doi:10.1038/nature13986

Supplementary Discussion

Previous comparisons of TADs and LADs with replication-timing data

TTRs are regions of reduced origin firing, within which replication forks progress in a predominantly unidirectional manner until they encounter a later firing cluster of origins[51,52]. We reasoned that the incomplete alignment between TADs and TTRs could stem from the fact that TTRs only form between adjacent replication domains with differential replication timing, whereas in any given cell type, many adjacent replication domains replicate at similar times[53,54] preventing the detection of a TTR between them[2]. Thus, a complete annotation of TTR positions requires their identification in many different cell types. Moreover, TTRs are functionally defined by the kinetics of replication and are much larger [median size = 300-500kb[4,52]] and variable in size than the boundaries of structurally defined TADs [median boundary size < 40 kb[8]] or LADs [boundary size ≅ 1 kb[40]]. Presumably, the true structural boundaries of replication domains are also much smaller than TTRs. Thus, localizing the replication domain boundary along the length of each TTR from many cell types was critical to evaluate the relationship of replication domains to TADs and LADs.

TAD and replication domain boundary alignment

We reasoned that detection of TAD boundaries at TTR-absent replication domain boundaries might require increased resolution relative to TTR-present replication domain boundaries, since adjacent TADs that replicate at the same time interact more frequently, obscuring the TAD boundary. To determine the extent to which resolution accounted for weaker TAD alignment to TTR-absent versus TTR-present boundaries, we used higher-resolution data[47] to call new TAD boundaries. An increased fraction of the "TTR-absent" replication domain boundaries aligned to these new calls while the percentage of TAD boundaries that aligned to replication domain boundaries remained significantly high (**Extended Data Fig. 4c**), demonstrating that the new TAD boundaries detected at higher resolution are located at replication domain boundaries and suggesting that even higher resolution data would yield a more complete list of TAD boundaries and further improve the alignment of replication domains and TADs.

Since many replication domain boundaries did not align even with the higher-resolution TAD boundary calls, we also searched for replication domain boundary subsets with properties that predicted TAD boundary alignment. In the course of this exercise, we discovered a general relationship between replication timing and TAD frequency, i.e. TADs in earlier-replicating regions were greater in number and therefore smaller than TADs in later-replicating regions (**Extended Data Fig. 5a**). Accordingly, replication timing of replication domain boundaries correlated with their alignment to TADs, but alignment remained significant for all replication domain boundary subsets (left plot in **Extended Data Fig. 5b**). Importantly, while the relationship between replication timing and TAD frequency was clear for regions with constitutive replication timing, intermediate TAD frequency was observed for regions with opposite timing

in H1-hESCs versus IMR90 cells with very little difference in TAD frequency between early and late regions (**Extended Data Fig. 5c**), again suggesting stable TAD structure. Still, TAD frequency mapped in cells where these switching regions replicated early was increased relative to TAD frequency for the exact same regions when mapped in cells where they are replicated late (**Extended Data Fig. 5c**). Although cell-type-specific changes to TAD structure cannot be ruled out, these results suggest late-replicating TAD boundaries are more difficult to detect using current mapping methods and are underrepresented in existing annotations.

Timing difference and genomic length across TTRs also correlated with TAD alignment, but, like replication timing, all subsets based on these properties aligned significantly to TADs (middle and right plots in **Extended Data Fig. 5b**). Strikingly, by considering developmental changes to replication timing or TTR size within all the cell types for which we data, clear thresholds emerged that distinguished replication domain boundaries that aligned with TADs from those that did not and the same thresholds applied both to TTR-present and TTR-absent boundaries (**Extended Data Fig. 5d**). Demonstrating that these thresholds were independent of TTR properties in IMR90, we directly compared replication domain boundaries whose TTR size in IMR90 was below the threshold and only observed TAD alignment to replication domain boundaries that exceeded the threshold in non-IMR90 cell types (**Extended Data Fig. 5e**). These results indicated that small timing transitions do not reliably reflect underlying TAD structure, but may be more closely associated with local chromatin features. Moreover, the influence of TTR properties in non-IMR90 cell types over the alignment of IMR90 TADs provided another line of evidence strongly suggesting TADs are stable during development.

Enriched marks at TTR-present replication domain boundaries

Since we did not observe strong enrichment of SINEs and CTCF at replication domain boundaries we reasoned that differential enrichment of features within TADs (e.g. early versus late) could have created the appearance of local enrichment at TAD boundaries. To test this, we reanalyzed TAD boundaries after orienting them such that the analyzed feature exhibited a decreasing density from left to right (**Extended Data Fig. 8a**). When TAD boundaries were oriented in this manner we found only minimal enrichment of SINEs at TAD boundaries in mouse (**Extended Data Fig. 8b**) and human (**Extended Data Fig. 8c**). Instead, we observed differential enrichment of SINEs among different TADs in both mouse and human, consistent with previous reports of SINE enrichment in early replication domains[55,56]. Moreover, examination of individual TAD boundaries confirmed local enrichment of SINEs was limited to a small fraction of TAD boundaries in both human and mouse (3-7%, **Extended Data Fig. 8b-c**). However, CTCF and marks associated with active transcription (e.g. H3K4me3, RNA polymerase II, H3K36me3) were still locally enriched on average at TAD boundaries in human and mouse (**Extended Data Fig. 8d-e**). Limitations of the current method for replication timing analysis and replication domain boundary calling may have prevented precise alignment to these marks.

Transcription factor prediction model

We assigned each TAD into data driven classes using different approaches. We found that a stack of denoising autoencoders[57], which is a powerful data reduction technique for unsupervised machine learning[58], gave the best clustering of our data over popular clustering and dimensionality reduction methods (**Extended Data Fig. 9**). The sizes of molecularly defined replication domains are consistent with cytogenetically defined, punctate sites of labeled DNA synthesis (replication foci), which retain their punctate structure when chased through many cell divisions, suggesting that replication foci are stable structural units of chromosomes and may represent replication domains measured molecularly[59,60]. Since cytogenetic patterns of replication suggest that S phase contains more than two stages, it may be possible to further subdivide these two high-level classes. However, previous correlations indicate relationships between chromatin composition and middle or late S phase can be cell-type specific[61]. Our findings support a model in which early replicating TADs correspond to the interiorly localized replication foci, while TADs encompassing TTRs or mid-late replicating regions correspond to mid- and late-replicating foci, interacting in different cell cycles with either the nuclear lamina, the nucleolus or other as yet undefined internally located heterochromatin clusters[19–21].

Additional References

51. Zhou, J., Ermakova, O. V., Riblet, R., Birshtein, B. K. & Schildkraut, C. L. Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells. *Mol. Cell. Biol.* **22,** 4876–4889 (2002).
52. Guilbaud, G. *et al.* Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* **7,** e1002322 (2011).
53. Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H. & Cardoso, M. C. DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. *Mol. Cell* **10,** 1355–1365 (2002).
54. Maya-Mendoza, A., Olivares-Chauvet, P., Shaw, A. & Jackson, D. A. S Phase Progression in Human Cells Is Dictated by the Genetic Continuity of DNA Foci. *PLoS Genet.* **6,** e1000900 (2010).
55. Holmquist, G. P. & Caston, L. A. Replication time of interspersed repetitive DNA sequences in hamsters. *Biochim. Biophys. Acta* **868,** 164–177 (1986).
56. Ashida, H., Asai, K. & Hamada, M. Shape-based alignment of genomic landscapes in multi-scale resolution. *Nucleic Acids Res.* **40,** 6435–6448 (2012).
57. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. in 1096–1103 (ACM Press, 2008). doi:10.1145/1390156.1390294
58. Bengio, Y., Courville, A. & Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35,** 1798–1828 (2013).
59. Sparvoli, E., Levi, M. & Rossi, E. Replicon clusters may form structurally stable complexes of chromatin and chromosomes. *J. Cell Sci.* **107 (Pt 11),** 3097–3103 (1994).
60. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140,** 1285–1295 (1998).
61. Wu, R., Terry, A. V., Singh, P. B. & Gilbert, D. M. Differential subnuclear localization and replication timing of histone H3 lysine 9 methylation states. *Mol. Biol. Cell* **16,** 2872–2881 (2005).