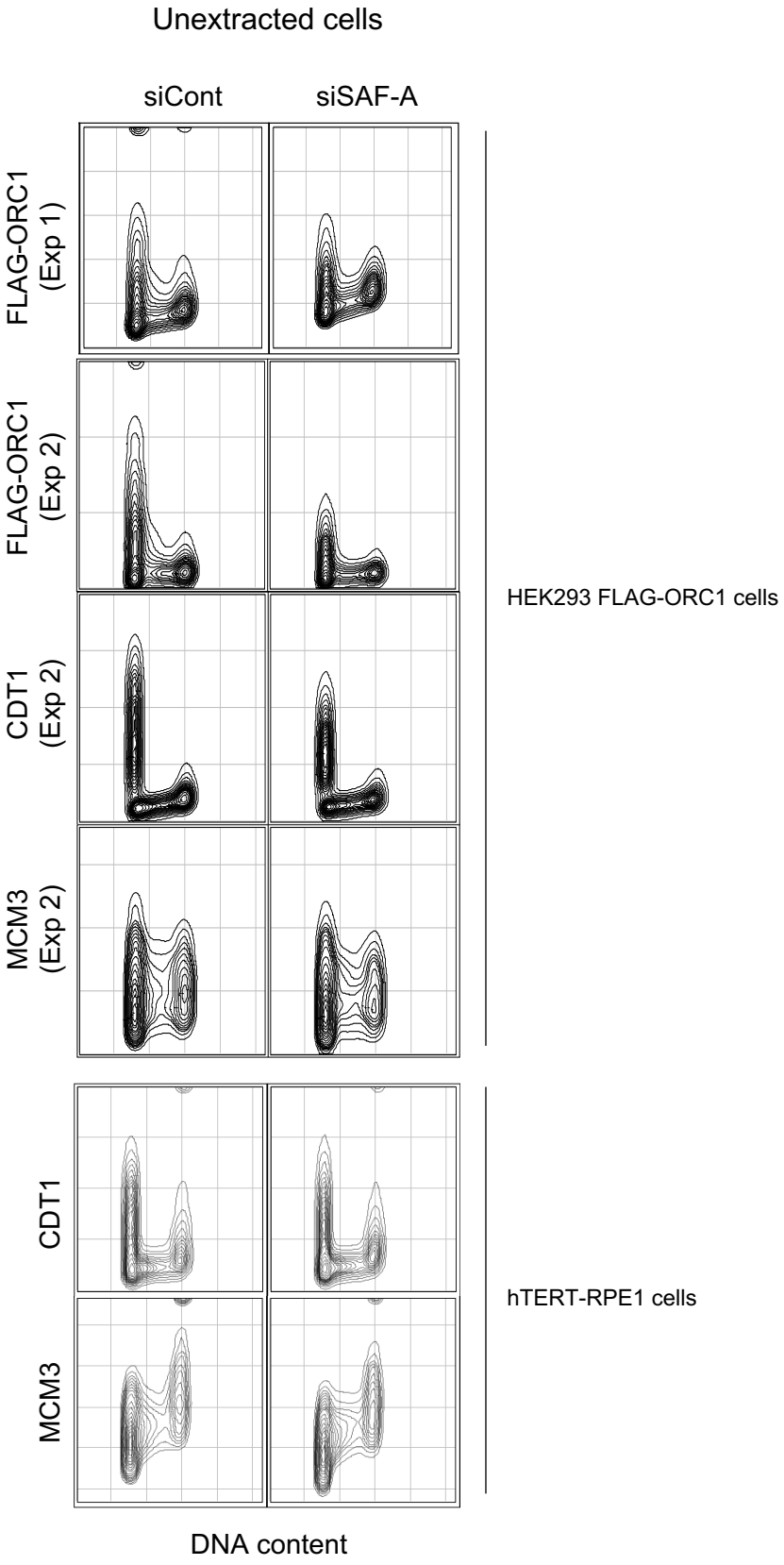
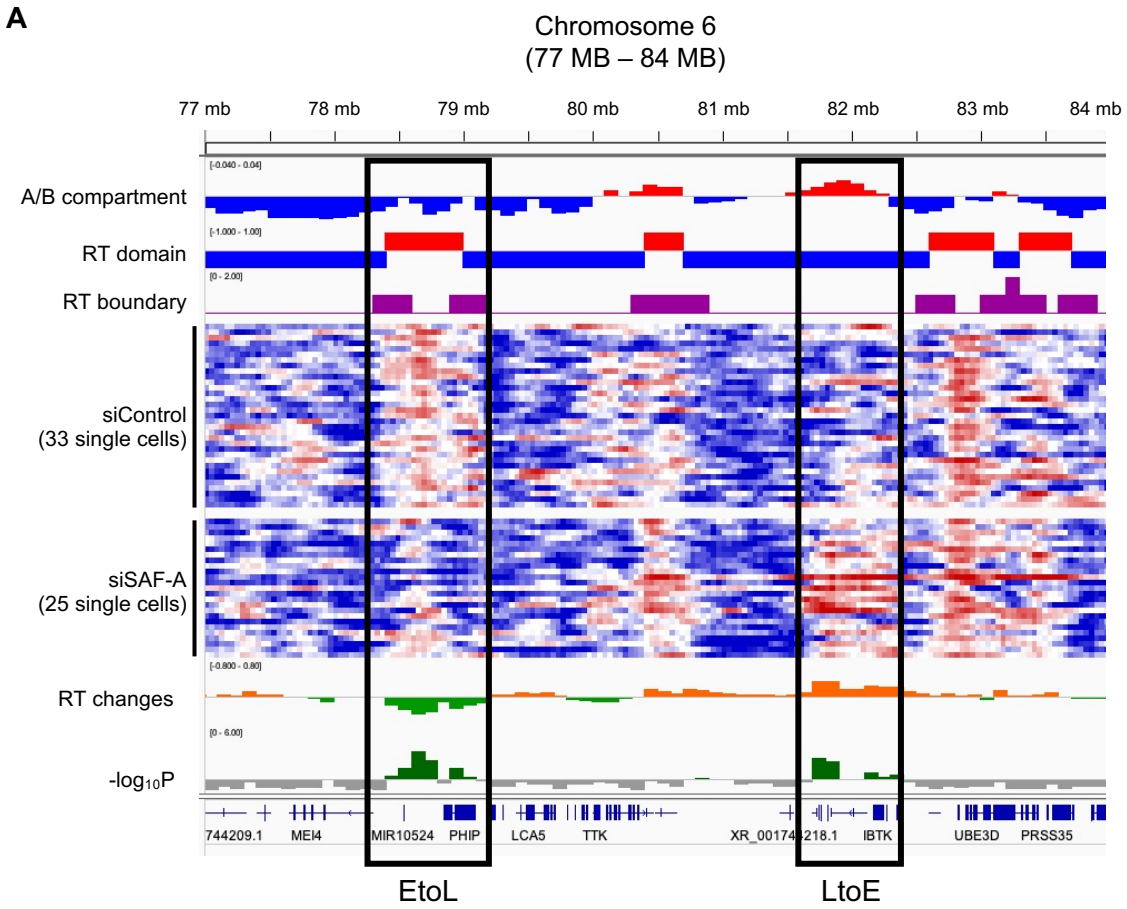

Unextracted cells
siCont
siSAF-A
FLAG-ORC1 (Exp 1)
FLAG-ORC1 (Exp 2)
CDT1 (Exp 2)
MCM3 (Exp 2)
HEK293 FLAG-ORC1 cells
CDT1
MCM3
hTERT-RPE1 cells
DNA content

Fig S4

Figure S4: Expression levels of licensing proteins in SAF-A depleted cells

HEK293 FLAG-ORC1 and hTERT-RPE1 cells were treated with siRNA as in Fig 2, but fixed with formaldehyde before permeabilization. Cellular levels of FLAG-ORC1 (in HEK293 FLAG-ORC1 cells only), CDT1, and MCM3 were analysed by flow cytometry. Contour line intervals are set to 5%. For FLAG-ORC1, results from two biological replicates are shown. For other proteins, results from one of two biological replicates are shown. Note that the impact of SAF-A depletion varies in the two experiments.


A
Chromosome 6
(77 MB – 84 MB)
77 mb
78 mb
79 mb
80 mb
81 mb
82 mb
83 mb
84 mb
A/B compartment
RT domain
RT boundary
siControl
(33 single cells)
siSAF-A
(25 single cells)
RT changes
-log10P
744209.1
MEI4
MIR10524
PHIP
LCA5
TTK
XR_001744218.1
IBTK
UBE3D
PRSS35
EtoL
LtoE

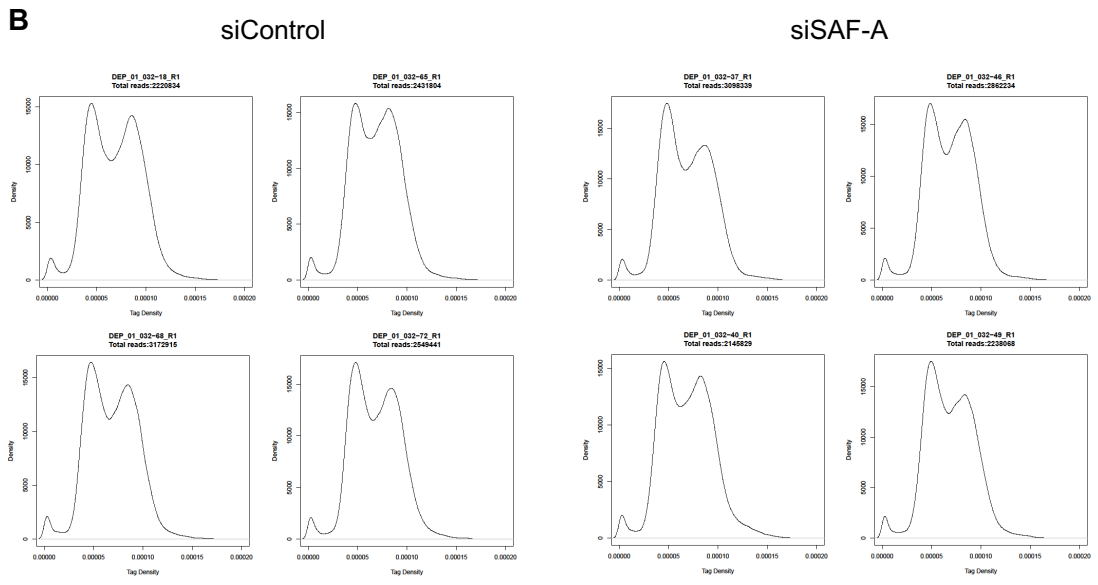

B
siControl
siSAF-A

Fig S5

Figure S5: Further replication timing analysis of SAF-A depleted cells

(A) Replication timing shift in siSAF-A cells at loci discordant for A/B compartment and replication timing. A/B compartment and replication timing properties at a region of chromosome 6 (77 MB - 84 MB) are shown as in Fig 5A. Two specimen loci with replication timing shift in siSAF-A cells are marked by black boxes. (B) Tag density distributions in 4 single siControl and 4 single siSAF-A cells. Tag densities were calculated for 200 kb sliding windows at 40 kb intervals over the genome for each cell, as in Fig 6A.

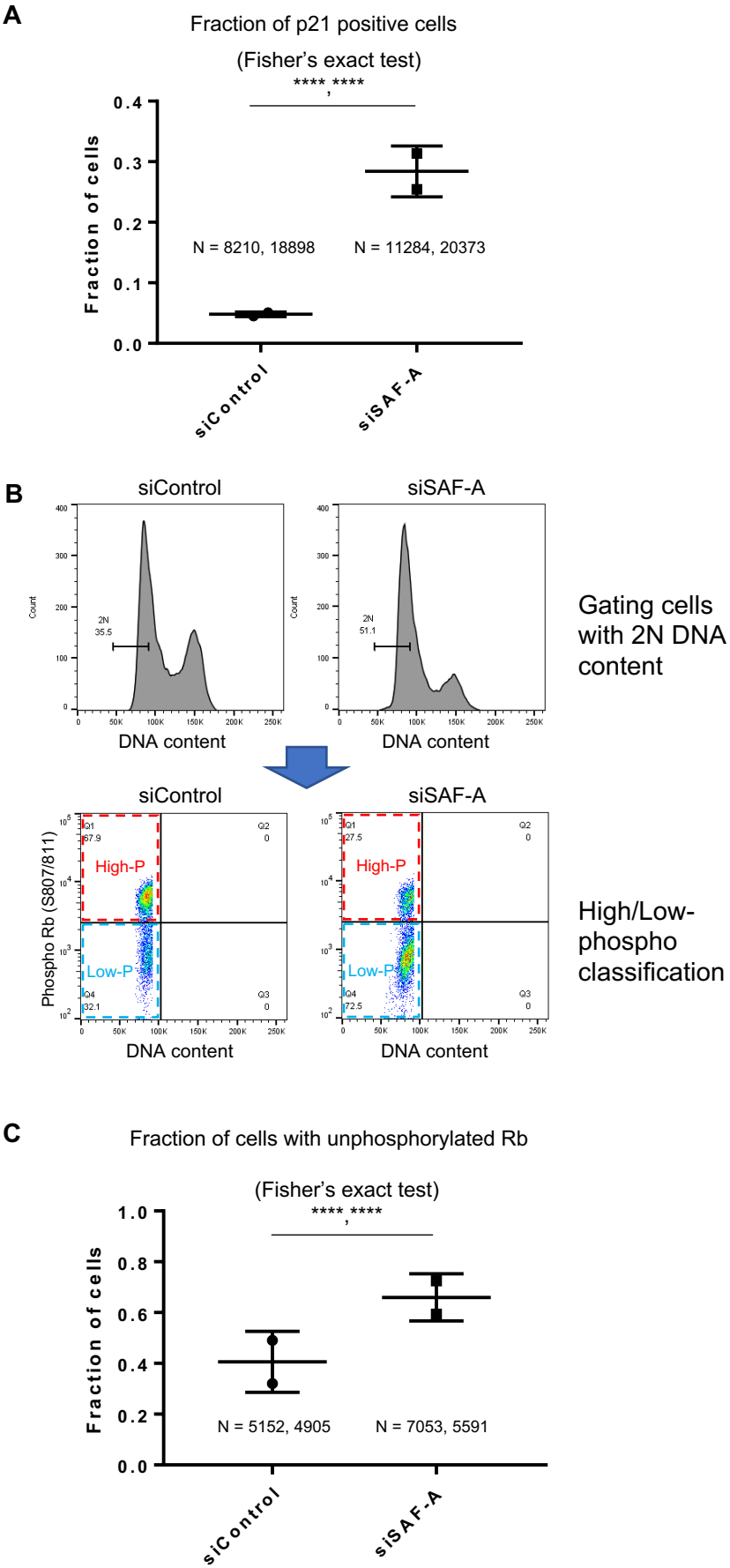

A
Fraction of p21 positive cells
(Fisher's exact test)
**** ****
Fraction of cells
0.4
0.3
0.2
0.1
0.0
N = 8210, 18898
N = 11284, 20373
siControl
siSAF-A
B
siControl
siSAF-A
Count
2N
35.5
2N
51.1
DNA content
DNA content
Gating cells with 2N DNA content
siControl
siSAF-A
Phospho Rb (S807/811)
High-P
Low-P
Q1
67.9
Q2
0
Q4
32.1
Q3
0
Q1
27.5
Q2
0
Q4
72.5
Q3
0
DNA content
DNA content
High/Low-phospho classification
C
Fraction of cells with unphosphorylated Rb
(Fisher's exact test)
**** ****
Fraction of cells
1.0
0.8
0.6
0.4
0.2
0.0
N = 5152, 4905
N = 7053, 5591
siControl
siSAF-A

Fig S6

Figure S6: SAF-A depleted cells tend to enter quiescence

(A) Analysis of p21-positive cells from two biological replicates. Fraction of p21-positive cells was determined in two biological replicates as in Fig 6B, and plotted. Distribution of cells between “p21-positive” and “p21-negative” cohorts was statistically tested for each replicate by Fisher’s exact test. (B) The gating strategy used to generate the data presented in Fig 6D. Cells analysed in Fig 6C were further gated as shown—that is, cells with 2N DNA content were analysed for phospho-Rb intensity, and separated into two populations. (C) Fractions of 2N cells with unphosphorylated Rb were determined using an approach similar to that in Fig S6B, from two biological replicates. Distribution of 2N cells to “Phospho-Rb positive” and “Phosoho-Rb negative” was statistically tested for each replicate by Fisher’s exact test. **** $p<0.0001$.

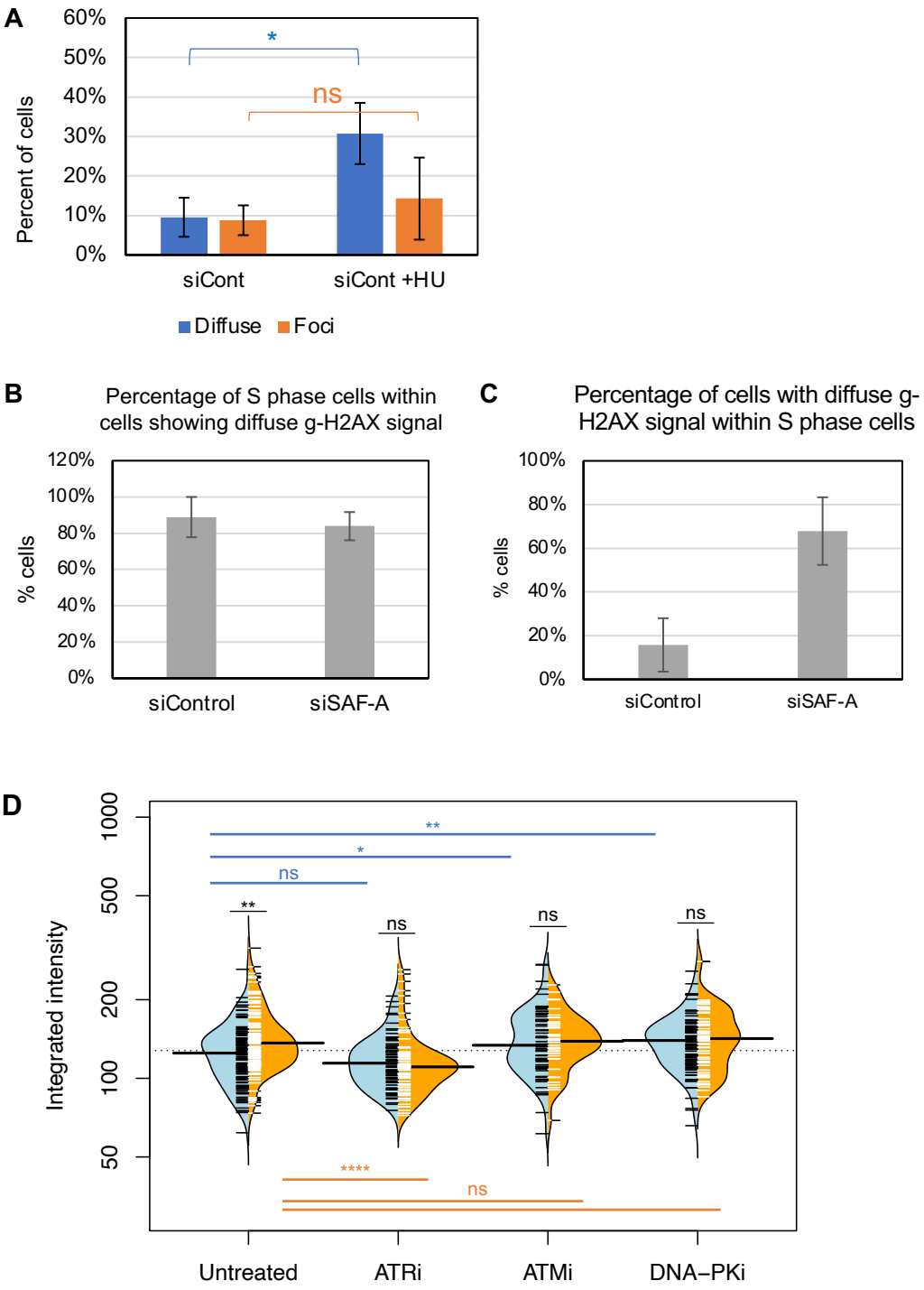

A
60%
50%
40%
30%
20%
10%
0%
Percent of cells
*
ns
siCont
siCont +HU
Diffuse
Foci
B
Percentage of S phase cells within cells showing diffuse g-H2AX signal
120%
100%
80%
60%
40%
20%
0%
% cells
siControl
siSAF-A
C
Percentage of cells with diffuse g-H2AX signal within S phase cells
100%
80%
60%
40%
20%
0%
% cells
siControl
siSAF-A
D
1000
500
200
100
50
Integrated intensity
**
*
ns
**
ns
ns
ns
****
ns
Untreated
ATRi
ATMi
DNA–PKi

Fig S7

Figure S7: Diffuse γ-H2AX signal is associated with replication stress

(A) Replication stress induces diffuse γ-H2AX signal. γ-H2AX was analysed in siControl cells without or with DNA replication stress (3 hr treatment with 1 mM HU), and cells with γ-H2AX signal scored as in Fig 7B. Averages and the standard deviations from 3 independent experiments are shown. The p-values were calculated by Student's t-test. (B) γ-H2AX signals are detected almost exclusively in S phase cells. Cells were pulse-labelled with EdU to detect S phase cells. γ-H2AX signals were detected as in Fig 7B, and the percentage of S phase (= EdU-positive) cells was calculated amongst cells with diffuse γ-H2AX signal. Averages and ranges from two independent experiments are shown. (C) The majority of SAF-A-depleted S phase cells are under replication stress. Percentage of cells with diffuse γ-H2AX signal was calculated within the S phase (=EdU-positive) population. Averages and ranges from two independent experiments are shown. (D) ATR activity is required for γ-H2AX formation in SAF-A depleted cells. A biological replicate for the date presented in Fig 7C. At least 40 EdU-positive cells were analysed for each condition.

ns: not significant; * $p<0.05$; ** $p< 0.01$; **** $<p<0.0001$