Eukaryotic chromosomes replicate in a temporal order known as the replication-timing program. In mammals, replication timing is cell-type-specific with at least half the genome switching replication timing during development, primarily in units of 400–800 kilobases (replication domains), whose positions are preserved in different cell types, conserved between species, and appear to confine long-range effects of chromosome rearrangements. Early and late replication correlate, respectively, with open and closed three-dimensional chromatin compartments identified by high-resolution chromosome conformation capture (Hi-C), and, to a lesser extent, late replication correlates with lamina-associated domains (LADs). Recent Hi-C mapping has unveiled substructure within chromatin compartments called topologically associating domains (TADs) that are largely conserved in their positions between cell types and are similar in size to replication domains. However, TADs can be further sub-stratified into smaller domains, challenging the significance of structures at any particular scale. Moreover, attempts to reconcile TADs and LADs to replication-timing data have not revealed a common, underlying domain structure. Here we localize boundaries of replication domains to the early-replicating border of replication-timing transitions and map their positions in 18 human and 13 mouse cell types. We demonstrate that, collectively, replication domain boundaries share a near one-to-one correlation with TAD boundaries, whereas within a cell type, adjacent TADs that replicate at similar times obscure replication domain boundaries, largely accounting for the previously reported lack of alignment. Moreover, cell-type-specific replication timing of TADs partitions the genome into two large-scale sub-nuclear compartments revealing that replication-timing transitions are indistinguishable from late-replicating regions in chromatin composition and lamina association and accounting for the reduced correlation of replication timing to LADs and heterochromatin. Our results reconcile cell-type-specific sub-nuclear compartmentalization and replication timing with developmentally stable structural domains and offer a unified model for large-scale chromosome structure and function.

Measurements of replication timing in human and mouse reveal chromosome segments with relatively uniform replication timing (constant timing regions, CTRs), mediated by clusters of near-synchronous initiation events that are heterogeneous in location from cell to cell and appear to fire through a stochastic mechanism. Despite stochastic origin firing, CTRs are interrupted at reproducible locations by transitions between early and late replication called timing transition regions (TTRs). We mapped TTRs in 35 mouse and 31 human data sets as part of the Mouse ENCODE project consortium. Replication timing of early TTR borders clustered better than late (Extended Data Fig. 1a), suggesting that initiation events defining early borders are coordinated, whereas events defining late borders are less synchronized, possibly resulting from passive fork fusion. To investigate a possible relationship between TTRs and TADs (Supplementary Discussion), we aligned mouse embryonic stem cell (mESC) TTRs (Fig. 1b) and compared them to the directionality index used to define TAD boundaries (transitions from upstream to downstream interaction bias). A single shift from upstream to downstream bias occurred within 500 kilobases (kb) of the average TTR, located near the aligned early border. Examination of individual TTRs indicated that TAD boundaries typically isolated early CTRs from TTRs, whereas TTRs and neighboring late CTRs predominantly belonged to the same TAD (Fig. 1c and Extended Data Fig. 1b, c). Similarly, transitions between Hi-C compartments exhibited preferential TAD boundary alignment to the border of the compartment associated with early replication (Fig. 1d). Hence, early TTRs separate TADs within compartment A from TADs within a compartment interaction gradient along TTRs, whereas late TTR borders have no detectable relationship to TAD structure.

Examination of replication timing across TADs (Fig. 1e) revealed, with few exceptions, that TADs were entirely early or late replicating, spanned all or part of a single TTR, or contained converging TTRs that constituted the previously described U-shaped replication-timing domains. Replication-timing patterns across LADs were remarkably similar except that LADs exclusively replicated during mid to late S phase (Fig. 1e), and TADs that replicated early versus late exhibited clearly distinct levels of lamina association (Extended Data Fig. 2a–c). Consistent with observations that TTRs associate with the nuclear lamina more frequently than CTRs with similar replication timing, we observed lamina association within late-replicating regions and TTRs (Extended Data Fig. 2d, e), explaining the modest correlation of LADs to replication timing. Although 30% of TTRs did not overlap with a computationally called LAD, these TTRs still associated with the nuclear lamina to some degree (Extended Data Fig. 2f) and may interact preferentially with other repressive sub-nuclear compartments. Together, these results revealed that TTRs resemble late-replicating regions with no discontinuity at late TTR borders, whereas early TTR borders are strong candidates for the structural boundaries of replication domains.

Localizing the replication domain boundary to early TTR borders (hereafter referred to as replication domain boundaries) prompted us to devise a more precise algorithm to map replication domain boundaries.
To determine the stability of replication domains during development, we generated a list of unique replication domain boundaries and classified each boundary as either 'TTR-present' or 'TTR-absent' in each available cell type (Fig. 2a). By examining the overlap of TAD boundaries with the compiled list of replication domain boundaries, we found that nearly all TAD boundaries corresponded to a replication domain boundary (Fig. 2b). Importantly, a majority corresponded to replication domain boundaries that were TTR-absent in cells where the TADs were mapped (IMR90 cells), supporting the conclusion that TADs are stable during development and function as replication domains. The fraction of TAD boundaries that did not align with any replication domain boundary is expected due to the portion of the genome with constitutive replication timing in the cell types for which data were available. Although nearly all TAD boundaries corresponded to replication domain boundaries, the reciprocal comparison indicated that many replication domain boundaries did not coincide with a corresponding TAD boundary (Extended Data Fig. 4). Although alignments of either TTR-present or TTR-absent replication domain boundaries to TAD boundaries were statistically significant (Fig. 2c), alignment to TTR-absent replication domain boundaries was not as strong (Fig. 2c), explained by incomplete TAD annotation and the observation that small TTRs lack a detectable relationship with replication domain boundaries (Extended Data Fig. 5 and Supplementary Discussion).

To corroborate TAD stability across cell types, we also compared TAD calls to high-resolution chromosome conformation capture-on-chip (4C) interaction frequency data across a replication domain that switches replication timing during mouse ESC differentiation to neural precursors. In ESCs, where TTRs flank this domain, TAD boundaries and marked decreases in 4C interaction frequency are apparent near both replication domain boundaries (ESC panels in Fig. 2d). However, in differentiated cells, where the replication domain is replicated at the same time as its neighbours, a TAD boundary is no longer called at the leftmost replication domain boundary, even though a sharp decrease in interaction frequency is detected by the higher-resolution 4C (NPC and cortex panels in Fig. 2d). Thus, the TAD boundary at this cell-type-specific TTR is stable during differentiation even though it is not identified as such by this Hi-C data set, providing additional evidence that TAD annotation is incomplete. To demonstrate the functional relationship between TADs and replication domains, we also compared the positions of TADs to replication-timing shifts observed previously at points of chromosome rearrangement. Figure 2e shows a rearrangement that joined...
otherwise early- and late-replicating regions. In this example, early replication appears to have spread into the late region up to a point that coincides with the nearest TAD boundary, where a new TTR was formed. Similar results were observed for additional examples (Extended Data Fig. 6). Taken together, these results provide compelling evidence that TADs act as stable units of replication-timing regulation during development.

To identify candidate factors involved in the developmental regulation of replication domains, we next compared replication domain boundaries to histone modifications, transcription factor binding sites, and DNase I hypersensitive sites (DHS) mapped by the ENCODE consortia. We aligned over 20 chromatin features to TTR-present replication domain boundaries in 7 mouse and 13 human cell types and found that only LAD boundaries were highly enriched in all the cell types where data were available (Fig. 3a, b and Extended Data Fig. 7). Notably, SUZ12 is a component of the Polycomb repressive complex 2 responsible for transcriptional repression of genes in a permissive chromatin environment. However, strong enrichment was not observed in the mouse embryonic stem cell (mESC) TAD boundaries8, but this apparent enrichment at boundaries was due to differential enrichment among TADs (Extended Data Fig. 8 and Supplementary Discussion). Similarly, densities of several DNA repeats and motifs were partitioned at replication domain boundaries and transitions in nucleotide skew (‘N-domain’ boundaries) were enriched near replication domain boundaries (Extended Data Fig. 7). Metazoan genomes have been segmented into a manually selected number of chromatin classes that correlate with replication timing29. By combining data for seven factors (CTCF, HCF1, MAFK, P300, RNA Pol II, ZC3H11A, and ZNF384), we assigned each TAD into classes using an unsupervised approach (Supplementary Discussion). We obtained two TAD classes, termed A and B, indicating the presence of clearly recognizable differences in the transcription factor composition of these classes, as well as to their proximal to early replication domains.

Our results support a unifying model in which TADs are stable regulatory units of replication timing (Fig. 4). In this ‘replication-domain model’, DNA synthesis begins within TADs that reside in the nuclear interior and contain features permissive for transcription. Meanwhile, replication gradually advances into adjacent later-replicating TADs that reside at the nucleoplasm or other repressive compartments and contain features associated with repression. This gradual progression forms a TTR that extends from the boundary separating early and late TADs to a context-dependent point (that is, independent of TAD structure, Extended Data Fig. 6a) determined by replication rate and time elapsed before replication origins throughout adjacent later-replicating TADs and the resulting forks merge. Similarly, TADs replicated by active origin firing in mid S phase form TTRs that extend into adjacent later-replicating TADs (Extended Data Fig. 6a). By contrast, timing transitions do not form at boundaries between adjacent
TADs residing in the same compartment due to coincidence of initiation events within their structural boundaries. Upon differentiation, TADs that switch replication timing acquire features associated with their new sub-nuclear compartment while their preexisting structural boundaries establish new compartment boundaries. The demonstration that TADs are units of regulation reveals an important organizational principle of mammalian genomes and represents a critical step towards understanding mechanisms regulating replication timing. Determining whether replication timing dictates chromatin structure within TADs to influence chromatin interactions or vice versa will be an important area of future investigation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February; accepted 22 October 2014.

1. Wright, M. L. & Grützner, F. in DNA Replication Current Advances (ed. Seligmann, H.) Ch. 20 (InTech, 2011).

2. Hiratani, I. et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* 6, e245 (2008).

3. Hansen, R. S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* 107, 139–144 (2010).

4. Ryba, T. et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 20, 761–770 (2010).

5. Yaffe, E. et al. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.* 6, e1001011 (2010).

6. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* http://dx.doi.org/10.1038/nature13992 (this issue).

7. Pope, B. D. et al. Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Hum. Mol. Genet.* 21, 4162–4170 (2012).

8. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012).

9. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome–nuclear lamina interactions during differentiation. *Mol. Cell* 38, 603–613 (2010).

10. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385 (2012).

11. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295 (2013).

12. Filippova, D., Patro, R., Duggal, G. & Kingsford, C. Identification of alternative topological domains in chromatin. *Algorithms Mol. Biol.* 9, 14 (2014).

13. Meuleman, W. et al. Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* 23, 270–280 (2013).

14. Rhind, N., Yang, S.-C.-H. & Bechhoefer, J. Reconciling stochastic origin firing with defined replication timing. *Chromosome Res.* 18, 35–43 (2010).

15. McQuiffe, S. R., Smith, D. J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* 50, 123–135 (2013).

16. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* 9, 999–1003 (2012).

17. Baker, A. et al. Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.* 8, e1002443 (2012).

18. Farkash-Amir, S. et al. Systematic determination of replication activity type highlights interconnections between replication, chromatin structure and nuclear localization. *PLoS ONE* 7, e48986 (2012).

19. Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178–192 (2013).

20. van Koningsbruggen, S. et al. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* 21, 3735–3748 (2010).

21. Németh, A. et al. Initial genomics of the human nucleolus. *PLoS Genet.* 6, e1000889 (2010).

22. Takebayashi, S., Dileep, V., Ryba, T., Dennis, J. H. & Gilbert, D. M. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proc. Natl Acad. Sci. USA* 109, 12574–12579 (2012).

23. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).

24. Shen, X. et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502 (2008).

25. Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* 23, 4061–4071 (2004).

26. Pasini, D. et al. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* 38, 4958–4969 (2010).

27. Audit, B. et al. Open chromatin encoded in DNA sequence is the signature of ‘master’ replication origins in human cells. *Nucleic Acids Res.* 37, 6064–6075 (2009).

28. Huvet, M. et al. Human gene organization driven by the coordination of replication and transcription. *Genome Res.* 17, 1278–1285 (2007).

29. Filion, G. J. et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* 143, 212–224 (2010).

30. Julienne, H., Zoufir, A., Audit, B. & Arneodo, A. Human genome replication proceeds through four chromatin states. *PLoS Comput. Biol.* 9, e1003233 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Laugesen, G. Andersen and K. Helin for providing Suz12 control and knockout naive mESCs lines. We thank F. Ay, M. Libbrecht, W. S. Noble, E. Besnard, J. M. LeMaitre, C. Cayrou, M. Mechali and J. Dekker for helpful discussions. This research was supported by NIH grants GM083337 and GM085354 to D.M.G., HG005602 to M.P.S., HG005573 and DK065806 to R.C.H. and HG03991 to B.R. B.D.P. is supported by the National Cancer Institute of the National Institutes of Health under award number F31CA165863.

Author Contributions B.D.P., T.R., M.P.S., J.A.S., J.T., R.C.H. and D.M.G. devised the study are available at the Mouse ENCODE portal website (http://mouseencode.org) and scripts are available at GitHub (https://github.com/popex/MCP05). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.M.G. (gilbert@bio.fsu.edu).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The image or other third party material in this article is included in the article’s Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-sa/3.0/
METHODS
Generation of replication timing data. Protocols for generating and quality control for replication timing data from microarray hybridization (Replic-chip) or sequencing (Replic-seq) were performed as previously described3,35, CH12, MEL, GM12878, GM12801, GM12812, GM12813, HeLa-S3, HepG2, HUVEC, IMR90, MCF-7, Sk-N-Sh and NHEK cells were obtained and grown according to standard ENCODE cell culture protocols36. Wild-type control and Suz12 knockout naive mESCs were derived from the previously described strain37 and obtained from Anne Lauegensen and Kristian Helin and cultured in 2i+LIF medium as previously described38. Previ-
ously published Replic-chip39,40 and Replic-seq data sets were also used in this study (Extended Data Table 1).
Analysis of aligned TTRs. Regions surrounding each of the 8,433 TTRs called in all mESC data sets were aligned by assigning a fixed number of evenly spaced windows from 100 kb downstream of the early boundary to 100 kb upstream of the late boundary. The number of windows depended on the type of data being analysed but was suited for 400 kb, in effect treating each TTR as if it were 600 kb (mean TTR length). Averaged mESC replication-timing profiles6 and lamina-association data9 were collected in 10-kb windows and DI data9 were collected in 40-kb windows surrounding called TTRs. Smoothed intervals containing 95% of the replication-timing data were plotted using custom R scripts. Analysis of aligned TADs and LADs. Averaged mESC replication-timing profiles6 were aligned across mESC TADs and LADs by assigning data into a fixed number of evenly spaced windows across their length, respectively suited for their median lengths of 800 kb (80 windows) and 450 kb (45 windows). Similarly, TIG3 lamina association data9 were assigned to evenly spaced windows across mESC TADs. IMR90 TADs were subdivided into 3 classes (that is, ‘early’, ‘TTR’, ‘late’) depending on both the means and standard deviations of IMR90 replication timing within each TAD. Early TADs had mean replication timing >0.9 and standard deviation <0.25, late TADs had mean replication timing <0.4 and standard deviation >0.25, and all others TADs were considered TTR.
Replication domain boundary identification and analysis. Replic-chip data were quantile normalized to the average distribution of data sets generated by the same microarray platform. Replic-seq weighted average data for 1-kb windows were centred about zero and initially normalized to an interquartile range of 1.59, which is equivalent to standard Replic-chip early/late CGH timing values (log2 Cy3/Cy5 enrichment) for individual probes. Replic-seq data sets were then quantile normalized to a combined pool of all Replic-seq points by random sampling. Loess smoothed replication-timing profiles were obtained from each quantile-normalized data set using a smoothing span of 35, 85, 115, and 345 data points for mouse 385k probe Replic-chip, mouse 720k probe Replic-chip, human 720k probe Replic-chip, and human 2M probe Replic-chip data sets respectively or 287–394 data points for human Replic-
seq data sets. Replication domain boundaries were identified as the early border of transitions between relatively earlier and later replicating regions in individual data sets with a slope above ±2.75 × 10−6 RT units per bp at the early border and with a sustained slope above ±1 × 10−6 RT units per bp for 200 kb–1 Mb and at least 0.55 RT units and 30, 60, 6.4, and 165 data points for mouse 385k probe Replic-chip, mouse 720k probe Replic-chip, human 720k probe Replic-chip, human 2M probe Replic-chip data sets, respectively, or 140–192 data points for human Replic-seq data sets. Initial boundary calls were filtered further by removing boundaries within 125 kb of gaps between adjacent data points spanning at least 80, 54.5, 77.5, and 80 kb for mouse 385k probe Replic-chip, mouse 720k probe Replic-chip, human 720k probe Replic-chip, and human 2M probe Replic-chip data sets respectively or 287–394 data points for human Replic-
seq data sets. Metro-provided genome boundaries were used to map the orientation of TAD boundaries, the combined lists of boundaries for each cell type was compared to transform all other samples. Model output was evaluated using gold standard TAD boundary identification. New annotation of IMR90 TADs was performed by calculating an interaction directionality index for IMR90 Hi-C data42. Briefly, the genome was divided into 40-kb windows and for each window the frequency of interaction within 2 Mb upstream of the window to the frequency of interaction within 2 Mb downstream of the window was compared as described previously43. A second directionality index was calculated for the surrounding 200 kb of the TAD for the frequency of interaction within 20 kb and only considering interactions within 1 Mb for each window.
Chromatin state analysis. ChromHMM44 was applied on pooled ChIP-seq mapped reads for each of four histone modifications (H3K4me3, H3K4me1, H3K36me3, and H3K27me3) in 15 mouse cell lines (G1E, G1E-ER4 + E2, erythro-
blasts, megakaryocytes, CH12, MEL, ESBruce4, thymus, kidney, small intestine, E14 whole brain, liver, spleen, heart, and Testis) and 3 human cell lines (GM12878, H1 hESC, and K562) to learn a multivariate HMM model for genome segmentation. Mapped reads were first processed into binarized data in 200-bp windows over the mapped genome, with ChIP input reads as the background control. To learn the model jointly from mouse and human, a pseudo genome table was first constructed by concatenating mouse (mm9) and human (hg19) tables. A model was then learned from all binarized data, giving a set of emission parameters and transition parameters, which was used to produce segmentations based on the most likely state assignment of the model. The eight-state model used in this paper appeared most parsimonious in the sense that all eight states had clearly distinct emission properties, while the interpretability of distinction between states in models with additional states was less clear.
TAD boundary orientation and analysis. For each analysed feature, data surrounding TAD boundaries were aligned in 10-kb windows from 500 kb upstream to 500 kb downstream of each TAD boundary. The considered regions were then inverted if the average feature density downstream of the TAD boundary was greater than the average feature density upstream of the TAD boundary, so that feature density decreased from left to right across each boundary. After orienting TAD boundaries in this manner, each window was averaged for all TAD boundaries. Local boundary enrichment was then evaluated by comparing the average density of windows within 20 kb of the TAD boundary to the average density of windows 140–430 kb away from the TAD boundary on the side with increased density.
Transcription factor combination model. Raw data from genome-wide ChIP-seq in vivo detection of transcription factor binding45 were processed in 200 bp bins over a 1 Mb window around the centre of each mESC TAD. Each sample can be thought as a 2-dimensional matrix with rows for seven transcription factor activity profiles and each row represented as a real vector with values for each bin in the 1 Mb window (number of bins = 5,000). The value X_kl represented the activity (that is, peak enrichment) of transcription factor j, k bins from the centre of TAD l. The model consists of a stack of 6 sigmoid denoising autoencoders, which define parameterized feature extraction (encoder) and reconstruction functions (decoder). The encoder projects the data into a smaller dimension feature space and the decoder reconstructs the input from the feature space. The parameters of the functions are optimized to give the smallest reconstruction error over all the training data samples. The training data is similar in principle to PCA (they both can be used as dimensionality reduction techniques), the denoising auto-encoder learns a nonlinear mapping between the input and its representation. Furthermore, constraints on the input and feature dimension sizes and the addition of noise to the input force it to learn important structure in the input. The stack reduces input dimensions gradually from (7 × 5,000), (7 × 10,000), (7 × 200), (7 × 60), (7 × 30), (7 × 20), to (7 × 10). Each autoencoder injects an additive binomial noise with a 20% corruption rate. We initialized weights at small random values with mean zero and used minibatch stoch-
istic gradient descent46 to minimize the mean squared input reconstruction error. We trained on 200 random samples for 500 epochs each layer then used the model to predict for all other samples. Model output was evaluated using gold standard labels based on both the means and standard deviations of mESC replication timing within each TAD. Early TADs had mean replication timing >0 and standard deviation <0.25, late TADs had mean replication timing <0 and standard deviation >0.25, and all others TADs were considered TTR. The programs we wrote to perform the analysis are available on the Bio-Infodev website (http://bitbucket.org/igelentina/dimer/commit/tags/TADS). To perform PCA, cluster the data, and evaluate models, we used scikitLearn Version 0.14- git (http://scikit-learn.org/).

31. Ryba, T., Battaglia, D., Pope, B. D., Hiratani, I. & Gilbert, D. M. Genome-scale analysis of replication timing: from bench to bioinformatics. Nature Protocols 6, 870–895 (2011).

32. Mouse ENCODE Consortium et al. An encyclopedia of mouse DNA elements (Mouse ENCODE), Genome Biol. 13, 418 (2012).
33. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
34. Hiratani, I. et al. Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. *Genome Res.* **20**, 155–169 (2010).
35. Ryba, T. et al. Abnormal developmental control of replication timing domains in pediatric acute lymphoblastic leukemia. *Genome Res.* **22**, 1833–1844 (2012).
36. Ryba, T. et al. Replication timing: a fingerprint for cell identity and pluripotency. *PLoS Comput. Biol.* **7**, e1002225 (2011).
37. Pope, B. D. et al. DNA replication timing is maintained genome-wide in primary human myoblasts independent of D4Z4 contraction in FSH muscular dystrophy. *PLoS ONE* **6**, e27413 (2011).
38. Chandra, T. et al. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* **47**, 203–214 (2012).
39. Lu, J., Li, F., Murphy, C. S., Davidson, M. W. & Gilbert, D. M. G2 phase chromatin lacks determinants of replication timing. *J. Cell Biol.* **189**, 967–980 (2010).
40. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
41. Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Chapman & Hall/CRC, 1998).
42. Cayrou, C. et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21**, 1438–1449 (2011).
43. Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Struct. Mol. Biol.* **19**, 837–844 (2012).
44. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
45. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
46. Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-3.0* (1996).
47. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
48. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215–216 (2012).
49. Visel, A. et al. ChiP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
50. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
Extended Data Figure 1 | Clustering of early replication-timing borders and TAD boundary alignment at TTRs and A/B compartment transitions.

a, Cumulative density plot showing clustering of timing values at the early and late side of timing transition regions. For each genomic orientation (forward and reverse are shown in right and left columns), timing values are more tightly distributed at the early border than the later border.

b, Directionality index data for individual IMR90 TTRs aligned at their early (left) or late (right) borders and arranged by TTR size. Solid black lines indicate the positions of early and late borders in each plot. c, Percentages of ESC TTR borders or random positions that align to TAD boundaries as a function of distance from the TTR centre. Boxplots indicate the positions of TTR borders. d, Percentages of mESC compartment A/B transitions that align to TAD boundaries as a function of distance from the A/B compartment threshold (eigenvector crosses zero). As observed for early TTR borders, more A/B transitions align with TAD boundaries on their compartment A side than on their compartment B side, which aligns with TAD boundaries at near random frequency.
Extended Data Figure 2 | TTRs and late-replicating regions associate with the nuclear lamina. 

a. Spearman correlations between genome-wide replication timing and lamina association (top) or observed changes between the indicated mouse cell types. b. Tig3 human fibroblast lamina association across individual IMR90 human fibroblast TADs with early (<0.5) or late (>−0.5) timing values in IMR90 human fibroblasts. c. Average levels of lamina association across the same early (red) and late (blue) TADs as in b. d. Lamina association in mESCs across individual mESC TTRs aligned as in Fig. 1b, c. TTRs were ordered in the heatmap by the distance between each early TTR border and the nearest downstream LAD. e. Heatmaps show lamina association and replication timing across aligned LADs ±400 kb. LADs were oriented with earlier replication timing to the left and ordered from top to bottom by the replication timing of the left LAD border. Averages for all LADs (grey) and the earliest (blue), middle (green), and latest (red) thirds are overlaid in the plots below. LADs on average replicate later than the surrounding genomic space and replication timing has little effect on the strength of lamina association. Interestingly, the left plot reveals a consistently sized gap ~100 kb in size separating neighbouring LADs. f. Average replication timing (blue or red) and lamina association (purple or green) are shown across aligned TTRs ±400 kb that either overlap with LADs (70%) or do not (30%). TTRs that do not overlap with called LADs still associate with the nuclear lamina to some degree.
Extended Data Figure 3 | Replication domain boundary calls are reproducible in replicate data sets. a, Histograms of distances between replication domain boundary calls and the nearest calls made in a separate data set from the same cell type. Four lymphoblastoid Repli-chip data sets (top left) were compared to each other and five lymphoblastoid Repli-seq data sets (top right) were compared to each other. The four lymphoblastoid Repli-chip data sets were then compared to the five Repli-seq data sets (bottom left). b, Boxplots for the same data from a are also shown.
Extended Data Figure 4 | Increasing the resolution of Hi-C analysis increases the number of called TAD boundaries and alignment with replication domain boundaries. a, Histograms of TAD sizes for original IMR90 calls (bottom, ref. 8) versus calls made using higher resolution IMR90 Hi-C data47 with 40 kb (middle) and 20 kb (top) directionality index bin sizes.

b, An example region of the IMR90 replication timing profile (grey) is shown with TTR-present and TTR-absent replication domain boundaries indicated by vertical blue lines. Directionality index plots for each of the Hi-C data sets from panel a are shown across the same region with the 5' TAD boundaries (start) indicated by solid red lines and the 3' TAD boundaries (end) indicated by dotted black lines. c, Overlap of TAD boundaries using original or higher resolution data with TTR-present (black) or all (grey) replication domain boundaries (top left) is shown within 175 kb. The reciprocal comparison is shown below. The percentage of replication domain boundaries that overlap with TAD boundaries increases when additional TAD boundaries are identified using higher resolution data, while the percentage of TAD boundaries that overlap with replication domain boundaries is unchanged. The overlap in each case is significant ($P < 10^{-77}$) relative to overlap with random positions (right).
Extended Data Figure 5 | Alignment of replication domain and TAD boundaries. a, IMR90 (original resolution in grey, high resolution with 20 kb bins in black) or H1 hESC (red) TAD boundary frequency for regions with the same replication timing in both cell lines. b, TAD boundary alignment to IMR90 replication domain boundary subsets based on IMR90 TTR properties. Random alignment was calculated based on the distribution of timing values within each subset. c, TAD boundary frequency for regions with different replication timing in IMR90 and H1 hESCs. d, Alignments for IMR90 replication domain boundaries as in b using TTR properties in all cell types. e, IMR90 TAD boundary probability density across small IMR90 TTRs that either do not (top) or do (bottom) coincide with larger TTRs (timing difference > 1.5) in other cell types. f, Histograms show the distribution of probability densities from Fig. 2c for TAD boundaries within 2 Mb of TTR-present (top) and TTR-absent (bottom) IMR90 replication domain boundaries (blue) or an equal number of random positions (grey). Vertical red lines mark the mean and three standard deviations from the mean random density. g, Percentages of TTR-present (top) and TTR-absent (bottom) IMR90 replication domain boundaries that aligned to TAD boundaries as a function of distance (red) are plotted with a random control (black). The significance of alignment is also shown (grey). The vertical dashed line indicates the distance at which alignment is most significant, while the vertical solid line indicates the distance at which alignment is most different from the control.
Extended Data Figure 6 | Replication-timing shifts at chromosome rearrangements are restrained by TAD boundaries. a, Distribution of early (blue) and late (grey) TTR borders within aligned, adjacent TADs for all TTRs (left), or TTRs that start in early (centre, timing < 0.5) or middle (right, timing ≥ 0.5) S phase. The right boundary of TAD 0 is nearest each early border and TADs 1–3 are neighbouring TADs in the direction of the timing transition (earlier to later from left to right). b, Histogram of replication domain sizes. c, Plots as in Fig. 2e show replication timing (red) across four rearrangement points (vertical green lines) that juxtapose otherwise early- and late-replicating regions on human chromosome 21 overlaid on the normal profile (black). Secondary rearrangement points (vertical grey lines) that joined regions with similar replication timing are also shown. The TAD boundary (vertical blue line) nearest to the fusion point is also indicated. In the examples at the top and bottom left, the shift forms a new TTR with its early border coinciding with the nearest TAD boundary. As in the other examples, the shift in the bottom right plot also does not extend beyond the nearest detected TAD boundary, but the TTR formed does not align with a called TAD boundary.
Extended Data Figure 7 | Alignment of TTR-present replication domain boundaries to chromatin features. Plots as in Fig. 3a, b show probability density functions (green curves) for selected chromatin features within 4 Mb of aligned replication domain boundaries in the indicated cell types. A vertical grey line indicates the replication domain boundary position and a vertical green line indicates the average position of maximum enrichment in replicate data sets, which is listed at the top left of each graph. Horizontal solid and dashed lines indicate the mean and three standard deviations from the mean probability density of each feature about an equal number of random positions.
Extended Data Figure 8 | Alignment of TAD boundaries to chromatin features. 

a, Approach for distinguishing local TAD boundary enrichment from differential enrichment among TADs in an aggregate analysis. Before averaging, TADs were oriented such that the analysed feature exhibited a decreasing density from left to right. 

b, c, SINE-B1 (b) or SINE-Alu (c) density across averaged TAD boundaries oriented (red) as in a or not (grey) are plotted (top). Individual TAD boundaries oriented (red) in a similar fraction exhibiting local enrichment indicated by blue brackets. 

d, Average CTCF peak intensity across boundaries from b (top) and c (bottom). 

e, Degree of local feature enrichment at TAD boundaries (see Methods).
Extended Data Figure 9 | Method comparison and summary of transcription factor prediction model. a, Precision (true positives / (true positives + false positives)), recall (true positives / (true positives + false negatives)), and the $f$-measure ($2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$) are plotted for $k$-means (top) or hierarchical (bottom) clustering of raw transcription factor composition data or of data mapped on reduced dimensions by principal component analysis or with denoising autoencoders (labelled as PCA and dAE $x$-dimensions, respectively). The metrics for each label were averaged and weighted by the number of true instances to account for label imbalance, thus the $f$-measure can give scores that are not between precision and recall. Clusters of low layer representations were as good as those of high layer representations. Since the first dAE layer is a nonlinear principal component analysis, we can say that higher layers of the stack do not affect the ability to separate the data while reducing dimensionality. b, Sums of the L2 distances between data points and the centre of their assigned $k$-means cluster are plotted. This is the same measure that was minimized by the clustering algorithm. The labels on the $y$ axis follow the convention used in a. Clustering the representations after each layer showed how the data became more and more separable at higher layers. c, The plot shows the distribution of the sum of the normalized transcription factor profile signal for each class assigned by the model.
## Extended Data Table 1 | Replication timing data sets used in this study

| Species          | Cell Type | Cell Line | Data Type | Replicate | Citation  |
|------------------|-----------|-----------|-----------|-----------|-----------|
| Human            | ESC       | BG01      | Repl-chip | 1         | 4         |
| Human            | ESC       | BG02      | Repl-chip | 1         | 4         |
| Human            | ESC       | BG02      | Repl-chip | 2         | 4         |
| Human            | ESC       | BG02      | Repl-seq  | 1         | 3         |
| Human            | ESC       | H1        | Repl-chip | 1         | RD*       |
| Human            | ESC       | H7        | Repl-chip | 1         | 4         |
| Human            | ESC       | H9        | Repl-chip | 1         | 4         |
| Human            | ESC       | iPSC5     | Repl-chip | 1         | 4         |
| Human            | ESC       | iPSC5     | Repl-chip | 1         | 4         |
| Human            | Endomesoderm | BG02   | Repl-chip | 1         | 36        |
| Human            | Endomesoderm | BG02   | Repl-chip | 2         | 36        |
| Human            | Endoderm  | BG02      | Repl-chip | 1         | 36        |
| Human            | Endoderm  | BG02      | Repl-chip | 2         | 36        |
| Human            | Mesoderm  | BG02      | Repl-chip | 1         | 36        |
| Human            | Mesoderm  | BG02      | Repl-chip | 2         | 36        |
| Human            | Smooth Muscle | BG02 | Repl-chip | 1         | 36        |
| Human            | Smooth Muscle | BG02  | Repl-chip | 2         | 36        |
| Human            | Lung Fibroblast | IMR90 | Repl-chip | 1         | 38        |
| Human            | Lung Fibroblast | IMR90 | Repl-seq  | 1         | This study|
| Human            | Foreskin Fibroblast | BJ | Repl-seq  | 1         | 3         |
| Human            | Cervical Carcinoma | HeLa_S3 | Repl-chip | 1         | RD*       |
| Human            | Cervical Carcinoma | HeLa_S3 | Repl-seq  | 1         | This study|
| Human            | Lymphoblastoid | GM06990 | Repl-chip | 1         | 36        |
| Human            | Lymphoblastoid | GM06990 | Repl-chip | 2         | 36        |
| Human            | Lymphoblastoid | GM06990 | Repl-seq  | 1         | 3         |
| Human            | Lymphoblastoid | GM12812 | Repl-seq  | 1         | This study|
| Human            | Lymphoblastoid | GM12813 | Repl-seq  | 1         | This study|
| Human            | Lymphoblastoid | GM12878 | Repl-seq  | 1         | This study|
| Human            | Lymphoblastoid | CO202    | Repl-chip | 1         | 4         |
| Human            | Lymphoblastoid | CO202    | Repl-chip | 2         | 4         |
| Human            | Leukemia   | K562      | Repl-seq  | 1         | 3         |
| Human            | T lymphocyte | TC       | Repl-chip | 1         | 35        |
| Human            | NPC        | BG01      | Repl-chip | 1         | 4         |
| Human            | NPC        | BG01      | Repl-chip | 2         | 4         |
| Human            | Neuroblastoma | SK-N-SH | Repl-seq  | 1         | This study|
| Human            | Myoblast   | CM1       | Repl-chip | 1         | 37        |
| Human            | Myoblast   | CM5       | Repl-chip | 2         | 37        |
| Human            | Keratinocyte | NHEK     | Repl-seq  | 1         | This study|
| Human            | Adenocarcinoma | MCF-7   | Repl-seq  | 1         | This study|
| Human            | Umbilical Vein Endothelium | HUVEC | Repl-seq  | 1         | This study|
| Human            | Hepatocellular Carcinoma | HepG2 | Repl-seq  | 1         | This study|
| Mouse            | ESC2i      | Suz12KO   | Repl-chip | 1         | This study|
| Mouse            | ESC2i      | Suz12wi   | Repl-chip | 1         | This study|
| Mouse            | B cell lymphoma | CH12 | Repl-chip | 1         | This study|
| Mouse            | Endoderm   | NA        | Repl-chip | 1         | 34        |
| Mouse            | Endoderm   | NA        | Repl-chip | 2         | 34        |
| Mouse            | EpiSC      | EpiSC7    | Repl-chip | 1         | 34        |
| Mouse            | EpiSC      | EpiSC7    | Repl-chip | 2         | 34        |
| Mouse            | Erythroleukemia | MEL    | Repl-chip | 1         | This study|
| Mouse            | Erythroleukemia | MEL    | Repl-chip | 2         | This study|
| Mouse            | ESC        | 46C       | Repl-chip | 1         | 2         |
| Mouse            | ESC        | D3        | Repl-chip | 1         | 2         |
| Mouse            | ESC        | D3        | Repl-chip | 2         | 2         |
| Mouse            | ESC        | TT2       | Repl-chip | 1         | 2         |
| Mouse            | ESC        | TT2       | Repl-chip | 2         | 2         |
| Mouse            | iPSC       | NA        | Repl-chip | 1         | 2         |
| Mouse            | iPSC       | NA        | Repl-chip | 2         | 2         |
| Mouse            | iPSC       | 1D4       | Repl-chip | 1         | 34        |
| Mouse            | iPSC       | 2D4       | Repl-chip | 1         | 34        |
| Mouse            | Lymphocytic Leukemia | L1210 | Repl-chip | 1         | 34        |
| Mouse            | Lymphocytic Leukemia | L1210 | Repl-chip | 2         | 34        |
| Mouse            | Mammary Carcinoma | C127  | Repl-chip | 1         | 39        |
| Mouse            | Mammary Carcinoma | C127  | Repl-chip | 2         | 39        |
| Mouse            | MEF        | F         | Repl-chip | 1         | 34        |
| Mouse            | MEF        | M         | Repl-chip | 1         | 34        |
| Mouse            | Mesoderm   | NA        | Repl-chip | 1         | 34        |
| Mouse            | Mesoderm   | NA        | Repl-chip | 2         | 34        |
| Mouse            | Myoblast   | J185a     | Repl-chip | 1         | 34        |
| Mouse            | NPC        | 46C       | Repl-chip | 1         | 2         |
| Mouse            | NPC        | TT2       | Repl-chip | 1         | 2         |
| Mouse            | NPC        | TT2       | Repl-chip | 2         | 2         |
| Mouse            | NPC        | D3        | Repl-chip | 1         | 2         |
| Mouse            | NPC        | D3        | Repl-chip | 2         | 2         |
| Mouse            | iPSC       | 1A2       | Repl-chip | 1         | 34        |
| Mouse            | iPSC       | 1B3       | Repl-chip | 1         | 34        |
| Mouse            | iPSC       | V3        | Repl-chip | 1         | 34        |

*RD indicates data set is publicly available at [http://www.replicationdomain.com](http://www.replicationdomain.com).