

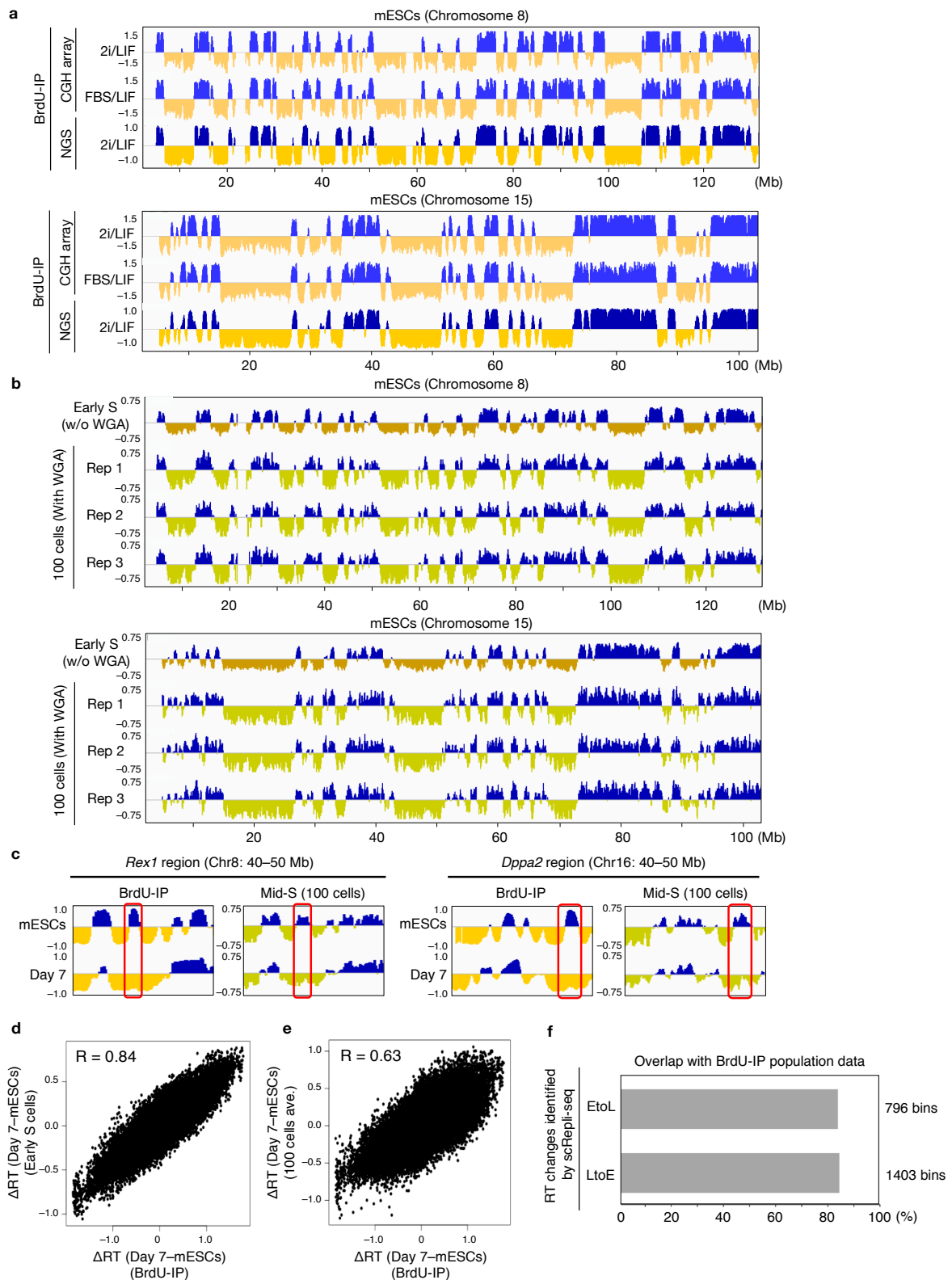
In the format provided by the authors and unedited.

Genome-wide stability of the DNA replication program in single mammalian cells

Saori Takahashi ^{1,2,6}, Hisashi Miura ^{1,2,6}, Takahiro Shibata^{3,4,6}, Koji Nagao ⁵, Katsuzumi Okumura⁴, Masato Ogata³, Chikashi Obuse⁵, Shin-ichiro Takebayashi ^{3*} and Ichiro Hiratani ^{1,2*}

¹Laboratory for Developmental Epigenetics, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan. ²RIKEN Center for Developmental Biology (CDB), Kobe, Japan. ³Department of Biochemistry and Proteomics, Graduate School of Medicine, Mie University, Tsu, Japan. ⁴Laboratory of Molecular & Cellular Biology, Graduate School of Bioresources, Mie University, Tsu, Japan. ⁵Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Japan. ⁶These authors contributed equally: Saori Takahashi, Hisashi Miura, Takahiro Shibata. *e-mail: stake@doc.medic.mie-u.ac.jp; ichiro.hiratani@riken.jp

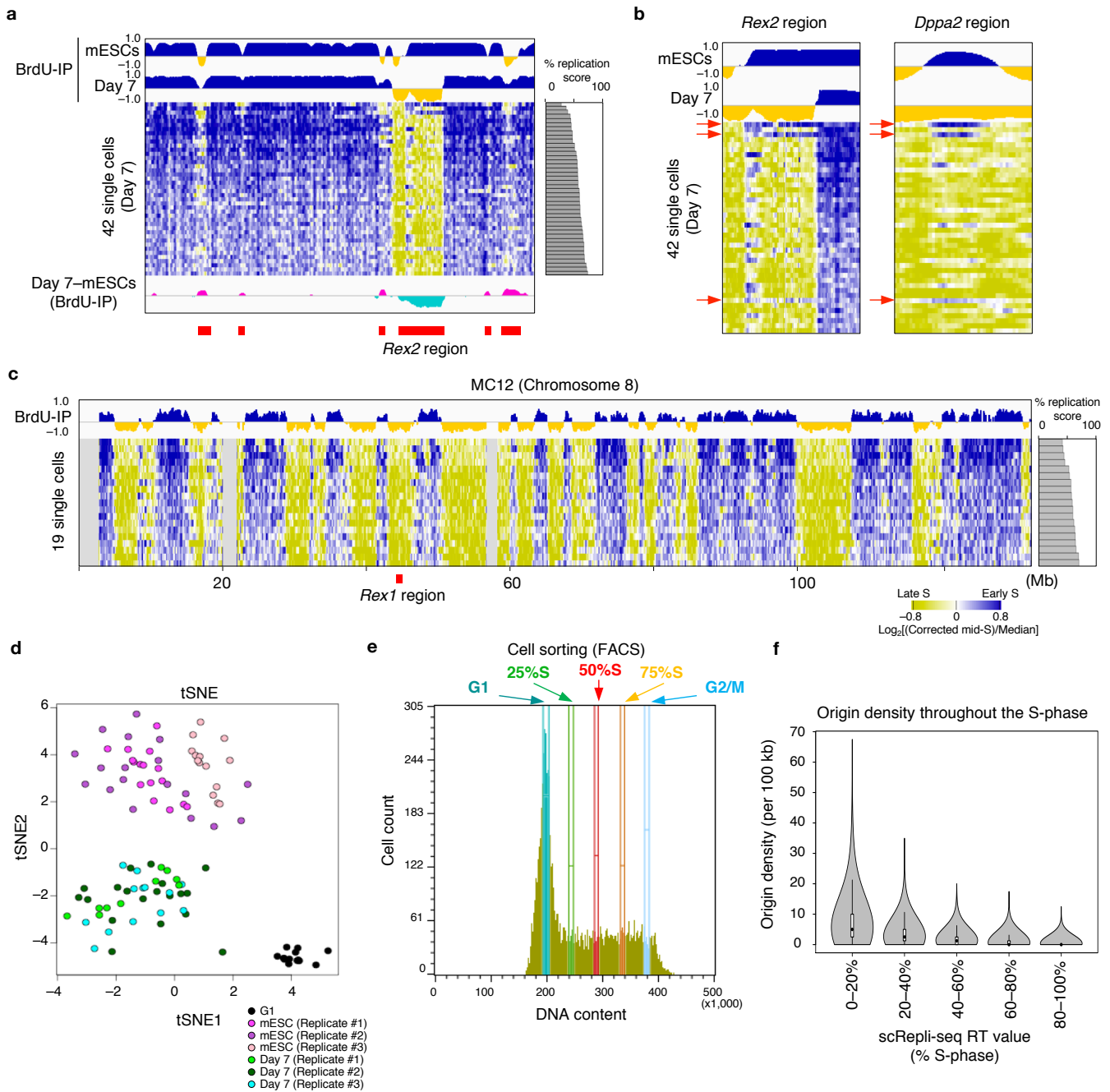
Supplementary Figure 1



Supplementary Figure 1 | A copy number method generates replication timing profiles similar to those of previous methods

(a) BrdU-IP population RT analysis of mESCs cultured under 2i/LIF or FBS/LIF conditions, using CGH microarray or NGS, generated comparable profiles. (b) A copy number-based RT analysis of 200,000 mESCs in the early half of S-phase without WGA and three 100 mid-S mESCs with WGA generated comparable profiles. Shown are three replicates of 100 cells along with the early-S population. (c) *Rex1* and *Dppa2* regions, which show EtoL changes upon mESC differentiation based on BrdU-IP analysis, also showed the expected EtoL changes by copy number RT analysis of 100 mid-S cells. (d, e) Scatter plots of ΔRT (Day 7–mESCs) values of BrdU-IP vs. Early S Repli-seq (d) and BrdU-IP vs. 100-cells Repli-seq. For the 100-cell Repli-seq data, we used the average of three independent experiments. The analysis was done using 40-kb bins throughout the genome. (f) More than 80% of genomic bins that changed RT (EtoL or LtoE) significantly during mESC differentiation by single-cell Repli-seq also changed RT significantly by the BrdU-IP population data. The analysis was done using 80-kb bins.

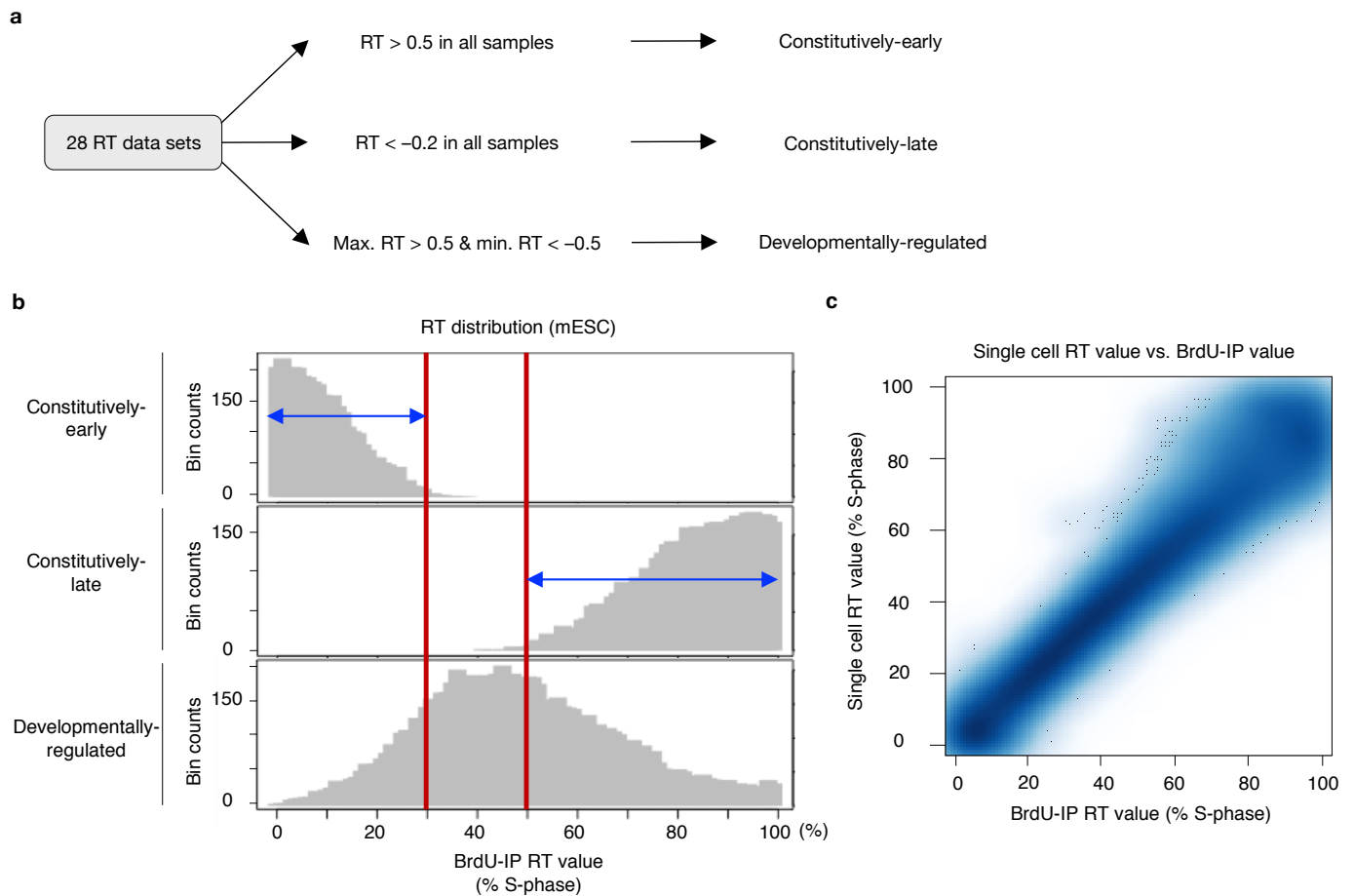
Supplementary Figure 2



Supplementary Figure 2 | scRepli-seq profile heterogeneity, single mESC sorting from S-phase windows other than mid S, and RT vs. origin density

(a) Magnified scRepli-seq view surrounding the *Rex2* locus on day 7 (see also Fig. 2a). The scRepli-seq raw data heatmap was sorted by the % replication scores of each cell. The bottom red lines show regions of significant RT difference between mESCs and day-7 cells by BrdU-IP. (b) Further magnified scRepli-seq views surrounding the *Rex2* and the *Dppa2* locus. Two out of 3 cells (red arrows) that exhibited unexpected early replication of *Rex2* also exhibited early replication of the *Dppa2* locus. (c) A scRepli-seq heatmap of MC12 embryonic carcinoma cells, which exhibit some similarity to epiblast cells. MC12 cells show late replication of EtoL-switching loci such as *Rex1* by the BrdU-IP assay. With MC12, 19 out of 19 cells exhibit late replication of *Rex1*, providing an example that replication timing of EtoL domains are not necessarily heterogeneous in differentiated cells. Data sets were sorted according to the % replication score of each cell. (d) tSNE analysis showing the distinct distribution of mESCs (n=48), day-7 differentiated cells (n=42), and G1 cells (n=15) using autosomal data. The plot is identical to the one in Fig. 2c but here the individual biological replicates are color-coded, allowing us to confirm data reproducibility. (e) A cell-cycle profile of cells stained with propidium iodide during FACS analysis is shown, along with the various gates used to collect cells throughout the S-phase shown in Fig. 3h. (f) Origin density (origin number per 100 kb) decrease with time during S-phase. A replication origin data set of mESCs generated by Cayrou et al.⁶¹ was used, while RT values were calculated from 131 scRepli-seq data from mESCs. The violin plot format is the same as in Fig. 4j.

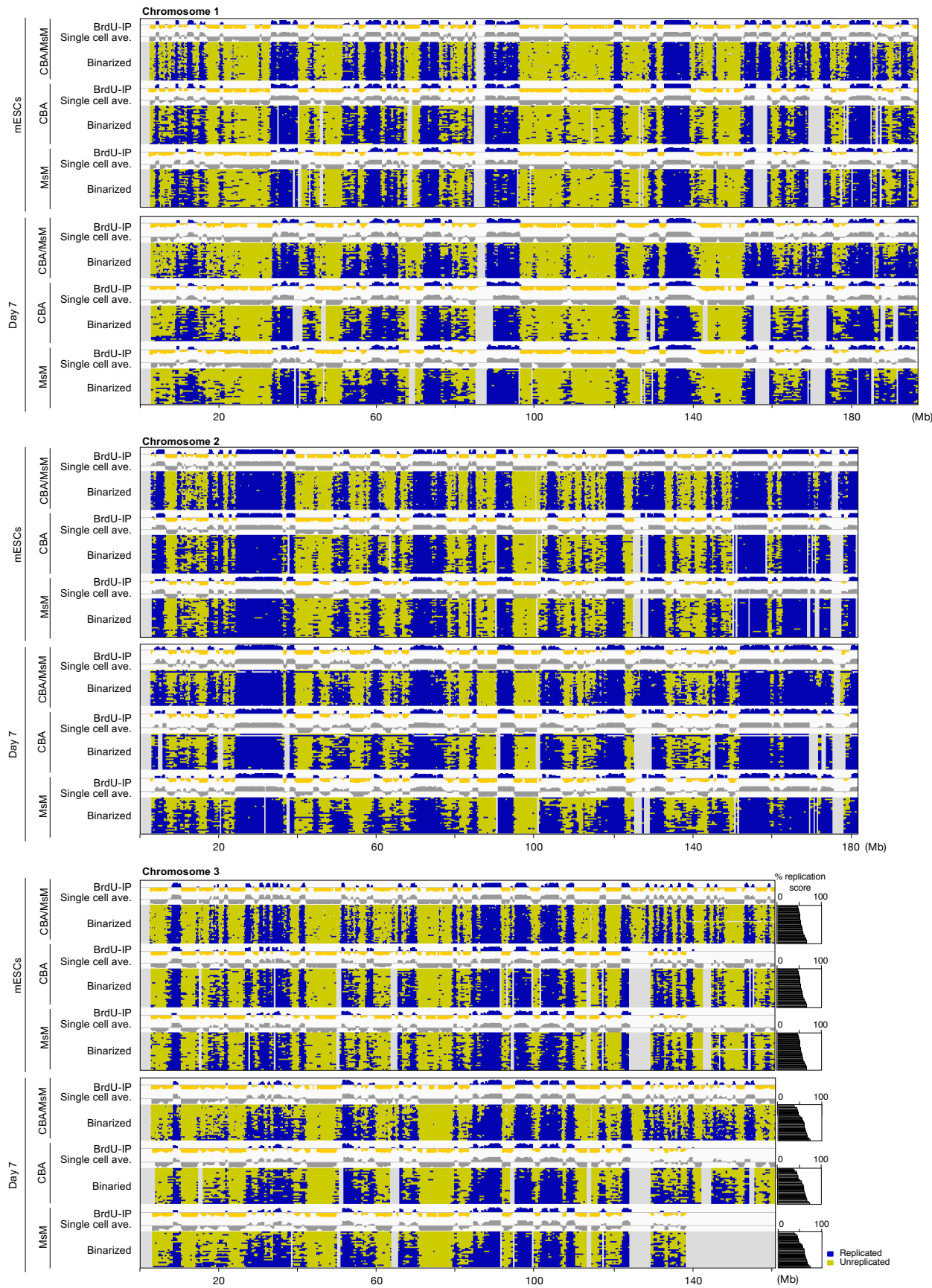
Supplementary Figure 3



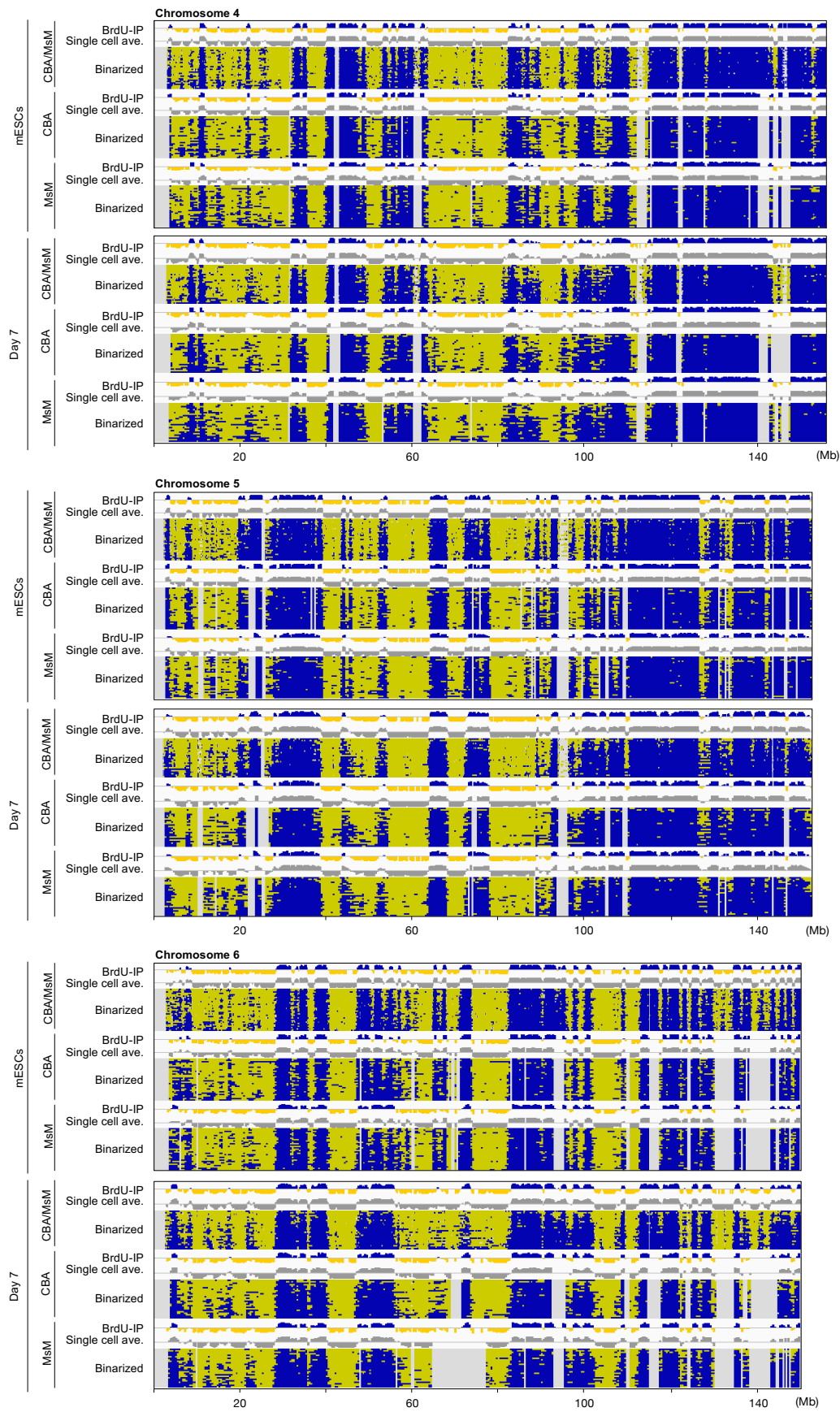
Supplementary Figure 3 | Definitions of constitutively-early, constitutively-late, and developmentally-regulated sequences

(a) BrdU-IP population RT data from 28 types of cells were analyzed as shown in the flowchart to define constitutively-early, constitutively-late, and developmentally-regulated regions in a manner similar to that described by Dileep et al²³. (b) RT distribution of each sequence categories based on BrdU-IP population RT data. For constitutively-early vs. developmentally-regulated and constitutively-late vs. developmentally-regulated comparisons, sequences with BrdU-IP population RT values within 0–30% and 50–100% S-phase ranges, respectively, were analyzed in Fig. 4. (c) A comparison of RT values derived from conventional BrdU-IP population data and scRepli-seq data of the entire S-phase by a smoothed scatter plot. The analysis was done using 400-kb bins.

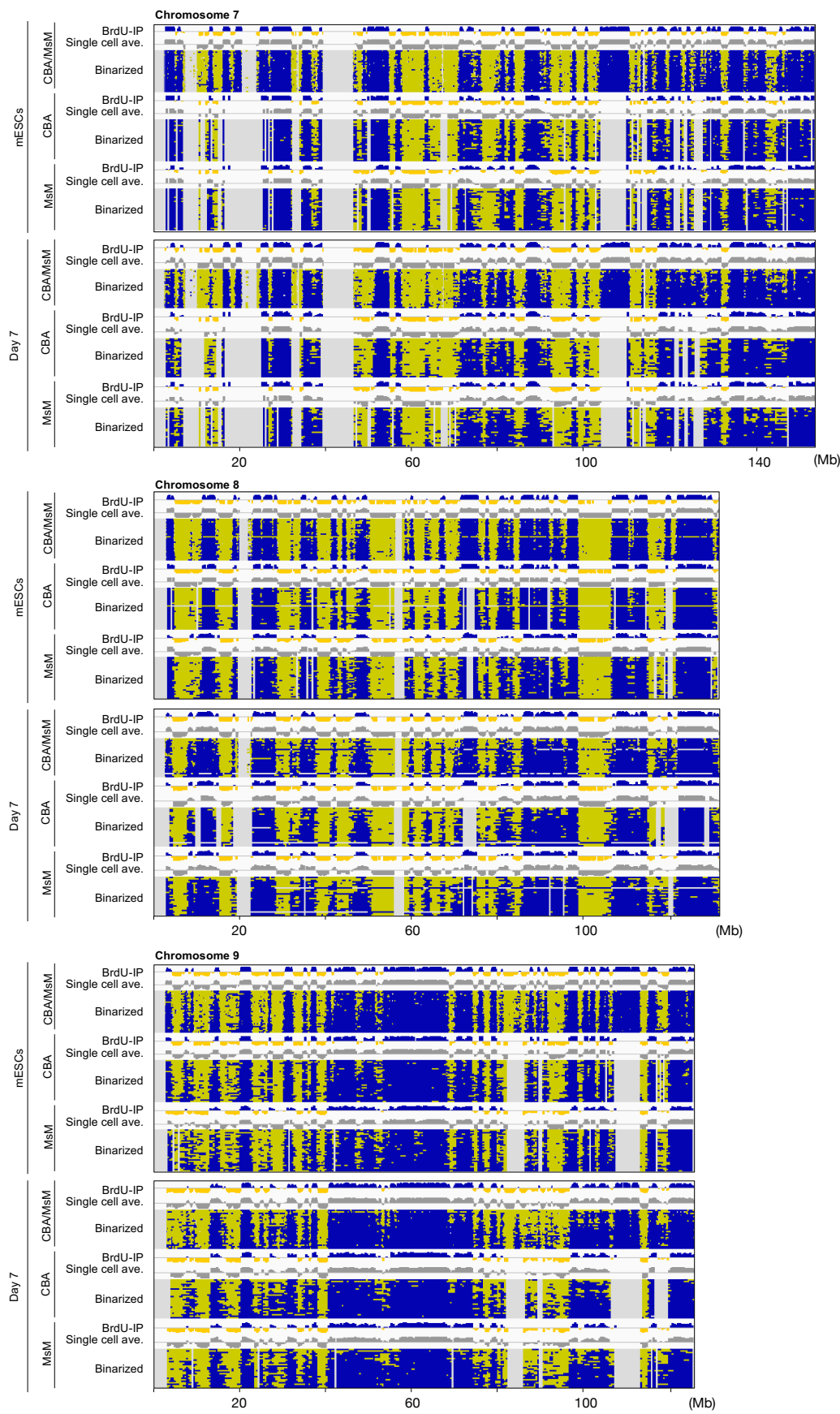
Supplementary Figure 4



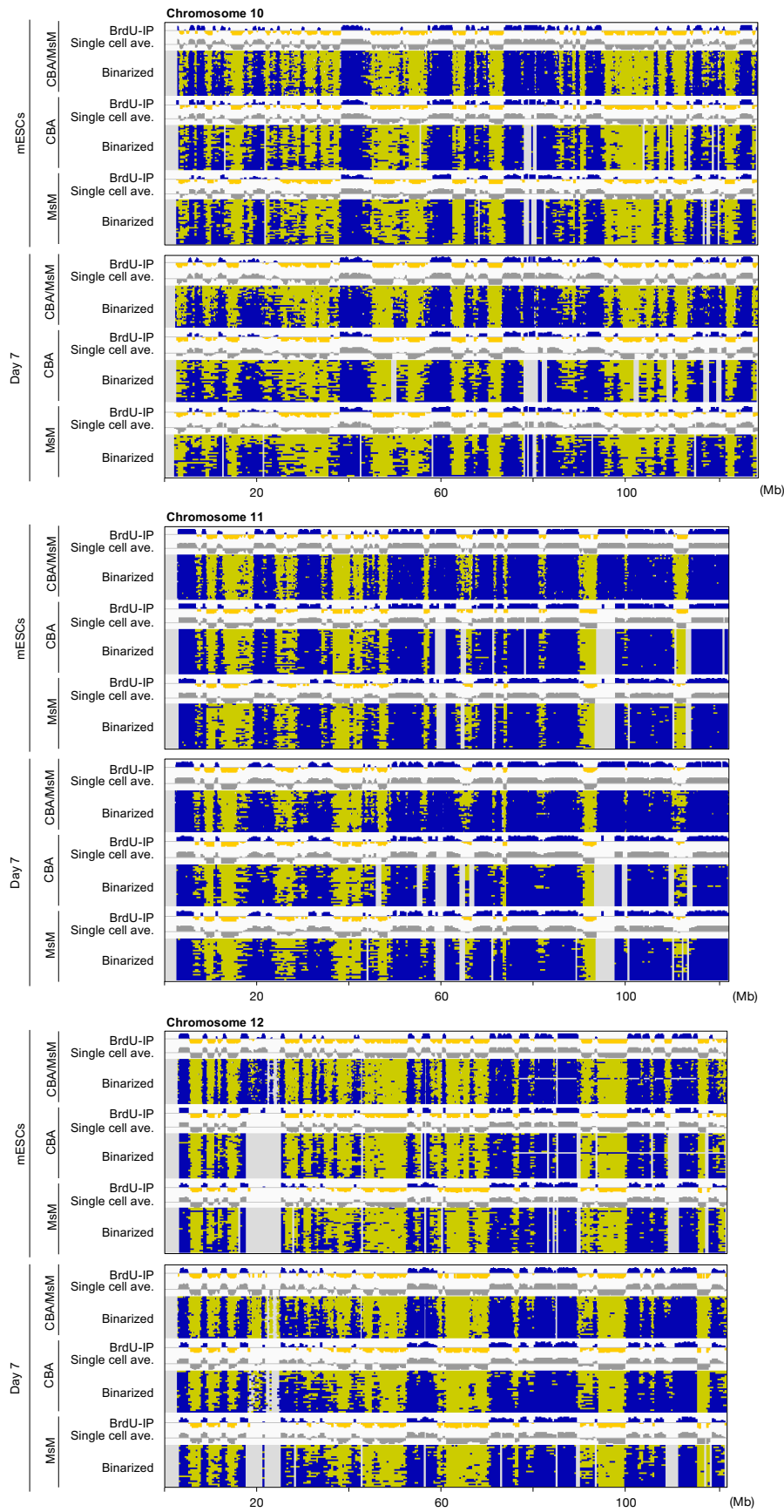
Supplementary Figure 4 (continued)



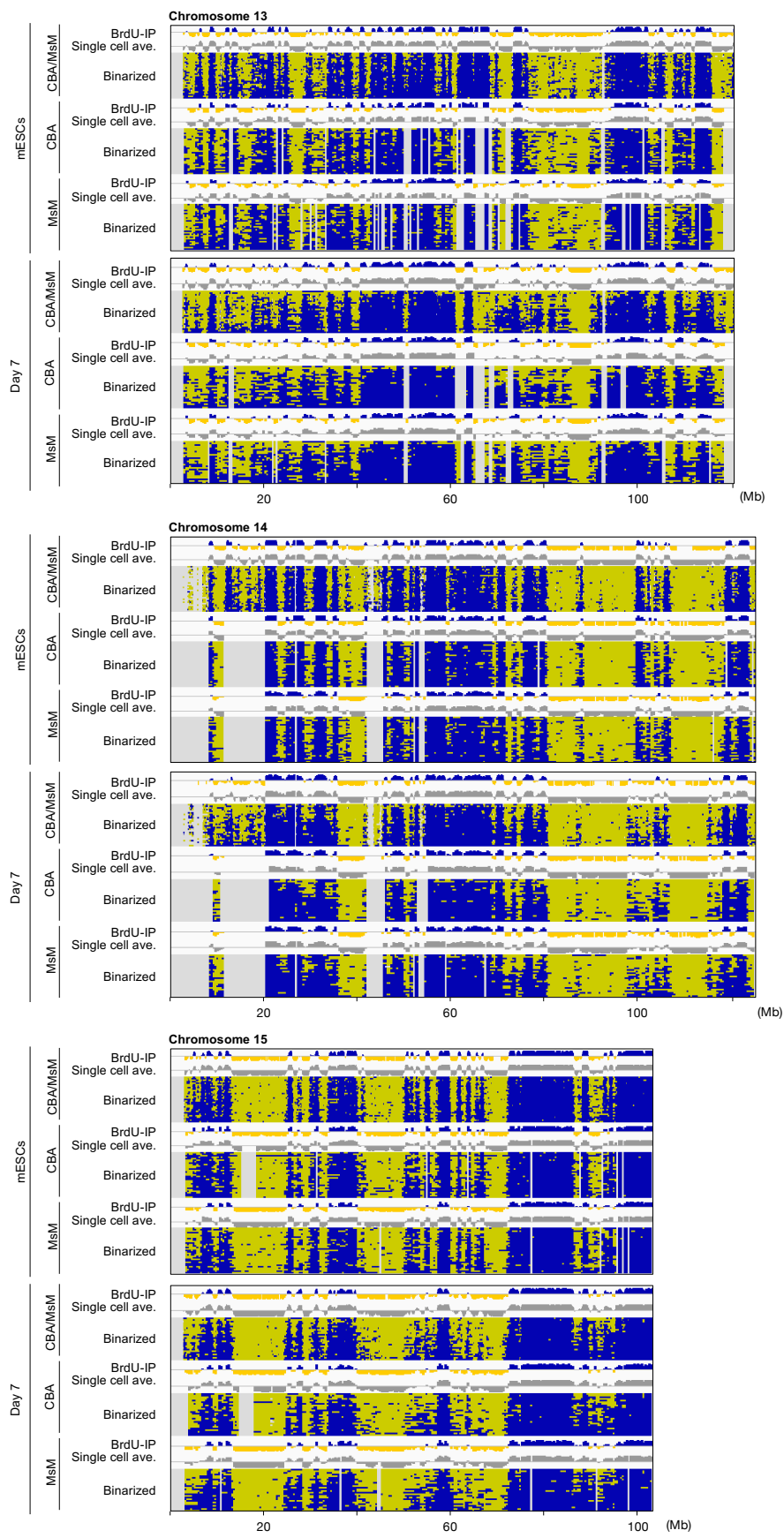
Supplementary Figure 4 (continued)



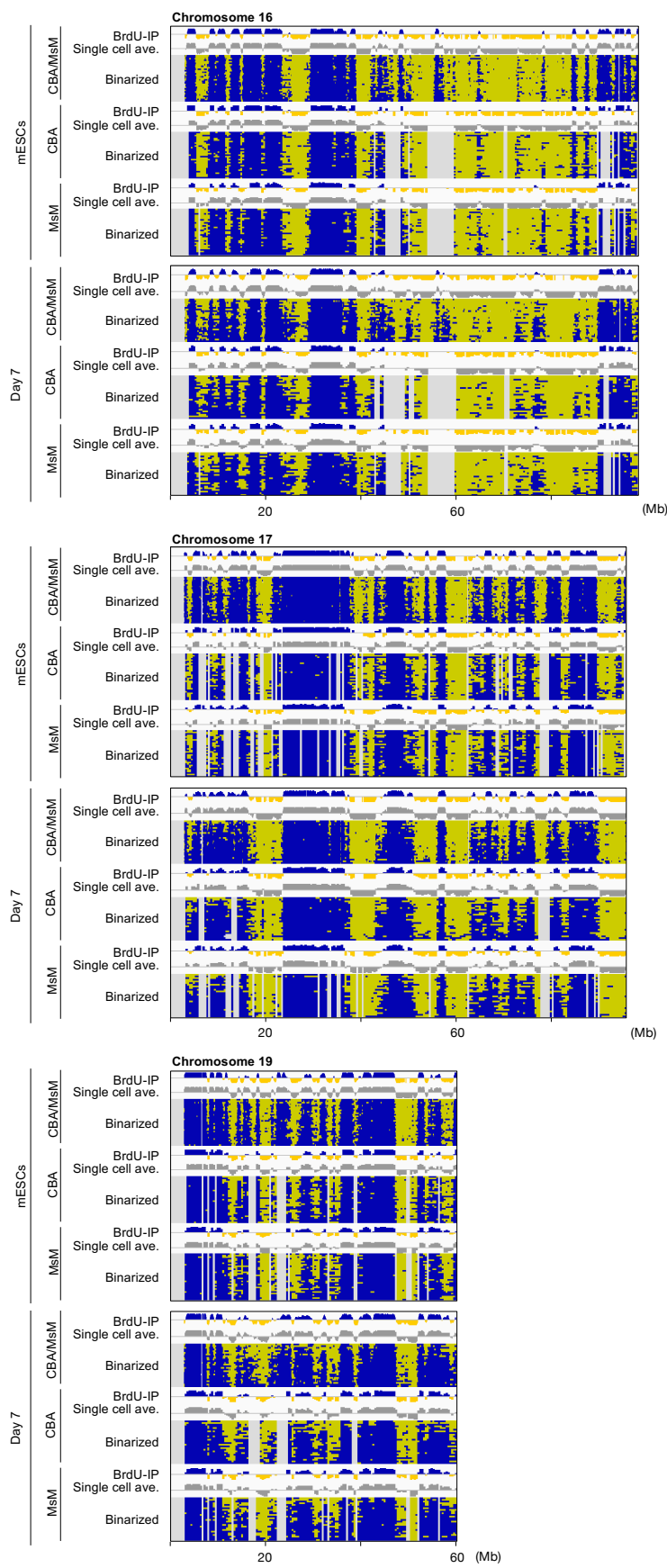
Supplementary Figure 4 (continued)



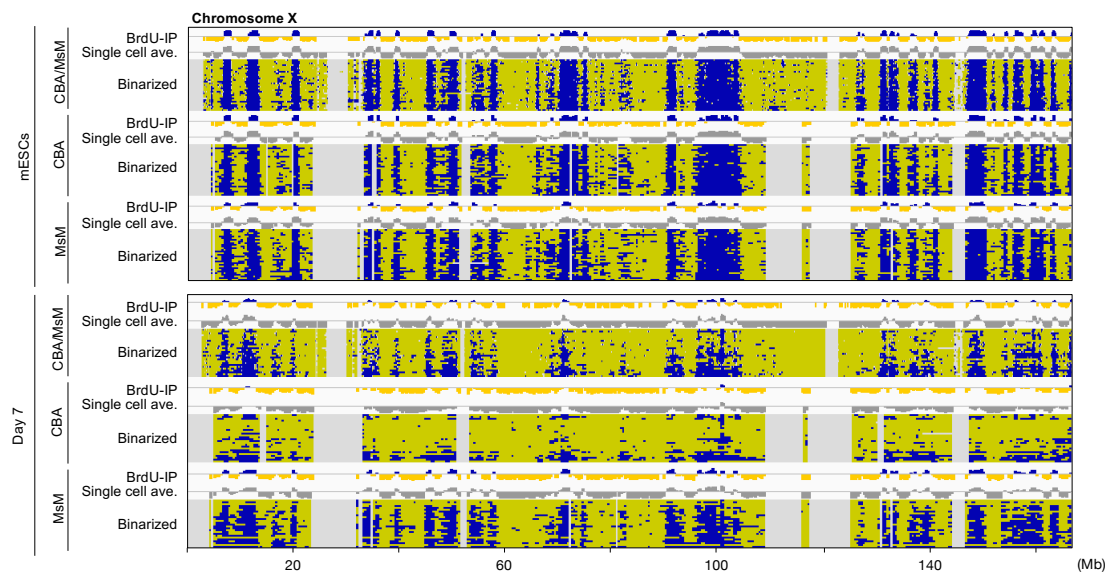
Supplementary Figure 4 (continued)



Supplementary Figure 4 (continued)

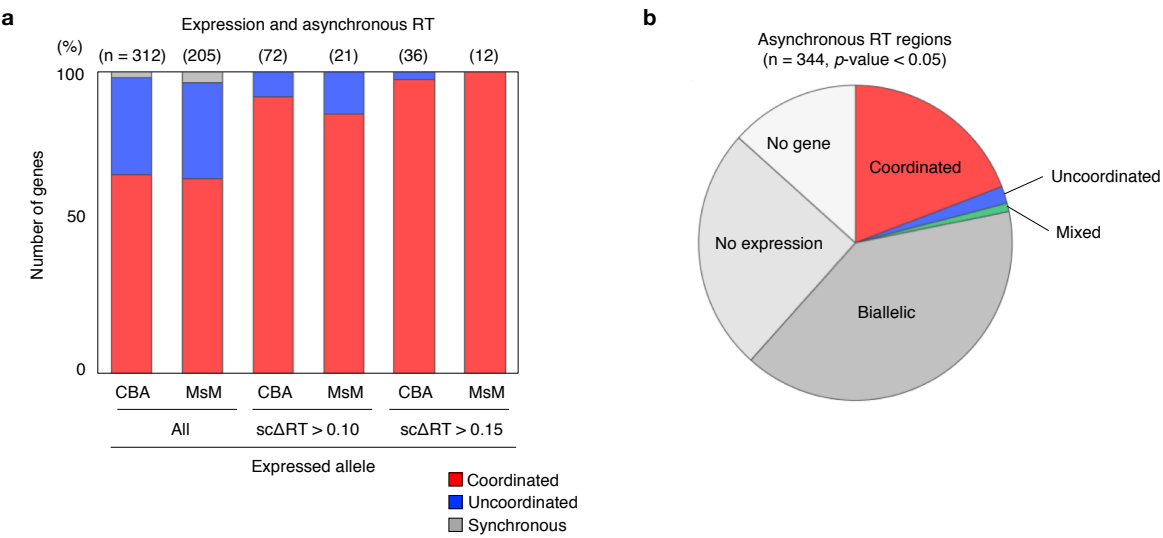


Supplementary Figure 4 (continued)



Supplementary Figure 4 | Binarized scRepli-seq profiles in mESCs and day-7 differentiated cells
Binarized mid-S scRepli-seq profiles with and without haplotype resolution on autosomes and the X chromosome, along with BrdU-IP population RT profiles and single-cell average profiles, are shown. Chr18 is not shown, due to variable chromosomal losses and gains. Gray areas represent mapping failures. Here, 31 mESCs and 26 day-7 cells are shown and are ordered by their % replication scores, which are shown next to chr3 profiles.

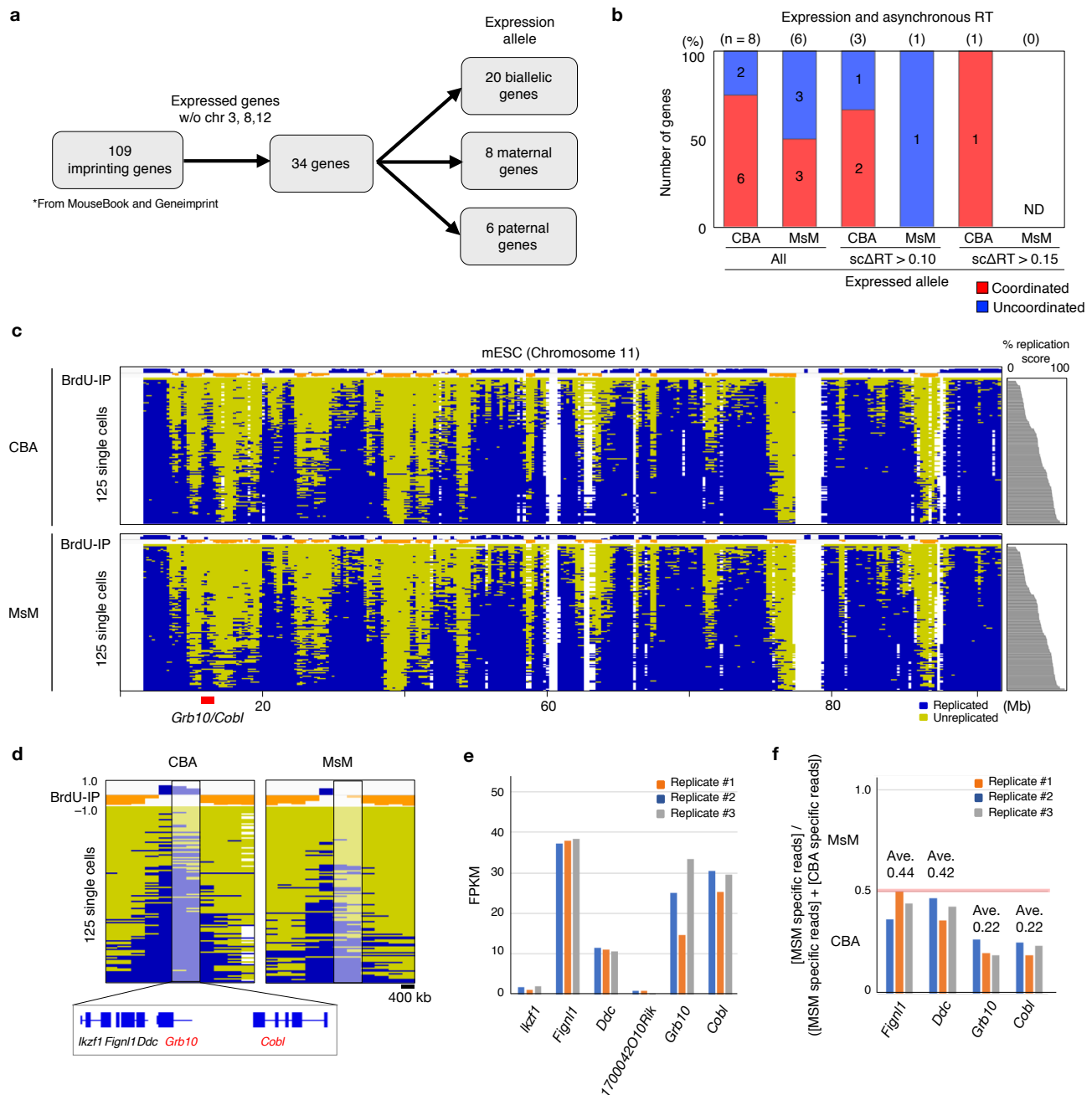
Supplementary Figure 5



Supplementary Figure 5 | Allelic expression imbalance and asynchronous replication

(a) Allelic expression imbalance was coordinated with replication timing asynchrony. The analysis was done in the same way as in Fig. 6j, except that the criteria for allelic expression imbalance was >70% expression from either allele by an allele-specific population RNA-seq (three independent RNA-seq data combined). 125 scRepli-seq data sets from mESCs were used for RT analysis. (b) Gene expression patterns in asynchronous RT regions. The analysis was done in the same way as in Fig. 6k, except that the criteria for allelic expression imbalance was >70% expression from either allele.

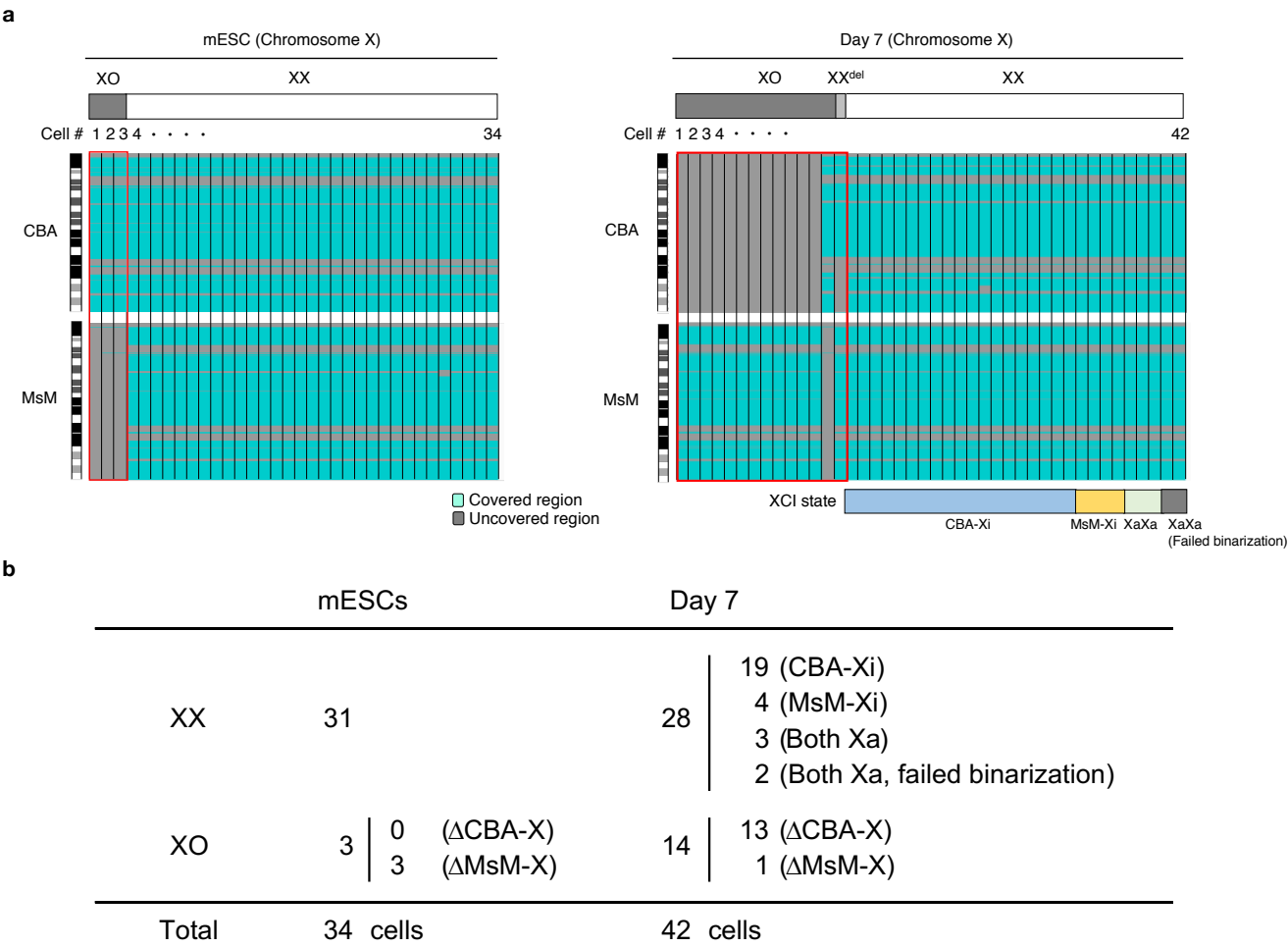
Supplementary Figure 6



Supplementary Figure 6 | Allelic expression imbalance and asynchronous RT in imprinting genes

(a) Imprinting genes that show allelic expression imbalance in mESCs. We found a total of 109 imprinting genes from the Mousebook and the Geneimprint databases. We excluded genes on chr3, 8, and 12 from the analysis because occasional gains/losses of all or portions of these chromosomes occurred in a subset of CBMS1 mESCs, which complicated interpretation of RNA-seq data. Imprinting genes with >10% of total cells missing binarized scRepli-seq data were also excluded. Among the 34 expressed genes (i.e. ≥ 10 total reads) in CBMS1, we found 8 maternally-expressed (CBA) and 6 paternally-expressed (MsM) genes, which were defined as those with >60% expression from either allele by allele-specific population RNA-seq. (b) Allelic expression imbalance of imprinting genes and RT asynchrony. The analysis was done as in Fig. 6j. While there could be a weak tendency toward a "coordinated" pattern, the numbers were too small to draw conclusions regarding statistical significance. (c) Haplotype-resolved, binarized scRepli-seq profiles of mESCs throughout the S-phase. Red bar, *Grb10/Cobl* maternal imprinting region. (d) Haplotype-resolved, binarized scRepli-seq profiles of mESCs surrounding *Grb10*. Gene names in red represent maternally-expressed genes, while those in black represent either biallelic or not expressed genes. Each pixel represents 400 kb. (e) FPKM values of genes in (d). Three independent population RNA-seq experiments were performed using mESCs. (f) Allelic expression imbalance of genes in the *Grb10* region analyzed by allele-specific RNA-seq. The y-axis shows the expression rate of the MsM allele. Genes with values lower than 0.5 are expressed more from the CBA allele.

Supplementary Figure 7



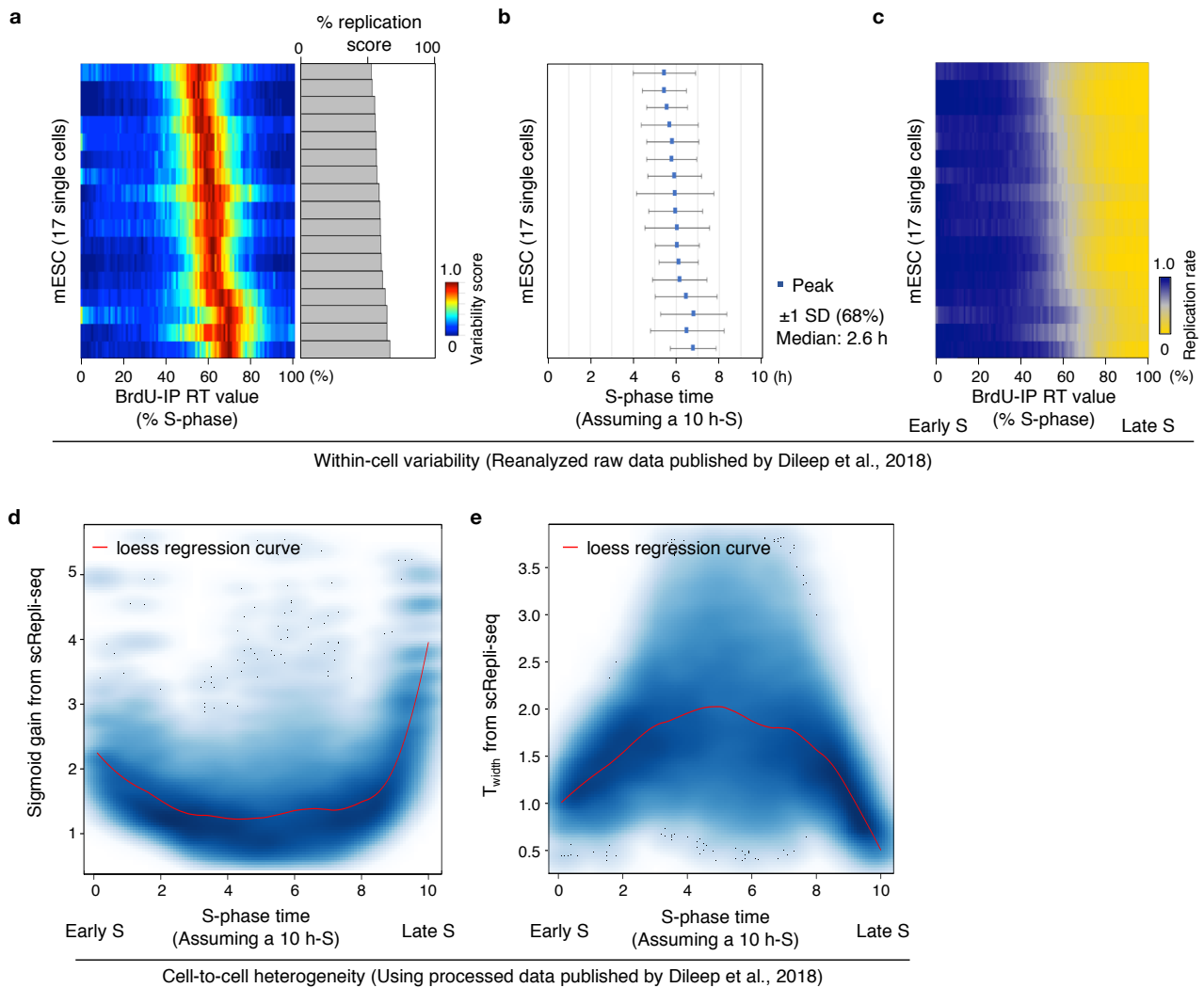
Supplementary Figure 7 | X-chromosome karyotypes in single cells

(a) Schematic diagrams of the X-chromosome karyotype in 34 mESCs and 42 day-7 cells. Turquoise and gray represent covered and uncovered genomic regions, respectively. There were 13 XO, 1 XX^{del} (distal deletion of CBA-X; counted as ΔCBA-X under XO for simplicity), and 28 XX cells. XCI states of day-7 cells are also shown; 'XaXa (failed binarization)' represents samples that exhibited an Xa/Xa state based on tSNE analysis but showed binarization failures. (b) A summary of the schematic diagrams in (a). While female mESCs are notorious for becoming XO (Robertson et al., 1983), 31/34 (91%) CBMS1 mESCs were XX even after several passages. The remaining three all lost the MsM-derived X (MsM-X). Later, we obtained 18 additional mid-S data from CBMS1 mESCs (third replicate) but excluded them from this table since we couldn't obtain XX G1 control cells. These G1 cells were XX^{del} with a distal deletion of CBA-X, precluding proper karyotype analysis. Many of the 18 additional mid-S cells may also be XX^{del}. Day-7 cells showed a frequent loss of CBA-X (13/14, 93%). Among the remaining 28 XX cells, five replicated both X chromosomes early, but three of them exhibited early replication of *Rex2* (which shows an EtoL switch upon differentiation), indicative of failed mESC differentiation. The remaining 23 had a late-replicating Xi, with 19 having a CBA-Xi and 4 having an MsM-Xi. This indicates that XCI was skewed (19/23=83%) in CBMS1, likely due to the *X-chromosome controlling element* (*Xce*) subtype combination effect⁵⁰. The frequent loss of CBA-X only in day-7 cells ($p=0.0059$, two-sided Fisher's exact test) suggests the possibility that loss of X occurs just as when it is undergoing XCI. 'Both Xa, failed binarization' represents samples that exhibited an Xa/Xa state based on tSNE analysis but showed binarization failures.

Reference:

Robertson, E. J., Evans, M. J. & Kaufman, M. H. X-chromosome instability in pluripotent stem cell lines derived from parthenogenetic embryos. *Journal of Embryology and Experimental Morphology* **74**, 297–309 (1983).

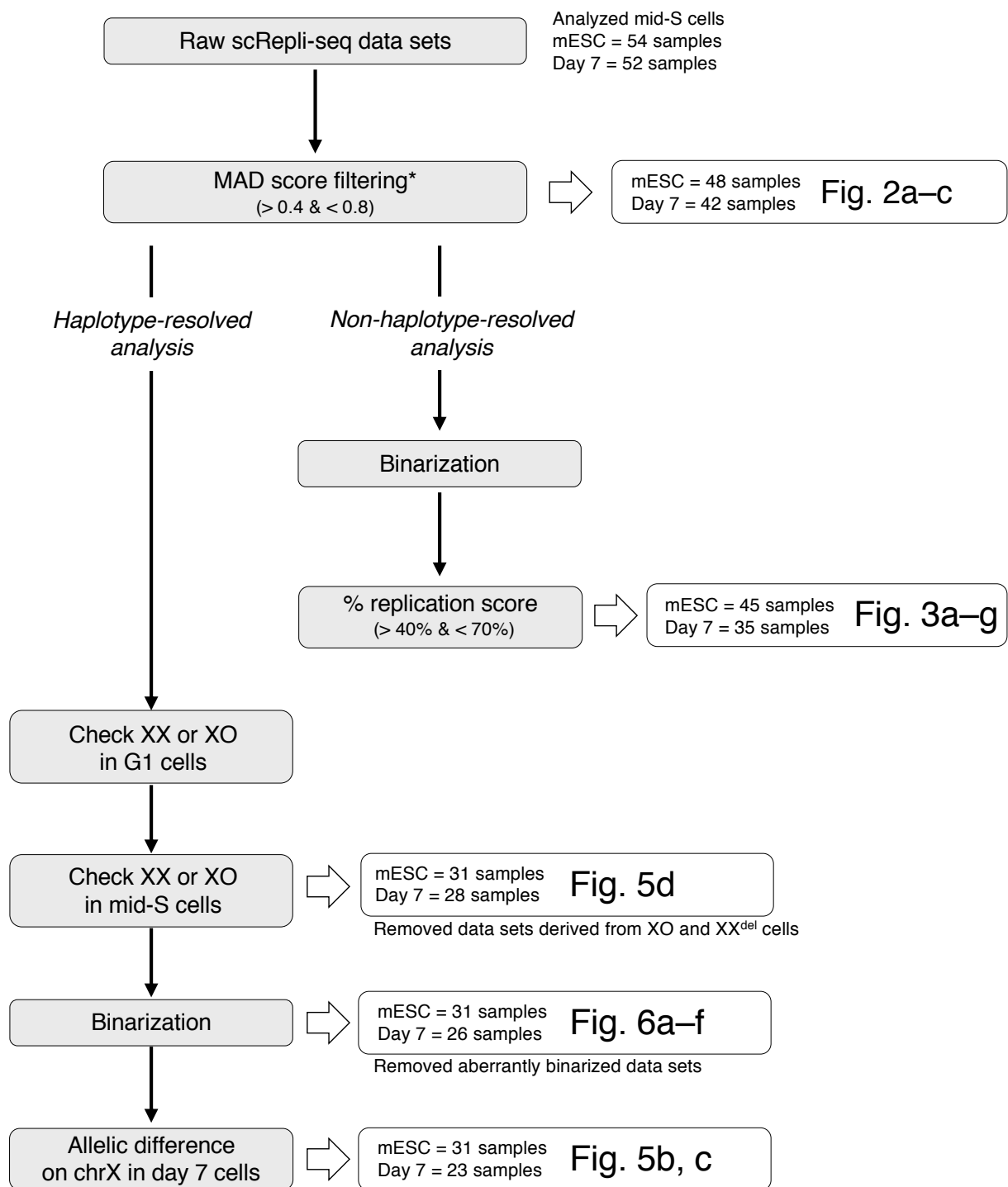
Supplementary Figure 8



Supplementary Figure 8 | Analysis of a published scRepli-seq data by Dileep et al.

(a) Analysis of within-cell variability of replication state using published diploid mESC mid-S scRepli-seq data sets by Dileep et al.¹³ as in Fig. 3d. Among the 96 raw mid-S data sets¹³, 17 data sets passed our filtering steps and were subject to further analysis by our hands. That is, 39 cells passed the MAD score filtering and among them, 21 had % replication scores in between 40% and 70%, and finally 17 cells were chosen as we removed 4 cells with low correlation (Pearson's $R < 0.5$). In brief, we subdivided the genome into one-percentile groups of bins with similar RT (% S-phase scores) and measured RT variability within each group in each cell. Each horizontal line in the heatmap represents a cell. (b) Within-cell variability peak and ± 1 SD range of the 17 cells from (a) is shown. Percent S-phase values were converted to hours, assuming a 10-h S-phase. (c) A heatmap showing the replication rate of each one-percentile group of bins with similar RT in the 17 cells. (d, e) Smoothed scatter plots of gain (d) and T_{width} (e) calculated from 217 processed scRepli-seq data sets derived from diploid and haploid mESCs throughout the S-phase by Dileep et al.¹³ by sigmoid model fitting, assuming a 10-h S-phase. Darker blue represents higher dot density. Red lines indicate loess regression curves (25% span). Consistent with our data, the degree of RT heterogeneity varies during S-phase, with the earliest and the latest replicating domains in S-phase exhibiting the least heterogeneity among cells. The degree of heterogeneity became higher and more variable during mid-early to mid-late S-phase.

Supplementary Figure 9



Supplementary Figure 9 | An overview of scRepli-seq analysis of mid-S cells from raw data to various processed forms

A flow chart showing the fate of raw scRepli-seq data from 54 mid-S mESCs and 52 mid-S day-7 cells during various analytical procedures. For intermediate stage data, corresponding figure numbers are shown. Raw scRepli-seq data sets underwent MAD filtering to avoid data sets containing data variability below or above certain thresholds within a sample. For analysis without haplotype resolution, data sets that passed the MAD filtering were subject to binarization and those that passed the % replication range shown were analyzed. Otherwise, the data sets were haplotype-resolved, binarized and subjected to the same % replication filtering used for non-haplotype analysis. For the inactive X-chromosome analysis shown in Fig. 5, only 23 cells with a late-replicating X chromosome were analyzed.