1   **Human ORC/MCM density is low in active genes and correlates with**

2   **replication time but does not solely define replication initiation zones**

3

4   Short title: Genome-wide human ORC/MCM distribution

5

6   Authors: Nina Kirstein[1], Alexander Buschle[2], Xia Wu[3], Stefan Krebs[4], Helmut Blum[4],

7   Wolfgang Hammerschmidt[2], Laurent Lacroix[3], Olivier Hyrien[3*], Benjamin Audit[5*],

8   Aloys Schepers[1,6*]

9

10   Authors affiliations:

11   [1] Research Unit Gene Vectors, Helmholtz Zentrum München (GmbH), German

12   Research Center for Environmental Health Marchioninistraße 25, 81377 Munich,

13   Germany; N.K.: current address: University of Miami, Miller School of Medicine,

14   Sylvester Comprehensive Cancer Center, Department of Human Genetics,

15   Biomedical Research Building, 1501 NW 10th Avenue, Miami, FL 33136, USA

16   [2] Research Unit Gene Vectors, Helmholtz Zentrum München (GmbH), German

17   Research Center for Environmental Health and German Center for Infection Research

18   (DZIF), Partner site Munich, Germany

19   [3] Institut de Biologie de l'ENS (IBENS), Département de Biologie, Ecole Normale

20   Supérieure, CNRS, Inserm, PSL Research University, F-75005 Paris, France;

21   X.W. current address: Zhongshan School of Medicine, Sun Yat-sen University, 74

22   Zhongshan Er Road, Guangzhou, Guangdong Province, China, 510080.

23   [4] Laboratory for Functional Genome Analysis (LAFUGA), Gene Center of the

24   Ludwig-Maximilians Universität (LMU) München, Feodor-Lynen-Str. 25, D-81377

25   Munich, Germany

26   [5] Université de Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS,

27   Laboratoire de Physique, F-69342, Lyon, France

28   [6] Current address: Monoclonal Antibody Core Facility, Helmholtz Zentrum München,

29   German Research Center for Environmental Health, Ingolstädter Landstraße 1, D-

30   85764 Neuherberg, Germany

31

32   Corresponding authors:

33   Aloys Schepers, Helmholtz Zentrum München, German Research Center for

34   Environmental Health (GmbH), Monoclonal Antibody Core Facility, Ingolstädter

35   Landstraße 1, D-85764 Neuherberg, Germany, +49 89 3187 1509,

36   schepers@helmholtz-muenchen.de

37   Benjamin Audit, Laboratoire de Physique, (UMR CNRS 5672), ENS de Lyon, 46

38   allée d'Italie, F-69364, Lyon cedex 07, France, +33 4 2623 3852,

39   benjamin.audit@ens-lyon.fr

40   Olivier Hyrien, Institut de Biologie de l'ENS (IBENS), Département de Biologie,

41   Ecole Normale Supérieure, CNRS, Inserm, PSL Research University, F-75005 Paris,

42   France, +33 1 44 32 39 20, hyrien@biologie.ens.fr

43

44

45   Keywords: ORC, MCM complex, ChIP-seq, DNA replication, OK-seq, replication

46   initiation, replication timing, transcription

2

Abstract

47

48   Eukaryotic DNA replication initiates during S phase from origins that have

49   been licensed in the preceding G1 phase. Here, we compare ChIP-seq profiles of the

50   licensing factors Orc2, Orc3, Mcm3, and Mcm7 with gene expression, replication

51   timing and fork directionality profiles obtained by RNA-seq, Repli-seq and OK-seq.

52   ORC and MCM are strongly and homogeneously depleted from transcribed genes,

53   enriched at gene promoters, and more abundant in early- than in late-replicating

54   domains. Surprisingly, after controlling these variables, no difference in ORC/MCM

55   density is detected between initiation zones, termination zones, unidirectionally

56   replicating and randomly replicating regions. Therefore, ORC/MCM density

57   correlates with replication timing but does not solely regulate the probability of

58   replication initiation. Interestingly, H4K20me3, a histone modification proposed to

59   facilitate late origin licensing, was enriched in late replicating initiation zones and

60   gene deserts of stochastic replication fork direction. We discuss potential mechanisms

61   that specify when and where replication initiates in human cells.

62

INTRODUCTION

DNA replication ensures exact genome inheritance. In human cells, replication initiates from 20,000 – 50,000 replication origins selected from a 5-10 fold excess of potential or "licensed" origins (Moiseeva & Bakkenist, 2018; Papior et al., 2012). Origin licensing, also called pre-replicative complex (pre-RC) formation, occurs in late mitosis and during the G1 phase of the cell cycle, when CDK activity is low. During this step, the origin recognition complex (ORC) binds DNA and together with Cdt1 and Cdc6, loads minichromosome maintenance (MCM) complexes, the core motor of the replicative helicase, as inactive head-to-head double hexamers (MCM-DHs) around double-stranded DNA (Bell & Kaguni, 2013; Evrin et al., 2009; Remus & Diffley, 2009). A single ORC reiteratively loads multiple MCM-DHs, but once MCM-DHs have been assembled, neither ORC, nor Cdc6, nor Cdt1 are required for origin activation (Hyrien, 2016; Powell et al., 2015). During S phase, CDK2 and CDC7 kinase activities in conjunction with other origin firing factors convert some of the MCM-DHs into pairs of active CDC45-MCM-GINS helicases (CMG) that nucleate bidirectional replisome establishment (Douglas et al., 2018; Moiseeva & Bakkenist, 2018). MCM-DHs that do not initiate replication are dislodged from DNA during replication.

In the unicellular *S. cerevisiae*, replication origins are genetically defined by specific DNA sequences (Marahrens & Stillman, 1992). In multicellular organisms, no consensus element for origin activity has been identified and replication initiates from flexible locations. Although mammalian origins fire at different times throughout S phase, neighbouring origins tend to fire at comparable times, which partitions of the genome into ~5,000 replication timing domains (RTDs), (Rivera-Mulia & Gilbert, 2016a). RTDs replicate in a timely coordinated and reproducible

4

88   order from early to late in S phase (Pope et al., 2014; Zhao et al., 2017). Different,

89   non-exclusive models exist for this temporal regulation. One model suggests that

90   RTDs are first selected for initiation followed by stochastic origin firing within

91   domains (Boulos et al., 2015; Pope et al., 2014; Rhind & Gilbert, 2013; Rivera-Mulia

92   & Gilbert, 2016b). The cascade or domino model suggests that replication first

93   initiates at the most efficient origins within RTDs and then spreads to less efficient

94   origins (Boos & Ferreira, 2019; Guilbaud et al., 2011). Nuclear processes such as

95   transcription have a major impact on replication profiles (Almeida et al., 2018; Chen

96   et al., 2019; Martin et al., 2011). It is believed that flexible chromatin features

97   including DNA and histone modifications and nucleosome dynamics contribute to

98   origin specification (Cayrou et al., 2015; Prioleau & MacAlpine, 2016; O. K. Smith &

99   Aladjem, 2014). For example, H4K20me3 has been proposed to support the licensing

100   of a subset of late replicating origins in heterochromatin (Brustel et al., 2017).

101   Different approaches have been developed to characterize mammalian origins,

102   at single-molecule level by optical methods or at cell-population level by sequencing

103   purified replication intermediates (Hulke et al., 2020). Sequencing of short nascent

104   strands (SNS-seq and INI-seq) identified initiation sites with high resolution,

105   correlating with transcriptional start sites (TSSs) and CG-rich regions enriched in G-

106   quadruplex motifs (G4s) and CpG-islands (Besnard et al., 2012; Cayrou et al., 2015;

107   Langley et al., 2016; Prioleau & MacAlpine, 2016). Sequencing of replication-bubble

108   containing restriction fragments (bubble-seq) identified origins enriched with DNase

109   hypersensitive regions and the 5'-end but not the body of active transcription units

110   (Mesner et al., 2013). INI-/SNS- and bubble-seq detected a higher origin density in

111   early RTDs than in mid-to-late RTDs (Besnard et al., 2012; Langley et al., 2016;

112   Mesner et al., 2013). Strand-oriented sequencing of Okazaki fragments (OK-seq)

113   revealed the population-averaged replication fork direction (RFD) allowing the

114   mapping of initiation and termination events (Chen et al., 2019; McGuffee et al.,

115   2013; Petryk et al., 2016; D. J. Smith & Whitehouse, 2012).

116   Consistently with bubble-seq (Mesner et al., 2013) and single molecule studies

117   (Demczuk et al., 2012; Lebofsky et al., 2006), OK-seq studies of human cells (Petryk

118   et al., 2016; Wu et al., 2018) demonstrated that replication initiates in broad but

119   circumscribed initiation zones (IZs) consisting of multiple inefficient sites. OK-seq

120   detected early-firing IZs precisely flanked on one or both sides by actively transcribed

121   genes and late-firing IZs distantly located from active genes. Early IZs are separated

122   from each other by short termination zones (TZs), which are enriched in transcribed

123   genes. Late IZs are separated by very broad, gene-poor TZs (Chen et al., 2019; Petryk

124   et al., 2016). OK-seq also identified unidirectionally replicating regions (URRs) that

125   sometimes separate IZs from TZs, as well as extended regions of null replication fork

126   directionality (NRRs) in gene deserts of uniformly late replication, consistent with

127   temporally late and spatially random initiation and termination (Petryk et al., 2016).

128   When comparing SNS-, bubble-, and OK-seq data, the highest concordance was

129   observed between IZs detected by OK- and bubble-seq (Petryk et al., 2016). Recently,

130   we found an excellent agreement between IZs determined by OK-seq and by EdUseq-

131   HU, which identified newly synthesized DNA in cells synchronously entering S phase

132   in the presence of EdU and hydroxyurea (Tubbs et al., 2018). Within the EdUseq-HU

133   zones, the most efficient sites were associated with poly(dA:dT) tracts but not any of

134   the GC-rich motifs found by SNS-seq (Tubbs et al., 2018). Repli-seq IZs, which are

135   highly consistent with OK-seq IZs, were also recently identified in high-resolution

136   replication timing (RT) profiles (Zhao et al., 2020).

137   Importantly, the number of IZs identified by OK-seq (5,000 - 10,000) (Petryk

138 et al., 2016) only account for a fraction of the 20,000 - 50,000 initiation events

139 estimated to take place in each S phase. This and replication kinetic considerations led

140 us to postulate that following efficient initiation at "master initiation zones" (Ma-IZs)

141 identified by OK-seq, replication proceeds by cascade activation of secondary zones,

142 which are too dispersed and inefficient to leave an imprint on population-averaged

143 profiles (Petryk et al., 2016). Consistently, single-molecule studies of yeast genome

144 replication by nanopore-sequencing revealed that 10-20% of initiation events occur

145 dispersedly and away from known, efficient origins, in a manner undetectable by cell

146 population methods (Hennion et al., 2020; Muller et al., 2019).

147   Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a

148 complementary method to map ORC and MCM chromatin binding. *Drosophila* ORC

149 ChIP-seq suggests a stochastic binding pattern often correlating with open chromatin

150 marks found at TSSs (MacAlpine et al., 2010). MCM ChIP-seq revealed that MCMs

151 are initially loaded at ORC binding sites in absence of Cyclin E/CDK2 activity, but

152 when Cyclin E/CDK2 activity rises in late G1, MCMs are more abundantly loaded

153 and redistributed, resulting in a loss of spatial overlap with ORC together with MCM-

154 DH displacement from actively transcribed genes (Powell et al., 2015). In human

155 cells, ChIP-seq experiments with single ORC subunits led to the identification of

156 13,000 to 52,000 potential ORC binding sites (Dellino et al., 2013; Miotto et al.,

157 2016). In a previous study, we used the Epstein-Barr virus (EBV) as model system to

158 compare ORC binding and replication initiation. During latency, replication of the

159 EBV genome is entirely dependent on the human replication initiation machinery. A

160 5-10-fold excess of potential origins are licensed per genome with respect to 1-3

161 initiation event(s) mapped (Norio & Schildkraut, 2001, 2004; Papior et al., 2012). A

162   recent genome-wide Mcm7 binding study in human HeLa cells proposed that MCM-

163   DHs bind in excess regardless of the chromatin environment, but that origin activation

164   preferentially occurs upstream of active TSSs (Sugimoto et al., 2018).

165   Here, we present the first comparative survey of four different pre-RC

166   components and replication initiation events in the human genome by combining

167   ChIP-seq and OK-seq analyses in the lymphoblastoid Raji cell line. We perform ORC

168   and MCM ChIP-seq in cell cycle synchronized pre-replicative (G1) chromatin, to

169   obtain a comprehensive picture of ORC/MCM behavior before replication. We find

170   ORCs and MCMs broadly distributed over the genome, where ORC density highly

171   correlates with early replication timing, an effect less prominently observed for

172   MCMs. We observe that active transcription locally influences ORC and MCM

173   distribution. ORC/MCM are displaced from actively transcribed gene bodies and

174   enriched at active gene promoters. ORC/MCM density is homogeneous over non-

175   transcribed genes and intergenic regions of comparable RT. Consequently, MCM

176   depletion characterizes genic borders of early IZs but not of other IZ borders.

177   Furthermore, URRs, which are refractory to replication initiation, show a similar

178   ORC/MCM density as IZs and TZs. NRRs do not show higher ORC/MCM densities

179   than IZs, TZs and URRs. These findings suggest that ORC/MCM densities do not

180   solely determine IZs and that a specific contribution of the local chromatin

181   environment is required. Indeed, we previously showed that IZs are enriched in open

182   chromatin marks typical of active or poised enhancers (Petryk et al., 2016). Here, we

183   further show that a subset of non-genic late IZs are enriched in H4K20me3,

184   confirming our previous finding that H4K20me3 enhances origin activity in certain

185   chromatin environments (Brustel et al., 2017; Shoaib et al., 2018).

8

186   Our findings support the cascade model for replication initiation: although, on

187   average, the entire genome (except transcribed genes) is licensed for replication

188   initiation, adjacent active transcription and internal epigenetic marks specify the

189   extent of Ma-IZs, from which diverging forks emanate before secondary origin

190   activation takes place.

Results

MODERATE AVERAGING IS THE BEST APPROACH FOR ORC AND MCM-DH

DISTRIBUTION ANALYSIS

To obtain a complete picture of ORC and MCM distributions prior to replication initiation, we cell-cycle fractionated human lymphoblastoid Raji cells by centrifugal elutriation into a pre-replicative G1 population, which is enriched for ORC/MCM bound chromatin (Papior et al., 2012). Propidium Iodide staining followed by FACS (Suppl. Fig. 1a) and Western blot analyses of cyclins A, B, and H3S10 phosphorylation (Suppl. Fig. 1b) confirmed the cell cycle stages. To ensure unbiased detection of ORC and MCM positions by ChIP-seq, we simultaneously targeted two members of both complexes: Orc2, Orc3, Mcm3 and Mcm7, using validated ChIP-grade antibodies (validated in: Papior et al., 2012; Ritzi et al., 2003; Schepers et al., 2001). ChIP efficiency and quality were measured using the Epstein-Barr virus latent origin *oriP* as reference (Suppl. Fig. 1c). The Raji cell line contains 50-60 EBV episomes (Adams et al., 1973). The viral protein EBNA1 recruits ORC to *oriP's* dyad symmetry element, followed by MCM-DH loading. We detected both ORC and MCM at the dyad symmetry element, as expected (Papior et al., 2012; Ritzi et al., 2003).

ChIP-sequencing of two ORC (Orc2, Orc3) and three MCM (Mcm3, Mcm7) replicates resulted in reproducible, but dispersed ChIP-seq signals as exemplified at the well-characterized replication origin Mcm4/PRKDC (Fig. 1a). We first employed the MACS2 peak-calling program (Feng et al., 2012; Zhang et al., 2008), but found that the obtained results were too dependent on the chosen program settings. At genome-wide levels, we found that ORC and MCM distributions were too dispersed

10

215 to be efficiently captured by peak calling (data not shown), requiring an alternative

216 approach.

217   Consequently, we summed up the reads of the ChIP replicates in 1 kb bins and

218 normalized the signals against the mean read frequencies of each ChIP sample and

219 against input, as is standard in most ChIP-seq analyses. However, in line with a

220 previous report (Teytelman et al., 2009), the input sequencing control was

221 differentially represented in DNase HS regions or regions of biological function. For

222 example, the input was significantly underrepresented in DNase HS regions (DNase

223 HS clusters retrieved from an ENCODE dataset comprising 125 cell lines, see

224 Material and Methods section; Suppl. Fig. 2a), TSSs (Suppl. Fig. 2b), and early RTDs

225 (Suppl. Fig. 2c). As sonication-hypersensitive regions correlate with DNase HS

226 regions (Schwartz et al., 2005), we carefully compared our input-normalized approach

227 versus no input normalization: some quantitative but no qualitative changes were

228 observed. We have included figures showing non-normalized data in the Supplement.

229 Furthermore, the 1 kb bin size was large enough to average out stochastic variations

230 of ORC/MCM, but small enough to detect significant local changes. For example,

231 ORC/MCM enrichments at the Mcm4/PRKDC origin were detected after binning

232 (Fig. 1b, Suppl. Fig. 3a without input division). The relative read frequencies of

233 Orc2/Orc3 (Fig. 1c, Suppl. Fig. 3b without input division) and Mcm3/Mcm7 (Fig. 1d,

234 Suppl. Fig. 3c) showed high Pearson correlation coefficients of $r = 0.866$ and $r =$

235 0.879, respectively. The correlations between ORC and MCM were only slightly

236 lower (Mcm3/Orc2/3: $r = 0.775/0.757$, Mcm7/Orc2/3: $r = 0.821/0.800$, Fig. 1e, Suppl.

237 Fig. 3d). Hierarchical clustering based on Pearson correlation distance between ChIP

238 profiles showed that ORC and MCM profiles clustered together. We conclude that

239 this binning approach is valid for analyzing our ChIP-seq data.

240
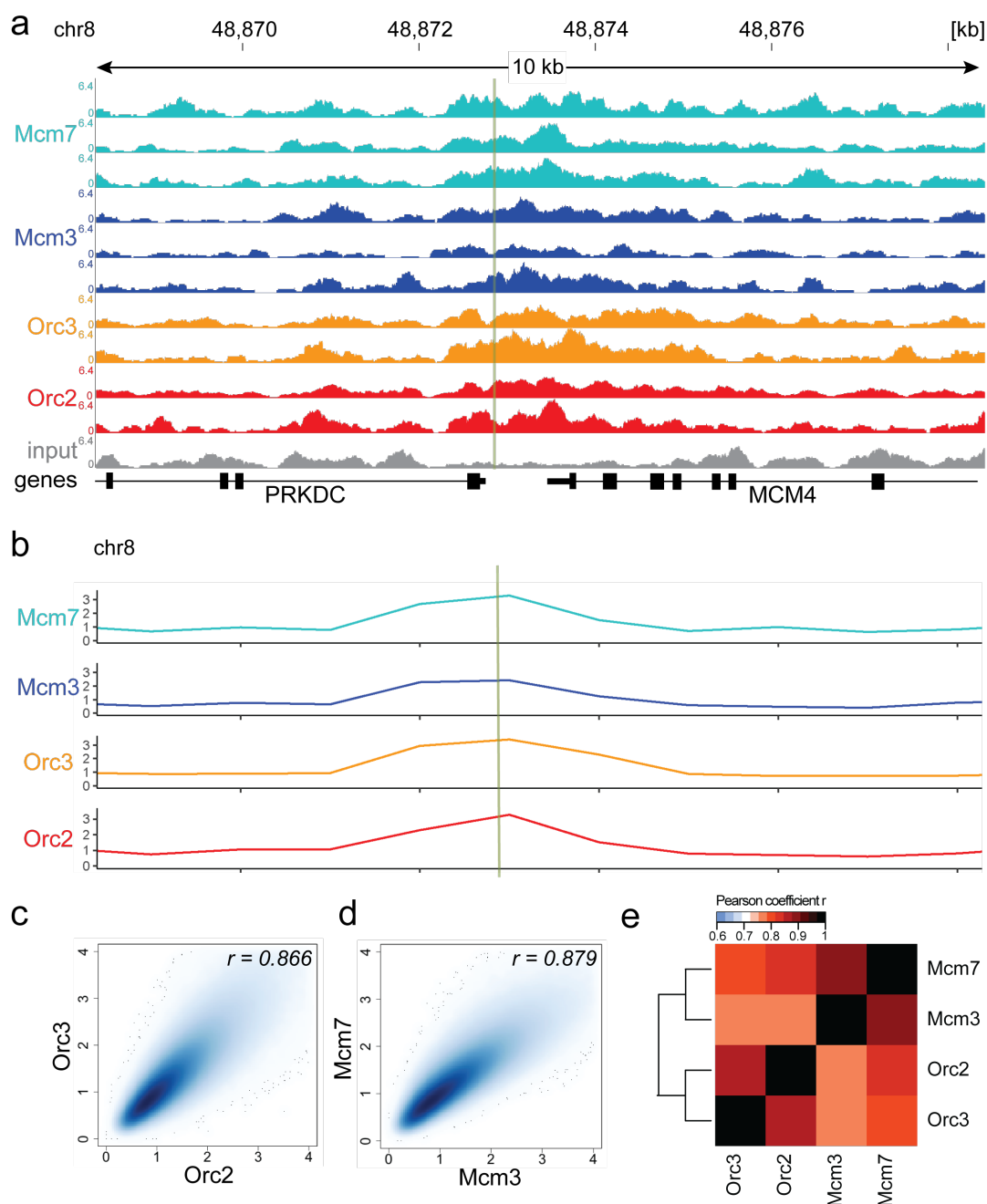
Figure 1: ORC/MCM ChIP-seq is best analyzed using a moderate averaging approach.
a) Sequencing profile visualization in UCSC Genome Browser (hg19) at the Mcm4/PRKDC origin after RPGC normalization: two samples of Orc2, Orc3 and three samples of Mcm3, Mcm7, plotted against the input. The profiles are shown in a 10 kb window (chr8: 48,868,314 - 48,878,313), the position of the origin is indicated as green line. b) The profile of ORC/MCM ChIP-seq after 1 kb binning at the same locus. The reads of replicates were summed and normalized by the total genome-wide ChIP read frequency followed by input division. Y-axis represents the resulting relative read frequency. c) Correlation plot between Orc2 and Orc3 relative read frequencies in 1 kb bins. d) Correlation plot between Mcm3 and Mcm7 relative read frequencies in 1 kb bins. e) Heatmap of Pearson correlation coefficients r between all ChIP relative read frequencies in 1 kb bins. Column and line order were

12

252 determined by complete linkage hierarchical clustering using the correlation distance (d = 1-
253 r).

254   Miotto et al. (2016) demonstrated that Orc2 positions highly depend on

255 chromatin accessibility and colocalize with DNase hypersensitive (HS) sites present

256 at active promoters and enhancers. Furthermore, Sugimoto et al. (2018) observed that

257 active origins, enriched with Mcm7 correlate with open chromatin sites. We indeed

258 found a significant enrichment of ORC/ MCM at DNase HS regions larger than 1 kb,

259 compared to regions deprived of DNase HS sites (Suppl. Fig. 3e), which further

260 validated our data.

261

262 ORC/MCM ARE ENRICHED IN INITIATION ZONES DEPENDENT ON TRANSCRIPTION

263   After confirming the validity of the ChIP experiments and establishing an

264 analysis approach based on moderate averaging, we compared the relative read

265 frequencies of ORC/MCM to active replication initiation units. Using OK-seq in Raji

266 cells (Wu et al., 2018), we calculated the RFD (see methods), and delineated zones of

267 preferential replication initiation as ascending segments (AS) of the RFD profile. OK-

268 seq detects RFD upshifts that define origins to kb resolution in yeast (McGuffee et al.,

269 2013), but in mammalian cells these transitions are more gradual, extending over 10-

270 100 kb (Chen et al., 2019; McGuffee et al., 2013; Petryk et al., 2016; Tubbs et al.,

271 2018; Wu et al., 2018). To assess ChIP signals within AS, we only kept ASs > 20 kb.

272 Using the RFD shift across the ASs (ΔRFD) as a measure of replication initiation

273 efficiency, we further required ΔRFD > 0.5 to make sure ASs corresponded to

274 efficient IZs. In total, we selected 2,957 ASs, with an average size of 52.3 kb, which

275 covered 4.9% (155 Mb) of the genome (Fig. 2a, green bars, Table 1). 2,451 (83%) of

276 all AS located close to genic regions (ASs extended by 20 kb on both sides

13

277   overlapped with at least one annotated gene). 673 ASs (22.8% of all AS) were flanked

278   by actively transcribed genes (TPM > 3; TPM: transcript per million) on both sides

279   (type 1 AS) with less than 20 kb between AS borders and the closest transcribed gene.

280   1,026 ASs (34.7%) had only one border associated to a transcribed gene with TPM >

281   3 (type 2 AS). 506 ASs (17.1%) were devoid of proximal genes (non-genic AS),

282   where 20 kb extended ASs did not overlap with any annotated gene (Table 1).

283   Although the slope did not change considerably in the different AS types, type 1 ASs

284   were on average the most efficient, while non-genic ASs were slightly less efficient

285   (Suppl. Fig. 4a). Furthermore, type 1 and type 2 ASs located to early replication

286   timing domains, while non-genic AS were predominantly late replicating (Suppl. Fig.

287   4b), which is in agreement with IZ previously described for GM06990 and HeLa

288   (Petryk et al., 2016).

289   Table 1: Characterization of different AS subtypes.
290   Only AS ≥ 20kb with ΔRFD>0.5 were considered. Genic ASs: ASs extended 20 kb on both
291   sides is overlapped by genic region(s) irrespective of transcriptional activity. Type 1 and type
292   2 AS: AS flanked by expressed genes (TPM ≥ 3) within 20kb on both sides (type 1) or one
293   side (type 2). Non-genic: no annotated gene ± 20kb of AS border.

|	Number	Genome coverage [%]	Average length [kb]
All AS	2,957	4.9	52.3
Genic AS	2,451	4.1	52.3
Type 1 AS	673	1.1	50.7
Type 2 AS	1,026	5.2	50.2
Non-genic AS	506	0.8	50.7

294   Replication can only be activated when functional pre-RCs are established in

295   the preceding G1 phase. We set our ORC/MCM ChIP-seq signals in relation to RFD

296   and computed the relative read frequencies of ORC/MCM around all AS aggregate

297   borders. Both ORC and MCM were enriched within ASs compared to flanking

298   regions (Fig. 2b, Suppl. Fig. 5a without input division).

299   To resolve the impact of transcriptional activity, we repeated this calculation

300   and sorted for type 1 ASs (Fig. 2c, Suppl. Fig. 5b), type 2 ASs (Fig. 2d, Suppl. Fig.

301   5c), and non-genic ASs (Fig. 2e, Suppl. Fig. 5d). Transcriptional activity in AS

302   flanking regions was associated with increased ORC/MCM levels inside ASs

303   (comparing Fig. 2b and Fig. 2c) and a prominent depletion of MCMs from transcribed

304   regions (Fig. 2c and 2d). In contrast, in type 2 ASs, ORC/MCM levels remained

305   elevated at non-transcribed AS borders (Fig. 2d, left border). No evident ORC/MCM

306   enrichments were detected within non-genic ASs (Fig. 2e).

307   AS borders associated with transcriptional activity were locally enriched in

308   ORC/MCM (Fig. 2c and 2d (right border)). This is in line with previously detected

309   Orc1 accumulation at IZ borders (Petryk et al., 2016). Reciprocally, non-genic AS

310   borders showed a local dip in ORC/MCM levels (Fig. 2d (left border) and 2e), but the

311   biological significance of these observations is unclear. A sequence analysis revealed

312   biased distributions of homopolymeric repeat sequences at AS borders (data not

313   shown). Such sequences may affect nucleosome formation and ORC binding, but may

314   also bias Okazaki fragment detection, hence border detection, at small scales, which

315   explains the local RFD peaks observed at AS borders (Suppl. Fig. 4a).

316
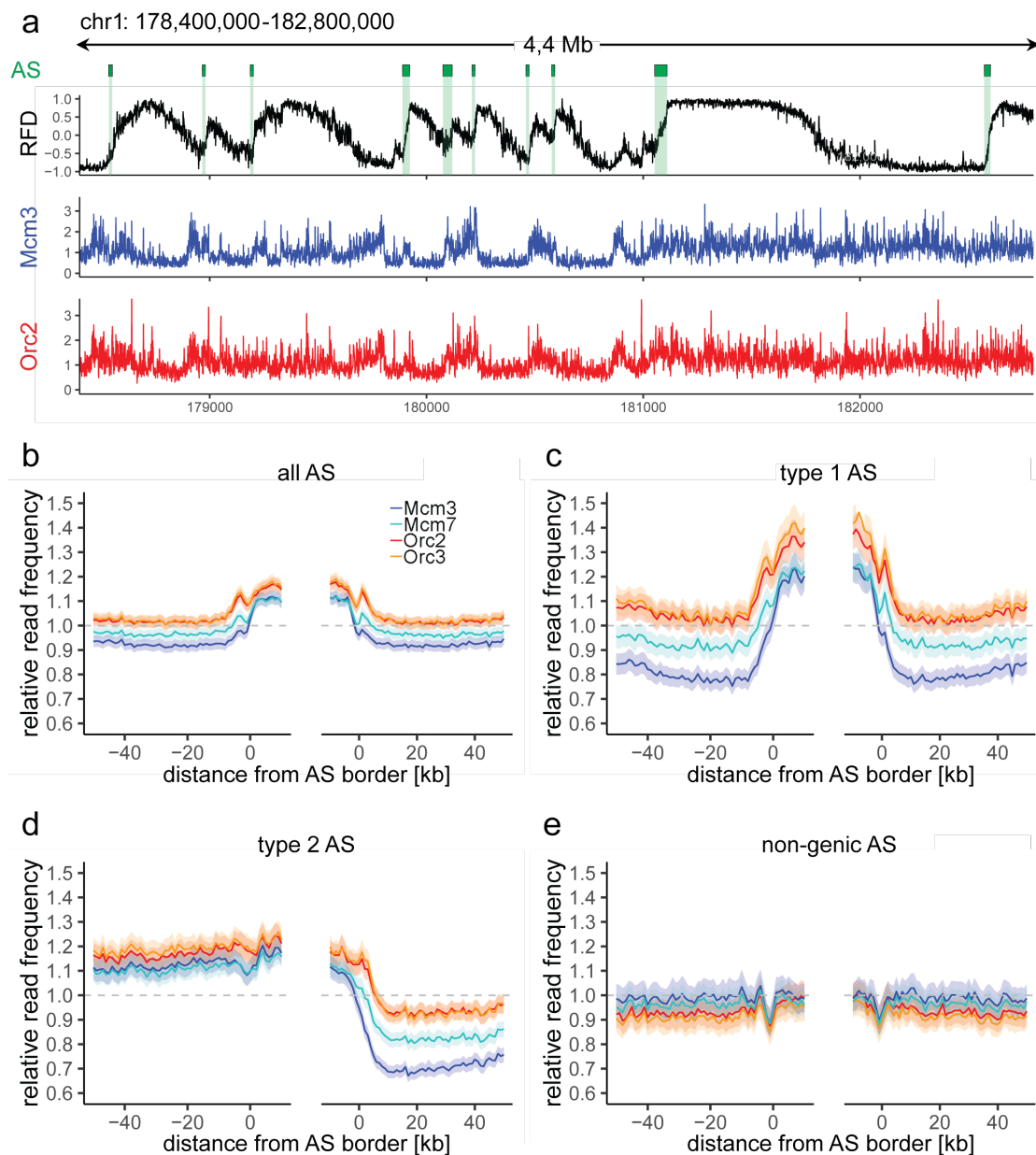
Figure 2: ORC/MCM enrichment within AS depends on active transcription. a) Top panel: Example of an RFD profile on chr1: 178,400,000 – 182,800,000, covering 4 Mb. Detected ASs are labeled by green rectangles. Middle and bottom panel: Representative Mcm3 (blue) and Orc2 (red) ChIP-seq profiles after binning for the same region. b-e) Average input-normalized relative ChIP read frequencies of Orc2, Orc3, Mcm3, and Mcm7 at AS borders of b) all AS (n = 2957), c) type 1 ASs with transcribed genes at both AS borders (n = 673), d) type 2 ASs oriented with their AS border associated to a transcribed genes at the right (n = 1026), and e) non-genic ASs in gene deprived regions (n = 506). The mean of ORC and MCM relative read frequencies are shown ± 2 x SEM (lighter shadows). The dashed grey horizontal line indicates relative read frequency 1.0 for reference.

329 ORC AND MCM ARE DEPLETED FROM TRANSCRIBED GENE BODIES AND ENRICHED

330 AT TSSs

331 Replication initiation and termination often correlate with active gene transcription

332 (Besnard et al., 2012; Cayrou et al., 2015; Cayrou et al., 2011; Picard et al., 2014).

333 Consistent with previous OK-seq studies (Chen et al., 2019; Petryk et al., 2016), we

334 find on average a strong ascending RFD signal upstream of TSSs and downstream of

335 TTSs of active genes, and a negative RFD slope across the active gene bodies (Suppl.

336 Fig. 6a). This behavior depends on transcriptional activity, as silent genes display an

337 overall flat RFD profile (Suppl. Fig. 6a). The ORC/MCM enrichment in type 1 and 2

338 ASs compared to flanking genic regions (Fig. 2c and d) argues for a major

339 contribution of active transcription to ORC/MCM positioning. To study this, we set

340 our ChIP-seq data in relation to transcription profiles obtained from asynchronously

341 cycling Raji cells. We analyzed ORC/MCM relative read frequencies around active

342 TSSs and transcriptional termination sites (TTSs) (Fig. 3, Suppl. Fig. 6b-e without

343 input normalization). ORC relative read distribution of G1-phased cells was

344 significantly enriched at active TSSs as already demonstrated in *Drosophila*

345 (MacAlpine et al., 2010) and human cells (Dellino et al., 2013; Miotto et al., 2016).

346 ORC levels were located moderately but significantly higher upstream of TSSs and

347 downstream of TTSs than within active genes (Fig. 3a, Suppl. Fig. 6b). The depletion

348 of ORC from gene bodies was statistically significant for approximately 45% of

349 actively transcribed genes (Suppl. Table 1). Compared to ORC, Mcm3 and Mcm7

350 enrichments at TSSs were less prominent, but depletion from gene bodies was more

351 pronounced (Fig. 3a, Suppl. Fig. 6b), with 75% and 58% of investigated transcribed

352 gene bodies significantly depleted from Mcm3 and Mcm7, respectively (Suppl. Table

353 1).

17
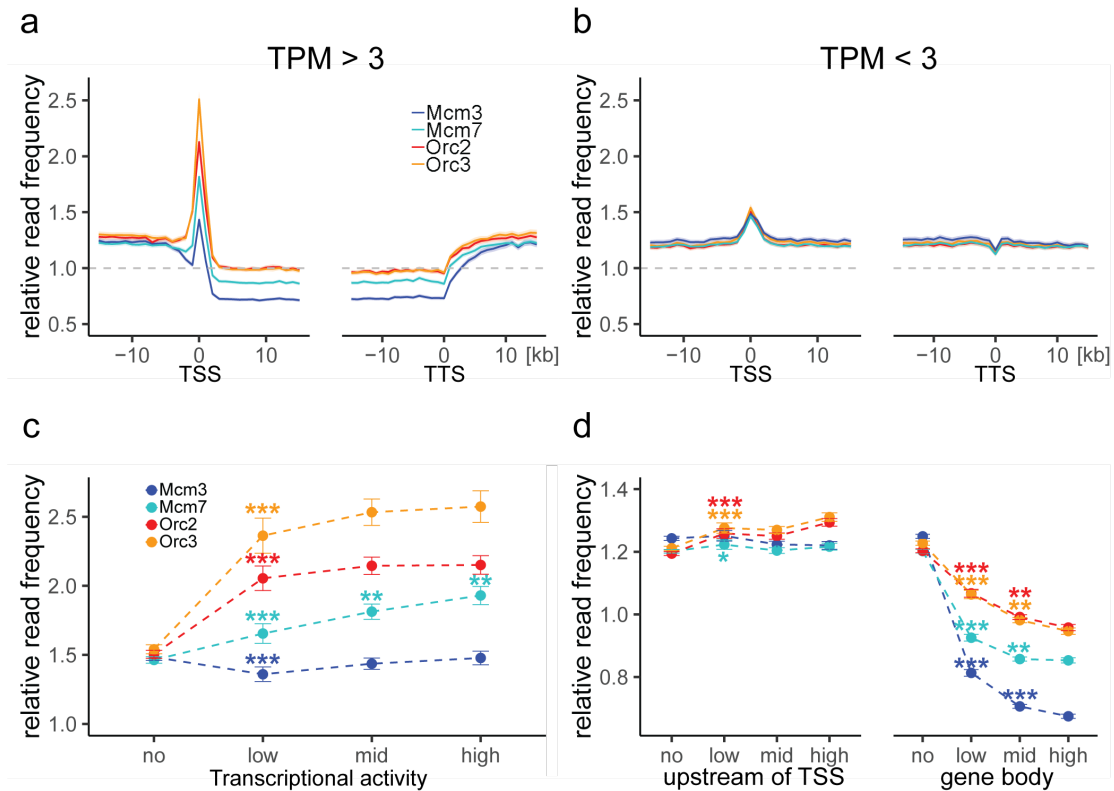
Figure 3: ORC is enriched at active TSSs while MCM is depleted from actively transcribed genes. a) - b) ORC/MCM relative read frequencies around TSSs or TTSs for a) active genes (TPM > 3) and b) inactive genes (TPM < 3). Only genes larger than 30 kb without any adjacent gene within 15 kb were considered. Distances from TSSs or TTSs are indicated in kb. Means of ORC and MCM frequencies are shown ± 2 x SEM (lighter shadows). The dashed grey horizontal line indicates relative read frequency 1.0 for reference. c) ORC/MCM relative read frequencies at TSSs dependent on transcriptional activity (± 2 x SEM). d) ORC/MCM relative read frequencies upstream of TSSs and within the gene body dependent on transcriptional activity (± 2 x SEM; TSSs ± 3 kb removed from analysis). Transcriptional activity was classified as: no (TPM < 3), low (TPM 3-10), mid (TPM 10-40), high (TPM > 40). Statistics were performed by one-way ANOVA followed by Tukey's post-hoc test. P-values are indicated always comparing to the previous transcriptional level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Depletion was strictly homogeneous from TSS to TTS, strongly suggesting that transcription itself displaces ORC and MCM-DH complexes (Fig. 3a). In contrast, at silent genes, ORC/MCM were hardly enriched at TSSs and were not depleted from gene bodies (Fig. 3b, Suppl. Fig. 6c without input-correction). This confirms that active transcription is required for TSS enrichment and gene body depletion of ORC/MCM. Nevertheless, we also observed that increasing transcriptional activity

18

374   (classified as follows: low: 3-10 TPM, mid: 10-40 TPM, high: > 40 TPM) did not

375   have any major impact on ORC/MCM enrichments at TSSs (Fig. 3c, Suppl. Fig. 6d

376   without input-correction). ORC/MCM depletion within gene bodies slightly increased

377   with transcription levels (Fig. 3d), but this was less convincing in the non-input-

378   normalized data (Suppl. Fig. 6e). Basal ORC/MCM levels upstream of TSSs and

379   downstream of TTSs were identical, indicating that the local ORC enrichment at TSSs

380   does not result in more MCM loading upstream than downstream of active genes.

381   Pre-replicative chromatin represents a cell cycle stage immediately prior to

382   replication initiation, with an excess of MCM-DH loaded onto chromatin (Powell et

383   al., 2015; Takahashi et al., 2005). Our observation that Mcm3 and Mcm7 are

384   significantly depleted from gene bodies is consistent with their active displacement by

385   transcription, as previously proposed in *Drosophila* (Powell et al., 2015).

386

387   ORC/MCM GENOMIC DISTRIBUTION IS BROAD AND CORRELATED WITH

388   REPLICATION TIMING BUT NOT INITIATION ZONES

389   In the previous sections, we observed a striking influence of transcriptional

390   activity on ORC/MCM distribution that contributes to circumscribing IZs. However,

391   transcriptionally active regions only account for a small subset of the genome.

392   Instead, at non-genic AS, a rather homogenous ORC/MCM pattern within and outside

393   AS is visible (Fig. 2e, Suppl. Fig. 5d without input-correction).

394   Replication timing (RT) is a crucial aspect of genome stability that has been

395   strongly correlated with gene expression and chromatin structure. RT is determined

396   by the timing of origin firing. In yeast, it has been proposed that the number of MCM-

397   DHs loaded at origins correlates with RT, suggesting how RT profiles can emerge

398   from stochastic origin firing (Das et al., 2015; Yang et al., 2010). In yeast, however,

19

399   MCM-DHs are only detectable at origins, whereas in human cells ORC/MCM are

400   broadly distributed within and outside IZs so that the presence of ORC/MCM does

401   not solely define RT.

402   To clarify the relationships between IZ location, IZ firing time and

403   ORC/MCM density in human cells, we used Raji Early/Late Repli-seq data from

404   Sima *et al.* (Sima et al., 2019) and related RT to ORC/MCM relative read frequencies

405   and RFD slope. Thus, we analyzed i) ascending RFD segments (ASs), corresponding

406   to IZs, ii) descending RFD segments (DSs), determined symmetrically to ASs and

407   representing predominant replication termination, iii) unidirectional replicating

408   regions (URRs; segments were $|RFD| > 0.8$ over > 300 kb), and iv) regions of null

409   RFD i.e. bidirectional replication (NRRs; segments were $|RFD| < 0.15$ over > 500 kb).

410   Example representations of these four categories are depicted in Suppl. Fig. 7a and

411   7b. In Fig. 4 we calculated relative Orc2 and Mcm3 (see Suppl. Fig. 7 d-e for Orc3

412   and Mcm7) read frequencies in 10 kb bins against RT in intergenic regions (left

413   column), silent (TPM < 3) genes (middle column), and active (TPM > 3) genes (right

414   column), while considering either all bins (top row) or bins corresponding to ASs,

415   DSs, URRs, and NRRs (following rows in descending order). ChIP frequencies are

416   normalized by column, i.e. each column is the probability density function of ChIP

417   frequency at a given RT bin. Intergenic regions (left column of each panel) show an

418   ORC/MCM distribution mostly independent from RFD slope, but depending on RT,

419   with higher ORC and MCM levels in early RTDs.
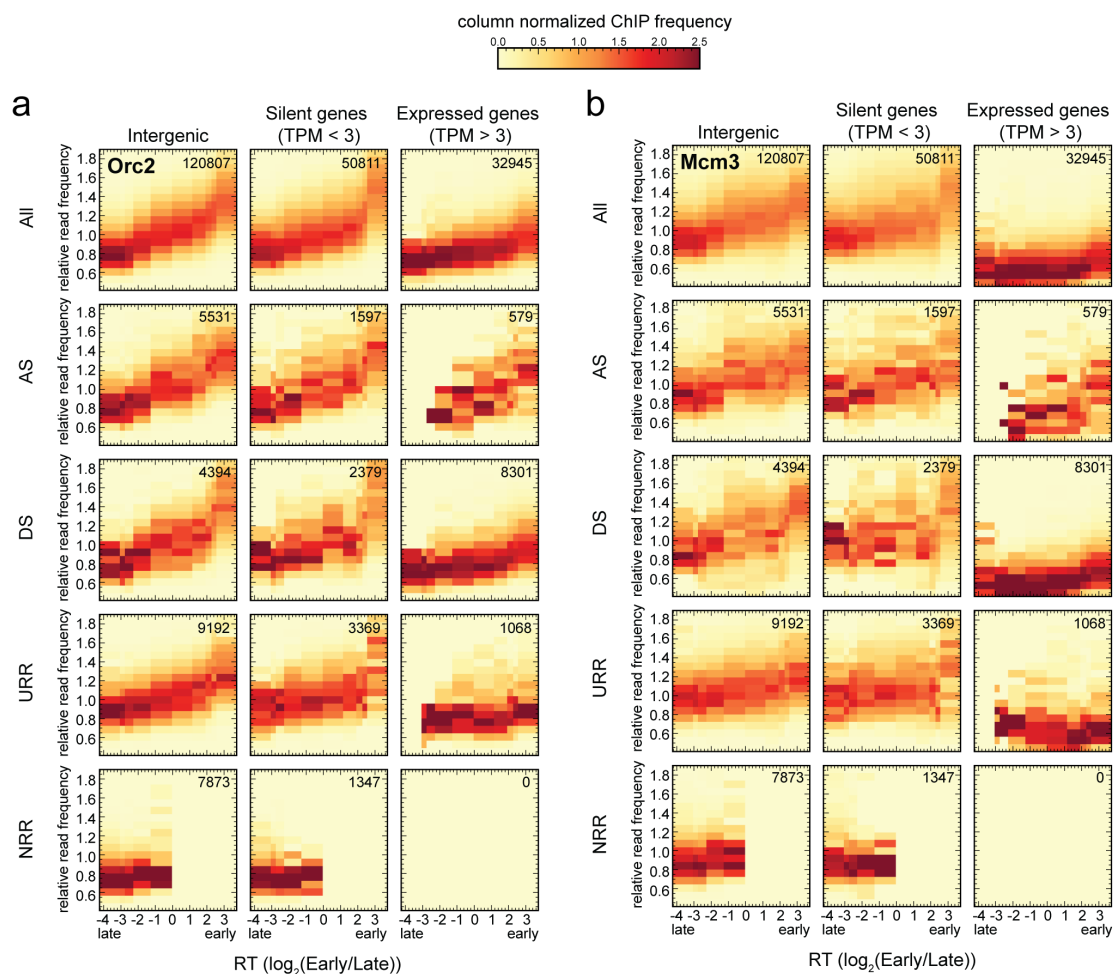
Figure 4: ORC/MCM levels correlate with RT and transcriptional activity but are otherwise homogeneously distributed along the genome and uncorrelated to RFD patterns (see also Suppl. Fig. 7). a-b) 3x5 panel of 2D histograms representing Orc2 (a) and Mcm3 (b) ChIP relative read frequency vs. RT (average log2(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis was performed in 10 kb windows. ChIP relative read frequencies are normalized by column and represent the probability density function of ChIP frequency at a given replication timing. The color legend is indicated on top. The columns of each panel represent only windows present in intergenic regions (left column), silent genes (TPM < 3, middle column), and expressed genes (TPM > 3, right column). TSSs and TTSs proximal regions were not considered (see Material and Methods). The rows show either all bins (top row), AS bins (predominant replication initiation, second row), DS bins (descending segment, predominant replication termination, third row), URR bins (unidirectional replication, no initiation, no termination, fourth row) and NRR bins (null RFD regions, spatially random initiation and termination, bottom row). The number of bins per panel is indicated in each panel. See Suppl. Fig. 7 for equivalent Orc3 and Mcm7 data. Refer to Suppl. Fig. 8a for statistical significances.

Silent genes (middle column) mirror the ORC/MCM pattern of intergenic regions.

Expressed genes (right column) show lower ORC/MCM frequencies than intergenic

439   regions and silent genes, as expected. Kolmogorov-Smirnov statistics (Suppl. Fig. 8a)

440   quantitatively show that in early- and mid-replicating regions, ORC/MCM relative

441   frequency distributions are substantially different between expressed genes and

442   intergenic regions as well as silent genes. In contrast, intergenic distributions in ASs

443   and DSs are not different in most RT bins. In expressed genes, the ChIP depletion is

444   more pronounced for MCM than ORC, as described in Fig. 3a. The dependence of the

445   residual signal on RT is much attenuated, particularly for MCM, as expected if

446   transcription completely removes this complex from active gene bodies.

447   In general, the global analysis of all windows, independent from RFD, demonstrates a

448   genome-wide, monotonous relationship between ORC/MCM binding and RT that is

449   only attenuated in transcribed genes. Furthermore, this analysis did not reveal any

450   convincing differences in the levels of ORC/MCM between non-transcribed ASs,

451   DSs, URRs, and NRRs when bins of similar RT were compared. The few (579) bins

452   corresponding to ASs in active genes are probably misleading, as they are mainly

453   attributable to annotation errors: the annotated gene overlapped the AS but the RNA-

454   seq signal was confined outside the AS (Petryk et al., 2016). Note that NRRs are

455   confined to late RT segments while URRs are enriched in mid-RT segments as

456   previously noted (Petryk et al., 2016) and confirmed in Suppl. Fig. 7c.

457   Strictly speaking, the slope of a RFD segment is proportional to the difference

458   between the density of initiation and termination events within the segment (Audit et

459   al., 2013). Therefore, we cannot exclude delocalized initiation events within DSs,

460   which explains why they were not significantly depleted in ORC/MCM compared to

461   ASs (Suppl. Fig. 8a). In contrast, we can mostly exclude initiation events within

462   URRs, although their ORC/MCM densities were not significantly lower than in ASs,

463 which are bona fide IZs, while NRRs, which presumably support spatially stochastic

464 initiation present a smaller ORC density (Suppl. Fig. 8a).

465   Taken together, these results suggest that the density of ORC/MCM is not a

466 reliable predictor of initiation probability, even though ORC density (and to a lesser

467 extent MCM density) is well correlated with RT. Our results suggest that potential

468 origins are widespread through the genome but that additional genetic or epigenetic

469 factors regulate whether and when they fire.

470

471   LATE REPLICATING NON-GENIC ASs and NRRs ARE CHARACTERIZED BY

472 H4K20ME3

473   The results above revealed a gradient of ORC/MCM according to RT. To

474 confirm this observation, we extracted early and late RTDs employing a threshold of

475 $\log_2$(Early/Late) > 1.6 for early RTDs and < -2.0 for late RTDs, which resulted in 302

476 early RTDs covering 642.8 Mb and 287 late RTDs covering 617.4 Mb of the genome.

477 Restricting the analysis to intergenic regions, we calculated the mean ORC/MCM

478 relative read frequencies in early compared to late RTDs. ORC was 1.4-times

479 enriched in early RTDs compared to late RTDs (Fig. 5a, Suppl. Fig. 8b without input-

480 correction, and Suppl. Table 2). Mcm3 and Mcm7 levels, although showing the same

481 tendencies, were less contrasted than ORC and more influenced by input

482 normalization (Fig. 5a, Suppl. Fig. 8b, Suppl. Table 2).

483
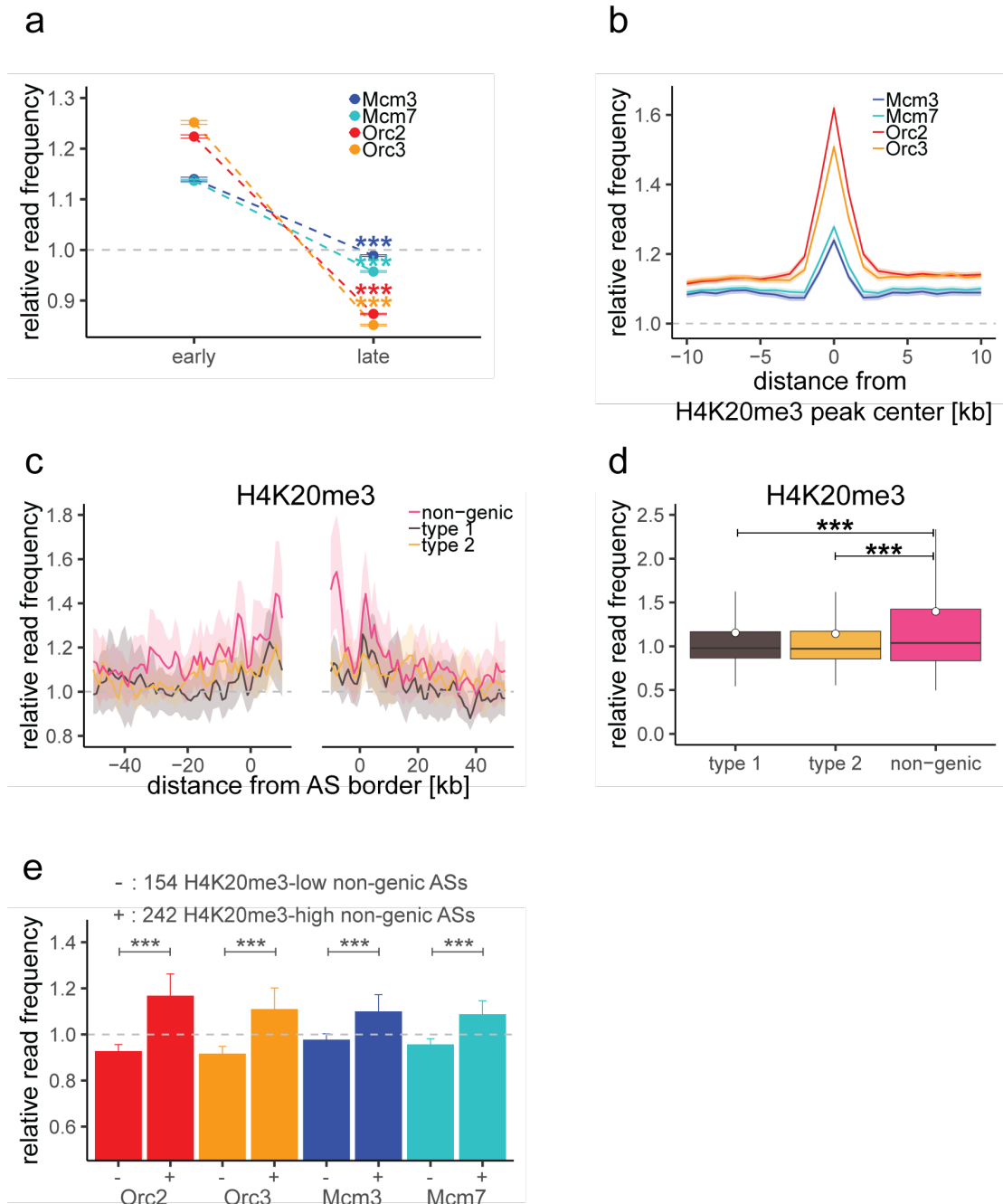
Figure 5: H4K20me3 selectively marks a subset of late replicating non-genic ASs.
a) ORC/MCM relative read frequencies (± 2 x SEM) in early or late RTDs. Early RTDs were defined as log2(Early/Late) > 1.6; late RTDs < -2.0. The analysis was performed in 10 kb bins. Any gene ± 10 kb was removed from the analysis. Statistics were performed using one-sided t-test. b) Average ORC/MCM relative read frequencies at H4K20me3 peaks (> 1 kb). c) H4K20me3 relative read frequencies at AS borders of the different AS types. Type 2 ASs are oriented with their AS borders associated to transcribed genes at the right. Means of H4K20me3 relative read frequencies are shown ± 2 x SEM (lighter shadows). d) Boxplot representation of H4K20me3 relative read frequencies within the different AS types. Boxplot represent the mean (circle), the median (thick line), the 1st and 3rd quartile (box), the 1st and 9th decile (whiskers) of the relative read frequencies in each AS type. Statistics were performed by one-way ANOVA followed by Tukey's post-hoc test. e) Histogram

496   representation of mean ± 2 x SEM of ORC/MCM relative read frequencies at 242

497   H4K20me3-low non-genic ASs and 154 H4K20me3-high non-genic ASs. Statistics were

498   performed using one-sided t-test. *** p < 0.001.

499   We recently demonstrated that H4K20me3 is involved in licensing a subset of

500   late replicating regions, often co-localizing with H3K9me3 (Benetti et al., 2007;

501   Brustel et al., 2017; Pannetier et al., 2008). Here, we looked further into the relation

502   between this epigenetic mark, ORC/MCM and replication initiation. We performed

503   ChIP for H4K20me3 and its precursor H4K20me1 in three replicates for pre-

504   replicative G1 chromatin, validated by qPCR (Suppl. Fig. 9a (H4K20me3) and 9b

505   (H4K20me1)). After sequencing, we performed MACS2 broad peak-detection and

506   kept only peaks overlapping in all three samples (16852 peaks for H4K20me3 and

507   12264 peaks for H4K20me1, see also Suppl. Table 3 for further characterization).

508   H4K20me3 peak sizes ranged from 200 bp to 105 kb (200 bp to 183 kb for

509   H4K20me1, Suppl. Table 3, Suppl. Fig. 9c). When calculating ORC/MCM coverage

510   at H4K20me3/me1 peaks > 1 kb (12251/6277 peaks, respectively), we predominantly

511   detected ORC, but also some MCM enrichment at H4K20me3 sites (Fig. 5b). By

512   contrast, H4K20me1 peaks were not enriched in both ORC and MCM (Suppl. Fig.

513   9d).

514   We calculated H4K20me3 coverage at the different AS types and specifically

515   detected an increased H4K20me3 signal in non-genic ASs, representing the first

516   histone modification characterizing late replicating ASs (Fig. 5c and 5d). Starting

517   from 506 non-genic ASs, we extracted two subsets of 154 and 242 non-genic ASs

518   where H4K20me3 relative read frequencies were respectively above the mean

519   genome value by more than 1.5x standard deviation or below the genome mean value.

520   We found that ORC/MCM were enriched at the H4K20me3-high subgroup compared

521   to the H4K20me3-low subgroup (Fig. 5e). These results suggest that the presence of

H4K20me3 at transcriptionally independent, non-genic ASs may contribute to specifying these regions as Ma-IZs at the origin-licensing step.
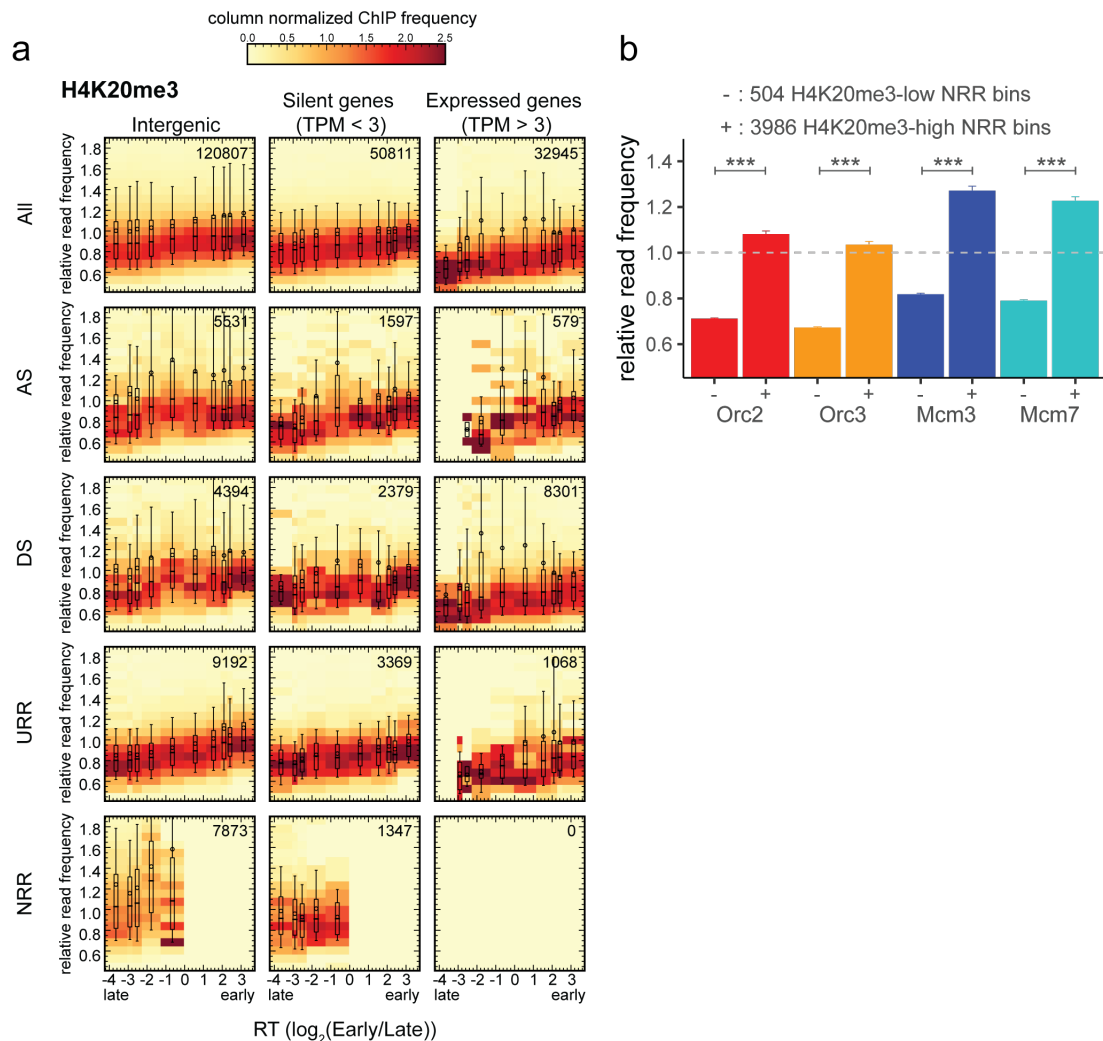
Figure 6: H4K20me3 is enriched in late replicating regions of null RFD (NRR).
a) 3x5 panel of 2D histograms representing H4K20me3 ChIP relative read frequencies vs. RT (average log2(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis was performed in 10 kb windows. H4K20me3 relative read frequencies are normalized by column and displayed for different window categories (columns: intergenic regions, silent genes, expressed genes; rows: all bins, AS-, DS-, URR-, NRR bins) as for ORC/MCM in Fig. 4. The number of bins per panel is indicated in each panel. Superimposed boxplots represent the mean (circle), the median (thick line), the 1st and 3rd quartile (box), the 1st and 9th decile (whiskers) of the relative read frequencies in each timing bins. Refer to Suppl. Fig. 8c for statistical significances. b) Histogram representation of mean ± 2 x SEM of ORC/MCM relative read frequencies at 504 H4K20me3-low NRR 10 kb bins and 3986 H4K20me3-high NRR 10 kb bins. Statistics were performed using one-sided t-test. *** p < 0.001.

539   To further explore the links between H4K20me3 and replication initiation

540 capability, we analyzed the density of this modification in genome segments of

541 various RT, gene activity and RFD patterns (Fig. 6a), as performed for ORC/MCM in

542 Fig. 4 and Suppl. Fig. 7d and 7e. Several interesting observations emerged from this

543 analysis. First, the H4K20me3 level was weakly but systematically more abundant in

544 early than in late replicating DNA. Second, the H4K20me3 level was slightly lower in

545 transcribed genes than in the non-transcribed rest of the genome (Fig. 6a and Suppl.

546 Fig 9e). Third, AS and DS bins showed comparable distributions of H4K20me3 levels

547 at comparable RT and gene expression status (Fig. 6a and Suppl. Fig 9e). These

548 results are, at first glance reminiscent of ORC/MCM distributions. However,

549 ORC/MCM distributions are more pronounced as the RT gradient is stronger and the

550 differences between expressed genes and non-transcribed intergenic regions and silent

551 genes are more salient (Fig. 4 and Fig. 6a). NRRs showed a very different, broader

552 distribution of H4K20me3 levels, including a higher proportion of windows with a

553 strong H4K20me3 signal, especially compared to URRs (compare boxplots in Fig. 6a,

554 Fig. 6b). A specific enrichment of H4K20me3 is therefore detected not only in late,

555 non-genic AS segments but also in late-replicating gene deserts of null RFD, which

556 presumably replicate by widespread and spatially random initiation.

557   The finding that the H4K20me3 levels are more broadly distributed in NRRs

558 suggests that ORC/MCM levels might also differ between H4K20me3-high and -low

559 NRRs. We repeated the analysis of ORC/MCM enrichment at H4K20me3-high and -

560 low 10 kb intergenic bins in NRRs. Similar to non-genic ASs, we also observed that

561 ORC/MCM is more abundant at H4K20me3-high than -low bins (Fig. 5e, and 6b),

562 supporting the hypothesis that this modification also facilitates origin licensing in

563 these heterochromatic segments.

564   **Discussion**

565   The study presented here provides a novel comprehensive genome-wide

566   analysis of multiple pre-RC proteins, replication initiation, and replication timing in

567   human cells. We find a widespread presence of ORC/MCM throughout the genome,

568   with variations that only depend on RT or active transcription. ORC/MCM are

569   depleted from transcribed genes and enriched at TSSs. ORC/MCM are more abundant

570   in early than in late RTDs. The even distribution of ORC/MCM observed within IZs

571   is consistent with OK-seq results suggesting a homogeneous initiation probability

572   from any site within any given IZ. However, when RT and transcriptional effects are

573   controlled, no significant differences in ORC/MCM densities are detected between

574   IZs, TZs, URRs and NRRs, which sustain different patterns of origin activity. We

575   consequently propose that potential origins, defined by loaded MCM-DHs, are

576   widespread through the genome, but their activation in S phase is regulated by

577   additional genetic and/or epigenetic factors. We previously reported that IZs are

578   enriched in open chromatin marks typical of active or poised enhancers (Petryk et al.,

579   2016), which suggests why IZs are more accessible to firing factors than flanking

580   segments with comparable MCM-DH density. We further show that a subset of non-

581   genic late IZs, and late, randomly replicating gene deserts, are enriched in

582   H4K20me3, which helps to recruit ORC/MCM.

583   Our data suggest that transcription has both positive and negative effects on

584   origin activity. We found that actively transcribed gene bodies flanking type 1 and

585   type 2 ASs are depleted of ORC/MCM (Fig. 2, and Fig. 3a). We propose that active

586   transcription removes ORC/MCM from transcribed gene bodies, which negatively

587   affects their replication initiation capacity. MCM-DH depletion from transcribed gene

588   bodies was previously reported in *Drosophila* (Powell et al., 2015). Experiments in

28

589   yeast have shown that RNA polymerases push MCM-DHs along the DNA and

590   redistribute them to shift replication initiation sites (Gros et al., 2015). A number of

591   previous studies have suggested that replication does not initiate within transcribed

592   genes (Hamlin et al., 2010; Hyrien et al., 1995; Knott et al., 2009; Macheret &

593   Halazonetis, 2018; Martin et al., 2011; Sasaki et al., 2006). Oncogene expression, by

594   abridging G1 phase, allows the ectopic activation of intragenic origins normally

595   suppressed by transcription in G1 resulting in genomic instability (Macheret &

596   Halazonetis, 2018). All these findings are consistent with an inhibitory effect of

597   transcription on local replication initiation that is important for genome stability. Why

598   forks emitted by ectopic, intragenic origins are more prone to genomic destabilisation

599   than forks emitted outside, but progressing into genes, is unclear at present.

600   On the other hand, ORC and to a lesser degree MCM, are enriched at active

601   TSSs. Active TSSs are regions of open chromatin structure characterized by high

602   DNase- or MNase hypersensitivity, a hallmark of Ma-IZs (Boulos et al., 2015; Papior

603   et al., 2012). ORC chromatin binding is known to favor open chromatin sites (Miotto

604   et al., 2016), situating active TSSs as hotspots of ORC binding. The slope of

605   aggregate RFD profiles around meta-TSSs is higher at TSSs than in upstream IZs,

606   suggesting a higher probability of initiation at TSSs (Chen et al., 2019). However, a

607   part of this effect may result from averaging multiple IZs that all end at a TSS but

608   initiate replication at different upstream distances from the respective TSS. When

609   individual IZs are examined, the RFD slope is not obviously increasing at the TSS

610   (data not shown). Furthermore, the most efficient initiation sites identified by

611   EdUseq-HU within IZs are associated with poly(dA:dT) tracts but are not enriched at

612   TSSs (Tubbs et al., 2018). We find here that MCMs, which mark potential initiation

613   sites, are less enriched at TSSs than ORC. Furthermore, MCMs are distributed fairly

614   evenly upstream and downstream of transcribed gene bodies (Fig. 3a), arguing against

615   a role for TSSs in directing MCM-DH loading specifically upstream of genes.

616   H4K20 methylation has multiple functions in ensuring genome integrity, such

617   as DNA replication (Beck et al., 2012; Picard et al., 2014; Tardat et al., 2010), DNA

618   damage repair, and chromatin compaction (Jorgensen et al., 2013; Nakamura et al.,

619   2019; Shoaib et al., 2018), suggesting that the different functions are context-

620   dependent and executed with different players. We previously demonstrated that

621   H4K20me3 provides a platform to enhance licensing in late replicating

622   heterochromatin (Brustel et al., 2017). Here, we detect both elevated ORC and MCM,

623   when selecting for H4K20me3-enriched, non-genic ASs and NRRs (Fig 5e and 6b).

624   Together with our previous observations (Brustel et al., 2017), we conclude that

625   H4K20me3 is important for the licensing of at least a subset of late replicating origins

626   in non-genic ASs and NRRs. It remains unclear whether additional chromatin features

627   are required for licensing the remaining, H4K20me3-low non-genic ASs.

628   In higher eukaryotes, it is controversially discussed whether replication timing

629   is determined by ORC (Dellino et al., 2013; Miotto et al., 2016) or MCM-DH (Das et

630   al., 2015; Hyrien, 2016) abundance. However, potential origins are defined by

631   assembled MCM-DHs, not by ORC. The weak correlation of MCM density with RT,

632   and the lack of correlation with initiation capability of ASs, DSs and URRs, appear

633   inconsistent with regulation of RT by MCM-DH density (Fig. 4b, Suppl. Fig. 7e, and

634   Fig. 5a). We conclude that MCM-DH density therefore is not a reliable predictor of

635   RT or RFD profiles and is unlikely to be a major determinant of RT itself. The

636   correlation of ORC density with RT is more convincing but not necessarily causative.

637   For example, open chromatin independently facilitates ORC binding in G1 phase and

638   access of firing factors to MCM-DHs in S phase, resulting in correlation of

639   ORC/MCM density with RT without implying any causal link.

640   The spatio-temporal replication program is relatively well conserved in

641   consecutive replication cycles for each cell type, differs only slightly between cell

642   lines and changes during differentiation (Hadjadj et al., 2016). Comparison with

643   chromatin conformation capture (Hi-C) data have shown that early and late RTDs

644   correspond to the more and less accessible compartments of the genome, respectively

645   (Ryba et al., 2010). A higher density of chromatin interactions characterizes less

646   accessible domains. Furthermore, both constitutive and developmentally regulated

647   RTD boundaries align to the boundaries of topological domains, which are

648   remarkably stable between cell types (Fragkos et al., 2015; Pope et al., 2014), and

649   enriched in IZs (Baker et al., 2012; Petryk et al., 2016). Recently, Sima et al. used the

650   CRISPR-Cas9 technology to identify three separate, cis-acting elements that together

651   control the early replication time of the pluripotency-associated Dppa2/4 domain in

652   mouse embryonic stem cells (mESCs) (Sima et al., 2019). Strikingly, these "early

653   replication control elements" (ERCEs) are enriched in CTCF-independent Hi-C

654   interactions and in active epigenetic marks (DNase1 HS, p300, H3K27ac, H3K4me1,

655   H3K4me3) previously observed at OK-seq IZs (Petryk et al., 2018; Petryk et al.,

656   2016). By mining mESC OK-seq data (Petryk et al., 2018), we found that the three

657   ERCEs of the Dppa2/4 domain indeed fall within IZs (Suppl. Fig. 10a). Furthermore,

658   the aggregate 1,835 ERCEs predicted genome-wide by Sima et al., from these and

659   additional epigenetic properties of mESCs, show a significant, positive RFD shift

660   indicative of efficient replication initiation (Suppl. Fig. 10b). This finding is

661   confirmed in proliferating PHA-stimulated primary splenic B cells, (Suppl. Fig. 10c),

662   attesting to the general validity of these observations.

663   Our data suggest that a higher ORC/MCM density *is not* a distinguishing

664   feature of IZs from the rest of the non-transcribed genome. IZ specification therefore

665   appears to occur at the origin activation rather than licensing step, which may be

666   explained if the open chromatin structure found at Ma-IZs and ERCEs (Petryk et al.,

667   2016; Sima et al., 2019) facilitates preferential accessibility to limiting firing factors

668   during S phase (Boos & Ferreira, 2019). This is in line with the previously proposed

669   cascade model in which replication of the human genome involves a superposition of

670   efficient initiation at Ma-IZs identified by RFD ASs, followed by a cascade of more

671   dispersive, less efficient origin activation along the intervening segments (Petryk et

672   al., 2016; Wu et al., 2018).

673

674   **CONCLUSION**

675   The mapping of ORC and MCM complexes reported here shows that in

676   human cells, most of the genome, except transcribed genes, is licensed for replication

677   during the G1 phase of the cell cycle. ORC/MCM are thereby more enriched in early

678   than in late RTDs (Fig. 7a). Only a fraction of MCM-DHs is selected for initiation

679   during S phase. Open chromatin marks define preferential IZs, often but not always

680   circumscribed by active genes (Fig. 7b). Once forks emanate from Ma-IZs within an

681   RTD, a cascade of replication activation may take place dispersedly between IZs due

682   to the omnipresence of MCM-DHs (Fig. 7b). The preferential binding of ORC in G1,

683   and preferential access of firing factors in S to open chromatin, appears sufficient to

684   explain why ORC/MCM levels correlate with RT. Replication licensing of

685   heterochromatic gene deserts is, for example facilitated by H4K20me3, which helps

686   ORC and MCM-DH recruitment and supports origin activity in these less accessible

687   chromatin segments (Fig. 7c). The identification of ERCEs supports the hypothesis

688   that combinations of additional chromatin and DNA features regulate the probability

689   of origin activation. This provides organizational links between active transcription

690   and replication initiation, operating during origin licensing and activation, which

691   facilitate the timely activation of appropriate replication origins for genome stability

692   during programmed development as well as altered gene expression patterns caused

693   by environmental cues.
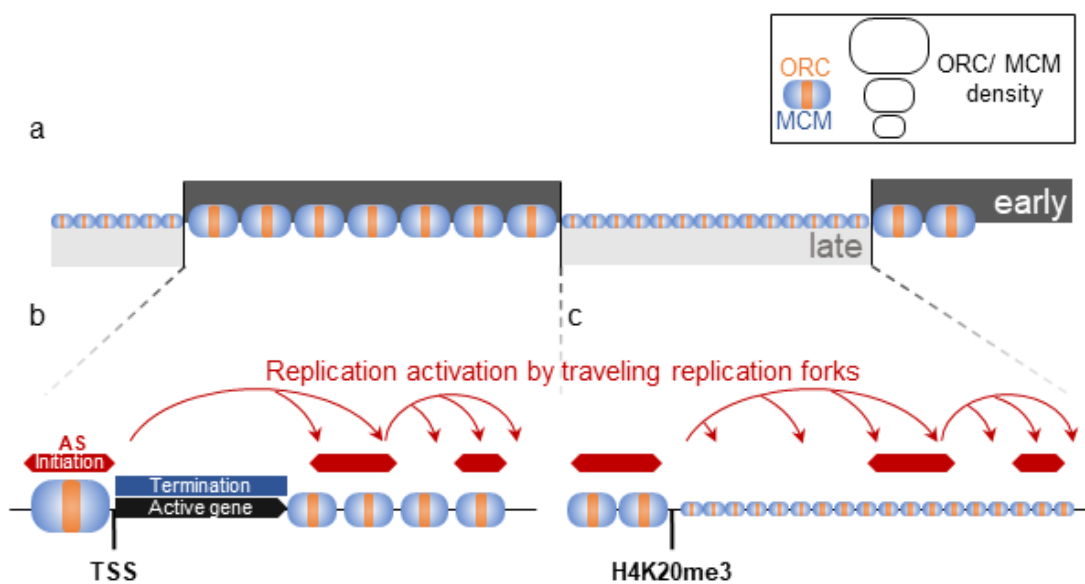
694

695   Figure 7: Model for replication organization in higher eukaryotes.
696   a) Replication is organized in large segments of constant replication timing (early RTD, dark
697   grey; late RTD, light grey), (Marchal et al., 2019). While we observe a homogeneous pattern
698   of ORC (orange) and MCM (blue) throughout the genome, the enrichment levels of
699   ORC/MCM were higher in early RTDs compared to late RTDs. b) Early RTDs are amongst
700   other characterized by active transcription. ORC/MCM are locally highly enriched at active
701   TSS. However, actively transcribed gene bodies (black) are deprived of ORC/MCM, often
702   correlating with replication termination (blue). Besides TSSs, we find ORC/MCM
703   stochastically distributed along intergenic regions. We hypothesize that traveling replication
704   forks trigger activation of replication in a cascade (red arrows). c) In gene deprived and
705   transcriptionally silent late replicating heterochromatin, we detected homogeneous
706   ORC/MCM distribution at generally lower levels. H4K20me3 is present at late replicating
707   non-genic ASs and NRRs and leads to enhanced ORC/MCM binding, linking this histone
708   mark to replication activation in heterochromatin.
709

Material and Methods

Cell culture

710   Raji cells (ATCC) were cultivated at 37°C and 5% CO2 in RPMI 1640
713 (Gibco, Thermo Fisher, USA) supplemented with 8% FCS (Lot BS225160.5,
714 Bio&SELL, Germany), 100 Units/ml Penicillin, 100 µg/ml Streptomycin (Gibco,
715 Thermo Fisher, USA), 1x MEM non-essential amino acids (Gibco, Thermo Fisher,
716 USA), 2 mM L-Glutamin (Gibco, Thermo Fisher, USA), and 1 mM Sodium pyruvate
717 (Gibco, Thermo Fisher, USA).

718

RNA extraction, sequencing, TPM calculation

720   RNA was extracted from 3×10^5 Raji cells using Direct-zolTM RNA
721 MiniPrep kit (Zymo Research) according to manufacturers' instructions. RNA quality
722 was confirmed by Bioanalyzer RNA integrity numbers between 9.8 and 10 followed
723 by library preparation (Encore Complete RNA-Seq Library Systems kit (NuGEN)).
724 Single-end 100 bp sequencing was performed by Illumina HiSeq 1500 to a
725 sequencing depth of 25 million reads. The reads were mapped to hg19 genome using
726 Tophat2 and assigned to annotated genes (HTSeq-count) (Anders et al., 2015; Kim et
727 al., 2013). TPM values (Transcripts per kilobase per million reads) were calculated
728 for each sample ($TPM_j = 10^6 \frac{n_j}{l_j} / \sum_i \frac{n_i}{l_i}$ where n_i is the number of reads that map to
729 gene i whose total exon length expressed in kb is l_i) as previously described (Wagner
730 et al., 2012).

731

Replication fork directionality profiling using OK-seq method in Raji

733   Raji OK-seq was recently published as part of Wu *et al*. and is available from
734 the European Nucleotide Archive under accession number PRJEB25180 (see data

735   access section) (Wu et al., 2018). Reads > 10 nt were aligned to the human reference

736   genome (hg19) using the BWA (version 0.7.4) software with default parameters (Li &

737   Durbin, 2009). We considered uniquely mapped reads only and counted identical

738   alignments (same site and strand) as one to remove PCR duplicate reads. Five

739   biological replicates were sequenced providing a total number of 193.1 million

740   filtered reads (between 19.1 and 114.1 million reads per replicate). RFD was

741   computed as $RFD = \frac{(R-F)}{(R+F)}$, where "R" (resp. "F") is the number of reads mapped to

742   the reverse (resp. forward) strand of the considered regions. RFD profiles from

743   biological replicates were highly correlated, with Pearson correlation computed in 50

744   kb non-overlapping windows with > 100 mapped reads (R+F) ranging from 0.962 to

745   0.993. Reads from the 5 replicate experiments were pooled together for further

746   analyses.

747

748   **Determining regions of ascending, descending and constant RFD**

749   RFD profiling of 2 human cell lines revealed that replication primarily

750   initiates stochastically within broad (up to 150 kb) zones and terminates dispersedly

751   between them (Petryk et al., 2016). These initiation zones correspond to quasi-linear

752   ascending segments (ASs) of varying size and slope within the RFD profiles. As

753   previously described for mean replication timing profiles analysis (Audit et al., 2013;

754   Baker et al., 2012), we determined the smoothed RFD profile convexity from the

755   convolution with the second derivative of the Gaussian function of standard deviation

756   32 kb. 4891 ASs were delineated as the regions between positive and negative

757   convexity extrema of large amplitude. The amplitude threshold was set in a

758   conservative manner in order to mainly detect the most prominent IZs described by

759   Petryk et al. (2016) and avoid false positives. Descending segments (DSs) were

760　　detected symmetrically to ASs, as regions between negative and positive convexity

761　　extrema using the same threshold. Noting *pos_5'* and *pos_3'* the location of the start

762　　and end position of an AS or DS segment, each segment was associated to its size

763　　*pos_3'-pos_5'* and the RFD shift across its length: $\Delta RFD = RFD\,(pos\_5') - RFD$

764　　(pos_3'). DS segments were less numerous (2477 vs 4891) and on average larger (126

765　　kb vs 38.8 kb) than AS segments as expected and presented a smaller average RFD

766　　shift ($|\Delta RFD|$=0.69 vs 0.83).

767　　Initial RFD profiling in human also revealed regions of unidirectional fork

768　　progression and regions of null RFD where replication is bidirectional. Unidirectional

769　　replicating regions (URRs) were delineated as regions where $|\Delta RFD|$>0.8

770　　homogeneously over at least 300 kb (401 regions of mean length 442 kb covering 177

771　　Mb). Null RFD regions (NRRs) were delineated as regions where $|\Delta RFD|$<0.15

772　　homogeneously over at least 500 kb (127 regions of mean length 862 kb covering 110

773　　Mb). Thresholds were set in a conservative manner to avoid false positive, in

774　　particular not to confuse RFD zero-crossing segments with NRR.

775

776　　**Centrifugal elutriation and flow cytometry**

777　　　　For centrifugal elutriation, 5×10^{9} exponentially growing Raji cells were

778　　harvested, washed with PBS and resuspended in 50 ml RPMI 1680, 8% FCS, 1mM

779　　EDTA, 0.25 U/ml DNaseI (Roche, Germany). Concentrated cell suspension was

780　　passed through 40 μm cell strainer and injected in a Beckman JE-5.0 rotor with a

781　　large separation chamber turning at 1500 rpm and a flow rate of 30 ml/min controlled

782　　by a Cole-Parmer Masterflex pump. While rotor speed was kept constant, 400 ml

783　　fractions were collected at increasing flow rates (40, 45, 50, 60, 80 ml/min).

784　　Individual fractions were quantified, 5×10^{6} cells washed in PBS, ethanol fixed,

36

785 RNase treated and stained with 0.5 mg/ml Propidium Iodide. DNA content was

786 measured using the FL2 channel of FACSCalibur^{TM} (BD Biosciences, Germany).

787 Remaining cells were subjected to chromatin cross-linking.

788

789 **Chromatin cross-linking with formaldehyde**

790   Raji cells were washed twice with PBS, resuspended in PBS to a concentration

791 of 2 x 10^7 cells/ml and passed through 100 μm cell strainer (Corning Inc., USA).

792 Fixation for 5 min at room temperature was performed by adding an equal volume of

793 PBS 2% methanol-free formaldehyde (Thermo Scientific, USA, final concentration:

794 1% formaldehyde) and stopped by the addition of glycine (125 mM final

795 concentration). After washing once with PBS and once with PBS 0.5% NP-40, cells

796 were resuspended in PBS containing 10% glycerol, pelleted and snap frozen in liquid

797 nitrogen.

798

799 **Cyclin Western Blot**

800   Cross-linked samples were thawed on ice, resuspended in LB3+ sonication

801 buffer containing protease inhibitor and 10 mM MG132. After sonicating 3 x 5 min

802 (30 sec on, 30 sec off) using Bioruptor in presence of 212-300 μm glass beads,

803 samples were treated with 50 U Benzonase for 15 min at room temperature and

804 centrifuged 15 min at maximum speed. 50 μg protein lysates were loaded on 10%

805 SDS-polyacrylamid gel (Cyclin A1/A2, Cyclin B1), or 12.5%-15% gradient gel

806 (H3S10P). Cyclin A1/A2 (Abcam, ab185619), Cyclin B1 (Abcam, ab72), H3S10P

807 (Cell signaling, D2C8) antibodies were used in 1:1000 dilutions, GAPDH (clone

808 GAPDH3 10F4, rat IgG2c; Monoclonal Antibody Core Facility, Helmholtz Center

809   München) was diluted 1:50. HRP-coupled secondary antibodies were used in 1:10000

810   dilutions. Detection was done using ECL on CEA Blue Sensitive X-ray films.

811

812

813   **Chromatin sonication**

814   Cross-linked cell pellets were thawed on ice, resuspended LB3+ buffer (25

815   mM HEPES (pH 7.5), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% Sarkosyl,

816   0.1% DOC, 0.5% Triton-X-100, 1X protease inhibitor complete (Roche, Germany))

817   to a final concentration of 2×10^7 cells/ml. Sonication was performed in AFA Fiber &

818   Cap tubes (12x12 mm, Covaris, Great Britain) at an average temperature of 5°C at

819   100W, 150 cycles/burst, 10% duty cycle, 20 min using the Covaris S220 (Covaris

820   Inc., UK) resulting in DNA fragments of 100-300bp on average.

821

822   **Chromatin immunoprecipitation and qPCR quality control**

823   Sheared chromatin was pre-cleared with 50 µl protein A Sepharose 4 Fast

824   Flow beads (GE Healthcare, Germany) per 500 µg chromatin for 2h. 500 µg

825   chromatin (or 250 µg for histone methylation) were incubated with rabbit anti-Orc2,

826   anti-Orc3, anti-Mcm3, anti-Mcm7 (Papior et al., 2012), mouse anti-H4K20me1

827   (Diagenode, MAb-147-100), rabbit anti-H4K20me3 (Diagenode, MAb-057-050), or

828   IgG isotype controls for 16h at 4°C. BSA-blocked protein A beads (0.5 mg/ml BSA,

829   30 µg/ml salmon sperm, 1X protease inhibitor complete, 0.1% Triton-X-100 in LB3(-

830   ) buffer (without detergents)) were added (50 µl/500 µg chromatin) and incubated for

831   at least 4h on an orbital shaker at 4°C. Sequential washing steps with RIPA (0.1%

832   SDS, 0.5% DOC, 1% NP-40, 50 mM Tris (pH 8.0), 1 mM EDTA) 150mM NaCl,

833   RIPA-300 mM NaCl, RIPA-250 mM LiCl buffer, and twice in TE (pH 8.0) buffer

834   were performed. Immunoprecipitated chromatin fragments were eluted from the

835   beads by shaking twice at 1200 rpm for 10 min at 65°C in 100µl TE 1% SDS. The

836   elution was treated with 80 µg RNAse A for 2h at 37°C and with 8 µg proteinase K at

837   65°C for 16h. DNA was purified using the NucleoSpin Extract II Kit. Quantitative

838   PCR analysis of the EBV *oriP* DS element (for pre-RC ChIP), or H4K20me1 and -

839   me3 positive loci were performed using the SYBR Green I Master Mix (Roche) and

840   the Roche LightCycler 480 System. Oligo sequences for qPCR were DS_fw:

841   AGTTCACTGCCCGCTCCT,   DS_rv: CAGGATTCCACGAGGGTAGT,

842   H4K20me1positive_fw: ATGCCTTCTTGCCTCTTGTC, H4K20me1positive_rv:

843   AGTTAAAAGCAGCCCTGGTG,   H4K20me3positive_fw:

844   TCTGAGCAGGGTTGCAAGTAC,   H4K20me3positive_rv:

845   AAGGAAATGATGCCCAGCTG. Chromatin fragment sizes were verified by

846   loading 1-2 µg chromatin on a 1.5% agarose gel. Samples were quantified using

847   Qubit HS dsDNA.

848

849   **ChIP-sample sequencing**

850   ChIP sample library preparations from > 4 ng of ChIP-DNA was performed

851   using Accel-NGS® 1S Plus DNA Library Kit for Illumina (Swift Biosciences). 50 bp

852   single-end sequencing was done with the Illumina HiSEQ 1500 sequencer to a

853   sequencing depth of ~ 70 million reads. Fastq-files were mapped against the human

854   genome (hg19, GRCh37, version 2009), extended for the EBV genome (NC007605)

855   using bowtie (v1.1.1), (Langmead et al., 2009). Sequencing profiles were generated

856   using deepTools' bamCoverage funtion using reads extension to 200 bp and reads per

857   genomic content (RPGC) normalization (Ramirez et al., 2014). Visualization was

858   performed in UCSC Genome Browser (http://genome.ucsc.edu) and the Integrated

859   Genome Browser (Kent et al., 2002).

860   For H4K20me1 and -me3 ChIP-seq data, MACS2 peak-calling (Zhang et al.,

861   2008) was performed using the broad setting and overlapping peaks in three replicates

862   were retained for further analyses.

863

864   **Binning approach and normalization**

865   All data processing and analysis steps were performed in R (v.3.2.3) and

866   numpy (v.1.18.5) python library, visualizations were done using the ggplot2 (v3.1.0)

867   package (R_Core_Team, 2018) and matplotlib (v.3.2.3) python library. The numbers

868   of reads were calculated in non-overlapping 1 or 10 kb bins and saved in bed files for

869   further analysis. To combine replicates, their sum per bin was calculated (= read

870   frequency). To adjust for sequencing depth, the mean frequency per bin was

871   calculated for the whole sequenced genome and all bins' counts were divided by this

872   mean value resulting in the normalized read frequency. To account for variations in

873   the input sample, we additionally divided by the normalized read frequency of the

874   input, resulting in the relative read frequency. When aggregating different loci, input

875   normalization was performed after averaging. This resulted in relative read frequency

876   ranging from 0 to ~30. Pair-wise Pearson correlations of ORC/MCM samples were

877   clustered by hierarchical clustering using complete linkage clustering.

878

879   **Relation of ChIP relative read frequencies to DNase hypersensitivity**

880   The ENCODE 'DNase clusters' track wgEncodeRegDnaseClusteredV3.bed.gz

881   (03.12.2017) containing DNase hypersensitive sites from 125 cell lines were retrieved

882 from (Thurman et al., 2012). Bins overlapping or not with HS sites larger than 1 kb

883 were defined and the respective ChIP read frequency assigned for comparison.

884

885 **Comparison of ChIP relative read frequencies to replication data**

886 ASs were aligned on their left (5') and right (3') borders. Mean and standard

887 error of the mean (SEM) of relative read frequencies of aligned 1 kb bins were then

888 computed to assess the average ChIP signal around the considered AS borders 50 kb

889 away from the AS to 10 kb within the AS. To make sure bins within the ASs were

890 closer to the considered AS border than to the opposite border, only ASs of size >20

891 kb were used (3247/4891). We also limited this analysis to ASs corresponding to

892 efficient initiation zones by requiring ΔRFD > 0.5, filtering out a further 290 lowly

893 efficient ASs.

894 In order to interrogate the relationship between ASs and transcription, we

895 compared the results obtained for different AS groups: 506 ASs were classified as

896 non-genic AS when the AS locus extended 20-kb at both ends did not overlap any

897 annotated gene; the remaining 2451 ASs were classified as genic ASs. From the latter

898 group, 673 ASs were classified as type 1 ASs when both AS borders where flanked

899 by at least one actively transcribed genes (distance of both AS borders to the closest

900 transcribed (TPM > 3) gene body was < 20 kb), and 1026 ASs were classified as type

901 2 ASs when only one AS border was associated to a transcribed gene (see also

902 Table 1).

903 In order to assess the role of H4H20me3 mark on AS specification, we also

904 classified non-genic ASs depending on their input normed H4K20me3 relative read

905 frequency. We grouped the non-genic ASs where the H4K20me3 relative read

906 frequency was above the genome mean value by more than 1.5 standard deviation

41

907   (estimated over the whole genome) and the non-genic ASs where the H4K20me3

908   relative read frequency was below the genome mean value. This resulted in 154 non-

909   genic ASs with H4K20me3 signal significantly higher than genome average and 242

910   non-genic ASs with H4K20me3 signal lower than genome average.

911   A similar selection was performed on fully intergenic 10 kb windows within

912   NRRs (as done above using the mean and standard deviation of H4K20me3 relative

913   read frequency estimated on all fully intergenic 10 kb windows). This resulted in 504

914   and 3986 windows with high and low H4K20me3 signal, respectively.

915

916   **Comparison of ChIP relative read frequencies to transcription data**

917   Gene containing bins were determined and overlapping genes removed from

918   the analysis. For cumulative analysis, we only worked with genes larger 30 kb, and

919   assigned the gene expression levels in TPM accordingly. Genes were either aligned at

920   their transcriptional start site (TSS) or their transcriptional termination site (TTS) and

921   the corresponding ChIP read frequencies were calculated in a 30 kb window around

922   the site.

923

924   **Comparison of ChIP relative read frequencies to replication timing**

925   For identification of RTDs in Raji cells, we used the early to late replication

926   timing ratio determined by Repli-seq (Sima et al., 2018). We directly worked from the

927   precomputed early to late log-ratio from supplementary file

928   GSE102522_Raji_log2_hg19.txt downloaded from GEO (accession number

929   GSE102522). The timing of every non-overlapping 10 kb bin was calculated as the

930   averaged $\log_2$(Early/Late) ratio within the surrounding 100 kb window. Early RTDs

931   were defined as regions where the average log-ratio > 1.6 and late RTDs as regions

42

932   where the average log-ratio < -2.0. These thresholds resulted in 1648 early RTDs,

933   ranging from 10 to 8940 kb in size, with a mean size of 591 kb, while we detected

934   2046 late RTDs in sizes from 10 to 8860 kb, averaging at 470 kb. These RTDs were

935   used to classify ChIP read relative frequencies calculated in 10 kb bins as early or late

936   replication timing. Bins overlapping gene extended by 10 kb on both sides were

937   removed from the analysis to avoid effects of gene activity on ChIP signals.

938

939   **Comparison of ChIP relative read frequencies distributions at different**

940   **replication timing depending on transcriptional and replicative status**

941   All non-overlapping 10 kb windows were classified as intergenic if closest

942   genes were more than 5 kb away, as belonging to a silent (resp. expressed) gene body

943   if the window was inside a gene with TPM<3 (resp. TPM>3) and at more than 3 kb of

944   gene borders, otherwise windows were disregarded. This made sure that specific ChIP

945   signal at gene TSS and TTS were not considered in the analysis. Using the 3 window

946   categories, we computed the 2D histograms of ChIP relative read frequencies vs

947   replication timing in intergenic, silent and expressed gene bodies. We used 10 timing

948   bins corresponding to the deciles of the whole genome timing distribution. For each

949   timing bin, the histogram counts were normalized so as to obtain an estimate of the

950   probability distribution function of the ChIP signal at the considered replication

951   timing. The analysis was reproduced after restricting for windows fully in (i) AS

952   segments (size > 20 kb, ΔRFD > 0.5), (ii) DS segments (size > 20 kb, ΔRFD < -0.5),

953   (iii) URRs and (iv) NRRs.

954

955

956

957 **Statistics**

958   Statistical analyses were performed in R using one-sided t-test with Welch

959 correction and 95% confidence interval or one-way ANOVA followed by Tukey's

960 multiple comparisons of means with 95% family-wise confidence level, if

961 appropriate. Comparison between ChIP signal distribution observed in two situations

962 were performed computing the 2 sample Kolmogorov–Smirnov statistics D_{KS} using

963 SciPy (v.1.5.0) statistical library and correcting for sample sizes by reporting

964 $Z_{KS} = D_{KS}\sqrt{\frac{nm}{n+m}}$, where n and m are the sizes of the two samples respectively.

965

966 **ERCE RFD profiles**

967   The position of the three genetically identified ERCEs in the mESC Dppa2/4

968 locus and of the 1,835 predicted mESC ERCEs were downloaded from Sima et al.

969 (Sima et al., 2019. The mESC OK-seq data were downloaded from Petryk et al.

970 (Petryk et al., 2018), (SRR7535256), and mapped to mm10 genome Petryk et al.

971 (Petryk et al., 2018). OK-seq data from cycling mouse B cells were downloaded from

972 Tubbs et al. (Tubbs et al., 2018) (GSE116319). The RFD profile was computed as in

973 Hennion et al., 2020 with 10kb binning steps. Predicted ERCE shuffling was

974 performed using a homemade function keeping the number of ERCE constant for

975 each chromosome and avoiding unmapped genome sequences (genome regions with

976 more than 20 consecutives N). Aggregated average RFD profiles were centered on the

977 ERCE and the profile's envelopes represent the CI95 based on the mean and standard

978 deviation at each position.

979

980

981

44

Data access

982

983   Data have been deposited to the European Nucleotide Archive (ENA,

984   https://www.ebi.ac.uk/ena). OK-seq data in Raji cells are available under the

985   accession numbers PRJEB25180 (study accession) and SAMEA104651899 (sample

986   accession, 5 replicates). Raji RNA-seq data are available under the accession number

987   PRJEB31867 (study) and SAMEA5537240, SAMEA5537246, and SAMEA5537252

988   (sample accession per replicate). Raji ChIP-seq data were deposited under the

989   accession number PRJEB32855.

990

991

1007

1008 **Author contributions**

1009   N.K. and A.S. designed and N.K. performed the majority of experiments;

1010 A.B. performed the RNA-seq experiment and TPM analysis; X.W. performed OK-seq

1011 experiments, S.K. and H.B. generated the sequencing library and sequencing, W.H.

1012 designed RNA-seq experiments; O.H. developed OK-seq, B.A supervised

1013 bioinformatic analyses; N.K. L.L. and B.A. performed bioinformatic analyses; A.S.

1014 and OH proposed and designed the project and experimental systems; N.K., O.H. and

1015 A.S. wrote the manuscript with comments from L.L. and B.A.; All the authors read

1016 and approved the manuscript.

1017

1018 **Competing Interests statement**

1019   The authors declare no competing interests.

References

Adams, A., Lindahl, T., & Klein, G. (1973). Linear association between cellular DNA and Epstein-Barr virus DNA in a human lymphoblastoid cell line. *Proc Natl Acad Sci U S A, 70*(10), 2888-92. doi:10.1073/pnas.70.10.2888

Almeida, R., Fernandez-Justel, J. M., Santa-Maria, C., Cadoret, J. C., Cano-Aroca, L., Lombrana, R., . . . Gomez, M. (2018). Chromatin conformation regulates the coordination between DNA replication and transcription. *Nat Commun, 9*(1), 1590. doi:10.1038/s41467-018-03539-8

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-9. doi:10.1093/bioinformatics/btu638

Audit, B., Baker, A., Chen, C. L., Rappailles, A., Guilbaud, G., Julienne, H., . . . Arneodo, A. (2013). Multiscale analysis of genome-wide replication timing profiles using a wavelet-based signal-processing algorithm. *Nat Protoc, 8*(1), 98-110. doi:10.1038/nprot.2012.145

Baker, A., Audit, B., Chen, C. L., Moindrot, B., Leleu, A., Guilbaud, G., . . . Arneodo, A. (2012). Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput Biol, 8*(4), e1002443. doi:10.1371/journal.pcbi.1002443

Beck, D. B., Oda, H., Shen, S. S., & Reinberg, D. (2012). PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev, 26*(4), 325-37. doi:10.1101/gad.177444.111

Bell, S. P., & Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect Biol, 5*(6). doi:10.1101/cshperspect.a010124

Benetti, R., Gonzalo, S., Jaco, I., Schotta, G., Klatt, P., Jenuwein, T., & Blasco, M. A. (2007). Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J Cell Biol, 178*(6), 925-36. doi:10.1083/jcb.200703081

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., . . . Lemaitre, J. M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol, 19*(8), 837-44. doi:10.1038/nsmb.2339

Boos, D., & Ferreira, P. (2019). Origin Firing Regulations to Control Genome Replication Timing. *Genes (Basel), 10*(3). doi:10.3390/genes10030199

Boulos, R. E., Drillon, G., Argoul, F., Arneodo, A., & Audit, B. (2015). Structural organization of human replication timing domains. *FEBS Lett, 589*(20 Pt A), 2944-57. doi:10.1016/j.febslet.2015.04.015

Brustel, J., Kirstein, N., Izard, F., Grimaud, C., Prorok, P., Cayrou, C., . . . Julien, E. (2017). Histone H4K20 tri-methylation at late-firing origins ensures timely heterochromatin replication. *EMBO J, 36*(18), 2726-41. doi:10.15252/embj.201796541

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J. C., . . . Mechali, M. (2015). The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res, 25*(12), 1873-85. doi:10.1101/gr.192799.115

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., . . . Mechali, M. (2011). Genome-scale analysis of metazoan replication origins

1068   reveals their organization in specific but flexible sites defined by conserved
1069   features. *Genome Res, 21*(9), 1438-49. doi:10.1101/gr.121830.111

1070   Chen, Y. H., Keegan, S., Kahli, M., Tonzi, P., Fenyo, D., Huang, T. T., & Smith, D. J.
1071   (2019). Transcription shapes DNA replication initiation and termination in
1072   human cells. *Nat Struct Mol Biol, 26*(1), 67-77. doi:10.1038/s41594-018-
1073   0171-0

1074   Das, S. P., Borrman, T., Liu, V. W., Yang, S. C., Bechhoefer, J., & Rhind, N. (2015).
1075   Replication timing is regulated by the number of MCMs loaded at origins.
1076   *Genome Res, 25*(12), 1886-92. doi:10.1101/gr.195305.115

1077   Dellino, G. I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., . . . Pelicci, P.
1078   G. (2013). Genome-wide mapping of human DNA-replication origins: levels
1079   of transcription at ORC1 sites regulate origin selection and replication timing.
1080   *Genome Res, 23*(1), 1-11. doi:10.1101/gr.142331.112

1081   Demczuk, A., Gauthier, M. G., Veras, I., Kosiyatrakul, S., Schildkraut, C. L.,
1082   Busslinger, M., . . . Norio, P. (2012). Regulation of DNA replication within
1083   the immunoglobulin heavy-chain locus during B cell commitment. *PLoS Biol,*
1084   *10*(7), e1001360. doi:10.1371/journal.pbio.1001360

1085   Douglas, M. E., Ali, F. A., Costa, A., & Diffley, J. F. X. (2018). The mechanism of
1086   eukaryotic CMG helicase activation. *Nature, 555*(7695), 265-68.
1087   doi:10.1038/nature25787

1088   Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., . . . Speck, C. (2009). A
1089   double-hexameric MCM2-7 complex is loaded onto origin DNA during
1090   licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A, 106*(48),
1091   20240-5. doi:0911500106 [pii]
1092   10.1073/pnas.0911500106

1093   Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying ChIP-seq
1094   enrichment using MACS. *Nat Protoc, 7*(9), 1728-40.
1095   doi:10.1038/nprot.2012.101

1096   Fragkos, M., Ganier, O., Coulombe, P., & Mechali, M. (2015). DNA replication
1097   origin activation in space and time. *Nat Rev Mol Cell Biol, 16*(6), 360-74.
1098   doi:10.1038/nrm4002

1099   Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., & Remus, D. (2015). Post-
1100   licensing Specification of Eukaryotic Replication Origins by Facilitated
1101   Mcm2-7 Sliding along DNA. *Mol Cell, 60*(5), 797-807.
1102   doi:10.1016/j.molcel.2015.10.022

1103   Guilbaud, G., Rappailles, A., Baker, A., Chen, C. L., Arneodo, A., Goldar, A., . . .
1104   Hyrien, O. (2011). Evidence for sequential and increasing activation of
1105   replication origins along replication timing gradients in the human genome.
1106   *PLoS Comput Biol, 7*(12), e1002322. doi:10.1371/journal.pcbi.1002322

1107   Hadjadj, D., Denecker, T., Maric, C., Fauchereau, F., Baldacci, G., & Cadoret, J. C.
1108   (2016). Characterization of the replication timing program of 6 human model
1109   cell lines. *Genom Data, 9*, 113-7. doi:10.1016/j.gdata.2016.07.003

1110   Hamlin, J. L., Mesner, L. D., & Dijkwel, P. A. (2010). A winding road to origin
1111   discovery. *Chromosome Res, 18*(1), 45-61. doi:10.1007/s10577-009-9089-z

1112   Hennion, M., Arbona, J. M., Lacroix, L., Cruaud, C., Theulot, B., Tallec, B. L., . . .
1113   Hyrien, O. (2020). FORK-seq: replication landscape of the Saccharomyces
1114   cerevisiae genome by nanopore sequencing. *Genome Biol, 21*(1), 125.
1115   doi:10.1186/s13059-020-02013-3

1116   Hulke, M. L., Massey, D. J., & Koren, A. (2020). Genomic methods for measuring
1117   DNA replication dynamics. *Chromosome Res, 28*(1), 49-67.
1118   doi:10.1007/s10577-019-09624-y
1119   Hyrien, O. (2016). How MCM loading and spreading specify eukaryotic DNA
1120   replication initiation sites. *F1000Res, 5*. doi:10.12688/f1000research.9008.1
1121   Hyrien, O., Maric, C., & Mechali, M. (1995). Transition in specification of embryonic
1122   metozoan DNA replication origins. *Science, 270*, 994-97.
1123   Jorgensen, S., Schotta, G., & Sorensen, C. S. (2013). Histone H4 lysine 20
1124   methylation: key player in epigenetic regulation of genomic integrity. *Nucleic*
1125   *Acids Res, 41*(5), 2797-806. doi:10.1093/nar/gkt012
1126   Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M.,
1127   & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res,*
1128   *12*(6), 996-1006. doi:10.1101/gr.229102
1129   Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013).
1130   TopHat2: accurate alignment of transcriptomes in the presence of insertions,
1131   deletions and gene fusions. *Genome Biol, 14*(4), R36. doi:10.1186/gb-2013-
1132   14-4-r36
1133   Knott, S. R., Viggiani, C. J., & Aparicio, O. M. (2009). To promote and protect:
1134   coordinating DNA replication and transcription for genome stability.
1135   *Epigenetics, 4*(6), 362-5. doi:10.4161/epi.4.6.9712
1136   Langley, A. R., Graf, S., Smith, J. C., & Krude, T. (2016). Genome-wide
1137   identification and characterisation of human DNA replication origins by
1138   initiation site sequencing (ini-seq). *Nucleic Acids Res*.
1139   doi:10.1093/nar/gkw760
1140   Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and
1141   memory-efficient alignment of short DNA sequences to the human genome.
1142   *Genome Biol, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25
1143   Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J., & Bensimon, A. (2006).
1144   DNA replication origin interference increases the spacing between initiation
1145   events in human cells. *Mol Biol Cell, 17*(12), 5337-45. doi:10.1091/mbc.e06-
1146   04-0298
1147   Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
1148   Wheeler transform. *Bioinformatics, 25*(14), 1754-60.
1149   doi:10.1093/bioinformatics/btp324
1150   MacAlpine, H. K., Gordan, R., Powell, S. K., Hartemink, A. J., & MacAlpine, D. M.
1151   (2010). Drosophila ORC localizes to open chromatin and marks sites of
1152   cohesin complex loading. *Genome Res, 20*(2), 201-11. doi:gr.097873.109 [pii]
1153   10.1101/gr.097873.109
1154   Macheret, M., & Halazonetis, T. D. (2018). Intragenic origins due to short G1 phases
1155   underlie oncogene-induced DNA replication stress. *Nature, 555*(7694), 112-
1156   16. doi:10.1038/nature25507
1157   Marahrens, Y., & Stillman, B. (1992). A yeast chromosomal origin of DNA
1158   replication defined by multiple functional elements. *Science, 255*, 817-23.
1159   Marchal, C., Sima, J., & Gilbert, D. M. (2019). Control of DNA replication timing in
1160   the 3D genome. *Nat Rev Mol Cell Biol, 20*(12), 721-37. doi:10.1038/s41580-
1161   019-0162-y
1162   Martin, M. M., Ryan, M., Kim, R., Zakas, A. L., Fu, H., Lin, C. M., . . . Aladjem, M.
1163   I. (2011). Genome-wide depletion of replication initiation events in highly
1164   transcribed regions. *Genome Res, 21*(11), 1822-32. doi:10.1101/gr.124644.111

McGuffee, S. R., Smith, D. J., & Whitehouse, I. (2013). Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell, 50*(1), 123-35. doi:10.1016/j.molcel.2013.03.004

Mesner, L. D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J. L., & Bekiranov, S. (2013). Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res, 23*(11), 1774-88. doi:10.1101/gr.155218.113

Miotto, B., Ji, Z., & Struhl, K. (2016). Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci U S A, 113*(33), E4810-9. doi:10.1073/pnas.1609060113

Moiseeva, T. N., & Bakkenist, C. J. (2018). Regulation of the initiation of DNA replication in human cells. *DNA Repair (Amst), 72*, 99-106. doi:10.1016/j.dnarep.2018.09.003

Muller, C. A., Boemo, M. A., Spingardi, P., Kessler, B. M., Kriaucionis, S., Simpson, J. T., & Nieduszynski, C. A. (2019). Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods, 16*(5), 429-36. doi:10.1038/s41592-019-0394-y

Nakamura, K., Saredi, G., Becker, J. R., Foster, B. M., Nguyen, N. V., Beyer, T. E., . . . Groth, A. (2019). H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat Cell Biol, 21*(3), 311-18. doi:10.1038/s41556-019-0282-9

Norio, P., & Schildkraut, C. L. (2001). Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science, 294*(5550), 2361-64.

Norio, P., & Schildkraut, C. L. (2004). Plasticity of DNA replication initiation in Epstein-Barr virus episomes. *PLoS Biol, 2*(6), e152.

Pannetier, M., Julien, E., Schotta, G., Tardat, M., Sardet, C., Jenuwein, T., & Feil, R. (2008). PR-SET7 and SUV4-20H regulate H4 lysine-20 methylation at imprinting control regions in the mouse. *EMBO Rep, 9*(10), 998-1005. doi:10.1038/embor.2008.147

Papior, P., Arteaga-Salas, J. M., Gunther, T., Grundhoff, A., & Schepers, A. (2012). Open chromatin structures regulate the efficiencies of pre-RC formation and replication initiation in Epstein-Barr virus. *J Cell Biol, 198*(4), 509-28. doi:10.1083/jcb.201109105

Petryk, N., Dalby, M., Wenger, A., Stromme, C. B., Strandsby, A., Andersson, R., & Groth, A. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science, 361*(6409), 1389-92. doi:10.1126/science.aau0294

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., . . . Hyrien, O. (2016). Replication landscape of the human genome. *Nat Commun, 7*, 10208. doi:10.1038/ncomms10208

Picard, F., Cadoret, J. C., Audit, B., Arneodo, A., Alberti, A., Battail, C., . . . Prioleau, M. N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet, 10*(5), e1004282. doi:10.1371/journal.pgen.1004282

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., . . . Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature, 515*(7527), 402-5. doi:10.1038/nature13986

Powell, S. K., MacAlpine, H. K., Prinz, J. A., Li, Y., Belsky, J. A., & MacAlpine, D. M. (2015). Dynamic loading and redistribution of the Mcm2-7 helicase

1214   complex through the cell cycle. *EMBO J, 34*(4), 531-43.
1215   doi:10.15252/embj.201488307

1216 Prioleau, M. N., & MacAlpine, D. M. (2016). DNA replication origins-where do we
1217   begin? *Genes Dev, 30*(15), 1683-97. doi:10.1101/gad.285114.116

1218 R_Core_Team. (2018). R: A language and environment for statistical computing. R
1219   Foundation for Statistical Computin. Retrieved from https://www.R-
1220   project.org

1221 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., & Manke, T. (2014). deepTools: a
1222   flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,*
1223   *42*(Web Server issue), W187-91. doi:10.1093/nar/gku365

1224 Remus, D., & Diffley, J. F. (2009). Eukaryotic DNA replication control: lock and
1225   load, then fire. *Curr Opin Cell Biol, 21*(6), 771-7. doi:S0955-0674(09)00152-5
1226   [pii]
1227 10.1016/j.ceb.2009.08.002

1228 Rhind, N., & Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Harb*
1229   *Perspect Biol, 3*(7), 1-26. doi:10.1101/cshperspect.a010132

1230 Ritzi, M., Tillack, K., Gerhardt, J., Ott, E., Humme, S., Kremmer, E., . . . Schepers, A.
1231   (2003). Complex Protein-DNA Dynamics at the Latent Origin of DNA
1232   Replication of Epstein-Barr Virus. *J Cell Sci, 116*, 3971-84.

1233 Rivera-Mulia, J. C., & Gilbert, D. M. (2016a). Replicating Large Genomes: Divide
1234   and Conquer. *Mol Cell, 62*(5), 756-65. doi:10.1016/j.molcel.2016.05.007

1235 Rivera-Mulia, J. C., & Gilbert, D. M. (2016b). Replication timing and transcriptional
1236   control: beyond cause and effect-part III. *Curr Opin Cell Biol, 40*, 168-78.
1237   doi:10.1016/j.ceb.2016.03.022

1238 Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., . . . Gilbert, D. M. (2010).
1239   Evolutionarily conserved replication timing profiles predict long-range
1240   chromatin interactions and distinguish closely related cell types. *Genome Res,*
1241   *20*(6), 761-70. doi:10.1101/gr.099655.109

1242 Sasaki, T., Ramanathan, S., Okuno, Y., Kumagai, C., Shaikh, S. S., & Gilbert, D. M.
1243   (2006). The Chinese hamster dihydrofolate reductase replication origin
1244   decision point follows activation of transcription and suppresses initiation of
1245   replication within transcription units. *Mol Cell Biol, 26*(3), 1051-62.
1246   doi:10.1128/MCB.26.3.1051-1062.2006

1247 Schepers, A., Ritzi, M., Bousset, K., Kremmer, E., Yates, J. L., Harwood, J., . . .
1248   Hammerschmidt, W. (2001). Human origin recognition complex binds to the
1249   region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J,*
1250   *20*(16), 4588-602.

1251 Schwartz, Y. B., Kahn, T. G., & Pirrotta, V. (2005). Characteristic low density and
1252   shear sensitivity of cross-linked chromatin containing polycomb complexes.
1253   *Mol Cell Biol, 25*(1), 432-9. doi:10.1128/MCB.25.1.432-439.2005

1254 Shoaib, M., Walter, D., Gillespie, P. J., Izard, F., Fahrenkrog, B., Lleres, D., . . .
1255   Sorensen, C. S. (2018). Histone H4K20 methylation mediated chromatin
1256   compaction threshold ensures genome integrity by limiting DNA replication
1257   licensing. *Nat Commun, 9*(1), 3704. doi:10.1038/s41467-018-06066-8

1258 Sima, J., Bartlett, D. A., Gordon, M. R., & Gilbert, D. M. (2018). Bacterial artificial
1259   chromosomes establish replication timing and sub-nuclear compartment de
1260   novo as extra-chromosomal vectors. *Nucleic Acids Res, 46*(4), 1810-20.
1261   doi:10.1093/nar/gkx1265

1262 Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K. N., Holcomb, N. P., . .
1263   . Gilbert, D. M. (2019). Identifying cis Elements for Spatiotemporal Control of

Mammalian DNA Replication. *Cell, 176*(4), 816-30 e18. doi:10.1016/j.cell.2018.11.036

Smith, D. J., & Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature, 483*(7390), 434-8. doi:10.1038/nature10895

Smith, O. K., & Aladjem, M. I. (2014). Chromatin structure and replication origins: Determinants of chromosome replication and nuclear organization. *J Mol Biol, 426*(20), 3330-41. doi:10.1016/j.jmb.2014.05.027

Sugimoto, N., Maehara, K., Yoshida, K., Ohkawa, Y., & Fujita, M. (2018). Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. *Nucleic Acids Res, 46*(13), 6683-96. doi:10.1093/nar/gky476

Takahashi, T. S., Wigley, D. B., & Walter, J. C. (2005). Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci, 30*(8), 437-44. doi:10.1016/j.tibs.2005.06.007

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., & Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat Cell Biol, 12*(11), 1086-93. doi:10.1038/ncb2113

Teytelman, L., Ozaydin, B., Zill, O., Lefrancois, P., Snyder, M., Rine, J., & Eisen, M. B. (2009). Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One, 4*(8), e6700. doi:10.1371/journal.pone.0006700

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., . . . Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature, 489*(7414), 75-82. doi:10.1038/nature11232

Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., . . . Nussenzweig, A. (2018). Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell, 174*(5), 1127-42 e19. doi:10.1016/j.cell.2018.07.011

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci, 131*(4), 281-5. doi:10.1007/s12064-012-0162-3

Wu, X., Kabalane, H., Kahli, M., Petryk, N., Laperrousaz, B., Jaszczyszyn, Y., . . . Hyrien, O. (2018). Developmental and cancer-associated plasticity of DNA replication preferentially targets GC-poor, lowly expressed and late-replicating regions. *Nucleic Acids Res, 46*(19), 10157-72. doi:10.1093/nar/gky797

Yang, S. C., Rhind, N., & Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol Syst Biol, 6*, 404. doi:10.1038/msb.2010.61

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Zhao, P. A., Rivera-Mulia, J. C., & Gilbert, D. M. (2017). Replication Domains: Genome Compartmentalization into Functional Replication Units. *Adv Exp Med Biol, 1042*, 229-57. doi:10.1007/978-981-10-6955-0_11

Zhao, P. A., Sasaki, T., & Gilbert, D. M. (2020). High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol, 21*(1), 76. doi:10.1186/s13059-020-01983-8

1314 **SUPPLEMENTARY MATERIAL**

1315 **Supplementary Figure 1**   53

1316 **Supplementary Figure 2**

1317 **Supplementary Figure 3**

1318 **Supplementary Figure 4**

1319 **Supplementary Figure 5**

1320 **Supplementary Figure 6**

1321 **Supplementary Figure 7**

1322 **Supplementary Figure 8**

1323 **Supplementary Figure 9**

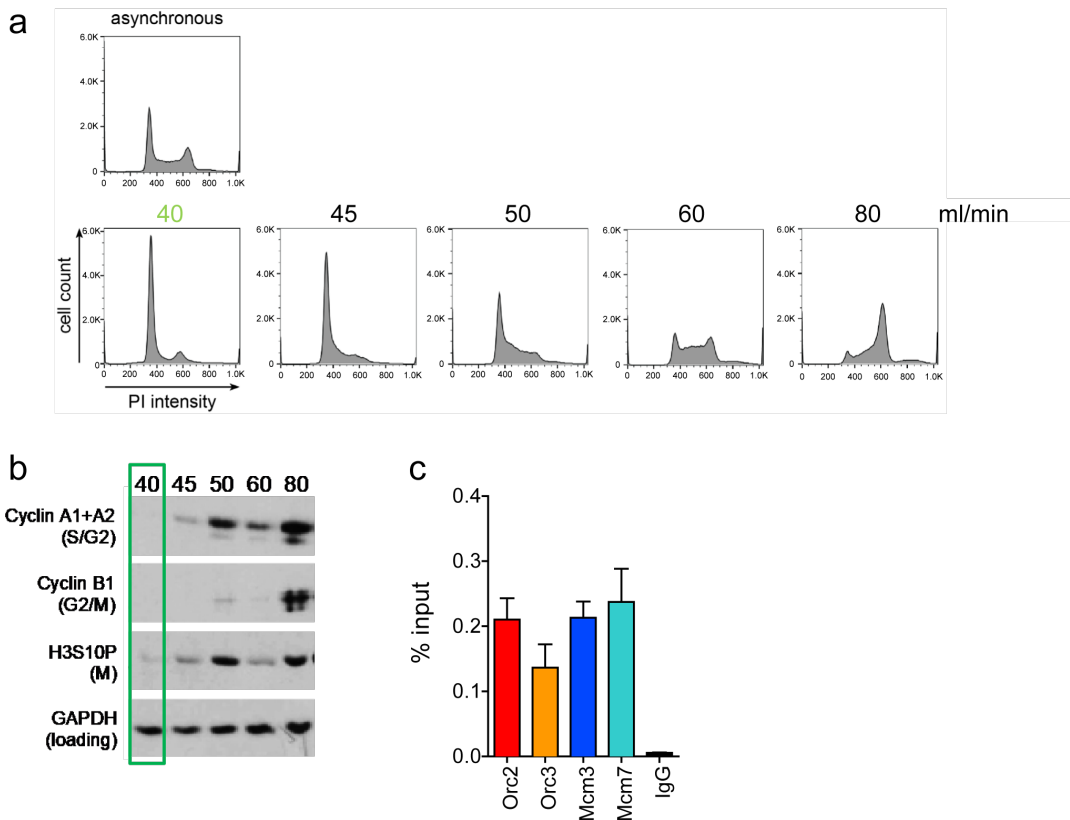1324 **Supplementary Figure 10**

1325

1326 **Supplementary Table 1**

1327 **Supplementary Table 2**

1328 **Supplementary Table 3**

1329   **Supplementary Figure 1**

1330

1331   Supplementary Figure 1: Experimental validation of cell cycle fractionation and ORC/MCM
1332   ChIP quality.
1333   a) Example DNA content (Propidium Iodide) staining followed by FACS of logarithmically
1334   growing Raji (top) cells after cell cycle fractionation by centrifugal elutriation (increasing
1335   counter flow rates indicated above each profile). b) Western Blot analyses of the single
1336   fractions detecting Cyclin A (S/G2), Cyclin B (G2/M), H3S10P (M) and GAPDH. c) qPCR
1337   validation of Orc2, Orc3, Mcm3 and Mcm7 enrichment at the EBV latent origin oriP dyad
1338   symmetry element. Representation in % input. Isotype IgG was used as control.
1339

1340 **Supplementary Figure 2**
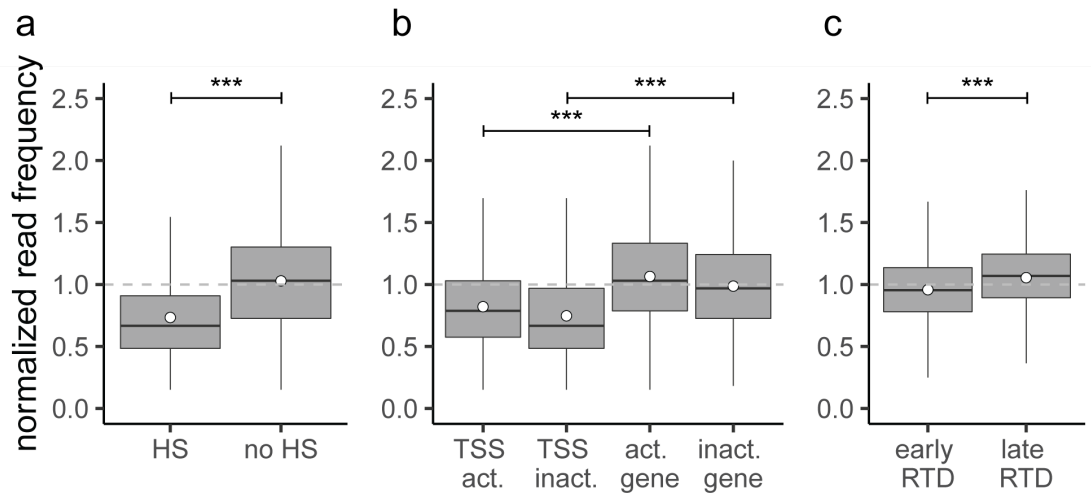
1341

1342   Supplementary Figure 2: The input sequencing control is differentially represented in regions
1343   of biological function.
1344   Boxplot of normalized input read frequencies in relation to a) DNase hypersensitivity. DNase
1345   HS clusters were obtained from 125 cell lines in ENCODE, only HS sites larger 1 kb were
1346   considered. b) transcription: TSSs and gene body of active (TPM > 3) and inactive (TPM < 3)
1347   genes, and c) early or late replication timing domains (RTDs). The dashed grey horizontal
1348   line indicates read frequency 1.0 for orientation. Boxplot represent the mean (circle), the
1349   median (thick line), the 1st and 3rd quartile (box), the 1st and 9th decile (whiskers) of the
1350   relative read frequencies, without representing outliers. Statistics were performed using one-
1351   sided t-test. * p < 0.05, ** p < 0.01, *** p < 0.001.
1352
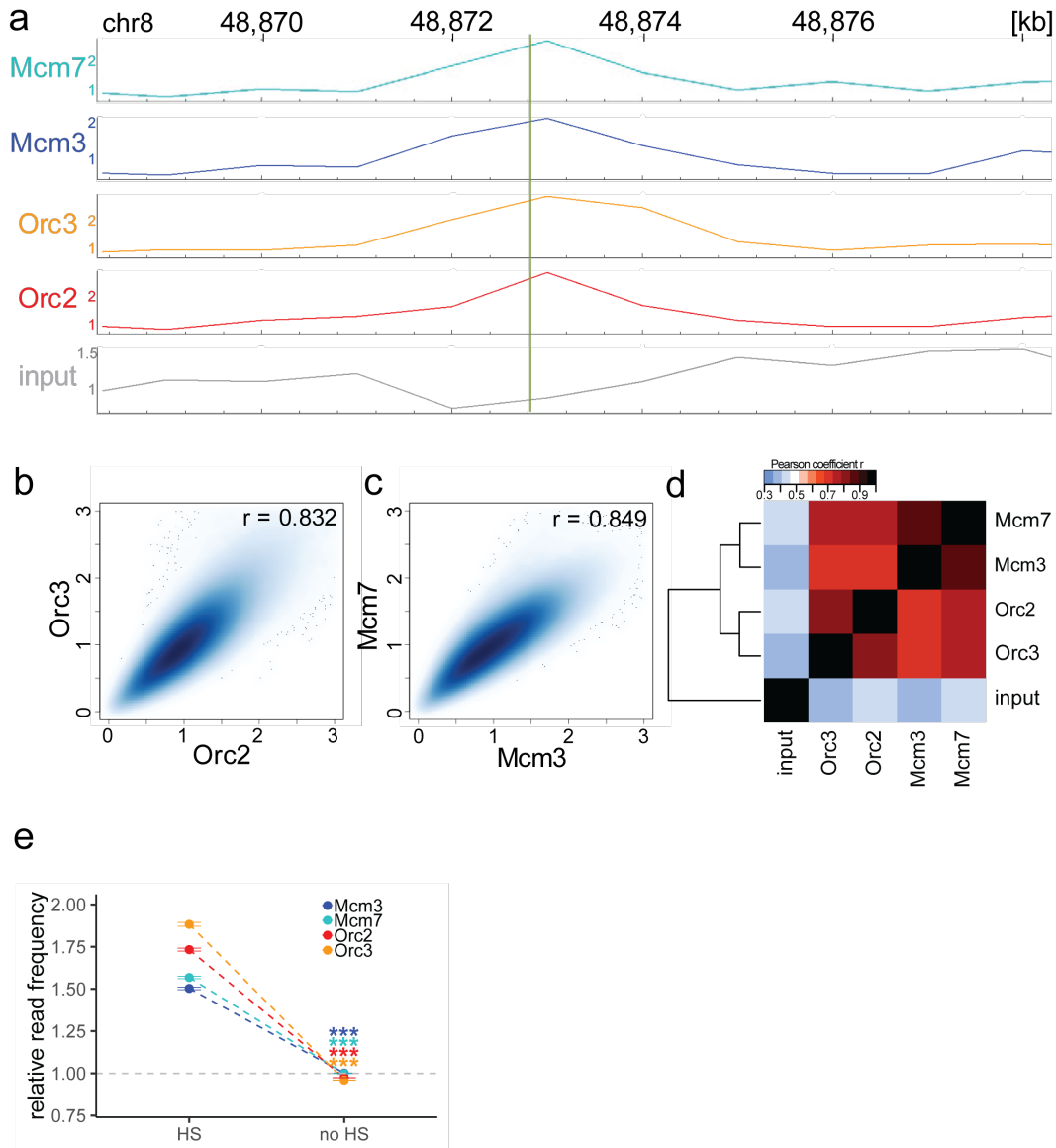
Supplementary Figure 3

Supplementary Figure 3: ORC/MCM enrichment at the MCM4/PRKDC origin persists without input normalization. a) The profile of ORC/MCM ChIP-seq after 1 kb binning in the same 10 kb window as Figure 1b (chr8: 48,868,314 - 48,878,313). The reads of replicates were summed and normalized by the total genome-wide ChIP read frequency. Y-axis represents the resulting normalized read frequency. b) Correlation plot between Orc2 and Orc3 normalized read frequencies in 1 kb bins. c) Correlation plot between Mcm3 and Mcm7 normalized read frequencies in 1 kb bins. d) Heatmap of Pearson correlation coefficients r between all ChIP normalized read frequencies including input in 1 kb bins. Column and line order were determined by complete linkage hierarchical clustering using the correlation distance (d = 1-r). e) ORC/MCM binding is confirmed at DNase HS sites: mean input-normalized ORC/MCM relative read frequencies (± 2 x SEM) in relation to DNase hypersensitivity. Only HS sites larger 1 kb were considered. Statistics were performed using one-sided t-test. *** $p < 0.001$.

56

1368 **Supplementary Figure 4**
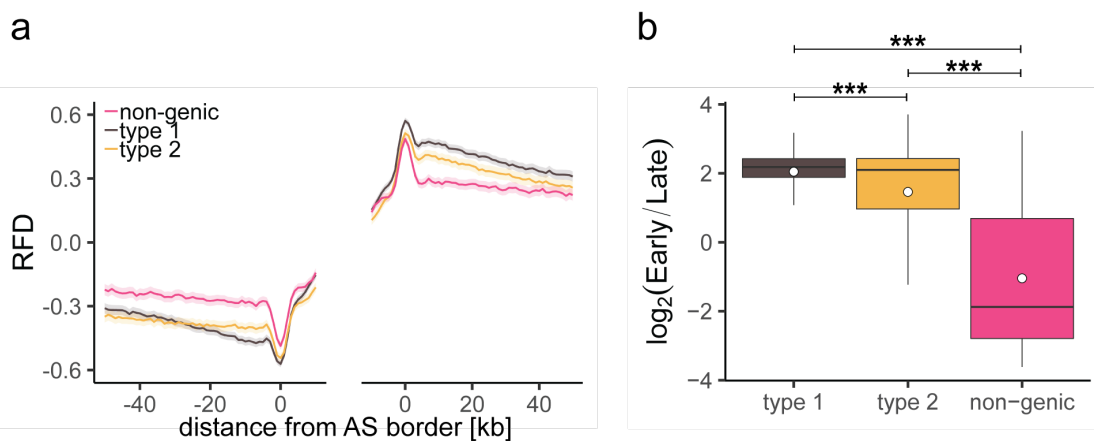
1369

1370 Supplementary Figure 4: Characterization of different AS types.
1371 a) Average RFD of different AS types plotted at AS borders ± 2 x SEM (lighter shadows). b)
1372 Replication timing ratio log2(Early/Late) was assigned to type 1, type 2, and non-genic AS
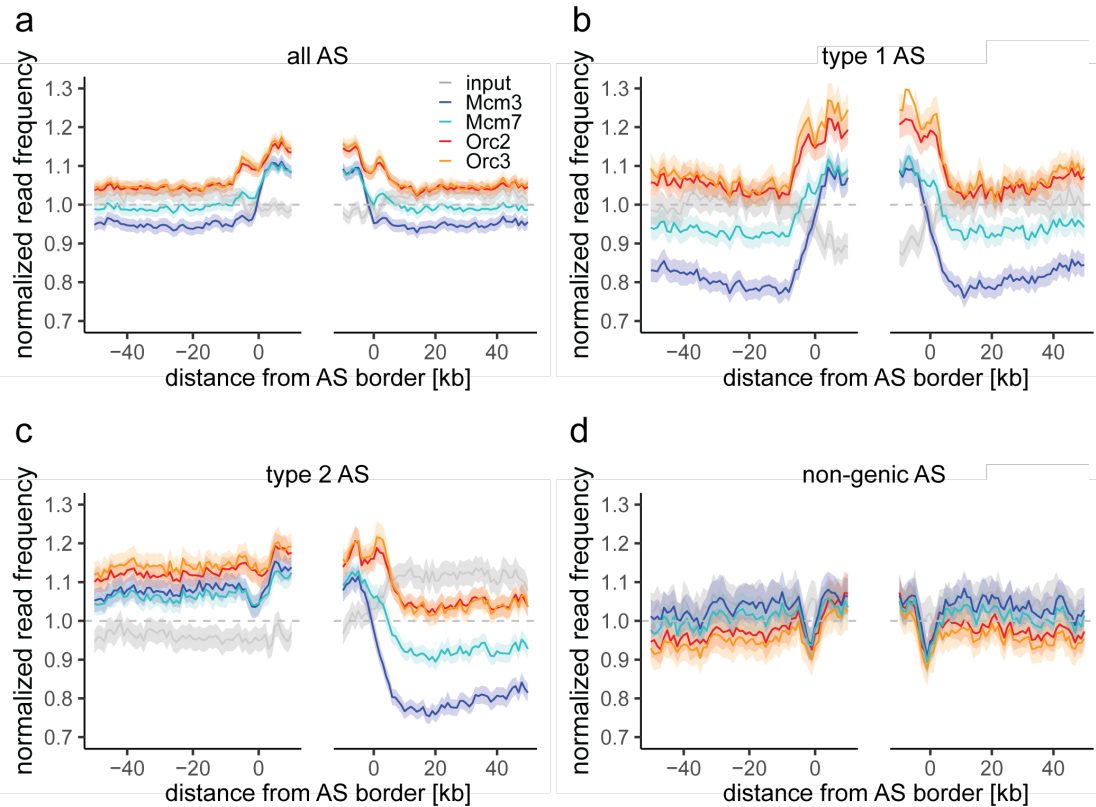1373 and represented as boxplot. Boxplot represent the mean (circle), the median (thick line), the
1374 1st and 3rd quartile (box), the 1st and 9th decile (whiskers) of the relative read frequencies,
1375 without representing outliers. Statistics were performed by one-way ANOVA followed by
1376 Tukey's post-hoc test. *** $p < 0.001$.
1377

1378   **Supplementary Figure 5**

1379

1380   Supplementary Figure 5: ORC/MCM enrichment within AS without input normalization.
1381   a-d) Average ChIP normalized read frequencies of Orc2, Orc3, Mcm3, Mcm7, and input at
1382   AS borders of b) all ASs (n = 2957), c) type 1 ASs with transcribed genes at both ASs borders
1383   (n = 673), d) type 2 ASs with transcribed genes oriented at their right ASs border (n = 1026),
1384   and e) non-genic ASs in gene deprived regions (n = 506). The mean of normalized read
1385   frequencies is shown ± 2 x SEM (lighter shadows). The dashed grey horizontal line indicates
1386   read frequency 1.0 for reference.
1387

1388   **Supplementary Figure 6**
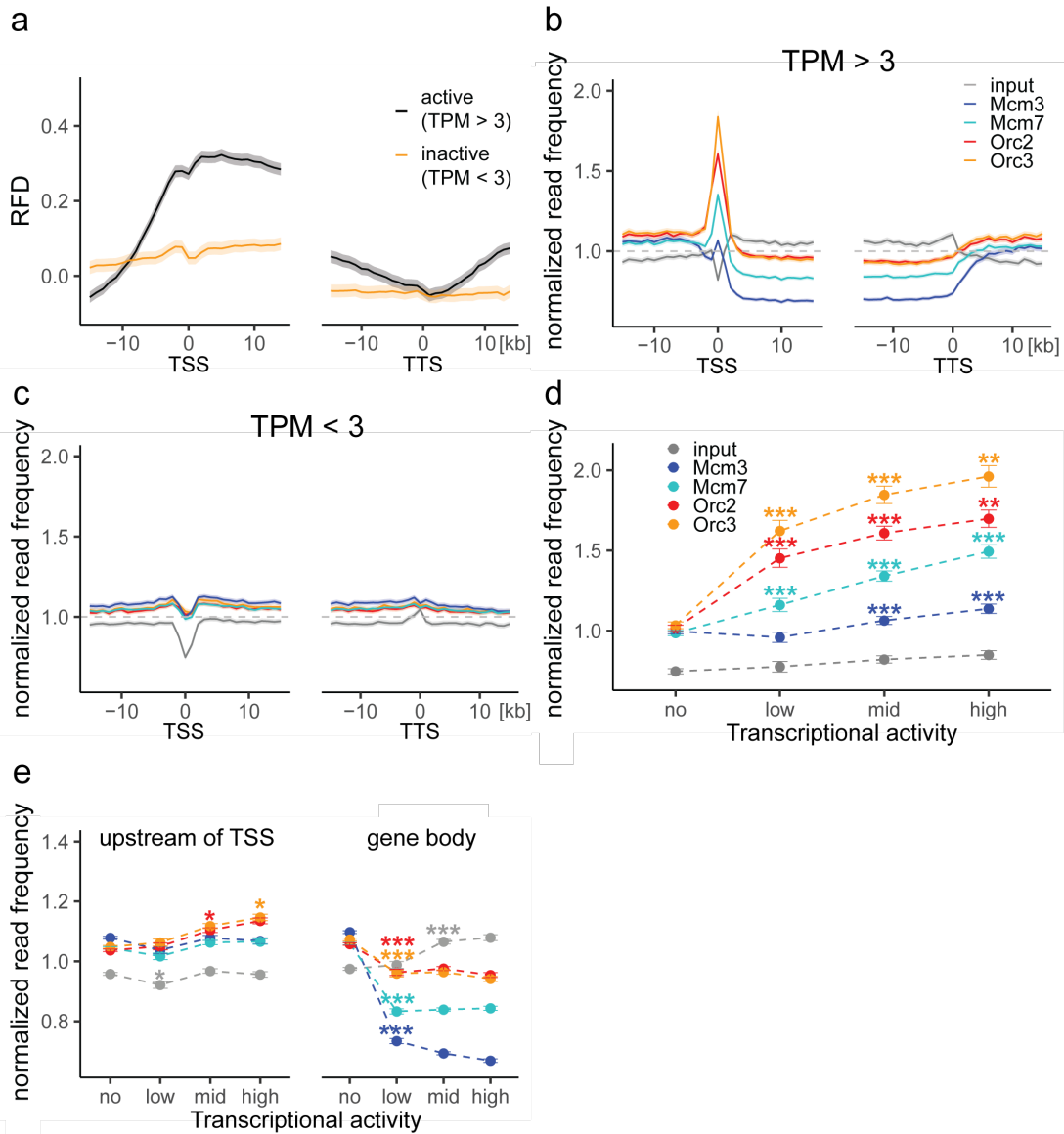
1389

1390   Supplementary Figure 6: ORC and MCM profiles at genes without input normalization.
1391   a) RFD around TSSs or TTSs of active genes (black) or inactive genes (yellow). Distances
1392   from TSSs or TTSs are indicated in kb. RFD means are shown ± 2 x SEM (lighter shadows).
1393   b - c) Normalized ORC/MCM/input read frequencies without input division around TSSs or
1394   TTSs for b) active genes (TPM > 3) and c) inactive genes (TPM < 3). Only genes larger than
1395   30 kb without any adjacent gene within 15 kb were considered. Distances from TSSs or TTSs
1396   are indicated in kb. Means of normalized read frequencies are shown ± 2 x SEM (lighter
1397   shadows). The dashed grey horizontal line indicates read frequency 1.0 for reference. d)
1398   Normalized ORC/MCM/input read frequencies at TSSs dependent on transcriptional activity
1399   (± 2 x SEM). e) Normalized ORC/MCM/input read frequencies upstream of TSSs and in the
1400   gene body dependent on transcriptional activity (± 2 x SEM; TSSs ± 3 kb removed from
1401   analysis). Transcriptional activity was classified as: no (TPM < 3), low (TPM 3-10), mid
1402   (TPM 10-40), high (TPM > 40). Statistics were performed by one-way ANOVA followed by
1403   Tukey's post-hoc test. P-values are indicated always comparing to the previous transcriptional
1404   level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

59

1405 **Supplementary Figure 7**
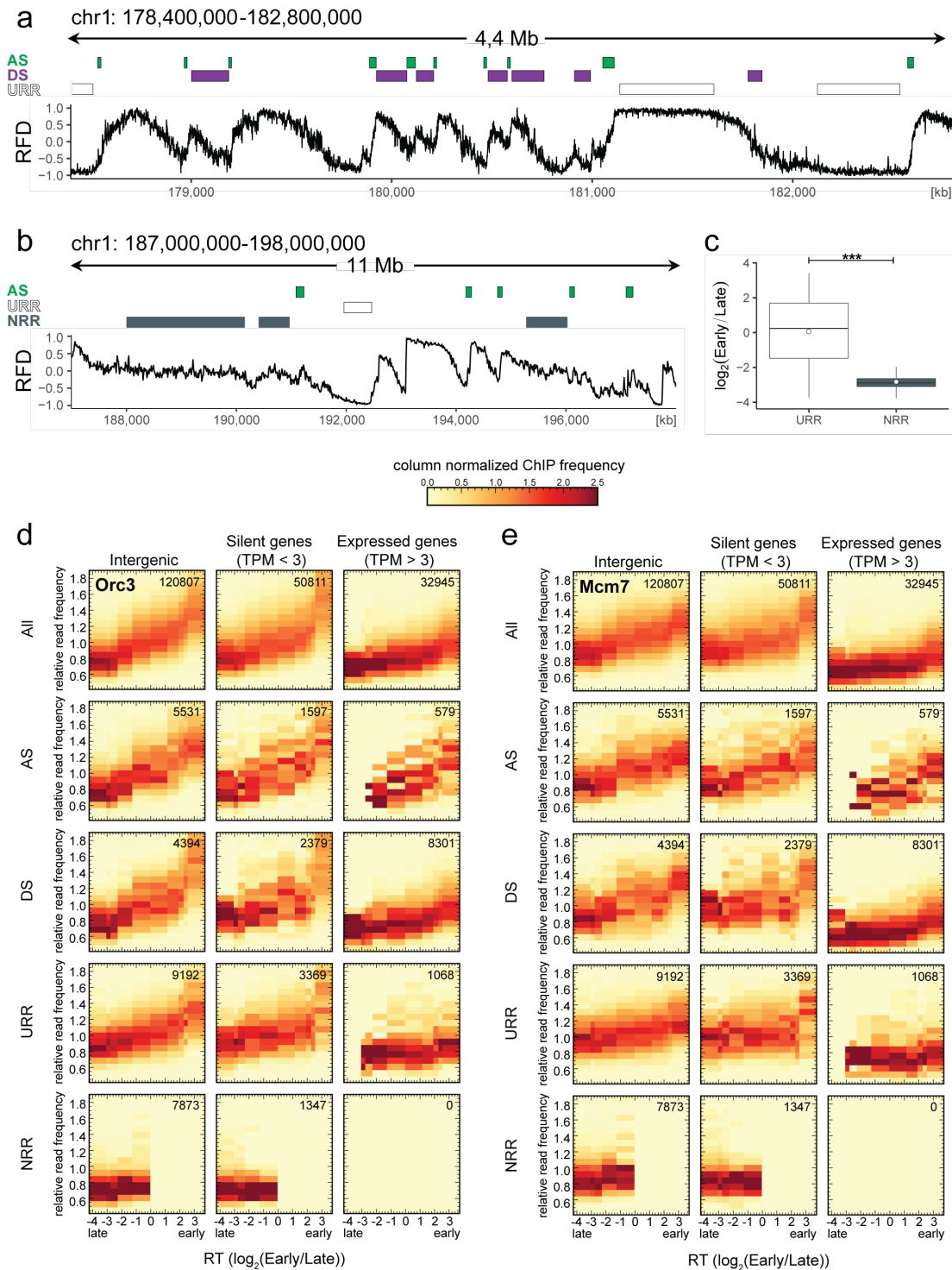
1406

1407 Supplementary Figure 7: ORC/MCM levels are correlated with RT and transcriptional
1408 activity but otherwise homogeneously distributed along the genome and uncorrelated to RFD
1409 patterns.
1410 a) Same RFD profile example as Fig. 2a (chr1: 178,400,000 – 182,800,000, covering 4 Mb)
1411 with indication of AS (green), DS (purple), and URR (white boxes) positions. b) RFD profile
1412 example on chr1: 187,000,000 – 198,000,000, covering 11 MB, with ASs (green), URRs
1413 (white boxes), and NRRs (dark gray) indicated. c) Replication timing ratio log2(Early/Late)

1414 was assigned to URRs and NRRs and represented as boxplot (mean (circle), the median (thick
1415 line), the 1st and 3rd quartile (box), the 1st and 9th decile (whiskers), without representing
1416 outliers). Statistics were performed using one-sided t-test. *** p < 0.001. d-e) 3x5 panel of
1417 2D histograms representing Orc3 (d) and Mcm7 (e) ChIP frequency vs. RT (average
1418 log2(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis
1419 was performed in 10 kb windows. ChIP relative read frequencies are normalized by column
1420 and represent the probability density function of ChIP frequency at a given replication timing.
1421 The color legend is indicated on top. The columns of each panel represent only windows
1422 present in intergenic regions (left column), silent genes (TPM < 3, middle column), and
1423 expressed genes (TPM > 3, right column). TSSs and TTSs proximal regions were not
1424 considered (see Material and Methods). The rows show either all bins (top row), AS bins
1425 (predominant replication initiation, second row), DS bins (descending segment, predominant
1426 replication termination, third row), URR bins (unidirectional replication, no initiation, no
1427 termination, fourth row) and NRR bins (null RFD regions, spatially random initiation and
1428 termination, bottom row). The number of bins per panel is indicated in each panel. See Fig. 4
1429 for equivalent Orc2 and Mcm3 data. Refer to Suppl. Fig. 8a for statistical significances.
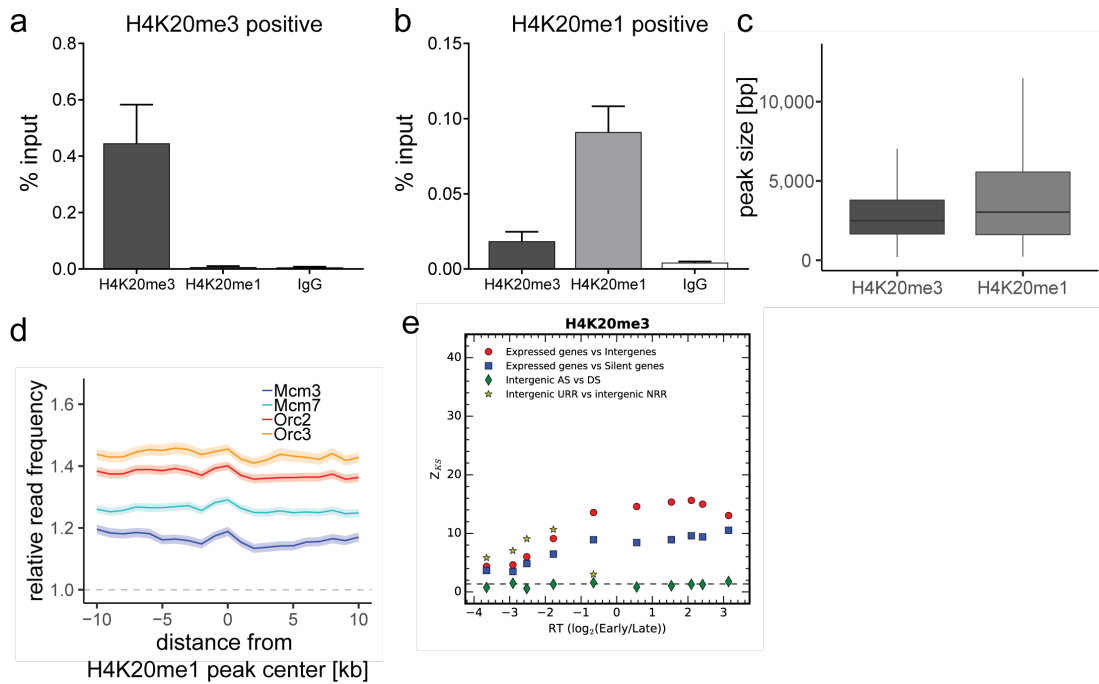1430

1431 **Supplementary Figure 8**
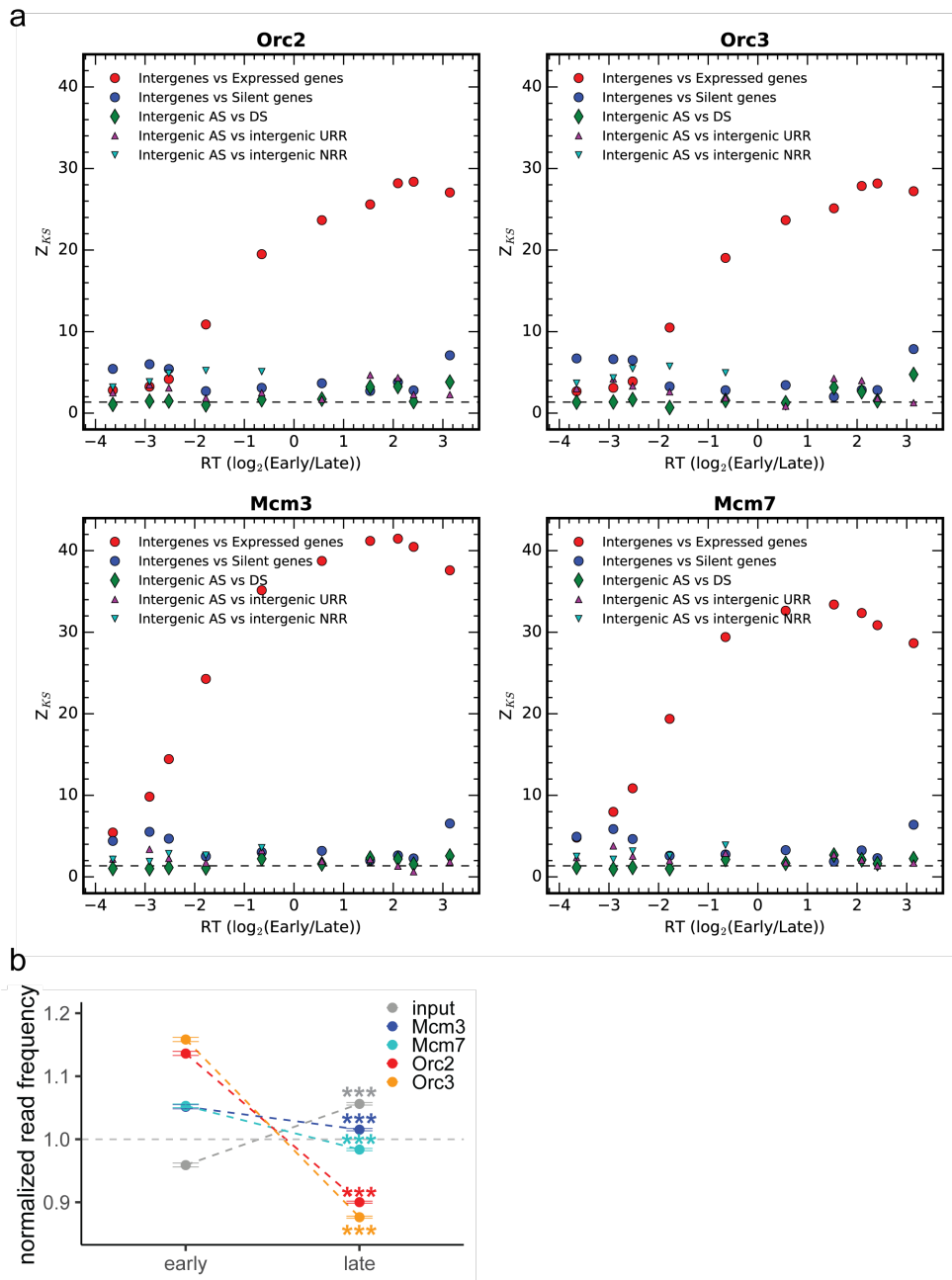
1432

1433 Supplementary Figure 8:
1434 a) Kolmogorov–Smirnov statistics between the ORC/MCM relative read frequency
1435 distributions in each replication timing bin (shown in Fig. 4 and Suppl. Fig. 7d and 7e) in
1436 intergenic vs expressed gene regions (red circles), in intergenic versus silent gene regions
1437 (blue circles) and between intergenic regions in AS versus DS (green diamonds), in AS
1438 versus URR (magenta triangles pointing up) and in AS versus NRR (cyan triangles pointing
1439 down). Z_KS is normalized for sample size. The horizontal dashed lines correspond to p-
1440 value = 5%. b) Normalized ORC/MCM/input read frequencies without input division (± 2 x
1441 SEM) in early or late RTDs. Early RTDs were defined as log2(Early/Late) > 1.6; late RTDs <
1442 -2.0. The analysis was performed in 10 kb bins. Any gene ± 10 kb was removed from the
1443 analysis. Statistics were performed using one-sided t-test. *** p < 0.001.
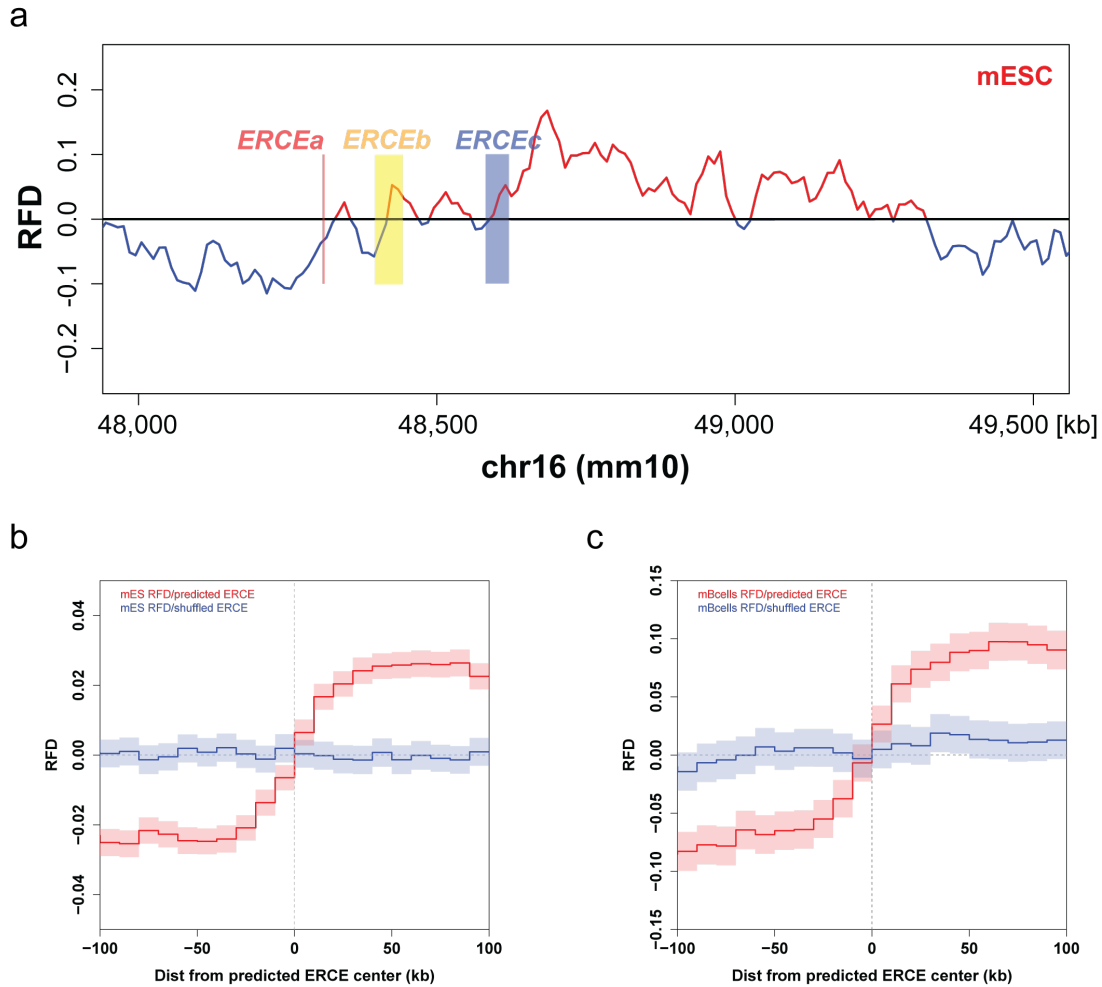
Supplementary Figure 9

Supplementary Figure 9: ORC/MCM is enriched in late replicating, H4K20me3-high non-genic AS and NRR windows.

a)-b) qPCR validation of H4K20me3 and H4K20me1 enrichment after ChIP at a) an H4K20me3 positive locus and b) an H4K20me1 positive locus. Representation in % input. Isotype IgG antibodies were used as control. c) Boxplot of H4K20me3 and H4K20me1 peak size (in bp) distribution. d) Average ORC/MCM relative read frequencies after input normalization at H4K20me1 peaks (> 1 kb). e) Kolmogorov–Smirnov statistics between the H4K20me3 relative read frequency distributions in each replication timing bin (shown in Fig. 6a): expressed gene versus intergenic (red circles) and silent gene (blue squares) regions as well as between intergenic regions in AS versus DS (green diamonds) and in URR versus NRR (yellow stars). Z_KS is normalized for sample size. The horizontal dashed lines correspond to p-value = 5%.

Supplementary Figure 10

Supplementary Figure 10: Early replication control elements (ERCEs) correlate with replication initiation.

a) mESC OK-seq RFD profile of the mouse Dppa2/4 locus (chr16: 48,000,000 – 49,500,000) with indicated ERCEs (ERCEa, ERCEb, ERCEc). ERCEa and ERCEc are located within ascending RFD segments (AS) in mESCs. ERCEb encompasses an entire AS. b) Mean mESC OK-seq RFD profile around the 1,835 mESC ERCEs predicted by (Sima et al., 2019) (red), and randomly shuffled ERCEs (blue). c) Mean mouse primary B-cell OK-seq RFD profile around the same mESC ERCE set (red) and randomly shuffled ERCEs (blue). mESC OK-seq data was obtained from (Petryk et al., 2018); mouse primary B cell OK-seq data was computed from Tubbs et al., 2018; mESC ERCEs were predicted by (Sima et al., 2019).

1471 **Supplementary Table 1**: Proportion of genes significantly depleted from

1472 ORC/MCM.

1473 A total of 1,941 genes met the criteria of transcriptional activity (TPM > 3), gene size

1474 larger 30 kb and no adjacent genes within 15 kb. We calculated the proportion of

1475 genes where the mean relative read frequency within the gene was significantly ($p <$

1476 0.05) reduced compared to the upstream region (excluding TSS +/- 3 kb).

1477

	Total genes	$p < 0.05$	%
Orc2		865	44.6
Orc3	1,941	850	43.8
Mcm3		1,460	75.2
Mcm7		1,131	58.3

1478

1479 **Supplementary Table 2:** Ratio of ChIP mean relative read frequencies in early vs.

1480 late RTDs.

1481 Calculated in 10 kb bins. All annotated genic regions were removed ± 10 kb.

	Mean relative read frequency ratio (early/late)
Orc2	1.40
Orc3	1.47
Mcm3	1.15
Mcm7	1.19

1482

1483

1484

1485 **Supplementary Table 3**: Characterization of H4K20me3 and H4K20me1 peaks

1486 determined by MACS2 broad peak calling.

1487

	Number of peaks (peaks > 1 kb)	Mean peak size [kb]	Peak size range [kb]
H4K20me3	16,852 (12,251)	3.5	0.2-105.1
H4K20me1	12,264 (6,277)	5.5	0.2-182.5

1488

1489