RESEARCH ARTICLE

Figures and figure supplements

Human ORC/MCM density is low in active genes and correlates with replication time but does not delimit initiation zones

Nina Kirstein *et al*

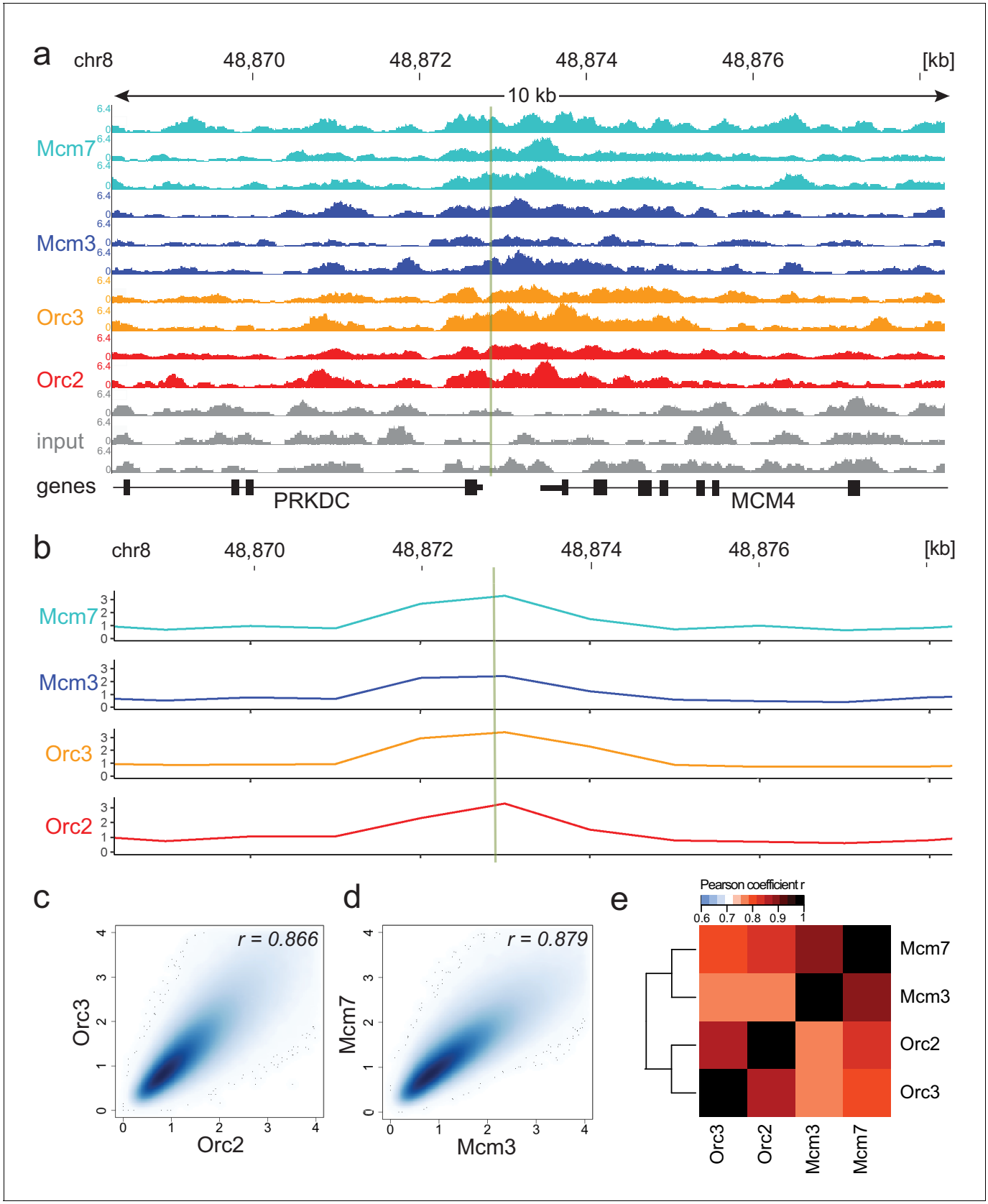

Figure 1. Moderate averaging represents a valid approach for origin recognition complex/minichromosome maintenance complex (ORC/MCM) chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis. (**a**) Sequencing profile visualization in UCSC Genome Browser (hg19) at the

Figure 1 continued on next page

Figure 1 continued

Mcm4/PRKDC origin after reads per genomic content normalization: two samples of Orc2 and Orc3, and three samples of Mcm3 and Mcm7, are plotted against the input in three replicates. The profiles are shown in a 10 kb window (chr8: 48,868,314–48,878,313); the mapped 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 determined by complete linkage hierarchical clustering using the correlation distance (d = 1 r). Refer to ***Figure 1—figure supplement 3*** for data representation without input division.

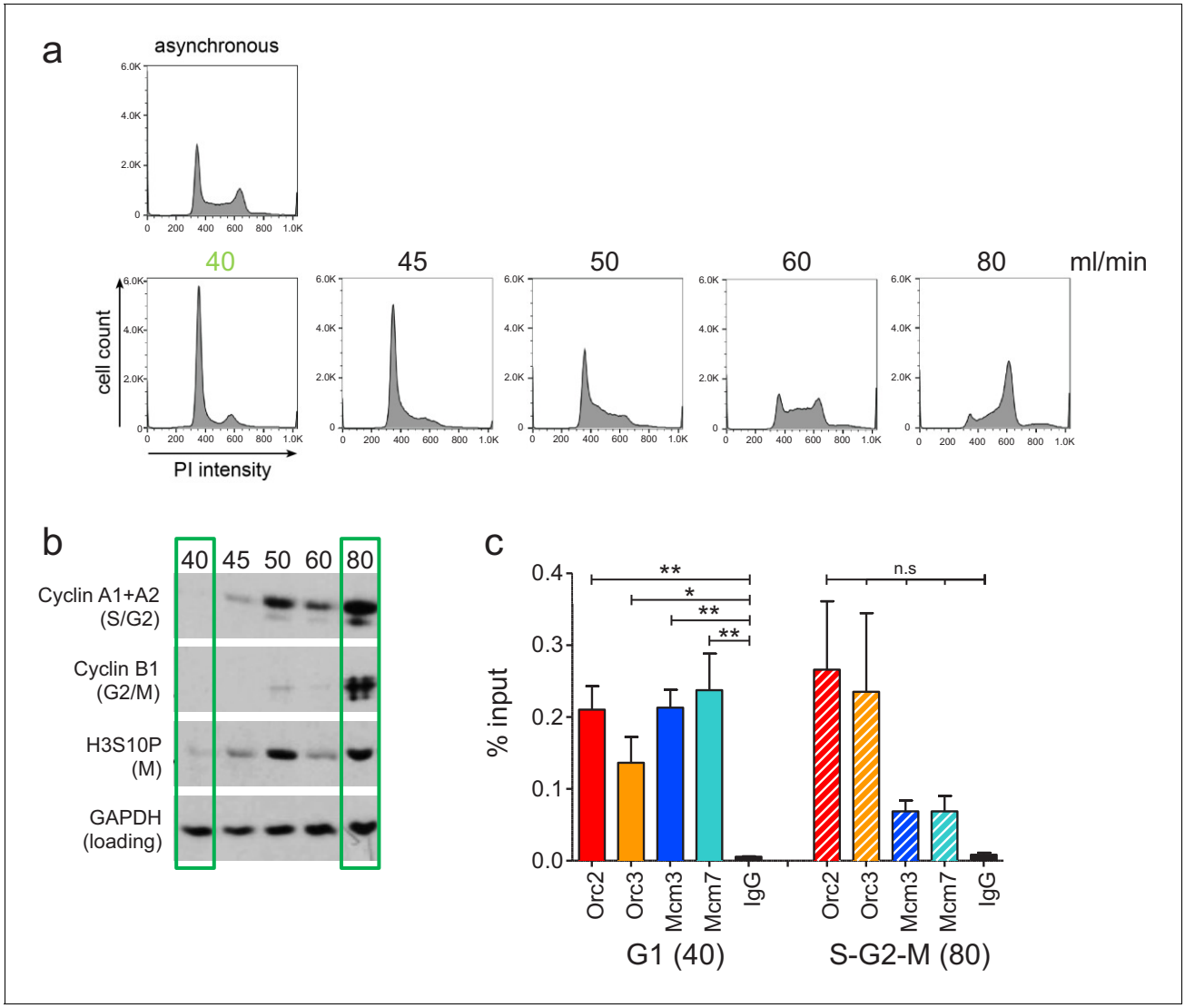

Figure 1—figure supplement 1. Experimental validation of cell cycle fractionation and origin recognition complex and minichromosome maintenance complex (ORC/MCM) chromatin immunoprecipitation followed by sequencing quality. (**a**) Example DNA content (propidium iodide) staining followed by FACS of logarithmically growing Raji (top) cells and after cell cycle fractionation by centrifugal elutriation (increasing counter flow rates indicated above each profile in ml/min). (**b**) Western blot analyses of the single fractions detecting cyclin A (S/G2), cyclin B (G2/M), H3S10P (M), and GAPDH. (**c**) qPCR validation of Orc2, Orc3, Mcm3, and Mcm7 enrichment at the Epstein–Barr virus latent origin *oriP* dyad symmetry element. Representation in % input. Isotype IgG was used as control. *p<0.05, **p<0.01.

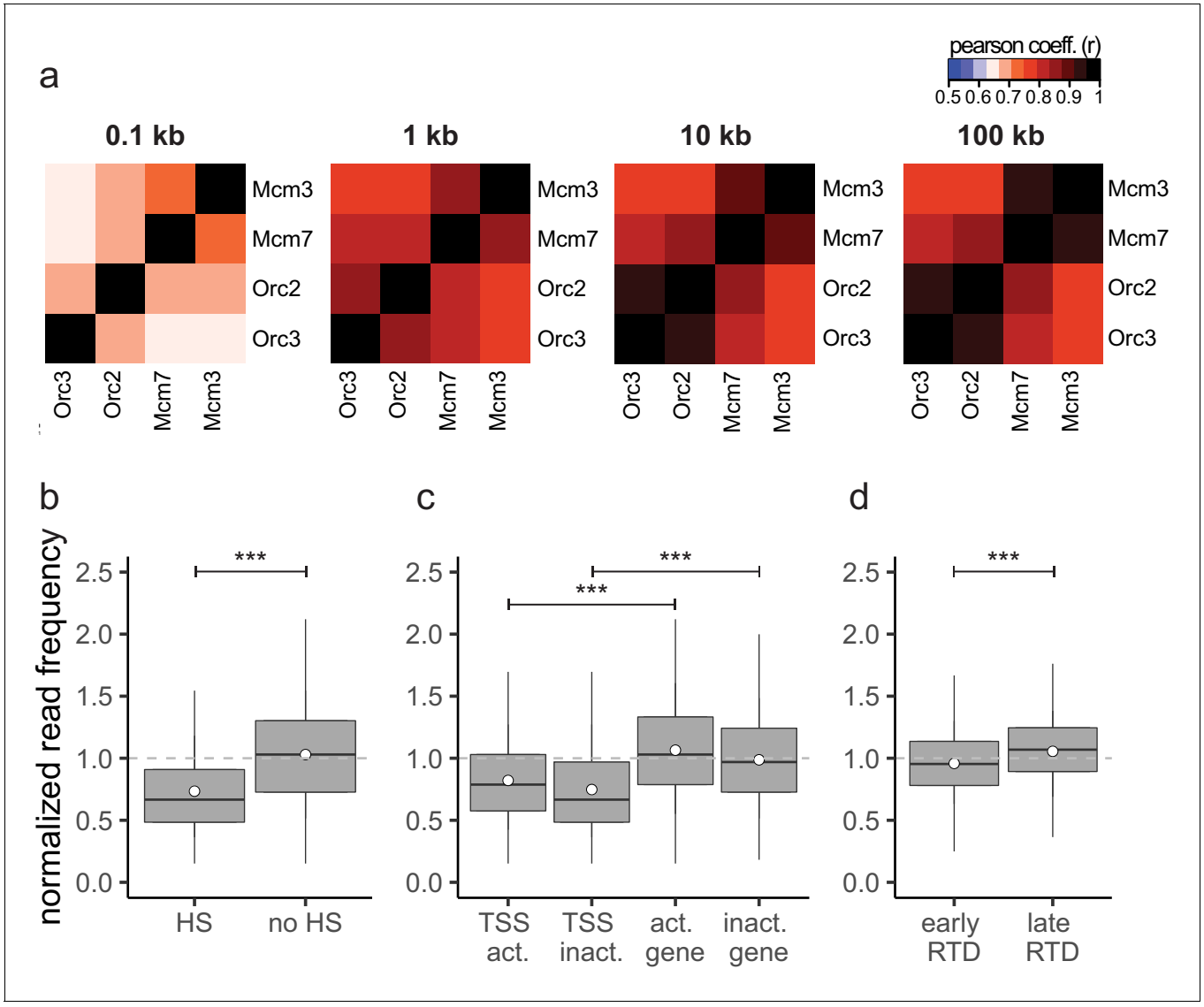

Figure 1—figure supplement 2. The input sequencing control is differentially represented in regions of biological function. (**a**) Heatmaps of Pearson correlation coefficients r between all chromatin immunoprecipitation relative read frequencies at different bin sizes: 100 bp, 1 kb, 10 kb, and 100 kb. (**b–d**) Boxplot of normalized read frequencies in relation to (**b**) DNase hypersensitivity. DNase hypersensitive (HS) clusters were obtained from 125 cell lines in ENCODE, only HS sites larger 1 kb were considered. (**c**) Transcription: transcription start sites and gene bodies of active (transcripts per kilobase per million [TPM] >3) and inactive (TPM <3) gene. (**d**) Early- or late-replication timing domains. The dashed gray horizontal line indicates read frequency 1.0 for orientation. Boxplot represents the mean (circle), median (thick line), first and third quartile (box), and first and ninth decile (whiskers) of the normalized read frequencies, without representing outliers. Statistics were performed using one-sided *t*-test. ***p<0.001.

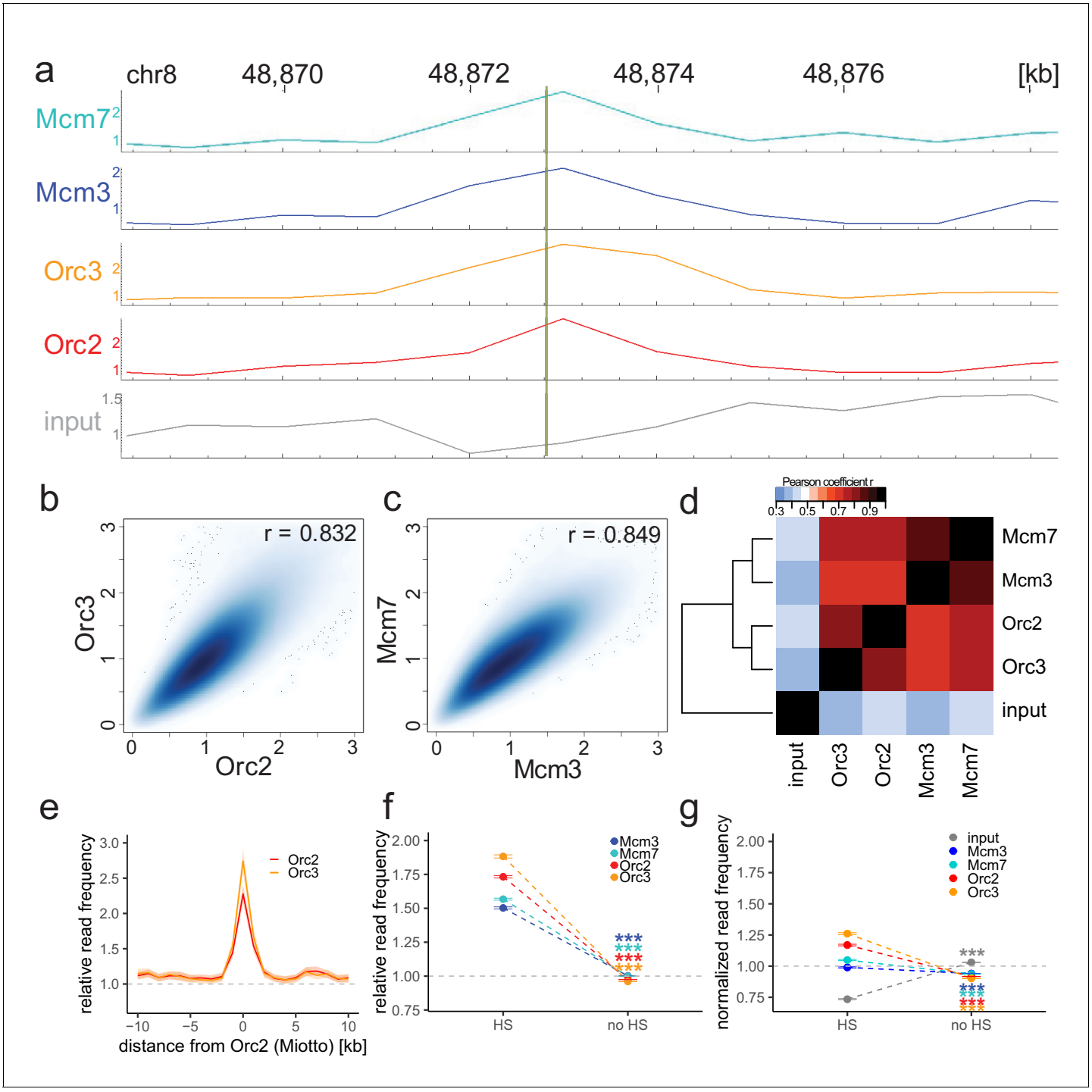

Figure 1—figure supplement 3. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) enrichments at the MCM4/PRKDC origin persists without input normalization. (**a**) The profile of ORC/MCM chromatin immunoprecipitation followed by sequencing (ChIP-seq) after 1 kb binning in the same 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 relative read frequencies at Orc2 peaks (>1 kb) retrieved from ***Miotto et al., 2016***. (**f, g**) ORC/MCM binding is confirmed at DNase hypersensitive (HS) sites. (**f**) Mean input-normalized ORC/MCM relative read frequencies (±2 × SEM) in relation to DNase hypersensitivity. (**g**) ORC/MCM-normalized read frequencies without input division (±2 × SEM) in relation to DNase hypersensitivity. Only HS sites larger than 1 kb were considered. Statistics were performed using one-sided *t*-test. ***p<0.001.

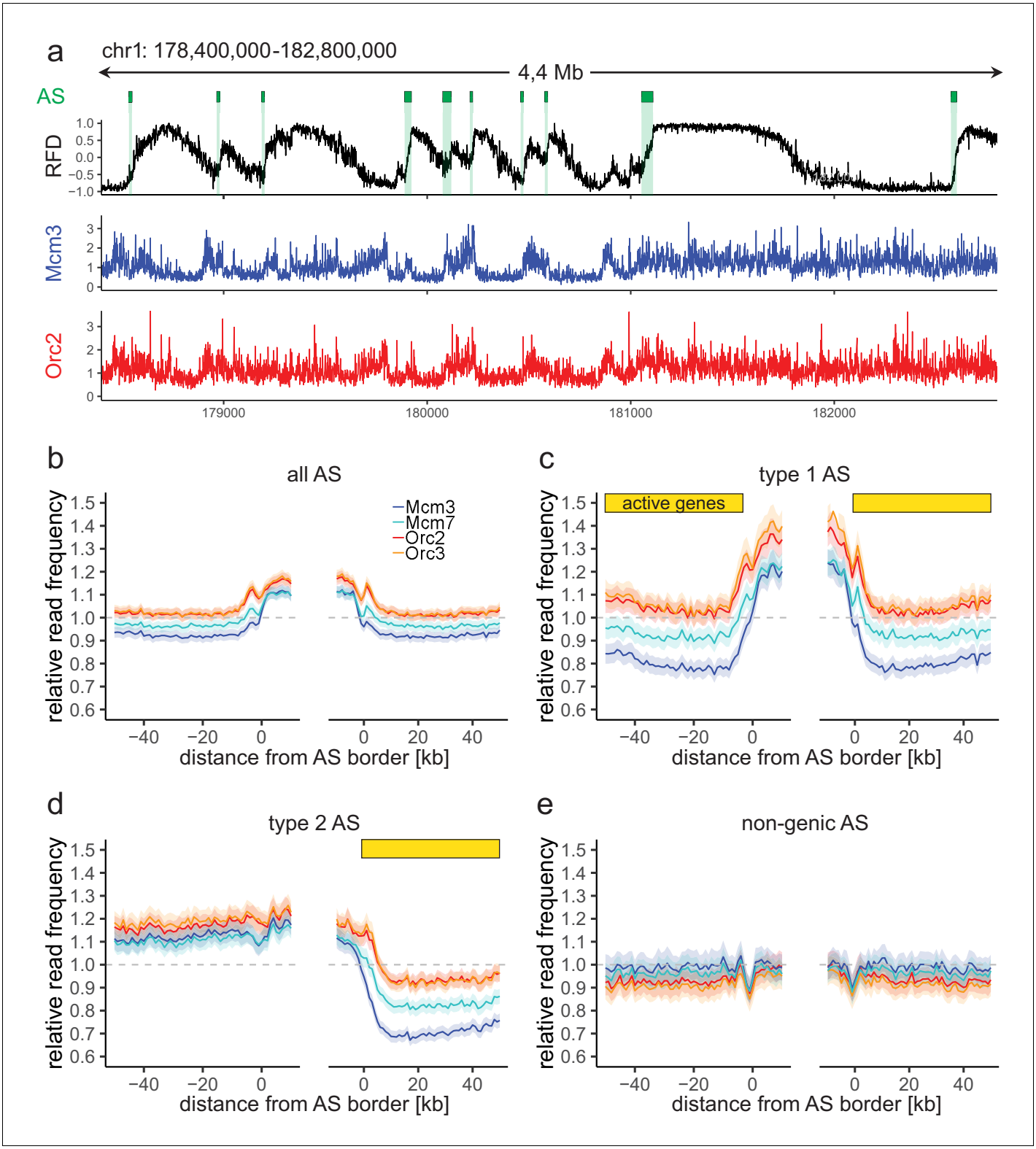

Figure 2. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) enrichment within ascending segments (ASs) depends on active transcription. (**a**) Top panel: example of an replication fork direction (RFD) profile on chr1: 178,400,000–182,800,000, covering 4 Mb. Detected ASs are labeled by green rectangles (irrespective of length and RFD shift). Middle and bottom panels: representative Mcm3 (blue) and Orc2 (red) chromatin immunoprecipitation followed by sequencing (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 (L > 20 kb and ΔRFD >0.5; n = 2957), (**c**) type 1 ASs with transcribed genes at

Figure 2 continued on next page

Figure 2 continued

both AS borders (n = 673), (**d**) type 2 ASs oriented with their AS border associated to 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 is shown ±2 × SEM (lighter shadows). The dashed grey horizontal line indicates relative read frequency 1.0 for reference. For type 1 and 2 ASs, yellow bars mark the AS borders associated to transcribed genes. Refer to ***Figure 2—figure supplement 2*** for analysis without input division.

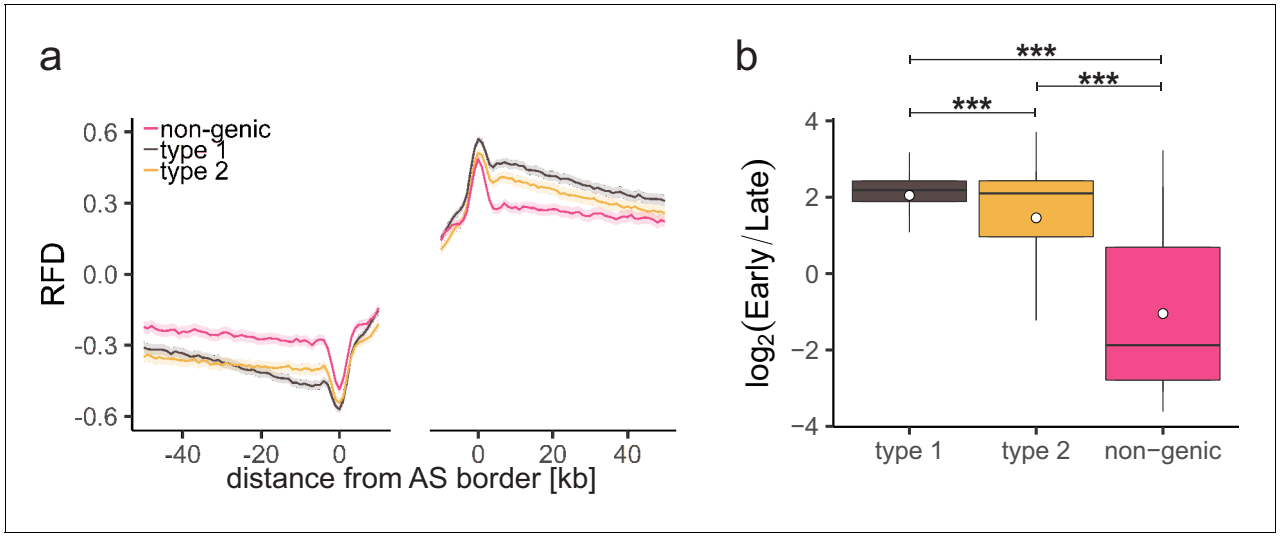

Figure 2—figure supplement 1. Characterization of different ascending segment (AS) types. (**a**) Average replication fork direction profiles (±2 × SEM; lighter shadows) for different AS types (***Table 1***) aligned at AS borders. (**b**) Replication timing ratio log_2(Early/Late) was assigned to type 1, type 2, and non-genic AS and represented as boxplot. Each boxplot represents the mean (circle), median (thick line), first and third quartile (box), and first and ninth decile (whiskers) of the relative read frequencies, without representing outliers. Statistics were performed by one-way ANOVA followed by Tukey's post-hoc test. ***p<0.001.

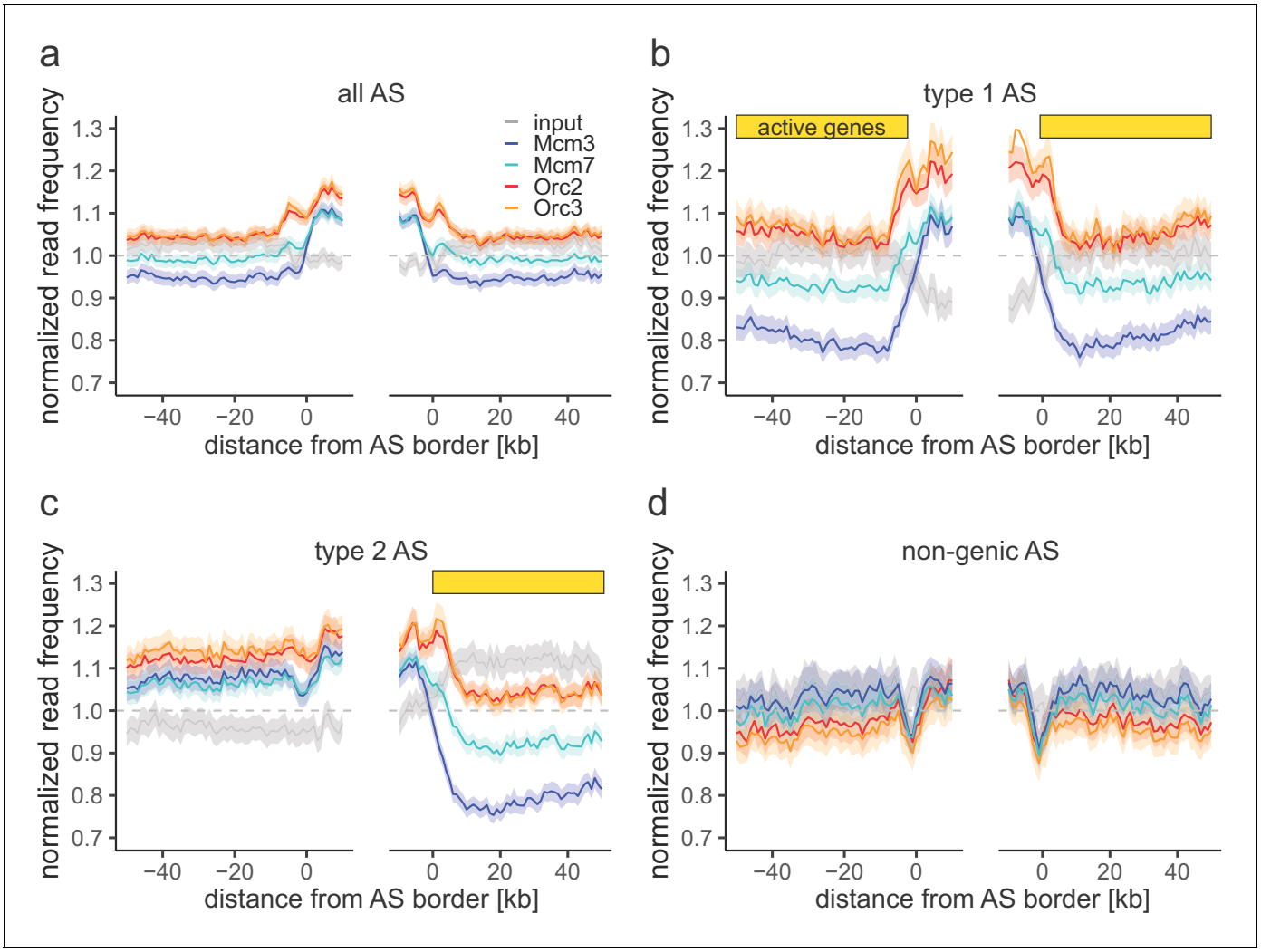

Figure 2—figure supplement 2. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) enrichments within ascending segments (ASs) without input normalization. (**a–d**) Average chromatin immunoprecipitation-normalized read frequencies of Orc2, Orc3, Mcm3, Mcm7, and input at AS borders of (**a**) all ASs (n = 2957), (**b**) type 1 ASs with transcribed genes at both ASs borders (n = 673), (**c**) type 2 ASs with transcribed genes oriented at their right ASs border (n = 1026), and (**d**) non-genic ASs in gene-deprived regions (n = 506). The mean of normalized read frequencies is shown ±2 × SEM (lighter shadows). The dashed grey horizontal line indicates read frequency 1.0 for reference. Yellow bars mark the AS borders associated to transcribed genes.

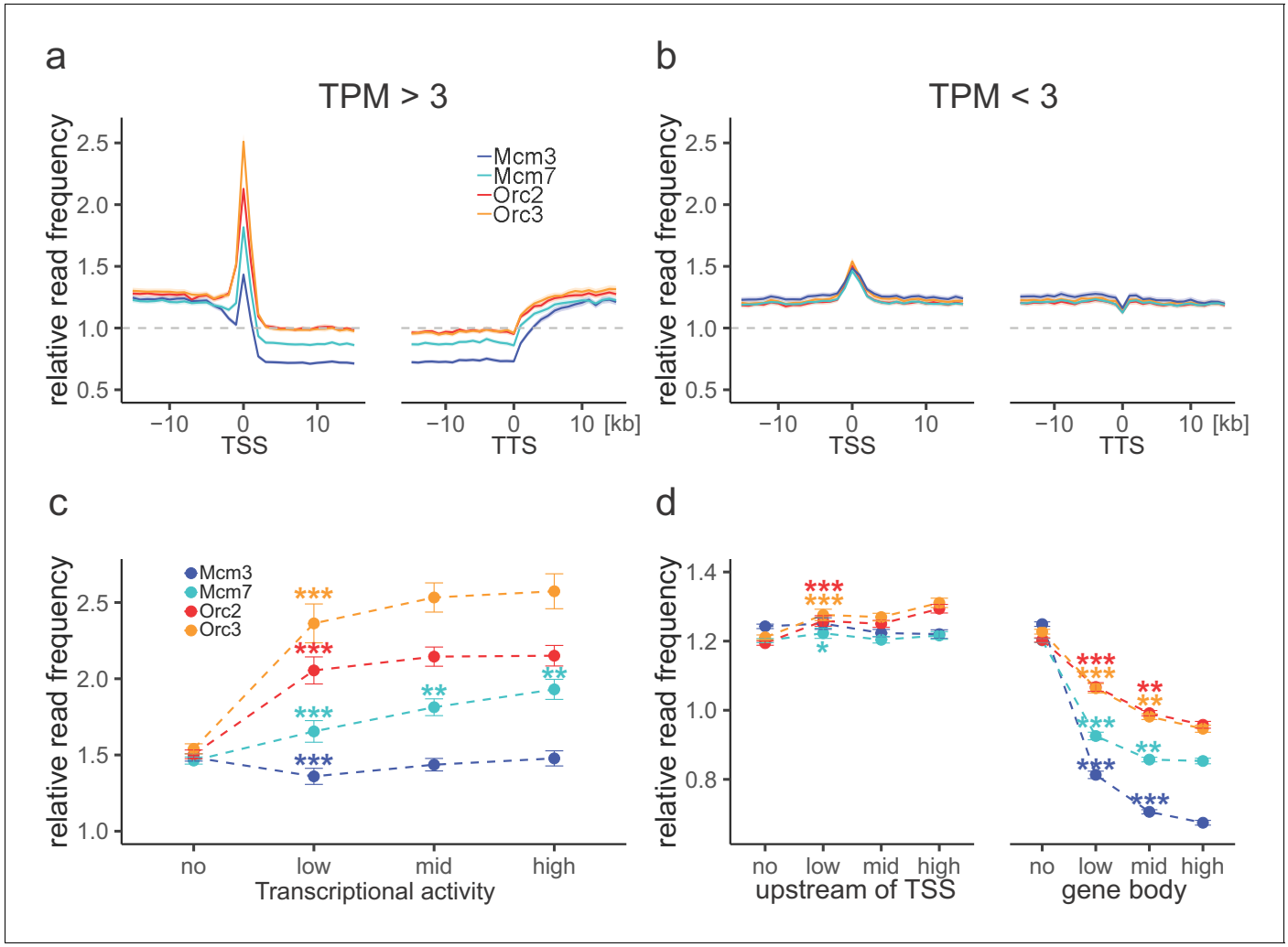

Figure 3. Origin recognition complex (ORC) is enriched at active transcription start sites (TSSs) while minichromosome maintenance complex (MCM) is depleted from actively transcribed genes. (**a, b**) ORC/MCM relative read frequencies around TSSs or transcriptional termination sites (TTSs) for (**a**) active genes (transcripts per kilobase per million [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 × 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 × SEM). (**d**) ORC/MCM relative read frequencies upstream of TSSs and within the gene body dependent on transcriptional activity (±2 × SEM; TSSs ± 3 kb removed from analysis). Transcriptional activity was classified as no (TPM <3), low (TPM 3–10), mid (TPM 10–40), and 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. Refer to ***Figure 3—figure supplement 1*** for analyses without input division.

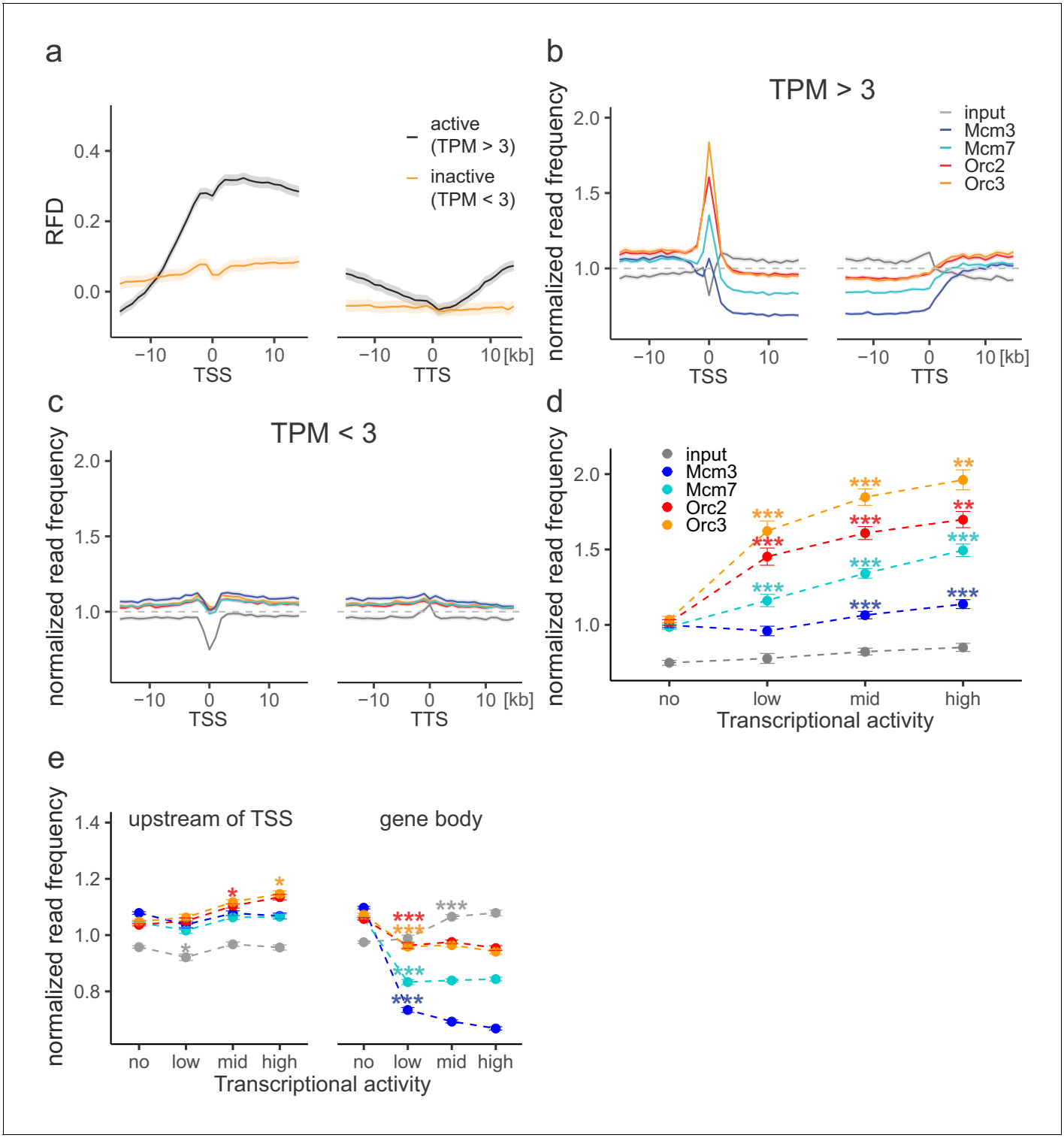

Figure 3—figure supplement 1. Replication fork direction (RFD) and origin recognition complex/minichromosome maintenance complex (ORC/MCM) profiles without input normalization at gene extremities. (**a**) Mean RFD profiles around transcription start sites (TSSs) or transcriptional termination sites (TTSs) of active genes (black) or inactive genes (yellow). Distances from TSSs or TTSs are indicated in kb. RFD means are shown ±2 × SEM (lighter shadows). (**b, c**) Normalized ORC/MCM/input read frequencies without input division around TSSs or TTSs for (**b**) active genes (TPM >3) and (**c**) 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 normalized read frequencies are shown ±2 × SEM (lighter shadows). The dashed grey horizontal line indicates read frequency 1.0 for reference. (**d**) Normalized ORC/MCM/input read frequencies at TSSs dependent on transcriptional activity (±2 × SEM). (**e**) Normalized ORC/MCM/input

Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

read frequencies upstream of TSSs and in the gene body dependent on transcriptional activity (±2 × SEM; TSSs ± 3 kb removed from analysis). Transcriptional activity was classified as no (TPM <3), low (TPM 3–10), mid (TPM 10–40), and 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.

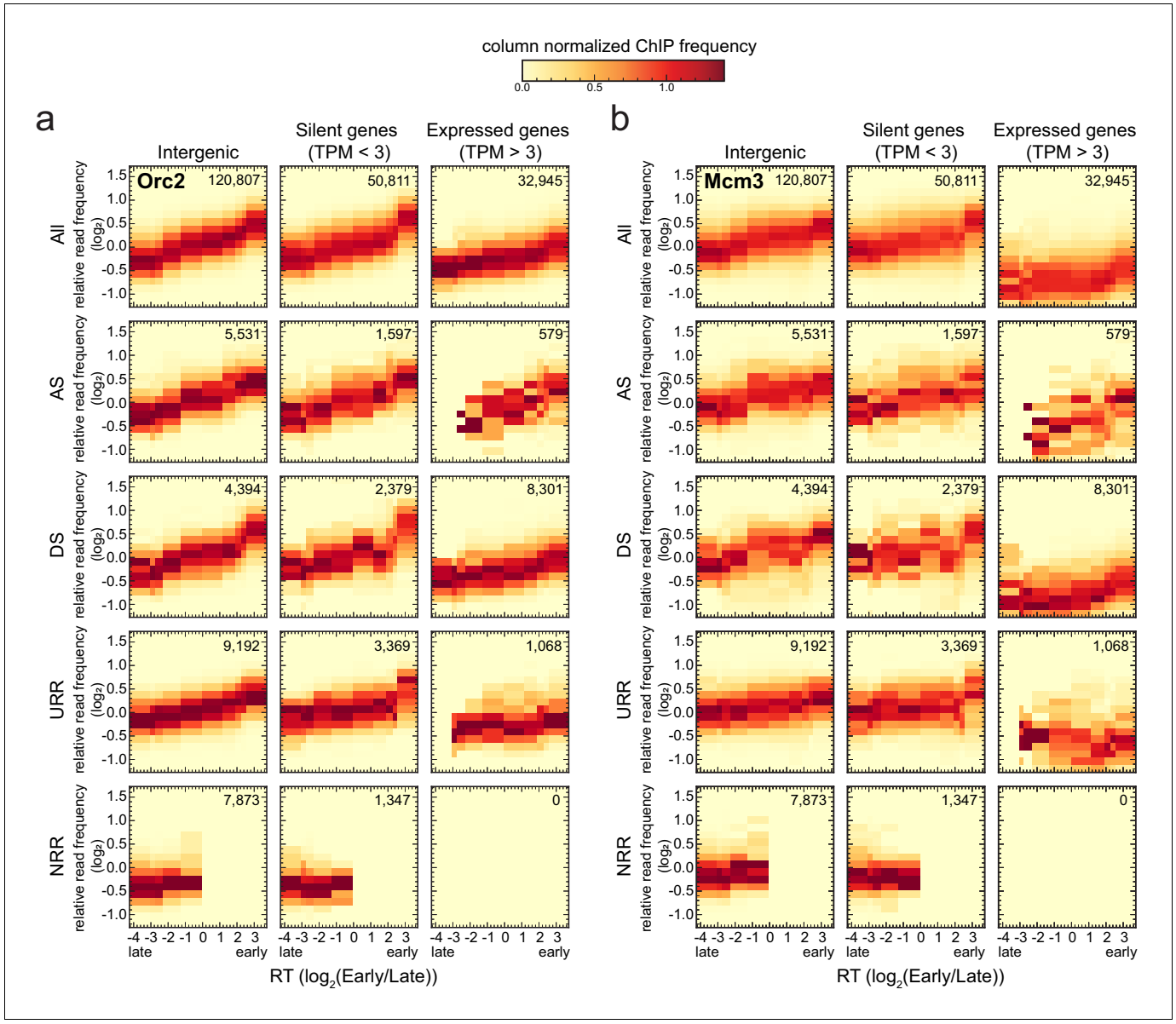

Figure 4. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) levels correlate with replication timing (RT) and transcriptional activity but are otherwise homogeneously distributed along the genome and uncorrelated to replication fork direction (RFD) patterns. (**a, b**) 3 × 5 panel of 2D histograms of Orc2 (**a**) and Mcm3 (**b**) chromatin immunoprecipitation (ChIP) relative read frequency versus RT (average $\log_2$(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis was performed in 10 kb bins. Histograms are normalized by column and represent the probability density functions of ChIP relative frequencies at a given replication timing. The color legend is indicated on top. 2D histograms are computed for intergenic regions (left column), silent genes (transcripts per kilobase per million TPM <3, middle column), and expressed genes (TPM >3, right column). Transcription start sites and transcriptional termination sites proximal regions were not considered (see Materials and methods). The rows show either all bins (top row) or restriction to ascending segment (AS) bins (predominant replication initiation, second row), descending segment (DS) bins (descending segment, predominant replication termination, third row), unidirectionally replicating region (URR) bins (unidirectional replication, no initiation, no termination, fourth row), and null RFD region (NRR) bins (null RFD regions, spatially random initiation and termination, bottom row). The number of bins per histogram is indicated in each panel. See ***Figure 4—figure supplement 1*** for equivalent Orc3 and Mcm7 analyses. Refer to ***Figure 4—figure supplement 2a*** for statistical comparisons.

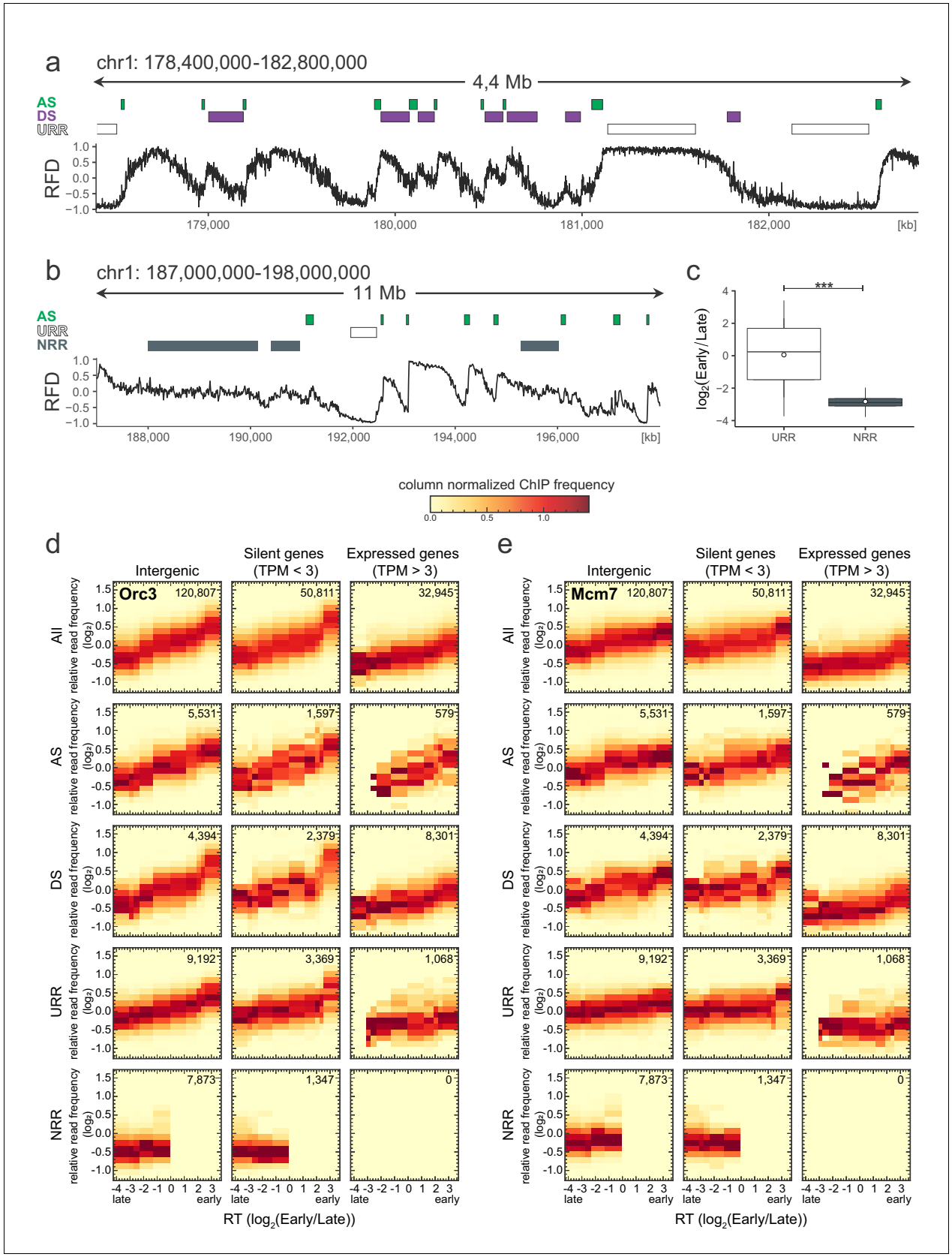

Figure 4—figure supplement 1. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) levels are correlated with replication timing (RT) and transcriptional activity but otherwise homogeneously distributed along the genome and uncorrelated to replication fork

Figure 4—figure supplement 1 continued on next page

Figure 4—figure supplement 1 continued

direction (RFD) patterns. (**a**) Same RFD profile example as ***Figure 2a*** (chr1: 178,400,000–182,800,000, covering 4 Mb) with indication of ascending segment (AS; green), descending segment (DS; purple), and unidirectionally replicating region (URR; white boxes) positions. (**b**) RFD profile example on chr1: 187,000,000–198,000,000, covering 11 MB, with ASs (green), URRs (white boxes), and null RFD regions (NRRs; dark gray) indicated. (**c**) RT ratio $\log_2$(Early/Late) was assigned to URRs and NRRs and represented as boxplot (mean [circle], median [thick line], first and third quartile [box], and first and ninth decile [whiskers], without representing outliers). Statistics were performed using one-sided *t*-test. ***p<0.001. (**d, e**) 3 × 5 panel of 2D histograms of Orc3 (**d**) and Mcm7 (**e**) chromatin immunoprecipitation (ChIP) frequency versus RT (average $\log_2$(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis was performed in 10 kb windows. Histograms are normalized by column and represent the probability density function of ChIP relative frequencies at given replication timings. The color legend is indicated on top. 2D histograms are computed for intergenic regions (left column), silent genes (transcripts per kilobase per million [TPM] <3, middle column), and expressed genes (TPM >3, right column). Transcription start sites and transcriptional termination sites proximal regions were not considered (see Materials 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 histogram is indicated in each panel. See ***Figure 4*** for equivalent Orc2 and Mcm3 data. Refer to ***Figure 4—figure supplement 2a*** for statistical comparisons.

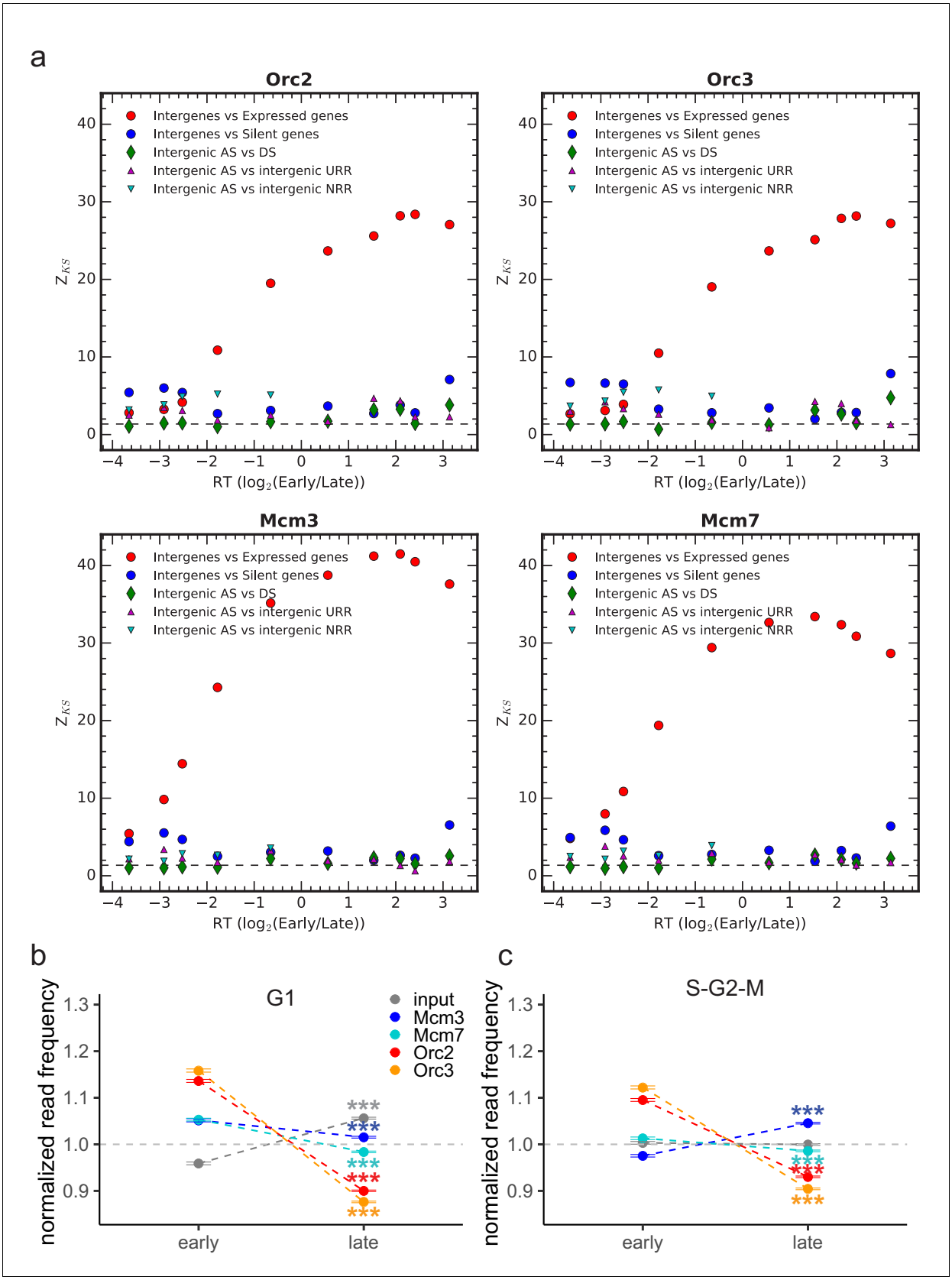

Figure 4—figure supplement 2. Kolmogorov–Smirnov statistics between the origin recognition complex/minichromosome maintenance complex (ORC/MCM). (**a**) Relative read frequency distributions in each replication timing bin (shown in ***Figure 4*** and ***Figure 4—figure***

Figure 4—figure supplement 2 continued on next page

Figure 4—figure supplement 2 continued

supplement 1d, e) in intergenic versus expressed gene regions (red circles), intergenic versus silent gene regions (blue circles) and between intergenic regions in ascending segments (AS) versus descending segments (DS; green diamonds), AS versus unidirectionally replicating regions (URR; magenta triangles pointing up), and AS versus null RFD regions (NRR; cyan triangles pointing down). Z_KS is normalized for sample size. The horizontal dashed lines correspond to p-value=5%. (**b**) Normalized ORC/MCM/input (G1) read frequencies without input division (±2 × SEM) in early or late replication timing domains (RTDs). Early RTDs were defined as $\log_2$(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. (**c**) ORC/MCM input normalized read frequencies (±2 × SEM) obtained from S-G2-M-phased chromatin in early or late RTDs using the same settings as in (**b**). Statistics were performed using one-sided *t*-test. ***p<0.001.

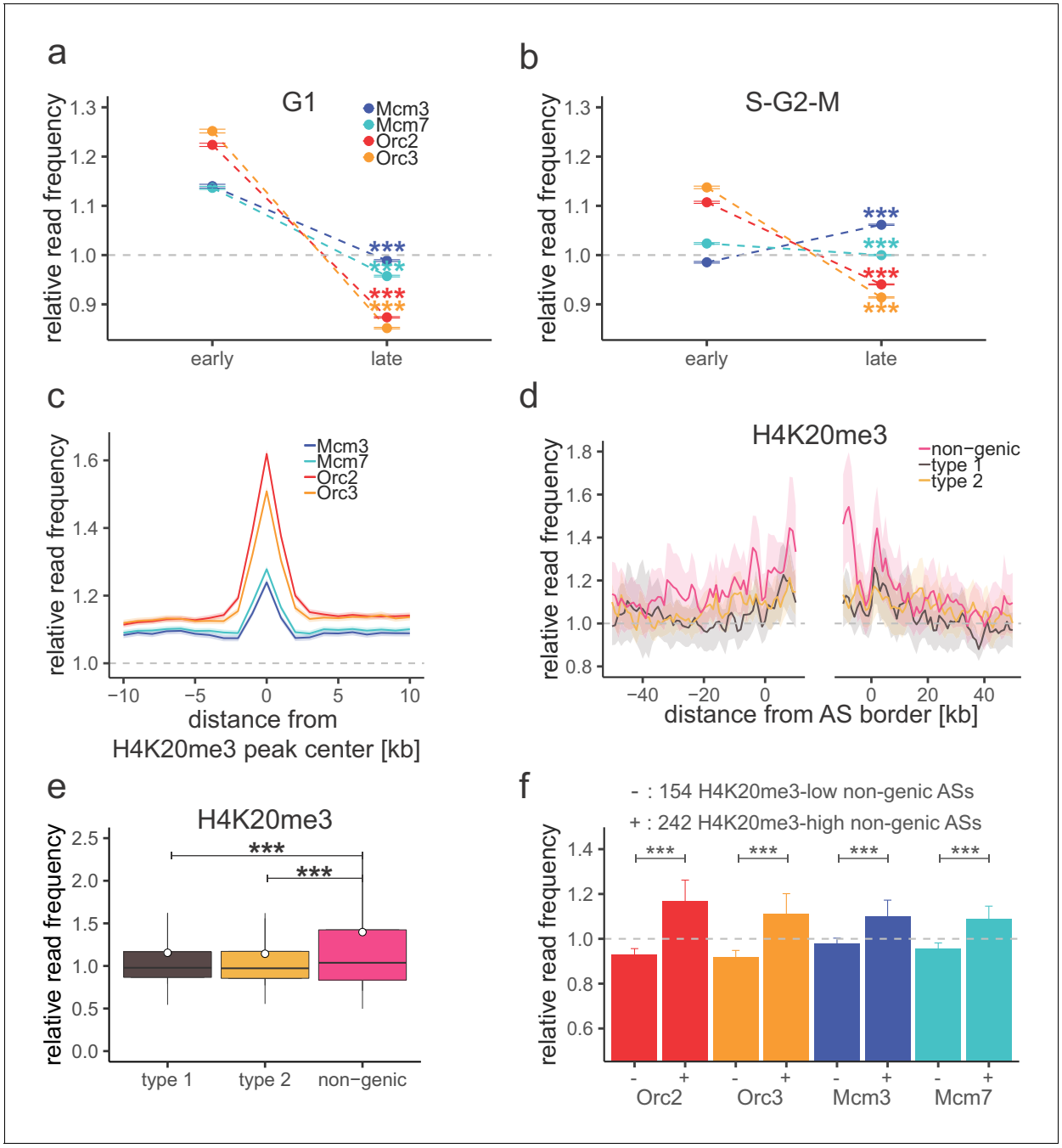

Figure 5. H4K20me3 selectively marks a subset of late-replicating non-genic ascending segments (ASs). (**a**) Origin recognition complex/minichromosome maintenance complex (ORC/MCM) G1 chromatin relative read frequencies (±2 × standard error of the mean [SEM]) in early or late replication timing domains (RTDs). Early RTDs were defined as $\log_2$(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**) ORC/MCM relative read frequencies (±2 × SEM) obtained from S-G2-M chromatin in early or late RTDs using the same settings as in (**a**). (**c**) Average ORC/MCM relative read frequencies at H4K20me3 peaks (>1 kb). (**d**) 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 × SEM (lighter shadows). (**e**) Boxplot representation of H4K20me3 relative read frequencies within the different AS types. Boxplot represents the mean (circle), median (thick line), first and third quartile (box), and first and ninth 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. (**f**) Histogram representation of mean ±2 × SEM of ORC/MCM relative read frequencies in G1 at 242 H4K20me3-low non-genic ASs and 154 H4K20me3-high non-genic ASs. Statistics were performed using one-sided *t*-test. ***p<0.001. Refer to ***Figure 5—figure supplement 1*** for validation of H4K20me3 chromatin immunoprecipitation.

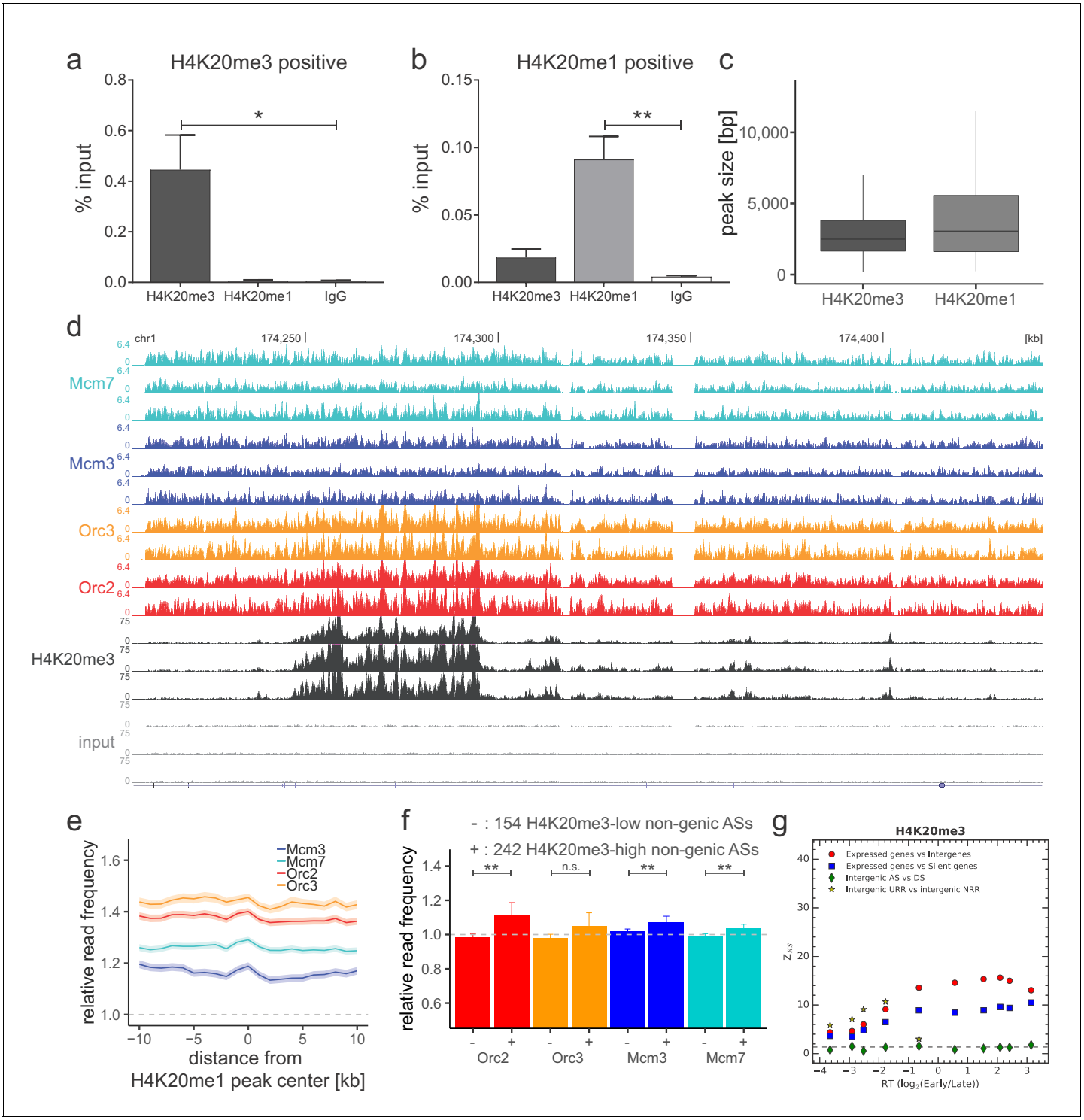

Figure 5—figure supplement 1. Origin recognition complex/minichromosome maintenance complex (ORC/MCM) is enriched in late-replicating, H4K20me3-high non-genic ascending segment (AS) and null RFD region (NRR) windows. (**a, b**) qPCR validation of H4K20me3 and H4K20me1 enrichment after chromatin immunoprecipitation at (**a**) an H4K20me3 positive locus and (**b**) an H4K20me1 positive locus. Representation in % input. Isotype IgG antibodies were used as control. *p<0.05, **p<0.01. (**c**) Boxplot of H4K20me3 and H4K20me1 peak size (in bp) distribution. (**d**) Sequencing profile visualization in UCSC Genome Browser (hg19) at chr1:174,205,071–174,444,322 after reads per genomic content normalization: three samples of H4K20me3 are plotted against the three input samples. Two samples of Orc2 and Orc3, and three samples of Mcm3 and Mcm7, are also visualized for comparison. (**e**) Average ORC/MCM relative read frequencies after input normalization at H4K20me1 peaks (>1 kb). (**f**) Histogram representation of mean ±2 × SEM of ORC/MCM relative read frequencies derived from S-G2-M chromatin at 242 H4K20me3-low non-genic ASs and 154 H4K20me3-high

Figure 5—figure supplement 1 continued on next page

Figure 5—figure supplement 1 continued

non-genic ASs. Statistics were performed using one-sided *t*-test. **p<0.01. (**g**) Kolmogorov–Smirnov statistics between the H4K20me3 relative read frequency distributions in each replication timing bin (shown in ***Figure 6a***): expressed gene versus intergenic (red circles) and silent gene (blue squares) regions as well as between intergenic regions in AS versus descending segments (DS; green diamonds) and unidirectionally replicating region (URR) versus null replication fork directionality region (NRR) (yellow stars). Z_KS is normalized for sample size. The horizontal dashed lines correspond to p-value=5%.

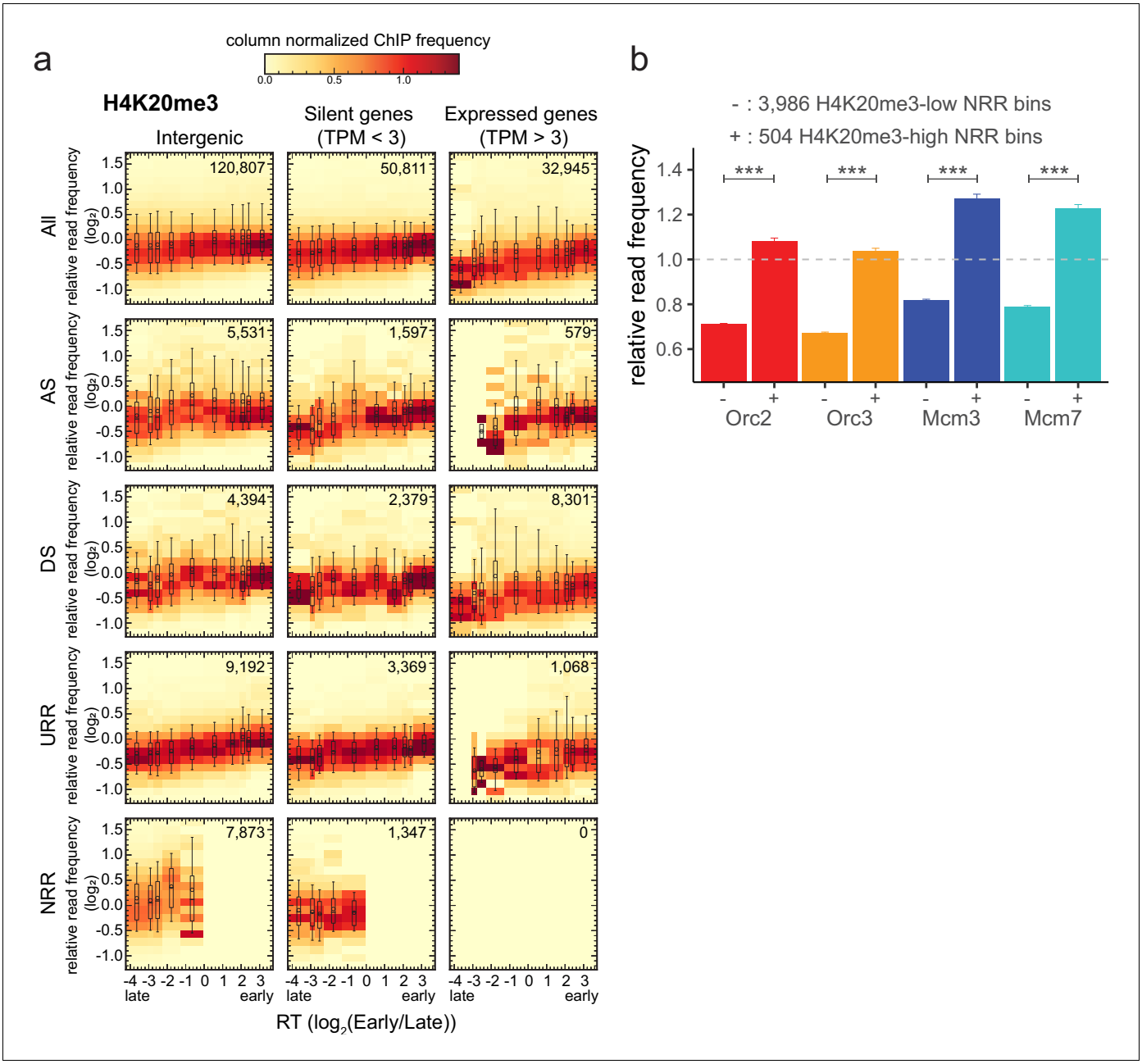

Figure 6. H4K20me3 is enriched in late-replicating null RFD regions (NRR). (**a**) 3 × 5 panel of 2D histograms of H4K20me3 chromatin immunoprecipitation relative read frequencies versus replication timing (RT) (average $\log_2$(Early/Late) over 100 kb binned according to the decile of RT distribution). The analysis was performed in 10 kb bins. Histograms are normalized by column and displayed for different bin categories (columns: intergenic regions, silent genes, expressed genes; rows: all bins, ascending segment [AS], descending segment [DS], unidirectionally replicating region [URR], null replication fork directionality region [NRR] bins) as for origin recognition complex/minichromosome maintenance complex (ORC/MCM) in ***Figure 4***. The number of bins per histogram is indicated in each panel. Superimposed boxplots represent the mean (circle), median (thick line), first and third quartile (box), and first and ninth decile (whiskers) of the relative read frequencies in each timing bins. Refer to ***Figure 5—figure supplement 1g*** for statistical comparisons. (**b**) Histogram representation of mean ±2 × SEM of ORC/MCM relative read frequencies at 3986 H4K20me3-low NRR 10 kb bins and 504 H4K20me3-high NRR 10 kb bins. Statistics were performed using one-sided *t*-test. ***p<0.001.

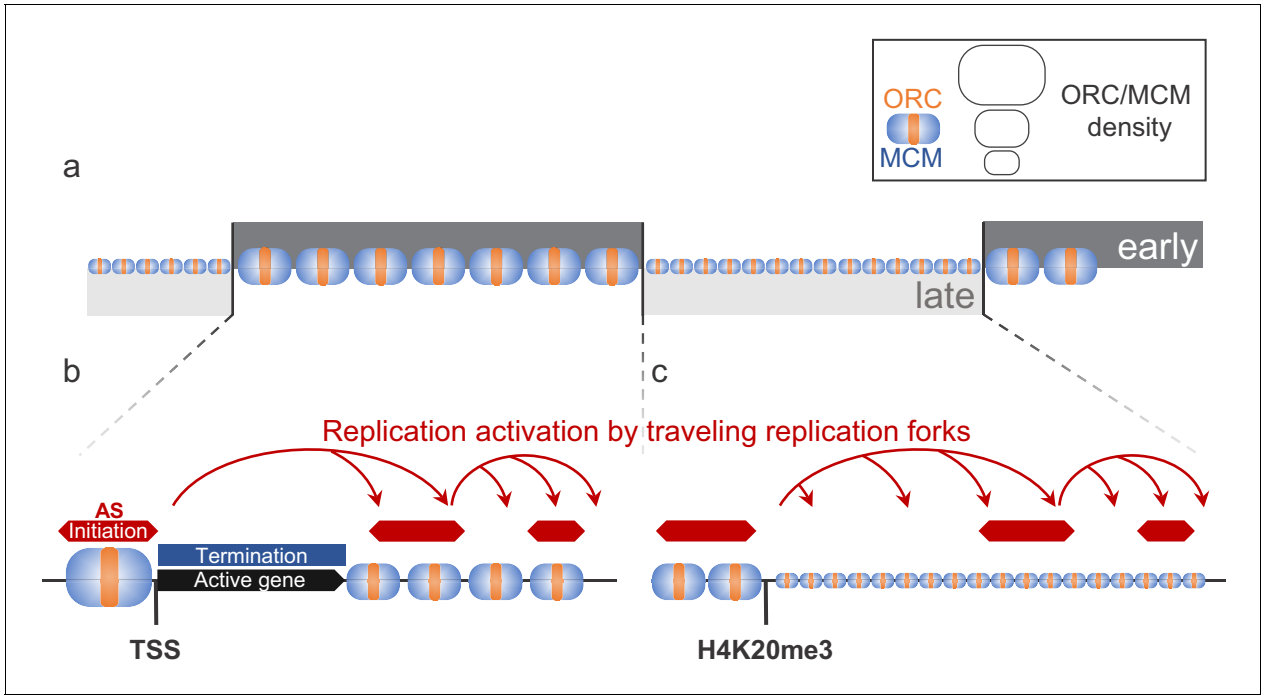

Figure 7. Model for replication organization in higher eukaryotes. (**a**) Replication is organized in large segments of constant replication timing (early replication timing domain [RTD], dark grey; late RTD, light grey) (***Marchal et al., 2019***). While we observe the ubiquitous presence of the origin recognition complex (ORC; orange) and the minichromosome maintenance complex (MCM; blue) throughout the genome, the enrichment levels of ORC/MCM were higher in early RTDs compared to late RTDs. (**b**) Early RTDs are among other characterized by active transcription. ORC/MCM are locally highly enriched at active transcription start site (TSS). However, actively transcribed gene bodies (black) are deprived of ORC/MCM, often correlating with replication termination (blue). Besides TSSs, we find ORC/MCM stochastically distributed along intergenic regions. We hypothesize that traveling replication forks trigger activation of replication in a cascade (red arrows). (**c**) In gene-deprived and transcriptionally silent late-replicating heterochromatin, we detected homogeneous ORC/MCM distribution at generally lower levels. H4K20me3 is present at late-replicating non-genic ascending segments (ASs) and null RFD regions (NRRs) and leads to enhanced ORC/MCM binding, linking this histone mark to replication activation in heterochromatin.

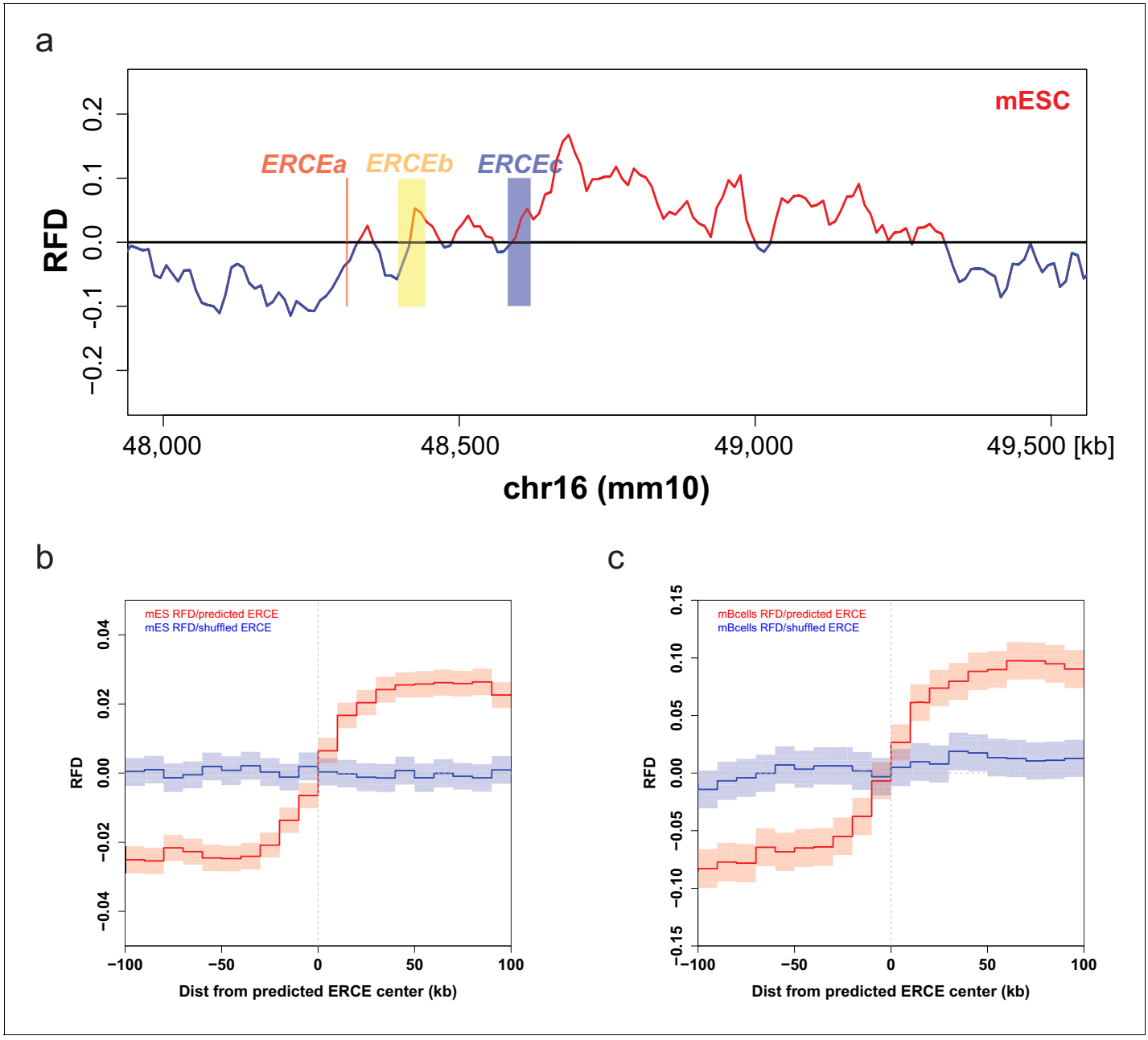

Figure 7—figure supplement 1. Early replication control elements (ERCEs) correlate with replication initiation. (**a**) Mouse embryonic stem cells (mESC) strand-oriented Okazaki-segment sequencing (OK-seq) replication fork direction (RFD) profile of the mouse Dppa2/4 locus (chr16: 48,000,000–49,500,000) with indicated ERCEs (ERCEa, ERCEb, ERCEc). ERCEa and ERCEc are located within ascending RFD segments (ascending segment [AS]) in mESCs. ERCEb encompasses an entire AS. (**b**) Mean mESC OK-seq RFD profile around the 1835 mESC ERCEs (red) predicted by ***Sima et al., 2019*** 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***.