An atlas of lamina-associated chromatin across thirteen human cell types reveals cell-type-specific and multiple subtypes of peripheral heterochromatin

Kathleen C. Keough[a,b]*, Parisha P. Shah[c]*, Nadeera M. Wickramasinghe[d], Carolyn E. Dundes[e], Angela Chen[e], Rachel E.A. Salomon[e], Sean Whalen[b], Kyle M. Loh[e], Nicole Dubois[d], Katherine S. Pollard[a,b,f]**, Rajan Jain[c]**

[a]University of California, San Francisco, CA 94117, USA; [b]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA 94158, USA;[c] Departments of Medicine and Cell and Developmental Biology, Penn CVI, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [d]Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; [e]Department of Developmental Biology and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305 USA; [f]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

Corresponding authors: Katherine S. Pollard and Rajan Jain

Katherine S. Pollard
Gladstone Institutes
1650 Owens Street
San Francisco, CA 94158
415-732-2711
kpollard@gladstone.ucsf.edu

Rajan Jain
09-102 Smilow TRC
3400 Civic Center Blvd
Philadelphia, PA 19104
215-573-3011
jainr@pennmedicine.upenn.edu

*These authors contributed equally to this work.
Running title: (above)
Keywords: Lamin-associated domains, peripheral chromatin organization, heterochromatin, chromatin compartmentalization, 3D genome, cellular differentiation and development.

Abbreviations

AIC: Akaike Information Criterion
CE: constitutive early replication domain
CL: constitutive late replication domain
CM: cardiomyocyte
CTI: cell-type-invariant
ESC: embryonic stem cell
Kb: kilobases (1000 basepairs)
KDD: K9-dimethyl domain, or H3K9me2-associated domain
LAD: lamina associated domain
LB1: LAMIN-B1
LTR: long terminal repeat
mESC: mouse embryonic stem cell
T1-LAD: Type 1 LAD
T1-LAD/KDDs: Type 1 LAD overlapping a KDD
T2-LAD: Type 2 LAD
T2-LAD/KDDs: Type 2 LAD overlapping a KDD
TE: transposable element

Abstract

Three-dimensional genome organization, specifically organization of heterochromatin at the nuclear periphery, coordinates cell type-specific gene regulation. While defining various histone modifications and chromatin-associated proteins in multiple cell types has provided important insights into epigenetic regulation of gene expression and cellular identity, peripheral heterochromatin has not been mapped comprehensively and relatively few examples have emerged detailing the role of peripheral heterochromatin in cellular identity, cell fate choices, and/or organogenesis. In this study, we define nuclear peripheral heterochromatin organization signatures based on association with LAMIN B1 and/or dimethylation of lysine 9 on H3 (H3K9me2) across thirteen human cell types encompassing pluripotent stem cells, intermediate progenitors and differentiated cells from all three germ layers. Genomic analyses across this atlas reveal that lamin-associated chromatin is organized into at least two different compartments, defined by differences in genome coverage, chromatin accessibility, residence of transposable elements, replication timing domains, and gene complements. Our datasets reveal that only a small subset of lamin-associated chromatin domains are cell type invariant, underscoring the complexity of peripheral heterochromatin organization. Moreover, by integrating peripheral chromatin maps with transcriptional data, we find evidence of cooperative shifts between chromatin structure and gene expression associated with each cell type. This atlas of peripheral chromatin provides the largest resource to date for peripheral chromatin organization and a deeper appreciation for how this organization may impact the establishment and maintenance of cellular identity.

Introduction

Adult human bodies are composed of trillions of cells, comprising more than 200 distinct cell types, which are faithfully established and maintained throughout a healthy lifespan. Identifying and understanding molecular mechanisms regulating cellular identity establishment and maintenance over time are areas of intense interest. In particular, it remains incompletely understood how coordinated genome-wide transcriptional regulation is achieved during development and/or physiologic tissue homeostasis.

Three-dimensional genome organization has emerged as a potential mechanism to coordinate cell type specific gene regulation and maintain cell type transcriptional fidelity (Rowley and Corces, 2018; Zheng and Xie, 2019). In particular, genome organization at the nuclear periphery may provide a key platform for cell type specific transcription (Buchwalter et al., 2019). The nuclear lamina is a filamentous network of LAMIN A/C, B1, and B2 proteins residing on the inner surface of the nuclear envelope (Burke and Stewart, 2006; Worman and Bonne, 2007). Approximately 30-40% of the genome in most cell types is localized to the lamina, in kilobase- to megabase-sized blocks termed lamina-associated domains (LADs) (Guelen et al., 2008). These loci are generally heterochromatic, and genes within LADs are frequently transcriptionally repressed and undergo active silencing, while genes away from the lamina are more often competent for transcriptional activation (Briand and Collas, 2020). Mutations in nuclear lamins disrupt LAD organization and can cause disease (Briand et al., 2018; Lee et al., 2019b; Vadrot et al., 2015; Worman, 2012). In rod photoreceptor cells, radial chromatin organization is naturally inverted, and this inversion can be generated by ablation of nuclear lamin genes, while

in Lamin B receptor-null thymocytes chromatin is only partially inverted, suggesting an active but complex role for lamins in anchoring LADs to the nuclear periphery (Falk et al., 2019; Solovei et al., 2013). The degree of conservation of LADs between cell types and species underscores the importance of these genomic regions (Guelen et al., 2008; Meuleman et al., 2013).

Spatial LAD positioning regulates organogenesis and transcriptional repression (Peric-Hupkes et al., 2010; Poleshko et al., 2017; Robson et al., 2016). In particular, subsets of LADs are repositioned away from or to the lamina during differentiation in a cell type-specific manner (Malik et al., 2010; Peric-Hupkes et al., 2010). In a mouse embryonic stem cell (mESC) neuronal differentiation model, key neuronal genes lose lamina occupancy in neuronal precursor cells (Meuleman et al., 2013). A similar phenomenon is observed during mESC cardiac differentiation, and preventing normal repositioning of LAD-bound chromatin results in precocious and aberrant cardiac differentiation (Poleshko et al., 2017). These studies demonstrate the biological significance of peripheral chromatin organization and changes therein, illustrating how changes to this organization can directly impact normal cellular behavior.

Previous studies suggest that a subset of LADs have varying characteristics, such as reduced lamin occupancy and increased gene density, indicating that LADs are heterogenous. Indeed, work in single cells has shown the contact frequencies of LADs with the nuclear lamina vary by locus, and correlate with gene density, suggesting a structural role for LADs with higher contact frequency (Kind et al., 2015; Meuleman et al., 2013). As described above, only a subset of LADs re-position away from the lamina during differentiation, and individual genomic regions

have varying probabilities of becoming re-localized to or from the nuclear lamina (Kind et al., 2013, 2015), suggesting that LADs may segregate into stable or “constitutive” versus dynamic or “facultative” subtypes. Moreover, chromatin at the nuclear periphery is frequently marked by the histone modification H3K9me2, and genomic loci enriched for this signal have been shown to have substantial overlap with LADs (Poleshko et al., 2017). Together, these studies raise the intriguing possibility that peripheral heterochromatin may comprise distinct types defined by unique characteristics. Integrative modeling of various chromatin marks has characterized multiple subtypes of chromatin, including heterochromatin, having varying gene density and expression (Filion et al., 2010; Wang et al., 2020). Defining various categories of peripheral chromatin domains across multiple human cell types, and characterizing distinct subtypes of chromatin, will provide critical knowledge about how nuclear organization regulates cellular identity.

Here, we have defined nuclear organization signatures based on the binding of LAMIN B1 (LB1) and histone H3 lysine 9 dimethylation (H3K9me2) occupancy across thirteen human cell types from all three germ layers and embryonic stem cells (ESCs). We identify two types of lamina-associated chromatin in all examined cell types, which have distinctive molecular features that suggest they are physically and functionally separate. By linking the organizational maps with transcriptional data, we identify cooperative shifts between chromatin structure and gene expression associated with each cell type. Overall, this work provides the largest atlas of human peripheral chromatin to date, and reveals critical insights into the complexity of peripheral heterochromatin organization.

Results

A 3-state Hidden Markov Model approach identifies two types of LADs

We generated ChIP-seq datasets for LB1 from human ESCs and thirteen ESC-derived cell types from the three germ layers (endoderm, mesoderm and ectoderm), representative of multiple early differentiation trajectories (Ang et al., 2018; Bardot et al., 2017; Loh et al., 2014a, 2016; Martin et al., 2020) (Fig. 1A; Table 1). These data were divided into two analysis tiers based on quality metrics including sequencing coverage and Spearman correlation between replicates, with tier one being highest quality (Table 1). Visual inspection of these data confirmed the presence of large, discrete domains of enrichment of LB1 signal consistent with the presence of LADs in all cell types investigated (Fig. 1B).

We implemented a Hidden Markov Model (HMM) trained on the tier one cell types in 20kb bins (Table 1) to identify LADs based on the LB1 ChIP-seq datasets. To our knowledge, this is the first use of HMMs to call LADs from ChIP-seq data. HMMs have previously been used to segment chromosomes into LAD versus non-LAD regions by summarizing diffuse DamID-generated LB1 binding profiles into two discrete states (Filion et al., 2010; Meuleman et al., 2013). Our key innovations were to include additional states and to apply this approach to LB1 ChIP-seq data. In order to minimize user-generated or cell type-specific bias, we initialized the state calls using k-means and began model training with a uniform transition matrix (see Methods). This is in contrast to other methods to identify LADs, such as broad peak callers like

Enriched Domain Detector (Lund et al., 2014), which generally require extensive user-defined parameter tuning that dramatically influences the resultant LADs identified. We found that a three-state HMM fit the data best compared to two-, four- and five-state models based on Akaike Information Criterion (AIC), visual inspection, region size and genome coverage (Fig. 2A, Table 2). States are of course a discretization of a quantitative signal, but this statistical support for three states suggests that LB1 binding in 20kb bins is generally at three different levels. We designated the three states as Type 1 LAD (T1-LAD), Type 2 LAD (T2-LAD) and nonLAD in order of descending LB1 signal (Fig. 2A, B, SFig. 1). Median T1-LAD sizes for tier one cell types ranged from 160-280kb, covering 21.4%-38.3% of the genome (Table 3). Compared to nonLADs, T1- and T2-LADs in each cell type demonstrated characteristic features including increased occupancy of the repressive histone mark H3K9me2, lower gene density, gene expression, GC content and accessibility via ATAC-seq, and greater B compartment overlap (Fig. 2B-G, SFig. 2-7). To expand these data, we also generated corresponding ChIP-seq datasets for H3K9me2, a histone modification enriched in chromatin at the nuclear periphery (Kind et al., 2015; Poleshko et al., 2017, 2019). As expected, the H3K9me2 data also showed broad domains of enrichment having a high degree of overlap with LB1 signal (discussed below). Overall, these results confirmed that an HMM-based analysis is able to define two types of LADs based on LB1 ChIP-seq data.

T1- and T2-LADs have distinct genomic features

T1-LADs appear to be distinct from T2-LADs, with T1-LADs showing the most LB1 and H3K9me2 enrichment, the lowest gene density, expression, GC content, and chromatin

accessibility, and greater B compartment overlap (Fig. 2B-G, SFig. 2-7). On aggregate, these differential characteristics suggest that T1-LADs may be organized separately from or differently than T2-LADs. Previous work has demonstrated that in each single cell nucleus, only ~30% of LADs are positioned at the periphery (Kind et al., 2013); in our data, across cell types a median of 30% of regions classified as LADs are T1-LADs. This led us to hypothesize that T2-LADs are more variable between single cells in a bulk population, thus driving their overall lower LB1 association in population data. We investigated this using previously published DamID data for LB1 in KBM7 human myeloid leukemia cells (Kind et al., 2015) and paraxial mesoderm, the most comparable cell type from our dataset. DamID probes that overlapped T1-LADs were associated with LB1 in the greatest percentage of KBM7 cells, followed by T2-LADs (Fig. 3A). This suggests that T2-LADs may be a more spatially dynamic or variable entity, within a population of cells.

It has been previously demonstrated that heterochromatin is generally correlated with later replication timing, and euchromatin with earlier replication timing (Hiratani et al., 2009; Ma and Duan, 2019; Rhind and Gilbert, 2013; Wang et al., 2020). We hypothesized that if T1- and T2-LADs are distinct sub-compartments, the more stable and repressed T1-LADs would replicate later than the more accessible T2-LADs. Thus, we measured the enrichment of replication timing domains defined by their consistency across cell types in T1- and T2-LADs: constitutive early (CE), constitutive late (CL), and switch domains, which have differing replication timing between different cell types (Dixon et al., 2018). We observed that T1-LADs are consistently and highly depleted for CE domains, slightly depleted for switch domains, and enriched for CL domains across cell types compared to random expectation (permutation test; Fig. 3B). In contrast, T2-LADs are generally depleted for both CE domains and CL domains and

slightly enriched for switch domains (Fig. 3B). Thus, the replication timing assessments indicate that T1-LADs likely replicate after T2-LADs across cell types and that T2-LADs are associated with replication timing switching between different cell types. In addition to nuclear organization and physical chromatin separation, replication timing is also associated with a host of functional implications, including copy number variation and transcriptional silencing (Marchal et al., 2019), suggesting that T1- and T2-LADs may be further distinguished by additional characteristics.

If T1- and T2-LADs are distinct subcompartments of peripheral heterochromatin, we hypothesized that they may be associated with different molecular mechanisms of establishment and/or maintenance. Previous work suggests that enrichment of the insulator factor CTCF at LAD boundaries (Guelen et al., 2008) and topologically-associated domain (TAD) boundaries is an important mechanism for maintenance of higher order genome organization (Dixon et al., 2012; Nora et al., 2012, 2017). We observed that T1-LADs and T1-LAD boundary regions have less CTCF binding than surrounding chromatin (Fig. 3C). T2-LADs showed similar levels of CTCF binding within the domains relative to the surrounding genome, but displayed peaks of CTCF binding at their boundaries (Fig. 3D). It has been shown that switch (aka “transition”) replication timing domains are enriched for CTCF (Pope et al., 2014). This result is concordant with the enrichment of switch domains in T2-LADs, and may reflect a role for CTCF in the generation or maintenance of T2- but not T1-LADs.

Transposable elements (TEs) have been shown to influence gene regulation via participation in genome folding and generation of regulatory elements, with expansions of TEs influencing evolution of gene expression across species (Sundaram and Wysocka, 2020). Specific TEs have previously been shown to be enriched within LADs, and LADs are postulated to be a key

element in multiple mechanisms that many species have evolved to silence and limit accessibility to highly repetitive DNA and other TEs (Hollister and Gaut, 2009; Meuleman et al., 2013). However, TEs have also been implicated in various genomic regulatory functions and are linked to species-specific transcription-factor binding sites in mammals (Bourque et al., 2008; Kunarso et al., 2010; Wang et al., 2007), as well as other speciation events via their impact on 3D genome folding (Choudhary et al., 2020; Fudenberg et al., 2019). TEs are further implicated in the genesis and regulation of long noncoding RNAs (Kapusta et al., 2013; Kelley and Rinn, 2012). Therefore, we postulated that variable enrichment of TEs in T1- and T2-LADs could reveal distinct, putatively functional characteristics.

We assessed the enrichment of various classes and families of TEs across T1- and T2-LADs (see Methods), and found that many types of TEs show divergent patterns of enrichment (Fig. 3E-F, SFig. 8). Notably, short interspersed nuclear elements (SINEs) are depleted in T1-LADs and enriched in T2-LADs; conversely, long interspersed nuclear elements (LINEs) are enriched in T1-LADs and depleted in T2-LADs, with other categories (such as DNA and simple repeats) also showing some distinction between T1- and T2-LADs (SFig. 8). We then assessed different TE types within these broader categories (SFig. 9). Of interest, L1-LINES, which are active but mostly silenced in the human genome, are enriched in T1- and depleted in T2-LADs (Fig. 3F). SVAs (non-autonomous retroelements within the primates) and Alu elements (a family of SINEs), both transcriptionally active in humans, are depleted in T1-LADs, suggesting that they are not regulated by general LAD-mediated sequestration and/or silencing.

We next assessed enrichment for various transcription factor binding motifs (TFBMs), in the context of the differential enrichment of TEs in T1-LADs versus T2-LADs, to determine whether

they were enriched in T1- or T2-LADs. Of interest, SINEs (enriched in T2-LADs) are known to be enriched for YY1 binding sites, a known architect of chromatin structure (Beagan et al., 2017; Harr et al., 2015a; Weintraub et al., 2017). Consistent with this, in ESCs and CMs we observed an enrichment of YY1 TFBMs in T2- compared to T1-LADs (Table 4). Alu elements (depleted in T1-LADs) are enriched for multiple TFBMs involved in maintaining and/or establishing cellular identity (Table 4), including motifs for NKX2-2, PITX1, TBX5, and NKX2-5 (Briscoe et al., 1999; Kathiriya et al., 2020; Luna-Zurita et al., 2016; Szeto et al., 1999; Targoff et al., 2013). Likewise, SVAs are enriched for TFBMs for SMAD2 and EAR2 (Table 4). This is in slight contrast to L1-LINEs which are enriched in T1 LADs. In our analyses, L1-LINEs are enriched for TFBMs of factors with known roles in stemness and genome structure, including MNT (Hurlin et al., 1997) and various SOX proteins (Hadjimichael et al., 2015), as well as multiple TFBMs of factors associated with cancer, including HOXA9 (Sun et al., 2018) and PU.1 (Song et al., 2015). Together, these differential TE enrichments and the associated TFBM assessments suggest that while T1-LADs may often act in a “classic” LAD mechanism to sequester and transcriptionally regulate TFBMs associated with development and cancer, there may also be an active structural genome regulation role for TEs in T1-LADs. Additionally, these results indicate a greater role in cell type-specific gene regulation by T2- compared to T1-LADs, supported by the enrichment of cell type-specific TFBMs in T2-LADs (Table 4). Taken together with the replication timing and CTCF binding analyses, these data show distinct genomic characteristics between T1- and T2-LADs, supporting a model in which peripheral heterochromatin across many cell types is sub-compartmentalized.

A 2-state Hidden Markov Model identifies H3K9me2 domains

H3K9me2 ChIP-seq also showed broad domains of enrichment, which had a high degree of overlap with LB1 signal. In order to define these domains, we trained another HMM on the H3K9me2 ChIP-seq tier one data in 20kb bins, finding a two-state model to be the best fit to the data based on AIC, avoiding artificial division of domains by visual inspection, and assessment of region size (Fig. 4A, Table 3,5). The state with higher H3K9me2 signal was assigned the label of H3K9me2-associated domain ("K9-dimethyl domain", abbreviated as "KDD"), and the other state "nonKDD" (Fig. 4A,C). Median KDD sizes for tier one cell types ranged from 380-2360kb, covering 44.2%-82.8% of the genome (Table 3). In addition to high levels of H3K9me2 occupancy, KDDs in each cell type demonstrated higher levels of association with LB1, lower gene density, generally lower gene expression, GC content, and overlap with ATAC-seq peaks, and greater overlap with the B compartment compared to nonKDDs (Fig. 4B-G, SFig. 10-16). These data confirm a high degree of similarity between KDDs and LADs across multiple cell types.

LADs and KDDs overlap but have distinct characteristics

Based on previous work demonstrating that LADs and KDDs are generally repressive peripheral chromatin compartments, and our initial observations of LAD and KDD similarity, we hypothesized that KDDs would be highly overlapping with T1- and T2-LAD compartments. In support of this, we found that the vast majority of KDDs (>99% for most cell types) overlap T1- or T2-LADs, and most T1-LADs (>90% for 7 of 8 tier one cell types) are in KDDs. Exceptions,

such as in the case of definitive ectoderm where only 10% of T1-LADs are in KDDs, likely stem from greater noise present in LB1 ChIP-seq data for some cell types, leading to instances where loci that are likely T1-LADs were instead annotated as T2-LADs. Interestingly, a median of only 47% of T2-LADs overlap KDDs across all cell types assessed in this study. Additionally, T1- and T2-LADs that overlap KDDs (T1-LAD/KDDs or T2-LAD/KDDs) trend larger than T1- and T2-LADs that do not overlap KDDs (Table 3). Together, these results suggest that while there is a high degree of overlap between LB1- and H3K9me2-associated peripheral chromatin, there are also regions associated with either LB1 or H3K9me2 alone, which may imply other subtypes of peripheral chromatin.

Given our observations that T1-LADs are the more repressive LAD subtype, we hypothesized that T1-LAD/KDDs may represent a particularly repressive peripheral chromatin compartment. Indeed, loci located in T1-LAD/KDDs have higher LB1 and H3K9me2 ChIP-seq signal compared to loci in T1-LADs without KDD overlap (T1-only LADs) and KDDs without LAD overlap (KDD-only) (Fig. 5A, B). Furthermore, T1-LAD/KDDs have lower gene density, lower proportion of expressed genes (Fig. 5C, D), and reduced accessibility (as measured by overlap with ATAC-seq peaks) (Fig. 5E, SFig. 17-19). However, T1-only LAD compartments demonstrated even lower gene density and ATAC overlap, and greater B compartment overlap relative to T1-LAD/KDDs and KDDs-only, suggesting a greater role for gene regulation in T1-LAD/KDDs and a different, perhaps structural, role for T1-LAD only domains (Fig. 4F, SFig. 17, 19-20).

Given the overlap between both types of LADs and KDDs, we hypothesized that KDDs, particularly T1-LAD/KDDs, would be enriched for CL domains. As expected, we found that

KDDs are generally depleted for CE domains (as noted generally for both T1- and T2-LADs); however, they showed mixed results for CL and switch domains (SFig. 21,22). Investigating further, we found that T1-LAD/KDDs and KDDs without T2-LAD overlap specifically are enriched for CL replication timing domains in the majority of cell types, but T2-LAD/KDDs and KDDs without T1-LAD overlap specifically are enriched for switch domains. These differences corroborate our data that T2-LADs are more dynamic than T1-LADs, and that T2-LAD/KDDs are an intermediate state between between early replicating open chromatin and later replicating T1-LAD/KDD chromatin. As above, we then further assessed CTCF occupancy across KDDs. KDDs demonstrate features of both T1- and T2-LADs with regards to CTCF. Overall, KDDs have less CTCF binding relative to the surrounding chromatin, similar to T1-LADs, and display a peak of binding at domain boundaries, similar to T2-LADs (SFig. 23,24). KDDs that overlap each type of LAD showed CTCF signal similar to that of each LAD type alone. Of note, the CTCF binding motif is cytosine-rich, and the regions surrounding KDDs with or without LAD overlap display increased CTCF binding relative to the regions surrounding T1-only or T2-only LADs, perhaps reflecting the AT-rich characteristic of LADs.

Taken together, our findings support a model in which T1-LADs are a highly repressed compartment while T2-LADs are a more dynamic and variable entity linked to cell-type-specifying processes, with KDDs linked to repression of genes in T1 and T2 LADs. Specifically, we observe that 1) KDDs overlap T1- and T2-LADs, 2) T1-LAD/KDDs comprise a particularly repressed peripheral chromatin sub-compartment relative to T1-LADs-only and KDDs-only, and 3) each of the three compartments (T1-LADs, T2-LADs, and KDDs) have distinct genomic characteristics, despite a high degree of overlap.

Cell type specificity in T1-LADs, T2-LADs and KDDs

Previous studies have also shown that many LADs are shared between cell types (Meuleman et al., 2013; Robson et al., 2016). However, these studies are often limited to comparisons between 2-3 cell types along a single differentiation pathway, and we hypothesized that the greater representation of cell types in our dataset would provide deeper insight into the proportion of LADs that are invariant (across all 13 cell types). We assessed the representation of individual genomic regions (in 20kb segments) across T1- and T2-LADs in all 13 cell types. We observed that the vast majority of genomic regions overlapping T1- or T2-LADs were not shared in all 13 cell types (Fig. 6A, B), indicating a lower degree of cell type invariant (CTI) LADs than previous studies have suggested, which may be attributed to the wide range of developmental cell types in this study. This assessment also revealed major differences in the distribution of genomic occupancy between the two LAD subtypes. First, approximately, 50% of the genome is not classified as a T1-LAD across all the 13 cell types (Fig. 6A). In contrast, a greater proportion of the genome is classified as a T2-LAD in at least one cell type (Fig. 6B). This suggests that T1-LADs, even in diverse cell types, appear restricted to approximately half of total genomic space, whereas T2-LADs can be found in most of the genome across this cohort of cell types. Second, more genomic regions are classified as T2-LADs in multiple cell types, compared to T1-LADs, which are more likely to be classified as such in a single cell type only. Thus, across the 13 cell types in our study, we find that the majority of LADs are not CTI, that T1-LADs are more restricted in terms of genomic occupancy, and that T1- and T2-LADs regions are shared to differing degrees across cell types.

In this context, we hypothesized that differences in gene content and transcription factor binding in LADs from different cell types may relate to cell type specificity, and that T1- and T2-LADs may be distinct in their genic complements. In order to test these hypotheses, we identified all genes in T1- and T2-LADs across all cell types and then compared how different subsets of genes were represented across these LAD subtypes in different cell types. Similar to the genomic region assessment, we found two distinct patterns of gene distribution in T1- versus T2-LADs (Fig. 6C, D; Table 6). In T2-LADs, we identified different subsets of genes (n = 500-5000) shared amongst multiple cell types. Though different subsets of genes (n = 75-800 genes) were shared across multiple cell types in T1-LADs, we found multiple examples of large subsets of genes which reside in T1-LADs of only one or two cell types. Thus, we conclude that T1-LADs encompass more “unique” gene subsets relative to T2-LADs, which show greater subsets of “shared” genes amongst multiple cell types. Our earlier characterizations indicated that T2-LADs are also more gene dense and more accessible than T1-LADs. In this context, the increased shared genomic region representation (Fig. 6A, B) and shared gene complement in T2- compared to T1-LADs (Fig. 6C, D) further underscores the distinction between these two categories of genomic loci, and indicates potentially different biological roles in cell type-specific gene regulation.

Ontology analysis of genes in T1- and T2-LADs provided additional support for the role of nuclear organization in cell type-specific regulation, revealing signatures of alternative fates in LADs (Table 7), such as enrichment of neuronal genes in unique cardiomyocyte (CM) T1-LADs. Likewise, unique T1-LAD ESC genes showed an enrichment for developmental signaling pathways and gene sets for specific cell types, such as muscle and cardiac development. The neuronal signature in unique CM LADs is even more pronounced when considering genes

bordering T1- and T2-LADs (Table 7). CTI-T2-LADs showed a strong enrichment for myriad immune categories, suggesting that these regions may not be “truly” invariant if our study had included immune-related cell types. The gene-centric analysis also further showed several examples of cell type-relevant genes relocalized into more permissive LAD categories in cell types where they are expressed. For instance, a gene expressed in multiple cardiac tissues, *ACE2* (Tucker et al., 2020), moves from KDD/T2-LAD in ESCs and all other cell types to nonLAD and nonKDD uniquely in CMs, in agreement with previous work demonstrating “release” of this gene from the nuclear periphery in CMs (Poleshko et al., 2017). *ACE2* is amongst a number of genes harbored in the LADs of one cell type that may underlie cellular identity in other cell types when they are shuffled out of peripheral heterochromatin. Such genes can be identified in pairwise comparisons, which are provided as a resource in Table 6.

To extend these observations, we assessed differential enrichment of TFBMs across various datasets to determine if this would reveal cell type-specific LAD organization, again hypothesizing that one potential function of LADs would be to sequester genes relevant to alternative fates. Relative to the union of T1-LADs from all cell types, CM T1-LADs were enriched for motifs for several transcription factors involved in differentiation and processes of cell types other than CMs, such as the intestinal epithelium gene *CDX2*, immune- and neuronal-related gene regulator *NKX2-2*, and the critical neural transcription factor *NPAS4* (Table 8). This supports the concept that some non-cardiac genes are silenced and located in T1-LADs in CMs. Similarly, midbrain T1-LADs were enriched for TFBMs relevant to non-neuronal genes, including the angiogenesis- and heart development-linked nuclear factor *COUP-TFII* (Pereira et al., 1999; Wu et al., 2013) and the pluripotency factor *OCT4*, amongst others (Table 8). Differential TFBMs in CM T1-LADs compared to ESC T1-LADs revealed

enrichment of *OCT4* binding sites, and included motifs of multiple genes implicated in neuronal differentiation or function (e.g., *EN1, NPAS4, OCT6*) (Bloodgood et al., 2013; He et al., 1989; Lin et al., 2008; Spiegel et al., 2014; Wurst et al., 1994; Zhu et al., 2014), and a general transcriptional activator and repressor (*FOXK1*) known to have the ability to translocate genes from the lamina to the nucleus (Wang et al., 2015). In comparison, ESC T1-LADs were enriched for several TFBMs not relevant to ESC identity, such as the vascular-related transcription factor *ETV2* (De Val et al., 2008; Lee et al., 2019a) (Table 8). These analyses also revealed enrichments for motifs of transcription factors known to preferentially bind methylated DNA and/or to function as repressors, which may contribute to target gene silencing or establishment of heterochromatin. Taken together, the gene and TFBM LAD assessments demonstrate likely transcriptional repression of genes regulated by cell-type-relevant transcription factors via sequestration in LADs. This may be mediated by factors that establish DNA methylation and heterochromatin-mediated silencing, a potentially critical mechanism for maintenance of cell type fidelity.

Discussion

In this study, we provide an in-depth view of the complex nature of peripheral heterochromatin, made possible by genome-wide measurements of LB1 and H3K9me2 DNA binding across multiple developmental trajectories, allowing for more robust unbiased LAD definitions and diverse cell type comparisons. We show that LADs are organized into at least two distinct subcompartments and that H3K9me2 contributes to both. We observe a high degree of LAD/KDD variance between cell types, and genes in LADs across cell types reflect signatures of alternative fates. Our collective assessments indicate that T1- and T2-LADs may function and/or be maintained distinctly. In sum, this study provides an expansive resource that further reveals the diversity of organization within peripheral heterochromatin, and suggests that multiple mechanisms may be employed within different peripheral chromatin compartments to achieve the transcriptional silencing that impacts cellular identity.

Peripheral chromatin is composed of distinct sub-compartments

In *Drosophila*, chromatin enriched in the linker histone H1 encompasses up to half of the genome and is organized into large domains similar to LADs (Grotthuss et al., 2010; van Bemmel et al., 2013). H1 is bound adjacent to the DNA entry/exit site of the nucleosome core particle, and its binding is linked to the compaction of nucleosomal arrays and nucleosomal stabilization (Robinson et al., 2008; Syed et al., 2010). Integrative analysis of genome-wide binding maps for various chromatin components defined five principal chromatin types in *Drosophila*, including three distinct subtypes of heterochromatin (including one subtype enriched

in H1), that may all possibly exist within the confines of the nuclear periphery (Filion et al., 2010). This segregation of *Drosophila* chromatin may be analogous to the findings in this study. Our novel and robust multi-state HMM analysis pipelines reveal two distinct LAD categories (T1- and T2-LADs) and KDDs. Furthermore, the KDDs demonstrate a high level of overlap with T1- and T2-LADs (Walker et al., 2015; Wen et al., 2009), suggesting that these domains all exist within the confines of the nuclear periphery.

Chromatin accessibility differences in T1- and T2-LADs are most obvious in the replication timing assessments, which showed an enrichment for late replicating domains in T1-LADs, compared to a depletion in T2-LADs. This stark distinction, evident across all the cell types examined, strongly supports a model where peripheral heterochromatin is segregated into at least two “layers.” The T1-LAD layer, with the most stable LB1 binding, highest level of LB1 occupancy, lowest density of genes, and lowest accessibility measure, is enriched for the latest replicating domains. Classification of genome occupancy in T1- or T2-LADs across cell types also revealed that T1-LADs appear restricted to only approximately half of the genome, even across diverse cell types, whereas a much higher proportion of the genome was represented in T2-LADs (Fig. 6 A,B). This observation provides support for the concept that some LADs from each category may be characterized as constitutive across multiple cell types, while others are more variable across cell types (Meuleman et al., 2013). The KDD analyses reveal further complexity to this layering model: T1-LADs that overlap KDDs are characterized by strong signals of genomic repression and association with the nuclear lamina, whereas regions occupied by KDD and not T1- or T2-LADs were distinct from T1-LADs, but not as clearly distinct from T2-LADs, suggesting that an “intermediate” peripheral heterochromatin chromatin layer with H3K9me2 may exist.

Future studies mapping additional structural elements, such as linker histone H1, histone variants associated with dynamic turnover (such as the histone variant histone H3.3 (Deaton et al.) or H2A.Z (Long et al., 2020), and poised RNA polymerase may provide an opportunity for even more in-depth analysis of this layered signature, possibly revealing even further sub-compartmentalization of peripheral heterochromatin beyond what we have described here. Likewise, super-resolution microscopy may provide the necessary resolution to assess the impact of this layering on chromatin accessibility differences between sub-compartments at the nuclear periphery (Otterstrom et al., 2019; Ricci et al., 2015), thereby providing greater insight into the layers of peripheral heterochromatin in single cells.

LAD compartments may be established and maintained by distinct mechanisms

Across all of our assessments, T2-LADs represent less repressive domains than T1-LADs. T2-LADs with and without KDD overlap have higher gene content and transcription compared to T1-LADs. One mechanism possibly driving T1- and T2-LAD differences could be variation in the frequency of lamina association between single cells in a bulk population, with T2-LADs representing a more dynamic feature, dissociating from the lamina in a greater proportion of cells than more static T1-LADs. This scenario could explain the concordance of our results with single cell DamID (Fig. 3A). Future studies will be needed to understand the differences in states of these LAD subtypes across additional cell types, and it will be of great interest to

understand the single cell variability in LAD distinction and underlying activity. This importance is reflected in recent findings in RNA sequencing in bulk populations compared with single cells, where an overall gene expression readout may be driven either by high expression in a subset of a cellular population or similar regulatory tuning across an entire cellular population. As such, future experiments assessing the frequency of T1- versus T2-LAD association with the nuclear lamina, possibly coupled with single-cell assessments of transcriptional activity (such as single cell RNA fluorescence *in situ* hybridization (FISH)) will greatly inform our understanding of functional behavior that drives the distinction of T1- and T2-LADs.

It is also of interest to consider that mechanisms for establishing or maintaining heterochromatin in T1- versus T2-LADs may be fundamentally different, which could drive LAD compartmentalization. The enrichment of CTCF at T2-LAD but not T1-LAD boundaries suggests that these two types of domains may be generated or maintained by fundamentally different mechanisms, further underscoring their distinction. The CTCF binding differences invoke potential distinction of T1- and T2-LADs by other higher order organization features, such as TADs, which have been shown to function as stable units for regulating replication timing (Pope et al., 2014). It has been shown that replication timing transition regions (the genomic regions between early and late replicating domains) share the same chromatin composition and overall lamin association as late replicating regions (Pope et al., 2014). Our results suggest that the replicating timing transition regions may encompass T2-LADs and do have some distinguishing features - higher gene density, higher transcription, lower LB1 - which were not robustly assessed prior to this study. Given the relatively higher transcription in T2-LADs, another intriguing model is that a threshold of transcription may be required to maintain the T2-LAD heterochromatin compartment, compared to the gene poor and highly silenced T1-LAD

compartment, which may not require such active mechanisms for establishment or maintenance of heterochromatin (Denholtz et al., 2020; Grewal and Elgin, 2007). Such dynamic mechanistic differences could have significant consequences on the accessibility, compaction, and gene regulation within T1- versus T2-LADs.

TE enrichment distinguishes LAD compartments and may contribute to functional differences

Differential enrichments of TEs support the hypothesis that different mechanisms may participate in the generation and maintenance of T1- versus T2-LADs. The majority of the human genome is made up of TEs, although fewer than 1% of these are capable of transposition (Sundaram and Wysocka, 2020). Many TEs contain TFBMs and participate in cell type- and species-specific genome regulation (Sundaram and Wysocka, 2020). SINE elements are enriched in T2-LADs but depleted in T1-LADs, and the opposite pattern is seen for LINEs (Fig. 3E, F). This mirrors findings in constitutive compared to facultative LADs (Meuleman et al., 2013). Of interest, SINEs often contain binding sites for YY1, a transcriptional repressor involved in the organization of heterochromatin that co-binds with CTCF (Donohoe et al., 2007; Schwalie et al., 2013). CTCF and YY1 co-occupancy has been associated with demarcating chromatin boundaries, a model which is supported by our findings for T2-LADs (Wang et al., 2012). Knockdown of YY1 leads to a loss of lamin association, and targeted recruitment of YY1 can generate ectopic LADs (Harr et al., 2015b). In agreement with these findings, TFBMs for

YY1 were enriched in T2- compared to T1-LADs in CMs and ESCs, suggesting that YY1 may play an active role in the establishment of T2-LADs but not T1-LADs.

Conversely, L1-LINEs are enriched in T1-LADs and depleted in T2-LADs, while L2-LINEs are depleted in T1- and T2-LADs; L1-LINEs therefore may be part of a mechanism generating and/or maintaining these compartments specifically. L1-LINEs actively transpose in the human genome, and are considered to be a driver of human genome evolution, with postulated roles in processes from embryogenesis to neuronal development to oncogenesis via generation of germline and somatic mosaicism (Kazazian, 2004; Vazquez et al., 2019). Interestingly, other TEs active in the human genome, Alu and SVA elements, are both depleted in T1-LADs and show mixed results in T2-LADs. *SIRT7* has been shown to participate in the epigenetic transcriptional repression of L1-LINEs and promotes association of these elements with the nuclear lamina (Vazquez et al., 2019). L1-LINEs are also actively repressed by piwi-RNAs, which are thought to mediate the generation of heterochromatin via phase separation (Ozata et al., 2019). Together, these results indicate that TEs may play an important role in T1- and T2-LAD generation and maintenance, and perhaps contribute to their gene regulatory roles.

LADs and KDDs are linked to cell type-specific transcription, and are diverse amongst cell types

Maintenance of chromatin structure via association with the nuclear periphery is amongst the many factors contributing to successful cell type specification and identity, and organogenesis (Meuleman et al., 2013; Peric-Hupkes et al., 2010; Poleshko et al., 2017; Robson et al., 2016).

LADs and KDDs have been previously shown to demonstrate heterogeneity among different cell types and between single cells (Kind et al., 2015; Meuleman et al., 2013; Wen et al., 2009). We uncovered broad evidence of cell type specificity being maintained by T1- and T2-LADs and KDDs. This included enrichment of motifs for other cell types in T1-LADs of a given cell type, suggesting repression of those genes to protect cell type identity. An interesting example is OCT4 TFBMs, which are enriched in T1-LADs from midbrain and CMs, but not ESCs, suggesting suppression of its targets in non-pluripotent cell types. Reduction of LAMIN A expression has been associated with increased expression of *OCT4* and more efficient reprogramming of somatic cells to an iPS-like state (Zuo et al., 2012). These results imply that the repression of *OCT4* and its associated network are mediated in part by association with the lamina, underscoring the importance of LADs in cell type differentiation and embryonic development. This is further supported by work showing that the *OCT4* promoter is silenced by H3K9 methylation via the H3K9 histone methyltransferase G9a (Epsztejn-Litman et al., 2008). Another intriguing example is *OCT6*, which has recently been implicated in a radical identity shift of mESCs undergoing directed cardiac differentiation towards a neuronal rather than cardiac fate (Hota et al., 2020). We found OCT6 TFBMs to be enriched in T1-LADs in CMs compared to ESCs, supporting a model in which during normal CM differentiation neuronal genes regulated by OCT6 are repressed in LADs, while in aberrant conditions OCT6-regulated genes are de-repressed, allowing the cells to differentiate towards a neuronal trajectory.This model represents just one example of hypothesis generation enabled by this collection of heterochromatin compartments across a diversity of human cell types. It will be exciting to experimentally test these ideas in future work.

The presence of multiple cell-type and pluripotency-relevant TFBMs in T2-LADs compared to T1-LADs may also reflect an important role for T2-LADs in cell type determination and fate. The lower LB1 signal in T2-LADs may indicate either less overall interaction with lamin at these genomic loci, or different proportions of a cell population having high and low levels of lamin interactions. Cell type specification and maintenance requires a complex stoichiometry of expression of many genes; perhaps the level or proportion of a cell population with a locus interacting with the lamina is yet another mechanism used to modulate levels of gene expression towards maintenance of pluripotency, differentiation, or maintenance of cell identity. Such study will require careful dissection of the localization and transcriptional status of various loci in T1- versus T2-LADs during differentiation. For example, recent data suggest LAMIN A/C filaments can act as transcriptional repressors (Buchwalter and Hetzer, 2017), and it is unclear if reduced lamin contact frequency of a particular locus at the nuclear periphery renders it susceptible to transcriptional activation first and then relocalization, or vice versa. This study has revealed several interesting candidates that can be used in future studies, perhaps in a hybrid single cell DNA- and RNA-FISH approach with and without super resolution microscopy, to reveal biological implications of T1- versus T2-LAD organization on cellular differentiation.

Conclusion

Overall, this work reveals critical details about peripheral heterochromatin, commonly viewed as a monolithic entity, specifically revealing consistent sub-compartmentalization of peripheral chromatin across multiple human cell types and a cell-type-specific set of overlapping but distinct peripheral chromatin domains. This work and these revelations follow the general trend

in genomics, in which no locus can be simply defined, as most have multiple functions depending on context (Halfon, 2019). These findings lay the groundwork for future studies aimed at defining the driving cause behind the difference between T1-and T2-LADs. What is the biological consequence of having different subtypes of peripheral heterochromatin? Is this subcompartmentalization conserved in other species? Does the dynamic range vary cell-to-cell or across populations? Is the enrichment of cell-type-specific TFBMs in T1-LADs, T2-LADs and KDDs a driving force in, or a consequence of, differentiation and cell identity? Finally, how does overall peripheral chromatin organization and sub-compartmentalization impact genomic and cell type stability? These questions are fundamental to gaining a greater understanding of the role of genome organization in establishing and maintaining cellular diversity.

Methods

Generation and maintenance of cell types

Human pluripotent stem cell maintenance

H9 human pluripotent stem cells (hPSCs) were maintained in E8 media and passaged every 4 days onto matrigel-coated plates (Roche). ESCs, cardiomyoytes, epicardium, and endothelium were H9 hESC-derived. The remainder of the cells were H7 hESC-derived. H7 (WiCell) hPSCs were maintained in feeder-free conditions using mTeSR1 media (StemCell Technologies) + 1% penicillin/streptomycin (Thermo Fisher), and fresh media was added daily. Cells were cultured on tissue culture plastics coated with Geltrex basement matrix (Thermo Fisher; which was diluted 1:100 in DMEM/F12 media [Thermo Fisher] before being used to coat culture plastics). Prior to reaching confluence, H7 hESCs were dissociated using either Accutase (Thermo Fisher) or Versene (Thermo Fisher), and then were passaged onto new plates.

For all H7-derived cells, differentiation was conducted in serum-free media, either Chemically Defined Medium 2 (CDM2) or Chemically Defined Medium 3 (CDM3). The composition of CDM2 basal medium (Loh et al., 2014a, 2016) is: 50% IMDM + GlutaMAX (Thermo Fisher, 31980-097) + 50% F12 + GlutaMAX (Thermo Fisher, 31765-092) + 1 mg/mL polyvinyl alcohol (Sigma, P8136-250G) + 1% v/v chemically defined lipid concentrate (Thermo Fisher, 11905-031) + 450 μM 1-thioglycerol (Sigma, M6145-100ML) + 0.7 μg/mL recombinant human insulin (Sigma, 11376497001) + 15 μg/mL human transferrin (Sigma, 10652202001) + 1% v/v penicillin/streptomycin (Thermo Fisher, 15070-063). Polyvinyl alcohol was brought into

suspension by gentle warming and magnetic stirring, and the media was sterilely filtered (through a 0.22 μm filter) prior to use.

The composition of CDM3 basal medium (Ang et al., 2018) is: 45% IMDM + GlutaMAX (Thermo Fisher, 31980-097) + 45% F12 + GlutaMAX (Thermo Fisher, 31765-092) + 10% KnockOut serum replacement (Thermo Fisher, 10828028) + 1 mg/mL polyvinyl alcohol (Sigma, P8136-250G) + 1% v/v chemically defined lipid concentrate (Thermo Fisher, 11905-031) + 1% v/v penicillin/streptomycin (Thermo Fisher, 15070-063). Polyvinyl alcohol was brought into suspension by gentle warming and magnetic stirring, and the media was sterilely filtered (through a 0.22 μm filter) prior to use.

hPSC CM differentiation

On Day 0 (start of differentiation) hPSCs were treated with 1mg/ml Collagenase B (Roche) for 1 hour, or until cells dissociated from plates, to generate embryoid bodies (EBs). Cells were collected and centrifuged at 300 rcf for 3 mins, and resuspended as small clusters of 50–100 cells by gentle pipetting in differentiation media containing RPMI (Gibco), 2 mM/L L-glutamine (Invitrogen), 4×10^4 monothioglycerol (MTG, Sigma-Aldrich), 50 μg/ml ascorbic acid (Sigma-Aldrich). Differentiation media was supplemented with 2ng/ml BMP4 and 3 μmol Thiazovivin (Milipore). EBs were cultured in 6cm dishes (USA Scientific) at 37°C in 5% CO^2, 5% O^2, and 90% N^2. On Day 1, the media was changed to differentiation media supplemented with 30 ng/ml BMP4 (R&D Systems) and 30 ng/ml Activin A (R&D Systems), 5ng/ml bFGF (R&D Systems) and 1 μM Thiazovivin (Milipore). On Day 3, EBs were harvested and washed once with DMEM (Gibco). Media was changed to differentiation media supplemented with 5 ng/ml VEGF (R&D Systems) and 5 μmol/L XAV (Stemgent). On Day 5, media was changed to

differentiation media supplemented with 5 ng/ml VEGF (R&D Systems). After Day 8, media was changed every 3-4 days to differentiation media without supplements.

CM dissociation

EBs were incubated overnight with 0.6mg/ml Collagenase Type II (Worthington) at 37ºC. Dissociated cells were harvested and washed with Wash media (DMEM, 0.1% BSA) + 1mg/ml DNase (VWR) twice and centrifuged at 300 rcf for 3 mins. Cells were resuspended in differentiation media supplemented with 1 μM Thiazovivin (Millipore) and filtered.

hPSC epicardium differentiation

The hPSC CM protocol was followed up to Day 3. At Day 3, EBs were dissociated with TrypLE Express (Gibco). Dissociated cells were washed with Wash media (DMEM, 0.1% BSA) + 1mg/ml DNase (VWR) twice and centrifuged at 300 rcf for 3 mins. Cells were resuspended in differentiation media supplemented with 1ng/ml BMP4 (R&D Systems), filtered and counted using a hemocytometer. Cells were plated onto a matrigel-coated 96 well plate at 80,000 cells per well. On Day 5, the media was changed to differentiation media supplemented with 5ng/ml VEGF (R&D Systems) and 1nM all-trans Retinoic Acid (Sigma-Aldrich). After Day 5, media was changed every 2 days with the same Day 5 differentiation media composition. On Day 11, the media was changed to differentiation media supplemented with 5ng/ml VEGF (R&D Systems) and cells were fed with the same differentiation media every 2 days until Day 15. On Day 15, cells were dissociated with 1mg/ml Collagenase B (Roche) for 1 hour, washed with Wash media (DMEM, 0.1% BSA) + 1mg/ml DNase (VWR) and centrifuged at 300 rcf for 3 mins. Cells were further dissociated with 3ml TrypLE Express, washed with Wash media (DMEM, 0.1% BSA) + 1mg/ml DNase (VWR) and centrifuged at 300 rcf for 3 mins. Cells were resuspended in differentiation media supplemented with 1 μM Thiazovivin (Millipore), filtered and counted using

a hemocytometer. Cells were plated in a matrigel-coated 6 well-plate at 100,000 cells per well. On Day 16, media was changed to differentiation media supplemented with 5ng/ml VEGF (R&D Systems) and cells were fed every 2 days until they reached confluence (approximately Day 22).

hPSC endothelial cell differentiation

The hPSC CM protocol was followed up to Day 5. At Day 5, EBs were dissociated with TrypLE Express (Gibco). Dissociated cells were washed with Wash media (DMEM, 0.1% BSA) + 1mg/ml DNase (VWR) twice and centrifuged at 300 rcf for 3 mins. Cells were resuspended in differentiation media supplemented with 100ng/ml VEGF (R&D Systems) and 50ng/ml bFGF (R&D Systems), filtered and counted using a hemocytometer. Cells were plated onto a matrigel-coated 96 well plate at 80,000 cells per well. Media was changed every 2 days using differentiation media supplemented with 100ng/ml VEGF (R&D Systems) and 50ng/ml bFGF (R&D Systems). Cells were harvested and sorted on Day 15.

Fluorescence-activated Cell Sorting (FACS)

Dissociated hPSC-derived cells were resuspended in differentiation media containing diluted antibodies (dilutions listed below) for 30mins on ice. Cells were washed with differentiation media and resuspended in differentiation media + DAPI (1.35µg/ml, Biolegend) for FACS (BD FACSAria). hPSC-derived CMs were sorted by gating for SIRPA+ (PE-Cy7 anti-human CD172a/b, Biolegend, 1:200) and CD90- (APC anti-human CD90 (Thy1) Antibody, Biolegend, 1:200) cells. PSC-derived endothelial cells were sorted by gating for CD31+ (PE anti-human CD31, Biolegend, 1:200) cells.

hPSC ectodermal differentiation

The day prior to beginning differentiation, H7 hESCs were dissociated with Accutase (Thermo Fisher) for 10 minutes at 37°C. Accutase was neutralized through the addition of excess DMEM/F12 media, and then hESCs were pelleted via centrifugation and the supernatant was aspirated. Pelleted hESCs were resuspended in mTeSR1 + 1% penicillin/streptomycin + 1 μM of the ROCK inhibitor Thiazovivin (Tocris) (henceforth referred to "cell-plating media"), and plated onto Geltrex-coated tissue culture plastics at a density of 4 x 10^5 cells/cm^2 (i.e., 2.1 x 10^6 cells per 10-cm dish). 24 hours after seeding, the cell-plating media was aspirated, and cells were briefly washed with DMEM/F12 to remove all traces of cell-plating media.

hESCs were then differentiated into definitive ectoderm through the addition of CDM2 basal medium supplemented with A8301 (1 μM; Tocris), LDN193189 (100 nM, Stemgent), C59 (1 μM; Tocris), and Trametinib (250 nM; Selleck Chem), which was added for 24 hours.

For border ectoderm induction, hESCs were differentiated into definitive ectoderm in 24 hours (as described above), and then definitive ectoderm was briefly washed (with DMEM/F12) and then treated with CDM2 basal media supplemented with BMP4 (30 ng/mL; R&D Systems), PD0325901 (100 nM; Tocris), CHIR99021 (3 μM; Tocris), and A8301 (1 μM; Tocris) for an additional 2 days, to generate Day 3 border ectoderm progenitors. Differentiation media was aspirated and added fresh every 24 hours.

For midbrain induction, hESCs were differentiated into definitive ectoderm in 24 hours (as described above), and then definitive ectoderm was briefly washed (with DMEM/F12) and then treated with CDM2 basal media supplemented with A8301 (1 μM; Tocris), LDN193189 (100 nM;

Stemgent), and Trametinib (250 nM; Selleck Chem) for an additional 2 days to generate neural progenitors. Day 3 neural progenitors were briefly washed (with DMEM/F12) and were then treated with CDM2 basal media supplemented with CHIR99021 (4 μM; Tocris), SB-505124 (2 μM; Tocris) and FGF2 (20 ng/mL; R&D Systems) for 2 days, to generate Day 5 midbrain progenitors. Differentiation media was aspirated and added fresh every 24 hours.

hPSC endodermal differentiation

The day prior to beginning differentiation, H7 hESCs were dissociated with Accutase (Thermo Fisher) at 37°C. Accutase was neutralized through the addition of excess DMEM/F12 media, and then hESCs were pelleted via centrifugation and the supernatant was aspirated. Pelleted hESCs were resuspended in cell-plating media, and plated onto Geltrex-coated tissue culture plastics at a 1:8-1:16 cell seeding ratio. 24 hours after seeding, the cell-plating media was aspirated, and cells were briefly washed with DMEM/F12 to remove all traces of cell-plating media.

hESCs were then differentiated into anteriormost primitive streak through the addition of CDM2 basal medium supplemented with Activin A (100 ng/mL; R&D Systems), CHIR99021 (3 μM; Tocris), FGF2 (20 ng/mL; R&D Systems), and PI-103 (50 nM; Tocris), which was added for 24 hours. Day 1 anteriormost primitive streak cells were briefly washed (with DMEM/F12) and then differentiated into Day 2 definitive endoderm through the addition of CDM2 basal medium supplemented with Activin A (100 ng/mL; R&D Systems), LDN-193189 (250 nM; Tocris), and PI-103 (50 nM; Tocris), which was added for 24 hours. Methods for anteriormost primitive streak and definitive endoderm formation have been described previously (Ang et al., 2018; Loh et al., 2014a, 2016; Martin et al., 2020).

For liver differentiation, Day 2 definitive endoderm cells were briefly washed (with DMEM/F12) and further differentiated into Day 3 posterior foregut through the addition of CDM3 base media supplemented with FGF2 (20 ng/mL; R&D Systems), BMP4 (30 ng/mL; R&D Systems), TTNPB (75 nM; Tocris), A8301 (1 μM; Tocris). Day 3 posterior foregut cells were briefly washed (with DMEM/F12), and then further differentiated on days 4–6 with CDM3 base media supplemented with Activin A (10 ng/mL; R&D Systems), BMP4 (30 ng/mL; R&D Systems), Forskolin (1 μM; Tocris) to generate Day 6 liver bud progenitors. Methods for liver bud progenitor formation have been described previously (Ang et al., 2018; Martin et al., 2020).

For mid/hindgut differentiation, Day 2 definitive endoderm cells were briefly washed (with DMEM/F12) and further differentiated into Day 6 mid/hindgut progenitors through the addition of CDM2 base media supplemented with FGF2 (100 ng/mL), BMP4 (10 ng/mL) and CHIR99021 (3 μM) for 4 days. Methods for mid/hindgut progenitor formation have been described previously (Ang et al., 2018; Loh et al., 2014a).

hPSC mesodermal differentiation

The day prior to beginning differentiation, H7 hESCs were dissociated with Accutase (Thermo Fisher) at 37°C. Accutase was neutralized through the addition of excess DMEM/F12 media, and then hESCs were pelleted via centrifugation and the supernatant was aspirated. Pelleted hESCs were resuspended in cell-plating media, and plated onto Geltrex-coated tissue culture plastics at a 1:8-1:16 cell seeding ratio. 24 hours after seeding, the cell-plating media was aspirated, and cells were briefly washed with DMEM/F12 to remove all traces of cell-plating media.

hESCs were then sequentially differentiated into anterior primitive streak, paraxial mesoderm and early somites as described previously (Loh et al., 2016). Briefly, hESCs were differentiated into anterior primitive streak through the addition of CDM2 basal medium supplemented with Activin A (30 ng/mL; R&D Systems), CHIR99021 (4 μM; Tocris), FGF2 (20 ng/mL; R&D Systems), and PIK90 (100 nM; Calbiochem), which was added for 24 hours, thus generating Day 1 anterior primitive streak (Loh et al., 2016).

For paraxial mesoderm induction, hESCs were differentiated into anterior primitive streak in 24 hours (as described above), and then anterior primitive streak was briefly washed (with DMEM/F12) and then treated with CDM2 basal media supplemented with A8301 (1 μM; Tocris), LDN193189 (250 nM; Tocris), CHIR99021 (3 μM, Tocris), and FGF2 (20 ng/mL; R&D Systems), which was added for 24 hours, thus generating Day 2 paraxial mesoderm (Loh et al., 2016).

For early somite induction, hESCs were differentiated into anterior primitive streak and then further differentiated into paraxial mesoderm (as described above). Day 2 paraxial mesoderm was briefly washed (with DMEM/F12) and then treated with CDM2 basal media supplemented with CDM2 base media supplemented with A8301 (1 μM; Tocris), LDN193189 (250 nM; Tocris), XAV939 (1 μM; Tocris), and PD0325901 (500 nM; Tocris) for 24 hours, thus generating Day 3 early somites (Loh et al., 2016).

hESCs were sequentially differentiated into mid primitive streak, dorsal lateral mesoderm and artery endothelial progenitors.

hESCs were differentiated into mid primitive streak through the addition of CDM2 basal medium supplemented with Activin A (30 ng/mL; R&D Systems), BMP4 (40 ng/mL; R&D Systems), CHIR99021 (6 μM; Tocris), FGF2 (20 ng/mL; Thermo Fisher) for 24 hours (with the optional addition of PIK90 (100 nM; Calbiochem)), as previously described (Loh et al., 2016).

For dorsal lateral mesoderm induction, hESCs were differentiated into mid primitive streak in 24 hours (as described above). Day 1 mid primitive streak cells were then briefly washed (DMEM/F12, Thermo Fisher) and then differentiated towards dorsal lateral mesoderm in CDM2 media supplemented with BMP4 (40 ng/mL; R&D Systems), GDC-0941 (2.5 μM; Cellagen Technology), Forskolin (10 μM; Tocris), SB-505124 (2 μM; Tocris), VEGF (100 ng/mL; R&D Systems), XAV939 (1 μM; Tocris) and ascorbic acid-2-phosphate (AA2P; 200 μg/mL, Sigma) for 24 hours, thus generating Day 2 dorsal lateral mesoderm.

For arterial endothelial progenitor induction, hESCs were differentiated into mid primitive streak and then dorsal lateral mesoderm (as described above). Day 2 dorsal lateral mesoderm cells were then treated with CDM2 supplemented with XAV939 (1 μM; Tocris), VEGF (100 ng/mL; R&D Systems), DMH1 (250 nM; Tocris), TTNPB (0.5 nM; Tocris), Activin A (15 ng/mL; R&D Systems), AA2P (200 μg/mL; Sigma), and GDC0941 (2.5 μM; Cellagen Technologies) for 48 hours, thus generating Day 4 arterial endothelial progenitors.

ChIP and ChIP-seq Library Preparation

Undifferentiated ESCs and all differentiated cell types were crosslinked in culture by addition of methanol-free formaldehyde (ThermoFisher, final 1% v/v), and incubated at room temperature for 10 minutes with gentle rotation. Crosslinking was quenched by addition of glycine (final

125mM) and incubated at room temperature for 5 minutes with gentle rotation. Media was discarded and replaced with PBS; cells were scraped and transferred to conical tubes and pelleted by centrifugation (250xg, 5 minutes at room temperature). Resulting pellets were flash frozen on dry ice and stored at -80C. For ChIP, 30uL protein G magnetic beads (per ChIP sample; Dynal) were washed 3 times in blocking buffer (0.5% BSA in PBS); beads were resuspended in 250uL blocking buffer and 2ug antibody (LaminB1: ab16048 [abcam]; H3K9me2: ab1220 [Abcam]) and rotated at 4C for at least 6 hours. Crude nuclei were isolated from frozen crosslinked cells as follows: cell pellet (from 10cm plate) was resuspended in 10mL cold Lysis Buffer 1 (50mM HEPES-KOH pH7.5, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100, and protease inhibitors), and rotated at 4C for 10 minutes, followed by centrifugation (250xg, 5 minutes at room temperature). Supernatant was discarded and the pellet was resuspended in 10mL cold Lysis Buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, and protease inhibitors), and rotated at room temperature for 10 minutes, followed by centrifugation (250xg, 5 minutes at room temperature). Supernatant was discarded and nuclei were resuspended/lysed in 1mL cold Lysis Buffer 3 (10mM Tris-HCl, pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, and protease inhibitors) and transferred to pre-chilled 1mL Covaris AFA tubes (Covaris). Samples were sonicated using a Covaris S220 sonicator (high cell chromatin shearing for 15 minutes; Covaris). Lysates were transferred to tubes and Triton X-100 was added (final 1%) followed by centrifugation (top speed, 10 minutes at 4C in microcentrifuge). Supernatant was transferred to a new tube; protein concentration was measured by Bradford assay. Antibody-conjugated beads were washed 3 times in blocking buffer, resuspended in 50uL blocking buffer and added to 500ug input protein for overnight incubation with rotation at 4C. 50ug lysate was aliquoted and stored at -20C for input. On day 2, beads were washed 5 times in 1mL RIPA buffer (50mM HEPES-KOH pH 7.5,

500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) with 2-minute incubation at room temperature with rotation for each wash. Beads were washed in 1mL final wash buffer (1xTE, 50mM NaCl) for 2 minutes with rotation at room temperature before final resuspension in 210uL elution buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS). To elute, beads were incubated with agitation at 65C for 30 minutes. 200uL eluate was removed to a fresh tube, and all samples (ChIP and reserved inputs) were reverse-crosslinked overnight at 65C with agitation for a minimum of 12 hours, but not more than 18 hours. 200uL 1xTE was added to reverse crosslinked DNA to dilute SDS, and samples were RNaseA treated (final 0.2mg/mL RNase; 37C for 2 hours) and Proteinase K (final 0.2mg/mL Proteinase K; 55C for 2 hours) before phenol:chloroform extraction and resuspension in 10mM Tris-HCl pH 8.0.

ChIP and input DNA was quantified by Qubit (ThermoFisher) before library preparation using the NEBNext Ultra II DNA library prep kit (NEB). Samples were indexed for multiplex sequencing. Library quality was analyzed by BioAnalyzer (Agilent Genomics) and quantified using qPCR (Kapa Biosystems). Libraries were pooled for multiplex sequencing, re-quantified, and sequenced on the Illumina NextSeq500 platform (vII; 75bp single-end sequencing; Illumina).

RNA Isolation and RNA-seq Library Preparation

Undifferentiated ESCs or differentiated cells were scraped from tissue culture plates with 1xPBS, and centrifuged at 1500g for 5 minutes at room temperature. After discarding supernatant, cell pellets were flash frozen in dry ice and stored at -80C until processing. RNA was isolated using QIAGEN RNeasy total RNA extraction kit (QIAGEN). RNA quality was analyzed by BioAnalyzer; samples with RIN scores >8 were chosen for further processing. RNA

libraries were prepared using the NEBNext Ultra II DNA Library Prep kit (NEB) with the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) to enrich for poly-A tailed RNA molecules. RNA-seq library quality was analyzed by BioAnalyzer (Agilent Genomics) and quantified using qPCR (Kapa Biosystems). Libraries were pooled for multiplex sequencing, re-quantified, and sequenced on the Illumina NextSeq500 platform (vII; 75bp single end sequencing; Illumina).

ChIP-sequencing data processing for LAMIN-B1 and H3K9me2

Adapters were trimmed using Trimmomatic [v0.39] (Bolger et al., 2014). Sequencing reads were aligned to human reference hg38 using BWA-MEM [v0.7.17] (Li and Durbin, 2010). Aligned reads were converted to BAM and sorted using Samtools [v0.1.19] (Li et al., 2009), with quality filter ("-F") set to 1804. Duplicates were removed using Picard [v2.18.7] MarkDuplicates. Sequencing reads from the ENCODE blacklist were removed using Bedtools [v2.29.0] (Quinlan and Hall, 2010). Each replicate had at least 1 million mapped sequencing reads. Data for both LB1 and H3K9me2 ChIP-seq were divided into higher quality ("tier one") and lower quality ("tier two") as assessed by replicate correlation values and visual assessment in order to generate stringent sets to train the LAD- and KDD-calling models. Spearman correlations between biological replicates was greater than 0.7 for all tier one cell types, and greater than 0.6 for tier two cell types assessed by comparing bigwig files for SES-normalized signal over controls generated with bamCompare using multiBigWigSummary and plotCorrelation from deepTools [v3.3.2] (Ramírez et al., 2014).

Identification of LADs

LB1 ChIP-seq signal was calculated and converted into BedGraph files using deepTools bamCompare [v3.3.2] (Ramírez et al., 2014) with 20kb bins, using the signal extraction scaling method (Diaz et al., 2012) for sample scaling. All HMMs were implemented using pomegranate [v0.11.1] (Schreiber, 2017). A 3-state HMM best fit data in the majority of cell lines according to AIC (Table 2) (Akaike, 1974). The HMM was initialized using a normal distribution and k-means with a uniform transition matrix. The Baum-Welch algorithm was then used to train the model, with tier one cell types (Table 1) used together in the model training (Figure 2A). Spearman correlation confirmed that replicates from the same cell type generally clustered together. The model was then applied to predict LAD state genome-wide per 20kb bins, using the median value from both replicates for each bin, for each cell type individually from both tier one and tier two, filtering regions in the ENCODE blacklist from consideration. States were labeled as T1-LAD, T2-LAD or non-LAD based on median LB1 signal for the bins with that state label, with the highest median LB1 signal being assigned T1-LAD, second highest T2-LAD, and lowest non-LAD.

Identification of KDDs

H3K9me2 ChIP-seq signal were calculated and converted into BedGraph files using deepTools bamCompare [v3.3.2] (Ramírez et al., 2014) with 20kb bins, using the signal extraction scaling method (Diaz et al., 2012) for sample scaling. Replicates from the same cell type generally clustered together. HMMs were implemented as described above. The best fitting model was a 2-state HMM according to AIC and visual inspection (Table 5). States were labeled as KDD or nonKDD based on median H3K9me2 signal for the bins with that state label, with the highest median H3K9me2 signal being assigned KDD and lowest nonKDD.

RNA-sequencing analysis

Transcriptome data were quantified using Kallisto [v0.44.0] quant with fragment length determined by BioAnalyzer, standard deviation of 10, and 30 bootstraps, assigning reads using the Ensembl [v96] genome annotation (Bray et al., 2016). TPM values were quantile-normalized between cell types. Differentially expressed transcripts (q<=0.01) between cell types were identified using Sleuth [0.30.0] (Pimentel et al., 2017).

Transcription factor binding motif analysis

Differential transcription factor binding was analyzed using Homer [v4.11.1] (Heinz et al., 2010).

Replication timing analysis

CL, CE and switch domains were obtained from (Dixon et al., 2018). The total number of occurrences of replication domains from each category in T1LADs, T2 LADs and KDDs was identified using pybedtools with option “-c”. This was compared to 100 permutations of random draws from the hg38 genomic background, excluding the ENCODE blacklist to calculate empirical p-values and z-scores.

Transposable element enrichment analysis

LADs and KDDs from each cell type relative to genomic background

The total number of occurrences of TEs from each category in T1LADs, T2 LADs and KDDs was identified using pybedtools with option “-c”. This was compared to 100 permutations of random draws from the hg38 genomic background, excluding the ENCODE blacklist to calculate empirical p-values and z-scores.

Cell Type Invariant LADs relative to LADs

This analysis was performed similarly to above, except the null distribution was 100 permutations of n draws from the union of LADs from all cell types (Type 1 or 2) where n is the number of CTI-LADs.

Comparison with single cell DamID

Single cell DamID data from 172 KBM7 cells from clone 5-5 was downloaded from the Gene Expression Omnibus (GSE68260) in hg19 and lifted over using pybedtools to hg38, removing regions that did not lift over. These data were intersected with T1 and T2LADs using pybedtools.

ATAC-seq analysis

ATAC peaks were downloaded as BED files from GEO, lifted over from hg19 to hg38, taking the intersection of 2 replicates for each cell type. Only the H9-derived data from (Liu et al., 2017) were used. Data are listed in Table 5.

CTCF analysis

Bigwig files were downloaded from GEO. CrossMap was used to lift over bigwigs from hg19 to hg38. Data are listed in Table 6.

Hi-C analysis

Hi-C data for CMs and ESCs were downloaded as Cooler files from the 4D Nucleome Data Portal (Zhang et al., 2019a). A and B compartments were called using cooltools [v0.3.0] (Abdennur and Mirny, 2020).

Supporting Analyses

Gene annotations used throughout are from Ensembl v96. The reference genome used was human hg38, downloaded from the UCSC Genome Browser. Plotting, statistical analyses and supporting analyses were conducted in Python [v3.6] with packages Jupyter, matplotlib (Hunter, 2007), seaborn (Michael Waskom et al., 2020), upsetplot (Lex et al., 2014), scikit-learn (Pedregosa et al., 2011), numpy (Walt et al., 2011), pybedtools (Dale et al., 2011; Quinlan and Hall, 2010) and deepTools [v3.3.2] (Ramírez et al., 2014).

Data Access

Data generated in this paper

RNAseq data from ventricular CMs, ESCs and Day 15 endothelial cells, ChIP-seq data for LB1 and H3K9me2 from all cell types in Table 1, LAD and KDD calls are available via GEO (GSE155244) .

Previously published data

RNA-seq

Data from Day 3 early somite and paraxial mesoderm cells were downloaded from SRP073808 (Koh et al., 2016). Data from mid-hindgut were downloaded from SRP033267 (Loh et al., 2014b). Data from neural ectoderm were downloaded from SRP113027 (Tchieu et al., 2017).

ATAC-seq

Data from ESCs, CMs and Day 2 mesoderm were downloaded from GSE85330 (Liu et al., 2017).

CTCF ChIP-seq

Data from ESCs were downloaded from GSM3263064 and GSM3263065, from Day 2 mesoderm from GSM3263066 and GSM3263067 and from Day 80 ventricular CMs from GSM3263074 (Zhang et al., 2019a).

Hi-C

Data from ESCs were downloaded from GSM3263064 and GSM3263065 and from Day 80 ventricular CMs from GSM3263074 (Zhang et al., 2019a).

Acknowledgements

We thank Geoffrey Fudenberg, Ashley Karnay, Wonho Kim, Andrey Poleshko, Cheryl Smith, and Jonathan A. Epstein for valuable conversations and critical review of the manuscript. We thank Kathryn Claiborn for copy editing this manuscript. K.C.K. was supported in part by a Discovery Fellowship. K.C.K. and K.S.P. were supported by NHLBI (UM1HL098179) and Gladstone Institutes. C.E.D. was supported by the Ford Foundation Predoctoral Fellowship and the Stanford Graduate Fellowship. A.C. and R.E.A.S. were supported by the California Institute for Regenerative Medicine (Bridges Program TB1-01195). K.M.L. is a Packard Foundation Fellow, Pew Scholar, Human Frontier Science Program Young Investigator, Baxter Foundation Faculty Scholar and The Anthony DiGenova Endowed Faculty Scholar. R.J was supported by a Burroughs Wellcome Career Award for Medical Scientists, NSF CMMI-1548571, a grant from the Pennsylvania Department of Health, funds from the American Heart Association and Allen foundation, and a NIH New Innovator Award (DP2, HL147123).

Disclosure Declaration

K.C.K., P.P.S., R.J., C.E.D., R.E.A.S. and K.M.L. have nothing to disclose. A.C. is currently an employee of Orca Bio. K.S.P. is a consultant for Tenaya Therapeutics.

References

Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control *19*, 716–723.

Ang, L.T., Tan, A.K.Y., Autio, M.I., Goh, S.H., Choo, S.H., Lee, K.L., Tan, J., Pan, B., Lee, J.J.H., Lum, J.J., et al. (2018). A Roadmap for Human Liver Differentiation from Pluripotent Stem Cells. Cell Rep. *22*, 2190–2205.

Bardot, E., Calderon, D., Santoriello, F., Han, S., Cheung, K., Jadhav, B., Burtscher, I., Artap, S., Jain, R., Epstein, J., et al. (2017). Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential. Nat. Commun. *8*, 14428.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. *27*, 1139–1152.

Bloodgood, B.L., Sharma, N., Browne, H.A., Trepman, A.Z., and Greenberg, M.E. (2013). The activity-dependent transcription factor NPAS4 regulates domain-specific inhibition. Nature *503*, 121–125.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Briand, N., and Collas, P. (2020). Lamina-associated domains: peripheral matters and internal affairs. Genome Biol. *21*, 85.

Briand, N., Guénantin, A.-C., Jeziorowska, D., Shah, A., Mantecon, M., Capel, E., Garcia, M., Oldenburg, A., Paulsen, J., Hulot, J.-S., et al. (2018). The lipodystrophic hotspot lamin A p.R482W mutation deregulates the mesodermal inducer T/Brachyury and early vascular differentiation gene networks. Hum. Mol. Genet. *27*, 1447–1459.

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T.M., Rubenstein, J.L.R., and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature *398*, 622–627.

Buchwalter, A., and Hetzer, M.W. (2017). Nucleolar expansion and elevated protein translation in premature aging. Nat. Commun. *8*, 328.

Buchwalter, A., Kaneshiro, J.M., and Hetzer, M.W. (2019). Coaching from the sidelines: the nuclear periphery in genome regulation. Nat. Rev. Genet. *20*, 39–50.

Burke, B., and Stewart, C.L. (2006). The Laminopathies: The Functional Architecture of the Nucleus and Its Contribution to Disease. Annu. Rev. Genomics Hum. Genet. *7*, 369–405.

Choudhary, M.N., Friedman, R.Z., Wang, J.T., Jang, H.S., Zhuo, X., and Wang, T. (2020). Co-opted transposons help perpetuate conserved higher-order chromosomal structures. Genome Biol. *21*, 16.

Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011). Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics *27*, 3423–3424.

De Val, S., Chi, N.C., Meadows, S.M., Minovitsky, S., Anderson, J.P., Harris, I.S., Ehlers, M.L., Agarwal, P., Visel, A., Xu, S.-M., et al. (2008). Combinatorial Regulation of Endothelial Gene Expression by Ets and Forkhead Transcription Factors. Cell *135*, 1053–1064.

Deaton, A.M., Gómez-Rodríguez, M., Mieczkowski, J., Tolstorukov, M.Y., Kundu, S., Sadreyev, R.I., Jansen, L.E., and Kingston, R.E. Enhancer regions show high histone H3.3 turnover that changes during differentiation. ELife *5*.

Denholtz, M., Zhu, Y., He, Z., Lu, H., Isoda, T., Döhrmann, S., Nizet, V., and Murre, C. (2020). Upon microbial challenge, human neutrophils undergo rapid changes in nuclear architecture and chromatin folding to orchestrate an immediate inflammatory gene program. Genes Dev. *34*, 149–165.

Diaz, A., Park, K., Lim, D.A., and Song, J.S. (2012). Normalization, bias correction, and peak calling for ChIP-seq. Stat. Appl. Genet. Mol. Biol. *11*.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardımcı, G.G., Chakraborty, A., Bann, D.V., Wang, Y., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet. *50*, 1388–1398.

Donohoe, M.E., Zhang, L.-F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf Cofactor, Yy1, for the X Chromosome Binary Switch. Mol. Cell *25*, 43–56.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., et al. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat. Struct. Mol. Biol. *15*, 1176–1183.

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., et al. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature *570*, 395–399.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell *143*, 212–224.

Fudenberg, G., Kelley, D.R., and Pollard, K.S. (2019). Predicting 3D genome folding from DNA sequence (Genomics).

Grewal, S.S., and Elgin, S.C.R. (2007). Transcription and RNAi in the formation of heterochromatin. Nature *447*, 399–406.

Grotthuss, M. von, Ashburner, M., and Ranz, J.M. (2010). Fragile regions and not functional constraints predominate in shaping gene organization in the genus Drosophila. Genome Res. *20*, 1084–1096.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Hadjimichael, C., Chanoumidou, K., Papadopoulou, N., Arampatzi, P., Papamatheakis, J., and Kretsovali, A. (2015). Common stemness regulators of embryonic and cancer stem cells. World J. Stem Cells *7*, 1150–1184.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015a). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. J. Cell Biol. *208*, 33–52.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015b). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. J. Cell Biol. *208*, 33–52.

He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W., and Rosenfeld, M.G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature *340*, 35–42.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hiratani, I., Takebayashi, S., Lu, J., and Gilbert, D.M. (2009). Replication timing and transcriptional control: beyond cause and effect--part II. Curr. Opin. Genet. Dev. *19*, 142–149.

Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. *19*, 1419–1428.

Hota, S.K., Blair, A.P., Rao, K.S., So, K., Blotnick, A.M., Desai, R.V., Weinberger, L.S., Kathiriya, I.S., and Bruneau, B.G. (2020). Chromatin remodeler Brahma safeguards canalization in cardiac mesoderm differentiation (Developmental Biology).

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95.
Hurlin, P.J., Quéva, C., and Eisenman, R.N. (1997). Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating cells and mediates repression at Myc binding sites. Genes Dev. *11*, 44–58.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLoS Genet. *9*, e1003470.

Kathiriya, I.S., Rao, K.S., Iacono, G., Devine, W.P., Blair, A.P., Hota, S.K., Lai, M.H., Garay, B.I., Thomas, R., Gong, H.Z., et al. (2020). Modeling human TBX5 haploinsufficiency predicts regulatory networks for congenital heart disease. BioRxiv 835603.

Kazazian, H.H. (2004). Mobile Elements: Drivers of Genome Evolution. Science *303*, 1626–1632.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. *13*, R107.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. Cell *153*, 178–192.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. Cell *163*, 134–147.

Koh, P.W., Sinha, R., Barkal, A.A., Morganti, R.M., Chen, A., Weissman, I.L., Ang, L.T., Kundaje, A., and Loh, K.M. (2016). An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. Sci. Data *3*, 1–15.

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Lee, D.H., Kim, T.M., Kim, J.K., and Park, C. (2019a). ETV2/ER71 Transcription Factor as a Therapeutic Vehicle for Cardiovascular Disease. Theranostics *9*, 5694–5705.

Lee, J., Termglinchan, V., Diecke, S., Itzhaki, I., Lam, C.K., Garg, P., Lau, E., Greenhaw, M., Seeger, T., Wu, H., et al. (2019b). Activation of PDGF pathway links LMNA mutation to dilated cardiomyopathy. Nature *572*, 335–340.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE Trans. Vis. Comput. Graph. *20*, 1983–1992.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lin, Y., Bloodgood, B.L., Hauser, J.L., Lapan, A.D., Koon, A.C., Kim, T.-K., Hu, L.S., Malik, A.N., and Greenberg, M.E. (2008). Activity-dependent regulation of inhibitory synapse development by Npas4. Nature *455*, 1198–1204.

Liu, Q., Jiang, C., Xu, J., Zhao, M.-T., Bortle, K.V., Cheng, X., Wang, G., Chang, H.Y., Wu, J.C., and Snyder, M.P. (2017). Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. Circ. Res.

Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., et al. (2014a). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237–252.

Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., et al. (2014b). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237–252.

Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451–467.

Long, H., Zhang, L., Lv, M., Wen, Z., Zhang, W., Chen, X., Zhang, P., Li, T., Chang, L., Jin, C., et al. (2020). H2A.Z facilitates licensing and activation of early replication origins. Nature *577*, 576–581.

Luna-Zurita, L., Stirnimann, C.U., Glatt, S., Kaynak, B.L., Thomas, S., Baudin, F., Samee, M.A.H., He, D., Small, E.M., Mileikovsky, M., et al. (2016). Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. Cell *164*, 999–1014.

Lund, E., Oldenburg, A.R., and Collas, P. (2014). Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. Nucleic Acids Res. *42*, e92.

Ma, J., and Duan, Z. (2019). Replication Timing Becomes Intertwined with 3D Genome Organization. Cell *176*, 681–684.

Malik, P., Korfali, N., Srsen, V., Lazou, V., Batrakou, D.G., Zuleger, N., Kavanagh, D.M., Wilkie, G.S., Goldberg, M.W., and Schirmer, E.C. (2010). Cell-specific and lamin-dependent targeting of novel transmembrane proteins in the nuclear envelope. Cell. Mol. Life Sci. *67*, 1353–1369.

Marchal, C., Sima, J., and Gilbert, D.M. (2019). Control of DNA replication timing in the 3D genome. Nat. Rev. Mol. Cell Biol. *20*, 721–737.

Martin, R.M., Fowler, J.L., Cromer, M.K., Lesch, B.J., Ponce, E., Uchida, N., Nishimura, T., Porteus, M.H., and Loh, K.M. (2020). Improving the safety of human pluripotent stem cell therapies using genome-edited orthogonal safeguards. Nat. Commun. *11*, 2713.

Meuleman, W., Peric-Hupkes, D., Kind, J., Beaudry, J.-B., Pagie, L., Kellis, M., Reinders, M., Wessels, L., and Steensel, B. van (2013). Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res. *23*, 270–280.

Michael Waskom, Olga Botvinnik, Joel Ostblom, Maoz Gelbart, Saulius Lukauskas, Paul Hobson, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, et al. (2020). mwaskom/seaborn: v0.10.1 (April 2020) (Zenodo).

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22.

Otterstrom, J., Castells-Garcia, A., Vicario, C., Gomez-Garcia, P.A., Cosma, M.P., and Lakadamyali, M. (2019). Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. Nucleic Acids Res. *47*, 8470–8484.

Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet. *20*, 89–108.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Pereira, F.A., Qiu, Y., Zhou, G., Tsai, M.-J., and Tsai, S.Y. (1999). The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. Genes Dev. *13*, 1037–1049.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol. Cell *38*, 603–613.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.

Poleshko, A., Shah, P.P., Gupta, M., Babu, A., Morley, M.P., Manderfield, L.J., Ifkovits, J.L., Calderon, D., Aghajanian, H., Sierra-Pagán, J.E., et al. (2017). Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. Cell *171*, 573-587.e14.

Poleshko, A., Smith, C.L., Nguyen, S.C., Sivaramakrishnan, P., Wong, K.G., Murray, J.I., Lakadamyali, M., Joyce, E.F., Jain, R., and Epstein, J.A. (2019). H3K9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. ELife *8*, e49278.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rhind, N., and Gilbert, D.M. (2013). DNA Replication Timing. Cold Spring Harb. Perspect. Biol. *5*.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell *160*, 1145–1158.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. J. Mol. Biol. *381*, 816–825.

Robson, M.I., de las Heras, J.I., Czapiewski, R., Lê Thành, P., Booth, D.G., Kelly, D.A., Webb, S., Kerr, A.R.W., and Schirmer, E.C. (2016). Tissue-Specific Gene Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis. Mol. Cell *62*, 834–847.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19*, 789–800.

Schreiber, J. (2017). Pomegranate: fast and flexible probabilistic modeling in python. J. Mach. Learn. Res. *18*, 5992–5997.

Schwalie, P.C., Ward, M.C., Cain, C.E., Faure, A.J., Gilad, Y., Odom, D.T., and Flicek, P. (2013). Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. Genome Biol. *14*, R148.

Solovei, I., Wang, A.S., Thanisch, K., Schmidt, C.S., Krebs, S., Zwerger, M., Cohen, T.V., Devys, D., Foisner, R., Peichl, L., et al. (2013). LBR and Lamin A/C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation. Cell *152*, 584–598.

Song, L.-J., Zhang, W.-J., Chang, Z.-W., Pan, Y.-F., Zong, H., Fan, Q.-X., and Wang, L.-X. (2015). PU.1 Is Identified as a Novel Metastasis Suppressor in Hepatocellular Carcinoma Regulating the miR-615-5p/IGF2 Axis. Asian Pac. J. Cancer Prev. *16*, 3667–3671.

Spiegel, I., Mardinly, A.R., Gabel, H.W., Bazinet, J.E., Couch, C.H., Tzeng, C.P., Harmin, D.A., and Greenberg, M.E. (2014). Npas4 Regulates Excitatory-Inhibitory Balance within Neural Circuits through Cell-Type-Specific Gene Programs. Cell *157*, 1216–1229.

Sun, Y., Zhou, B., Mao, F., Xu, J., Miao, H., Zou, Z., Phuc Khoa, L.T., Jang, Y., Cai, S., Witkin, M., et al. (2018). HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. Cancer Cell *34*, 643-658.e5.

Sundaram, V., and Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos. Trans. R. Soc. B Biol. Sci. *375*, 20190347.

Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S. (2010). Single-base resolution mapping of H1–nucleosome interactions and 3D organization of the nucleosome. Proc. Natl. Acad. Sci. *107*, 9620–9625.

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev. *13*, 484–494.

Targoff, K.L., Colombo, S., George, V., Schell, T., Kim, S.-H., Solnica-Krezel, L., and Yelon, D. (2013). Nkx genes are essential for maintenance of ventricular identity. Development *140*, 4203–4213.

Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. Cell Stem Cell *21*, 399-410.e7.

Tucker, N.R., Chaffin, M., Bedi, K.C., Papangeli, I., Akkad, A.-D., Arduini, A., Hayat, S., Eraslan, G., Muus, C., Bhattacharyya, R., et al. (2020). Myocyte Specific Upregulation of ACE2 in Cardiovascular Disease: Implications for SARS-CoV-2 mediated myocarditis (Cardiovascular Medicine).

Vadrot, N., Duband-Goulet, I., Cabet, E., Attanda, W., Barateau, A., Vicart, P., Gerbal, F., Briand, N., Vigouroux, C., Oldenburg, A.R., et al. (2015). The p.R482W substitution in A-type lamins deregulates SREBP1 activity in Dunnigan-type familial partial lipodystrophy. Hum. Mol. Genet. *24*, 2096–2109.

van Bemmel, J.G., Filion, G.J., Rosado, A., Talhout, W., de Haas, M., van Welsem, T., van Leeuwen, F., and van Steensel, B. (2013). A Network Model of the Molecular Organization of Chromatin in Drosophila. Mol. Cell *49*, 759–771.

Vazquez, B.N., Thackray, J.K., Simonet, N.G., Chahar, S., Kane-Goldsmith, N., Newkirk, S.J., Lee, S., Xing, J., Verzi, M.P., An, W., et al. (2019). SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina. Nucleic Acids Res. *47*, 7870–7885.

Walker, M., Billings, T., Baker, C.L., Powers, N., Tian, H., Saxl, R.L., Choi, K., Hibbs, M.A., Carter, G.W., Handel, M.A., et al. (2015). Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. Epigenetics Chromatin *8*, 31.

Walt, S. van der, Colbert, S.C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. Comput. Sci. Eng. *13*, 22–30.

Wang, J., Lunyak, V.V., and Jordan, I.K. (2012). Genome-wide prediction and analysis of human chromatin boundary elements. Nucleic Acids Res. *40*, 511–529.

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc. Natl. Acad. Sci. *104*, 18613–18618.

Wang, W., Li, X., Lee, M., Jun, S., Aziz, K.E., Feng, L., Tran, M.K., Li, N., McCrea, P.D., Park, J.-I., et al. (2015). FOXKs Promote Wnt/β-Catenin Signaling by Translocating DVL into the Nucleus. Dev. Cell *32*, 707–718.

Wang, Y., Zhang, Y., Zhang, R., van Schaik, T., Zhang, L., Sasaki, T., Hupkes, D.P., Chen, Y., Gilbert, D.M., van Steensel, B., et al. (2020). SPIN reveals genome-wide landscape of nuclear compartmentalization (Bioinformatics).

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588.e28.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

Worman, H.J. (2012). Nuclear lamins and laminopathies. J. Pathol. *226*, 316–325.

Worman, H.J., and Bonne, G. (2007). “Laminopathies”: A wide spectrum of human diseases. Exp. Cell Res. *313*, 2121–2133.

Wu, S., Cheng, C.-M., Lanz, R.B., Wang, T., Respress, J.L., Ather, S., Chen, W., Tsai, S.-J., Wehrens, X.H.T., Tsai, M.-J., et al. (2013). Atrial Identity Is Determined by a COUP-TFII Regulatory Network. Dev. Cell *25*, 417–426.

Wurst, W., Auerbach, A.B., and Joyner, A.L. (1994). Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. Dev. Camb. Engl. *120*, 2065–2075.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019a). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019b). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. *20*, 535–550.

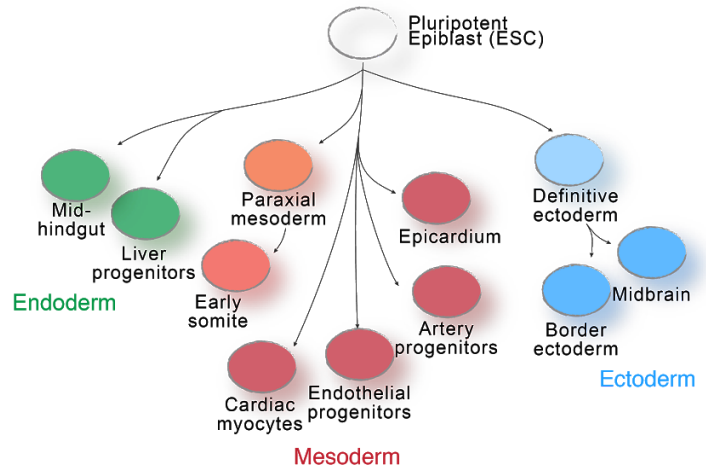
Zhu, Q., Song, L., Peng, G., Sun, N., Chen, J., Zhang, T., Sheng, N., Tang, W., Qian, C., Qiao, Y., et al. (2014). The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. ELife *3*.

Zuo, B., Yang, J., Wang, F., Wang, L., Yin, Y., Dan, J., Liu, N., and Liu, L. (2012). Influences of lamin A levels on induction of pluripotent stem cells. Biol. Open *1*, 1118–1127.

Figures and Figure Legends

Figure 1:

A



B

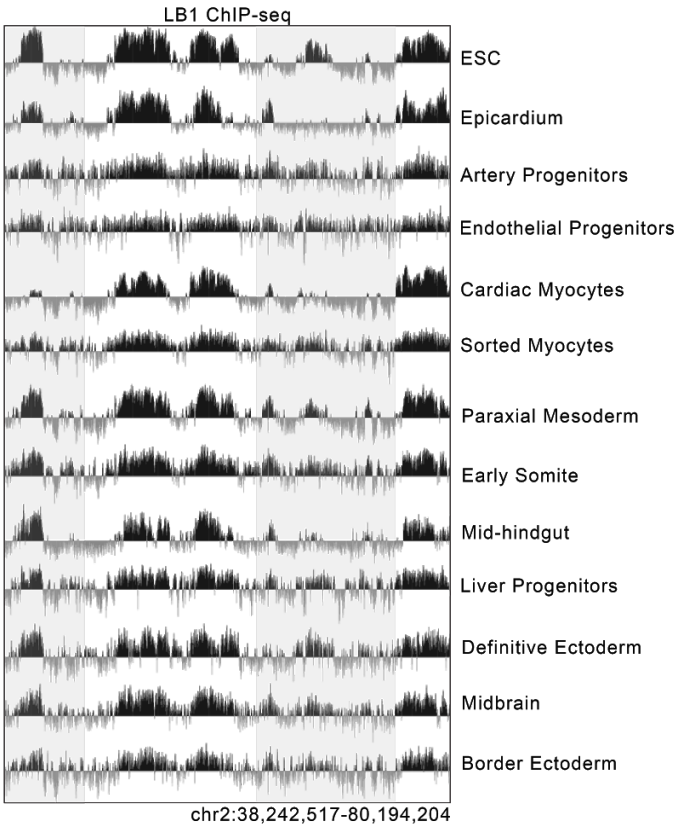


Figure 1: Mapping peripheral chromatin in human cell types. A. Illustration of the cell types and associated developmental lineages in this study (see also Table 1). B. Genome browser view of LB1 ChIPseq data across all thirteen cell types in a 41Mb region of Chromosome 2. Shaded gray boxes highlight LB1 regions that are gained and lost across different cell types versus domains which appear to be stably maintained throughout development.

Figure 2:

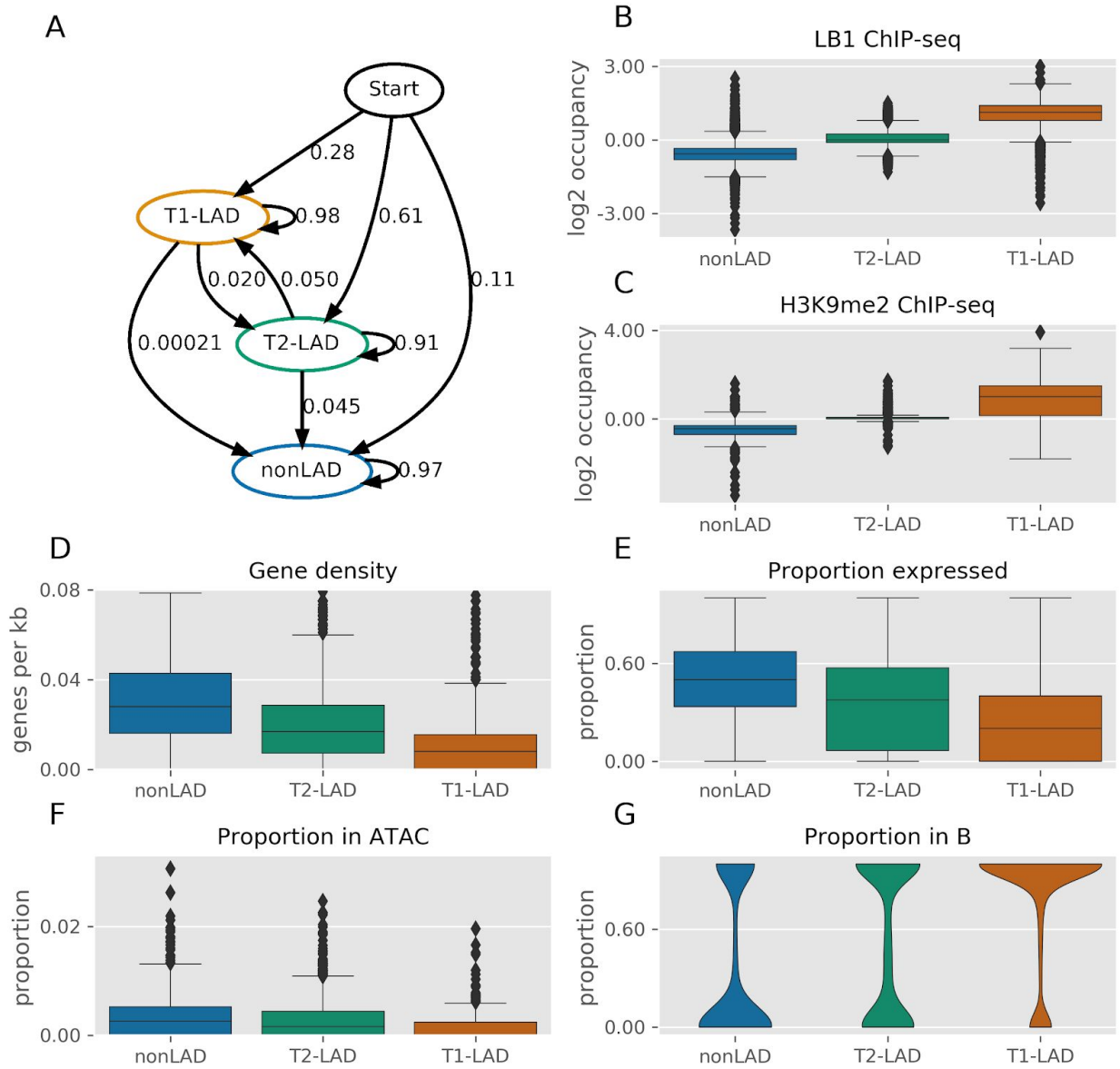


Figure 2: HMM to identify and characteristic features of T1- and T2-LADs in ESCs (see Supplementary Figs. 1-7 for other cell types). A: Schematic diagram of the 3-state HMM. States correspond to T1-LADs (orange), T2-LADs (green) and non-LADs (blue), as well as the model's initial state ("Start"), which is included for technical completeness. Lines between states are labeled with transition probabilities between those states. B-G: LB1 occupancy, H3K9me2 occupancy, gene density, proportion of genes expressed (quantile-normalized TPM < 0.0), proportion of domains in ATAC peaks and B compartment for T1- and T2-LADs.

Figure 3:

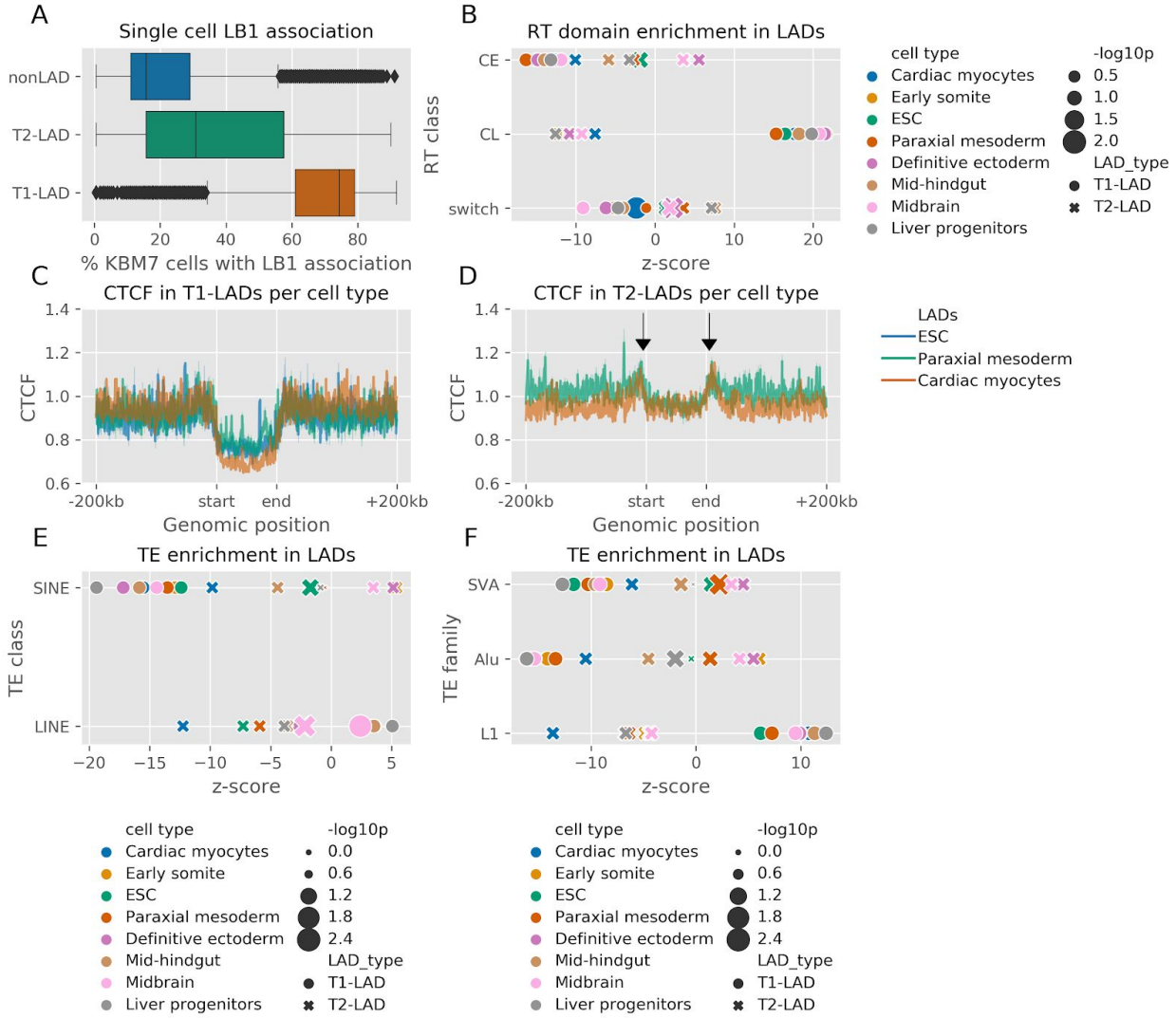


Figure 3: Mechanistic features of T1- and T2-LADs. A: Percentage of cells from previously published single-cell DamID in KBM7 human myeloid leukemia cells (Kind et al., 2015) with LB1 association for each domain category with LADs from paraxial mesoderm cells. B: Enrichment of constitutive early (CE), constitutive late (CL) and switch replication timing domains. C, D: CTCF binding (RPKM) in 500 bp bins (Zhang et al., 2019a) in T1- and T2-LADs using matched cell type data. E, F: Enrichment of transposable elements (TEs) classes (E) and families (F) from RepeatMasker downloaded from the UCSC Genome Browser.

Figure 4:

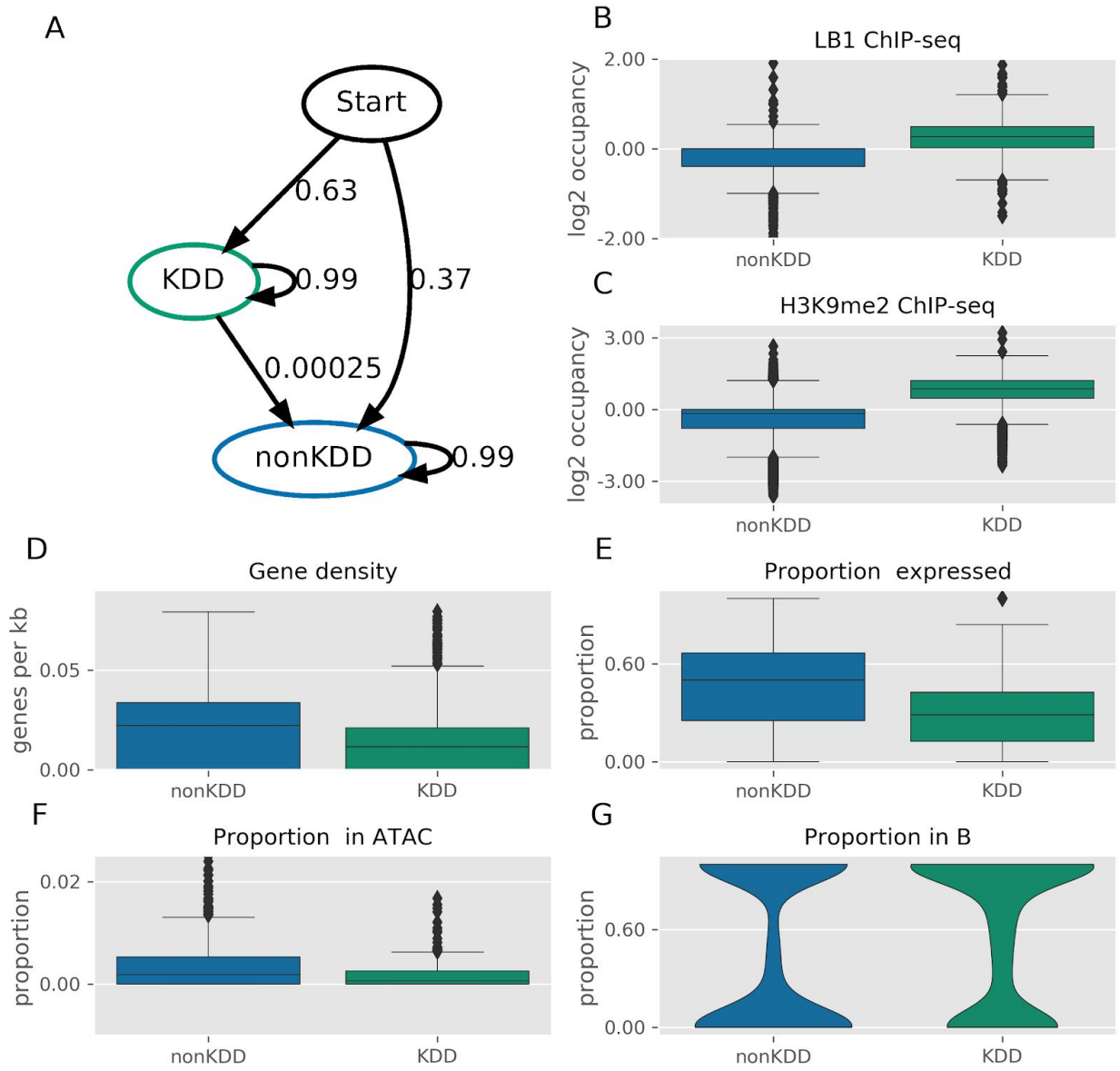


Figure 4: HMM to identify and characteristic features of KDDs in ESCs (See Supplementary Figs. 10-16 for other cell types). A: Schematic diagram of the 2-state HMM. States correspond to KDDs (green) and non-KDDs (blue), as well as the model's initial state ("Start"), which is included for technical completeness. Lines between states are labeled with transition probabilities between those states. B-G: LB1 occupancy, H3K9me2 occupancy, gene density, proportion of genes expressed (quantile-normalized TPM < 0.0), proportion of domains in ATAC peaks and B compartment.

Figure 5:

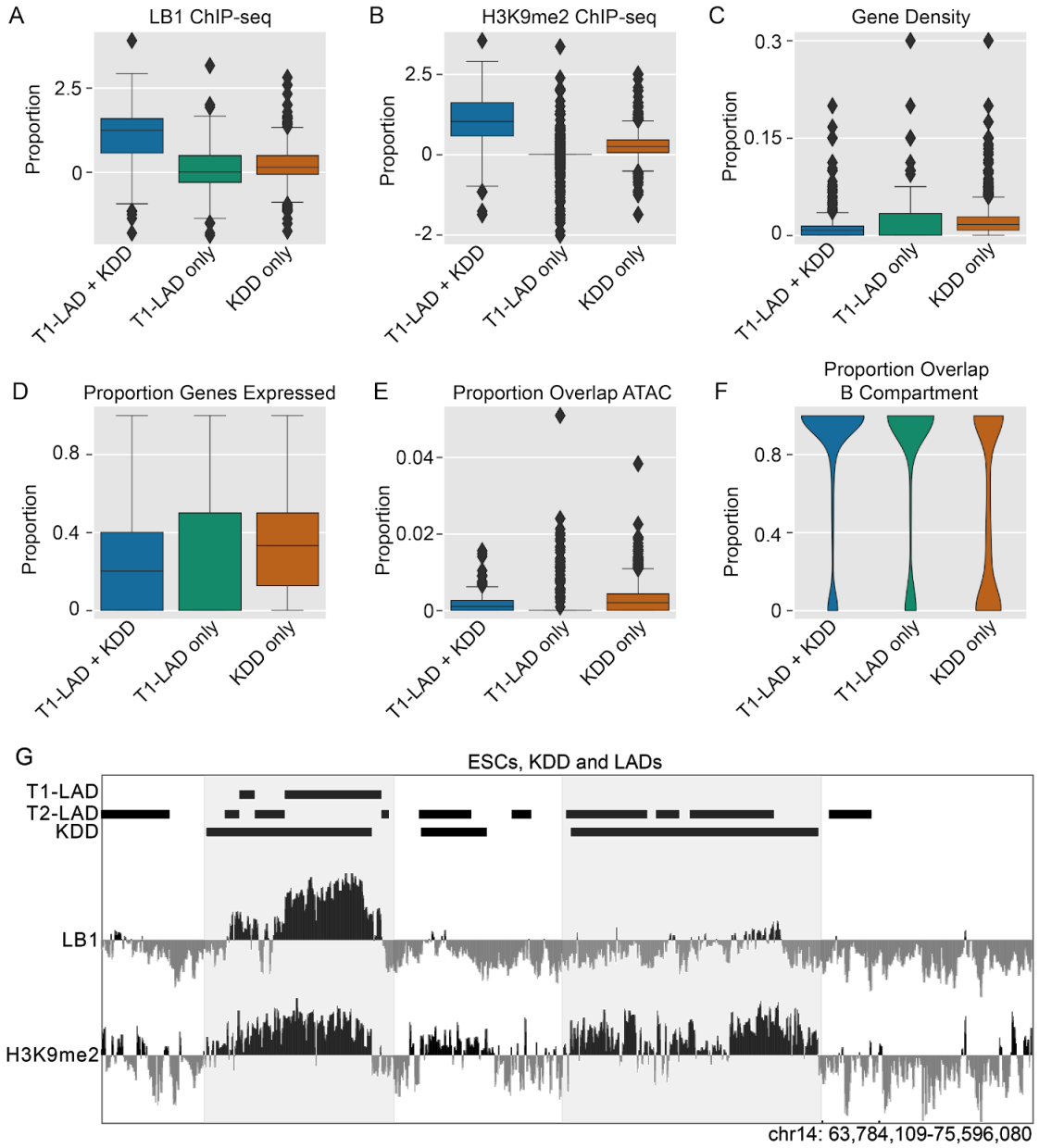


Figure 5: KDDs compared with T1-LADs in ESCs. A, B: LB1 and H3K9me2 binding. C: Gene density. D: Proportion genes expressed (quantile-normalized TPM > 0.0). E, F: Proportion overlap with ATAC-seq peaks matched based on cell type (Liu et al., 2017) and the B compartment (Zhang et al., 2019a). See SFigs. 17-20 for other cell types. G. Genome browser view of LB1 and H3K9me2 ChIPseq with accompanying KDD, T1- and T2-LAD calls (top) across an 11Mb region of Chromosome 14. Shaded gray boxes highlight a T1-LAD/KDD region flanked by a prominent KDD-only and KDD/T2-LAD region.

Figure 6:

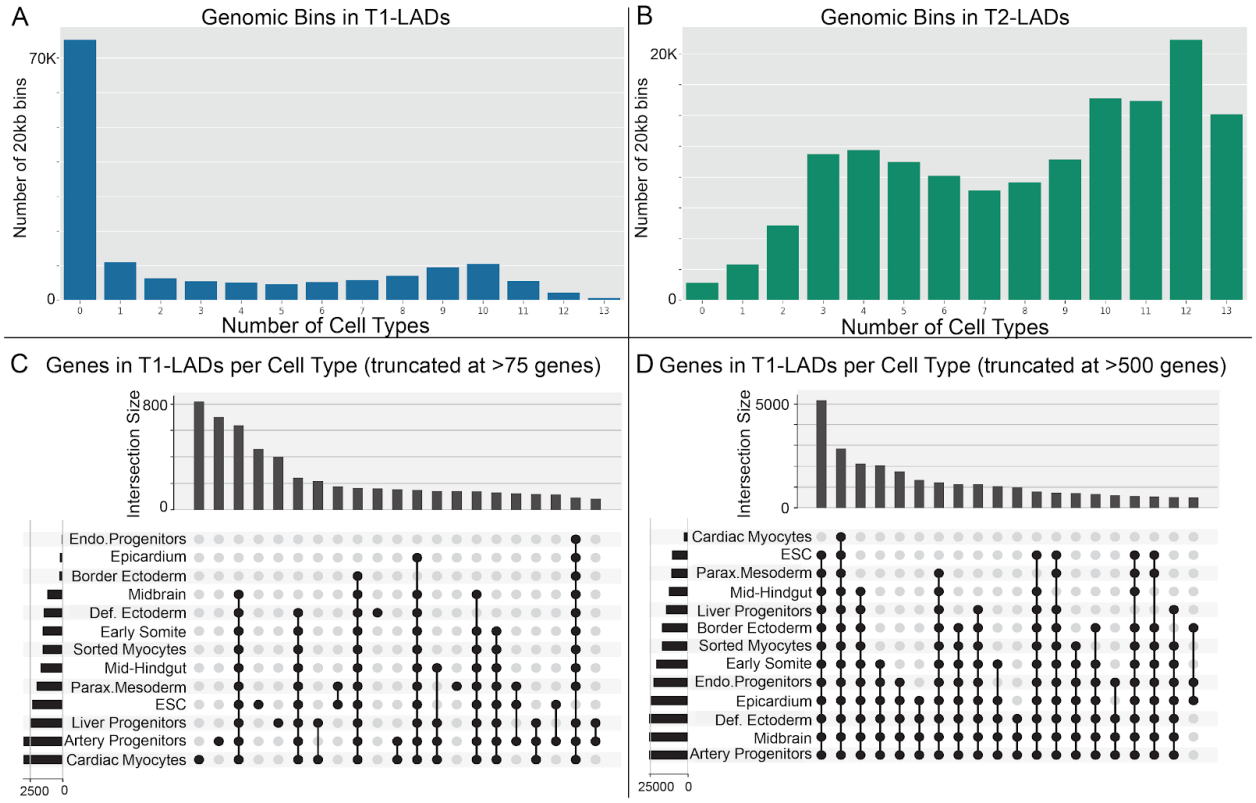


Figure 6: T1- and T2-LADs across cell types. A, B: Number of cell types for which each 20kb genomic bin is classified as T1- (A) or T2- (B) LAD. C, D: Genes in T1- (C) and T2- (D) LADs per cell type (clipped at smaller intersections for clarity). Any amount of gene overlap was considered (many genes are on the boundaries of LAD elements). Vertical bar plots represent the number of genes in the cell types indicated by the upset plot below the vertical bar. Number of genes in each T1 or T2 LAD are indicated in horizontal bar plots next to cell type.

Tables and Table Legends

Table 1:

Cell type	Germ Layer	Analysis Tier
Definitive Ectoderm	Ectoderm	1
Midbrain	Ectoderm	1
ESCs	Pluripotent	1
Mid-Hindgut	Endoderm	1
Liver Progenitors	Endoderm	1
Cardiac Myocytes	Mesoderm	1
Early Somite	Mesoderm	1
Paraxial Mesoderm	Mesoderm	1
Border Ectoderm	Ectoderm	2
Sorted Cardiac Myocytes	Mesoderm	2
Epicardium	Mesoderm	2
Artery Progenitors	Mesoderm	2
Endothelial Progenitors	Mesoderm	2

Table 1: LAMIN-B1 and H3K9me2 ChIP-seq data.

Cell types for which ChIP-seq data were generated with assigned quality tiers, with tier one being higher quality compared to tier two.

Table 2:

Cell type	AIC: 2 states	AIC: 3 states	AIC: 4 states	AIC: 5 states
Cardiac myocytes	389256.9	275060.7	299659.2	364001.4
Early somite	158571.2	99780.94	123336.9	136621
ESC	169847.1	99201.25	109000.9	114512.6
Paraxial mesoderm	232794.7	137728	180225.2	207855.9
Definitive ectoderm	31520.96	6536.492	148.8388	463.9411
Mid-hindgut	56455.64	40894.81	948.636	4801.792
Midbrain	44849.58	13976.73	-21631.2	-19878.3
Liver progenitors	141712.4	102587	68983.76	73101.81
Sorted cardiac myocytes	37390.8	5034.415	8502.368	7872.165
Border ectoderm	45887.52	23004.08	12933.4	13927.93
Epicardium	145791.9	99558.66	116515.3	139602.8
Artery progenitors	111366.6	78018.7	41764.8	44536.94
Endothelial progenitor	59377.17	38122.65	40137.52	40356.52

Table 2: Akaike Information Criterion for various numbers of HMM states trained on LAMIN-B1 ChIP-Seq data.

AIC statistics for various HMM states incorporated towards the decision of how many states to use for the final LAD model.

Table 3:

T1-LAD	T2-LAD	KDD	Region size range (kb)
+	-	+/-	80-280
+	-	+	180-830
+	-	+	260-1980
+/-	+/-	+	280-2460
-	+	+	140-3109
-	+	+/-	100-380
-	+	+	250-2360

Table 3: Size ranges LADs, KDDs.

Size ranges of T1- and T2-LADs and KDDs, considering all cell types.

Table 4:

Table 4: TFBM enrichments T2 vs T1 LADs (supplemental file)

TFBM enrichment results for T2 vs T1 LADs for CMs and ESCs (including SVAs, Alus, and T1-SINEs).

Table 5:

Cell type	AIC 2 states	AIC 3 states	AIC 4 states	AIC 5 states
Cardiac myocytes	666188.5	664856.1	627631.4	635808.5
Early somite	206873.5	236616.9	206254.6	206512.7
ESC	572211.2	602511.6	544268.2	553732.8
Paraxial mesoderm	169183.2	149286.7	105163.6	95363.34
Definitive ectoderm	204973.8	197252.4	121449	120097.8
Mid-hindgut	217423.6	217743	166534.1	165673.2
Midbrain	200791.9	205578.5	178946.8	176755.5
Liver progenitors	191137.8	179924.6	129030.4	126295.8
Sorted cardiac myocytes	195437.7	188457.2	101029.5	99702.61
Border ectoderm	170970.1	152679.6	44552.64	41451.46
Epicardium	297894.7	338171.2	291762.4	291932.1
Artery progenitors	171150.1	152808	104667.1	96831.84
Endothelial progenitors	174392.7	171664.1	3698.507	4664.605

Table 5: Akaike Information Criterion for various numbers of HMM states trained on H3K9me2 ChIP-Seq data.

AIC statistics for various HMM states incorporated towards the decision of how many states to use for the final KDD model.

Table 6:

Table 6: LAD and KDD category per gene (supplemental file)

LAD and KDD category for all genes per cell type.

Table 7:

Table 7: GO Analysis comparing gene signatures of different LADs (supplemental file)

GO analyses of gene sets in CM, ESC, and CTI LAD genes.

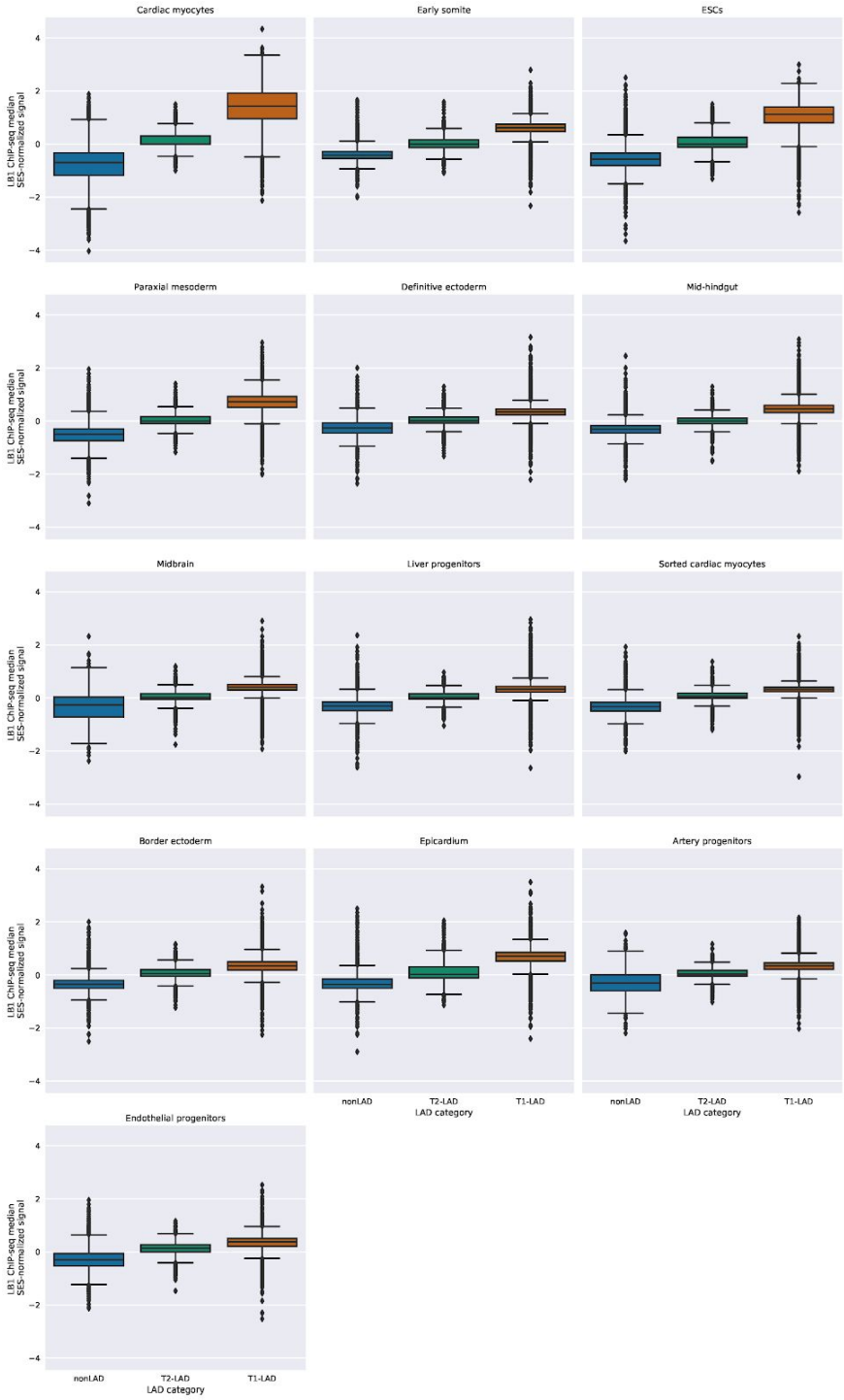
Table 8:

Table 8: TFBM enrichments cell-type-specific T1-LADs (supplemental file)

TFBM enrichment results for T1-LADs versus the union of T1-LADs for all cell types for CMs, midbrain and ESCs.

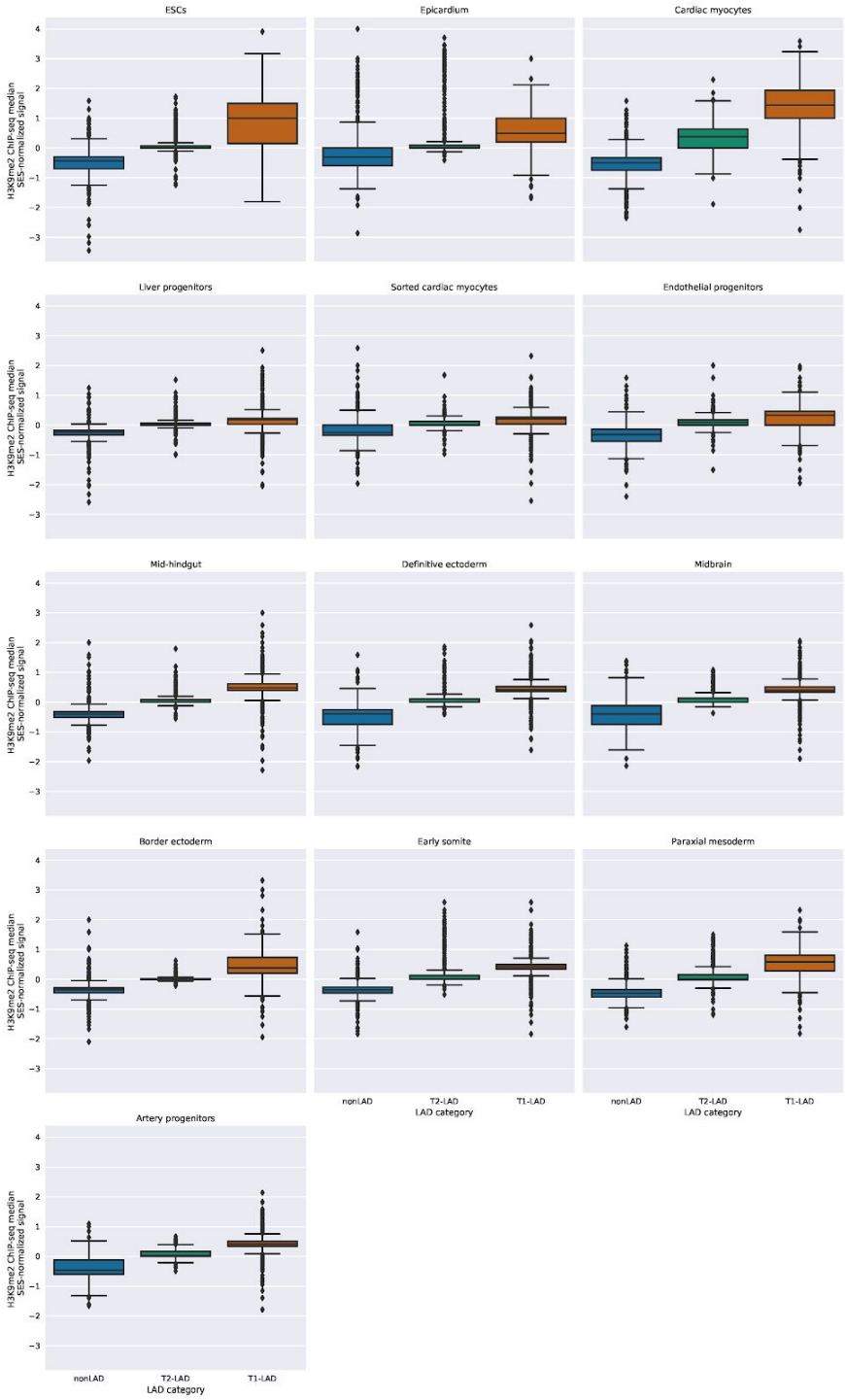
Supplementary Figures

SFig 1: LB1 occupancy in T1- and T2-LADs in all cell types



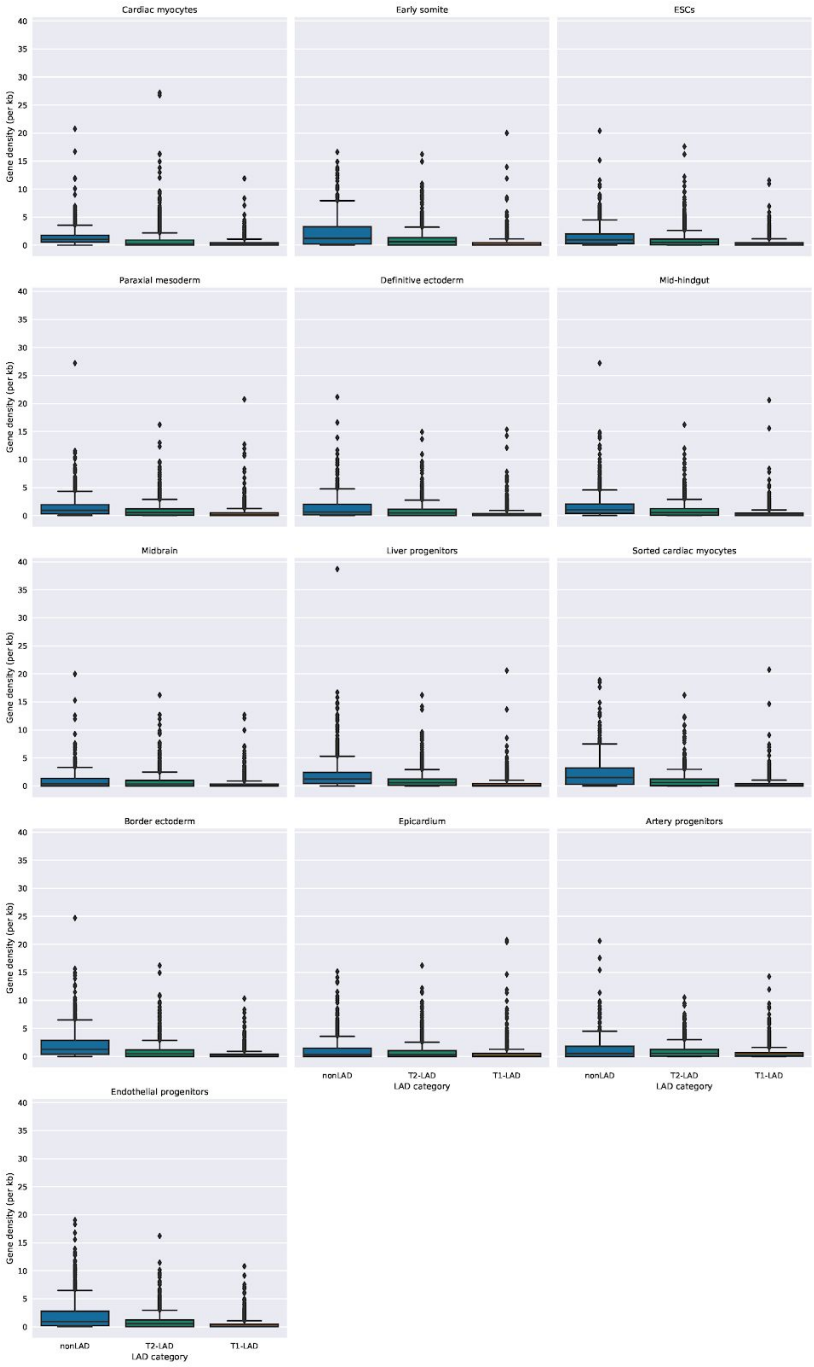
SFig. 1: LB1 occupancy in T1-, T2- and nonLADs across all cell types assayed in this study.

SFig. 2: H3K9me2 occupancy in T1- and T2-LADs in all cell types



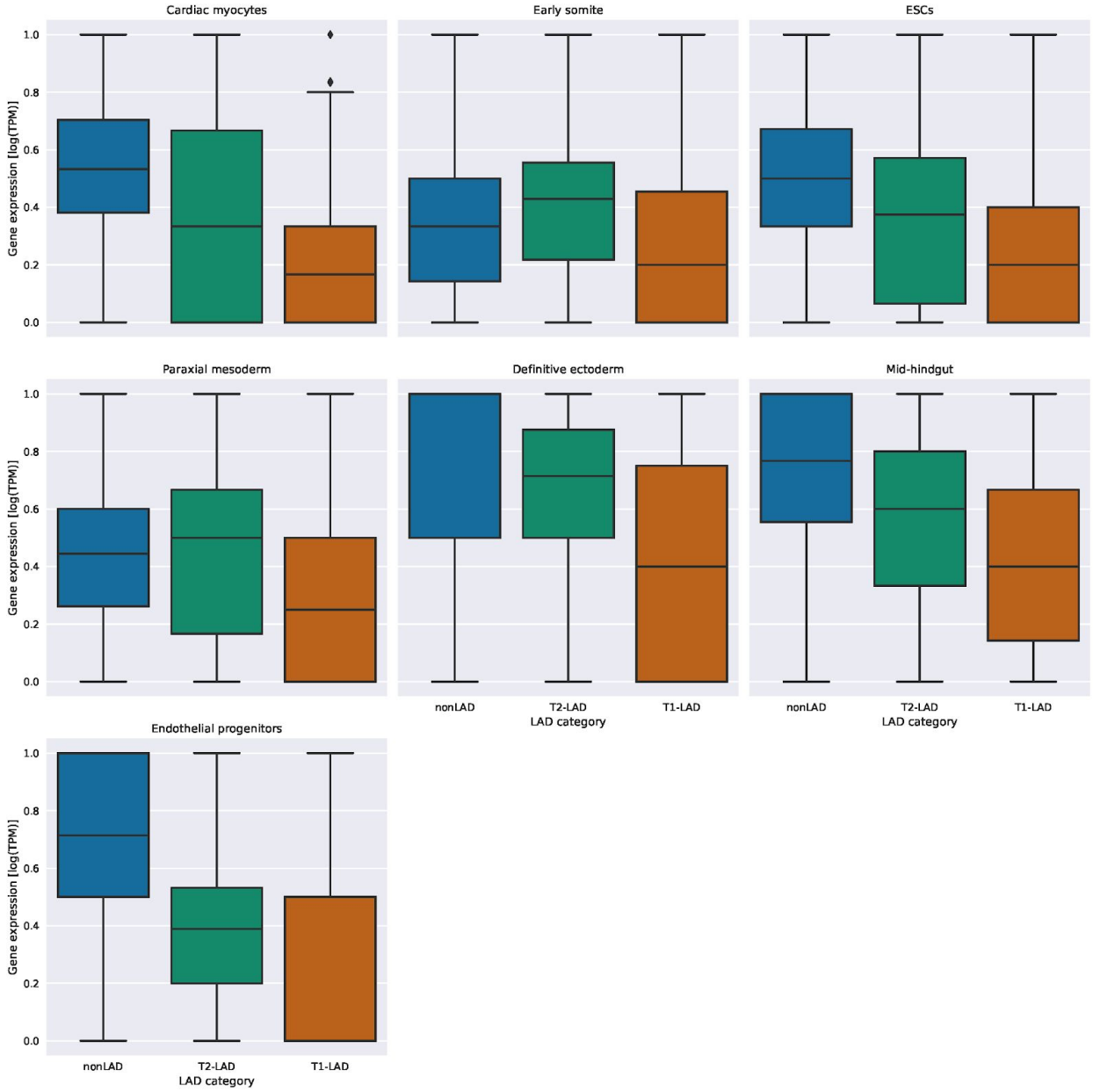
SFig. 2: H3K9me2 occupancy in T1-, T2- and nonLADs across all cell types assayed in this study.

SFig. 3: Gene density in T1- and T2-LADs in all cell types



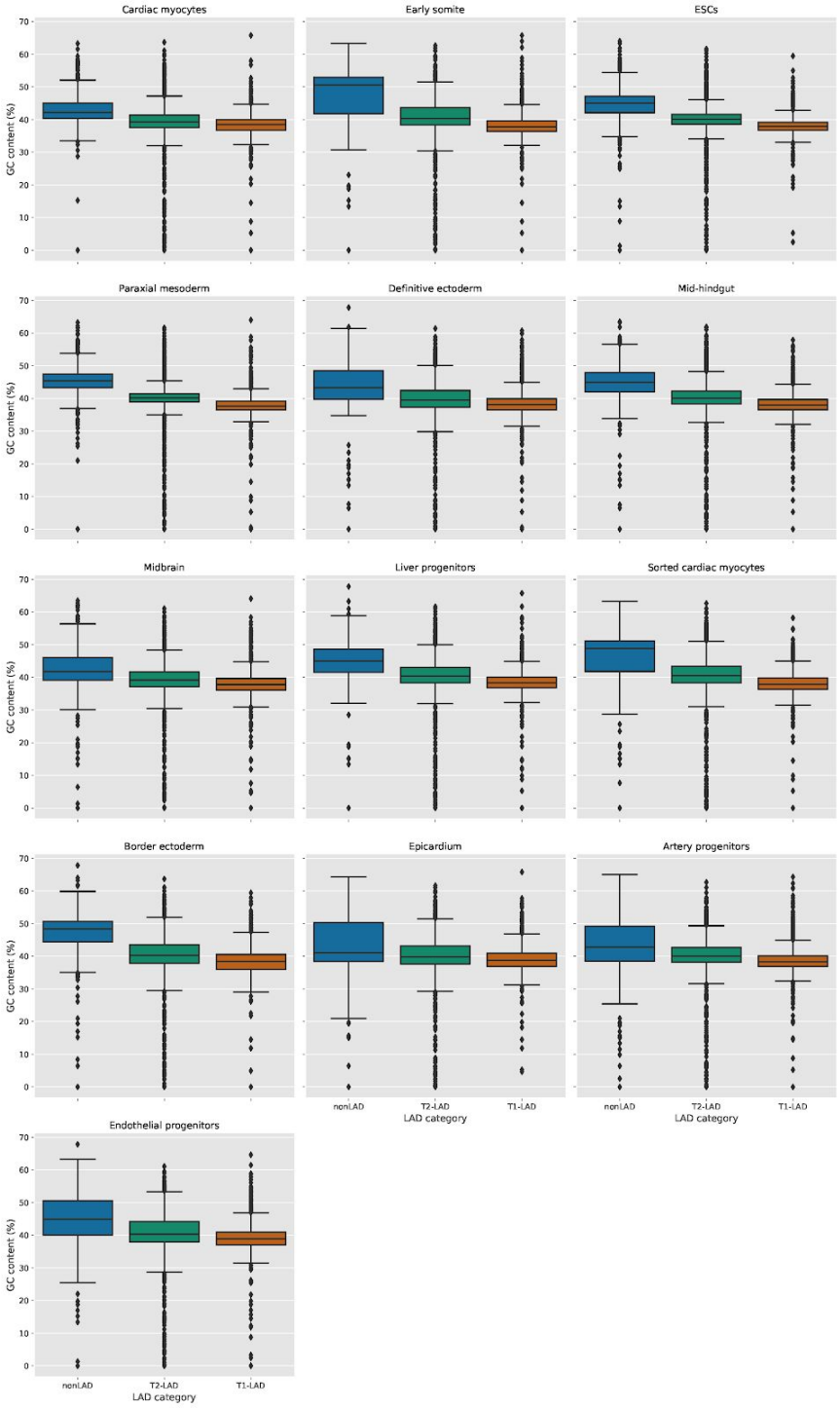
SFig. 3: Gene density in T1-, T2- and nonLADs across all cell types assayed in this study.

SFig. 4: Gene expression in T1- and T2-LADs across cell types



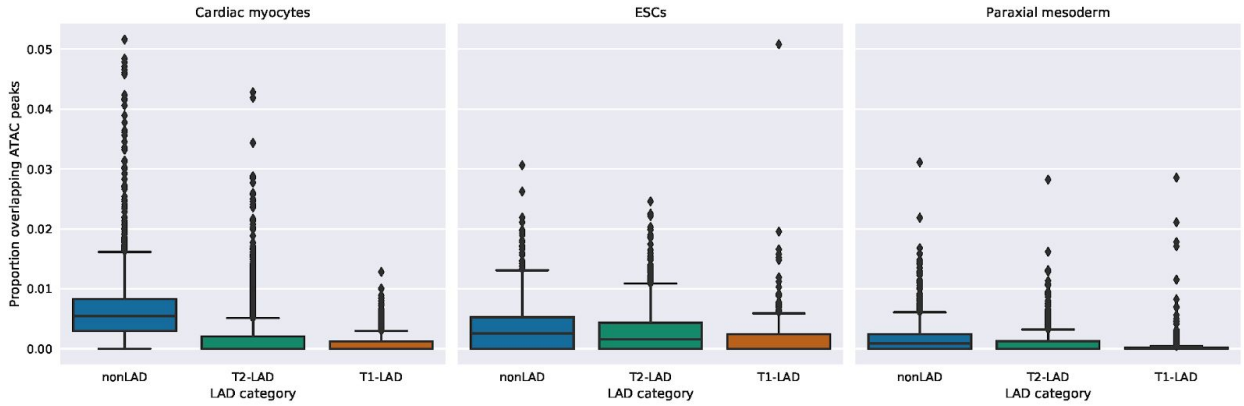
SFig. 4: Expression of genes in T1-, T2- and nonLADs across all cell types assayed in this study.

SFig. 5: GC content in T1- and T2-LADs across cell types



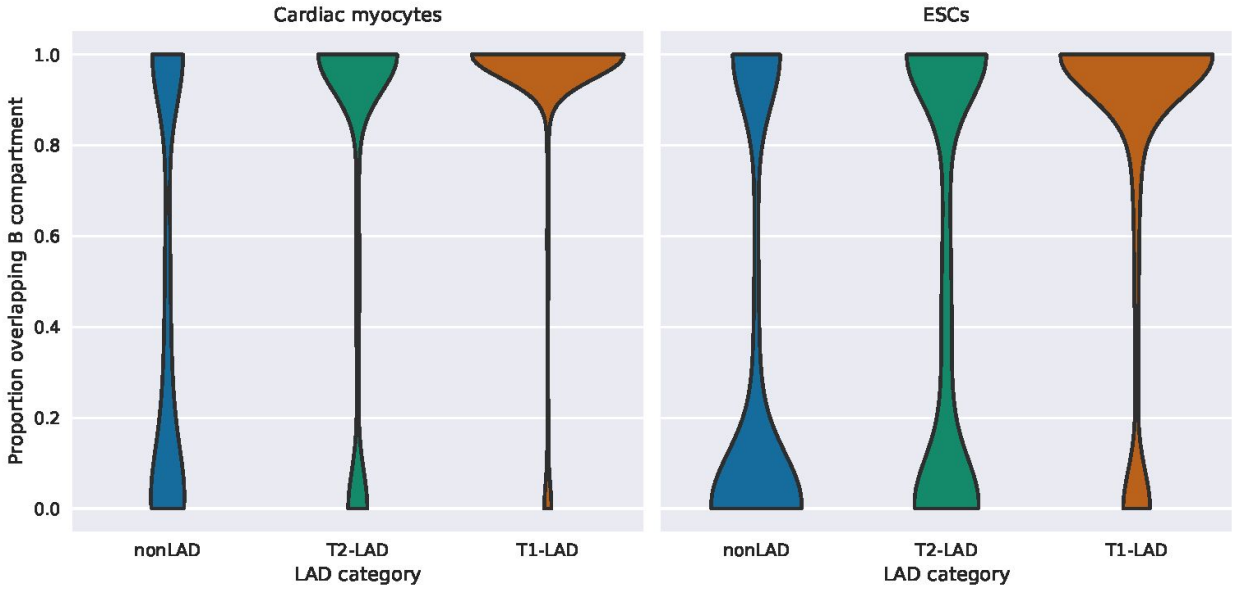
SFig. 5: GC content in T1-, T2- and nonLADs across all cell types assayed in this study.

SFig. 6: ATAC-seq peak overlap in T1- and T2-LADs across cell types



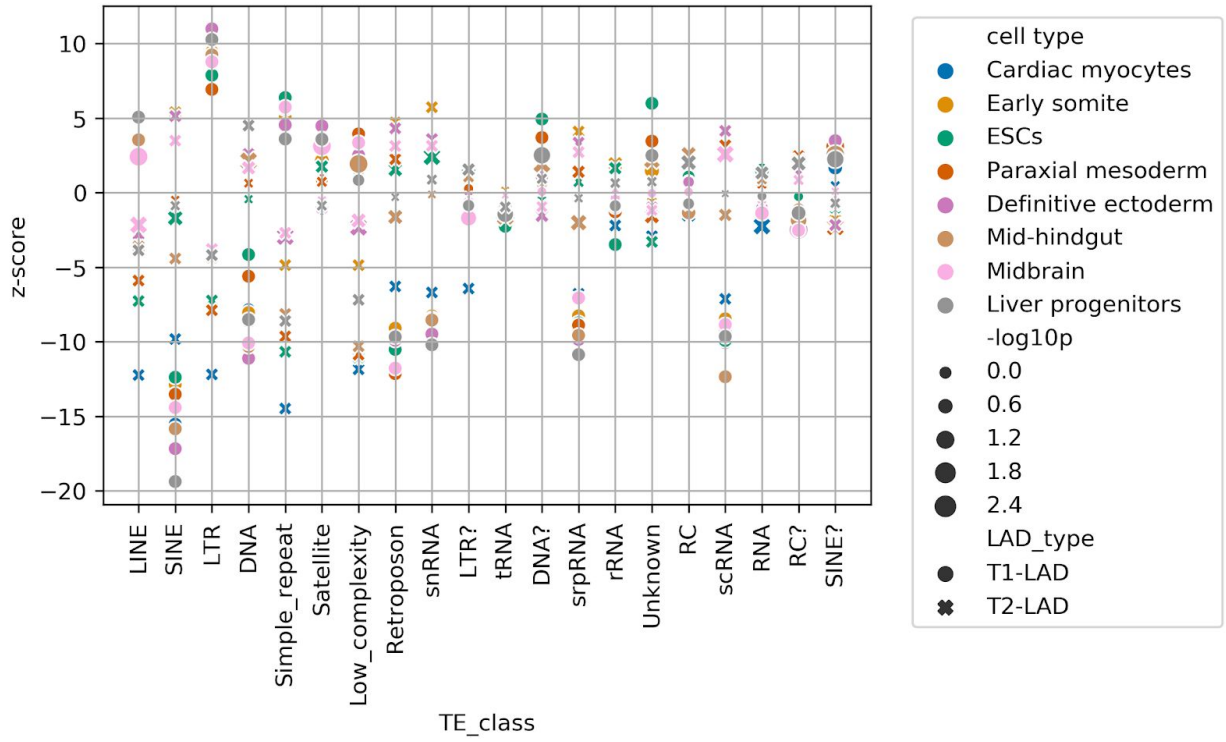
SFig. 6: ATAC-seq peak overlap from previous work (Zhang et al., 2019a) in T1-, T2- and nonLADs across matched cell types from this study.

SFig. 7: B compartment overlap in T1- and T2-LADs across cell types



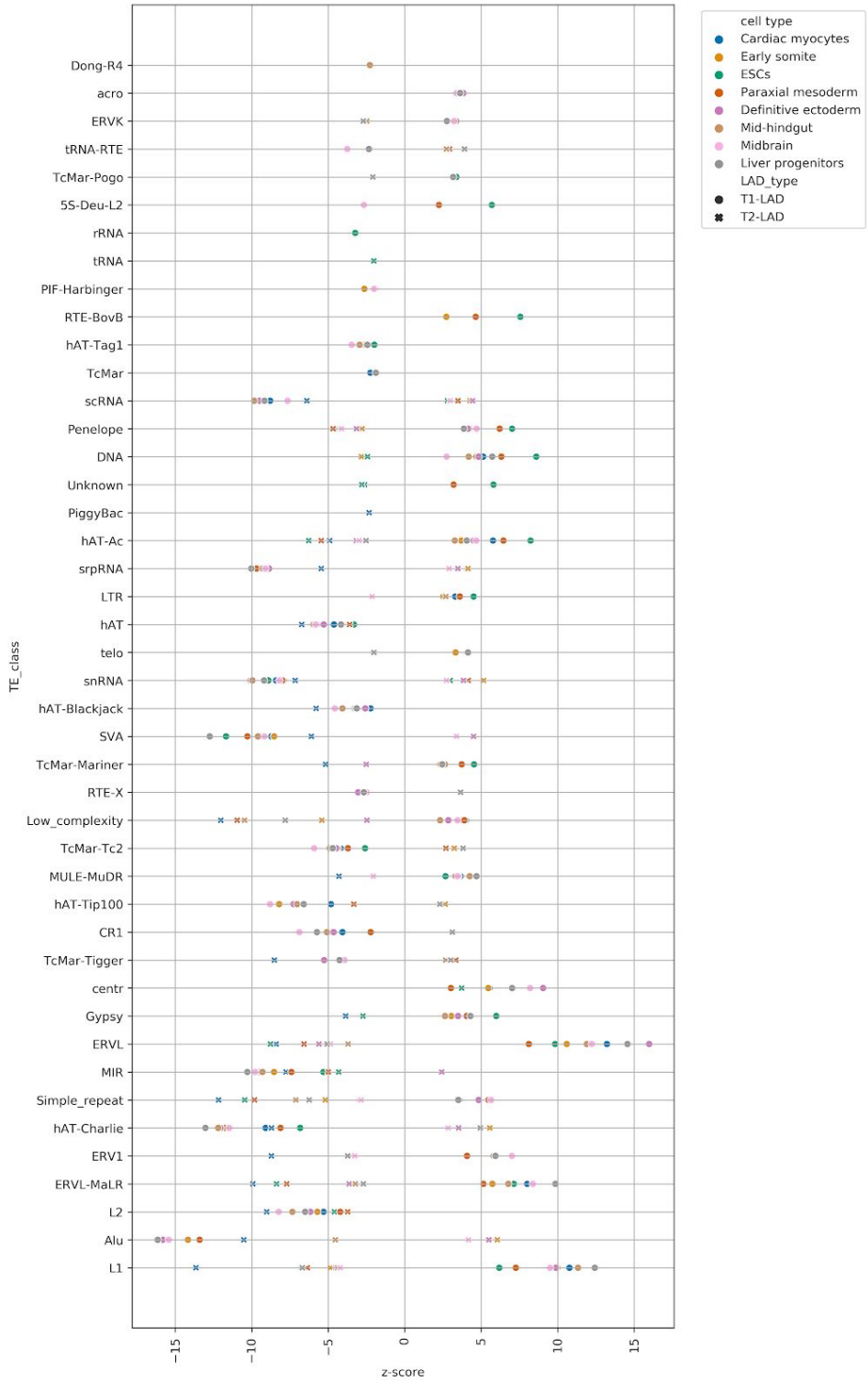
SFig. 7: B compartment overlap derived from Hi-C data generated in previous work (Zhang et al., 2019a) in T1-, T2- and nonLADs across matched cell types from this study.

SFig. 8: Enrichment of TE groups in T1- and T2-LADs across cell types



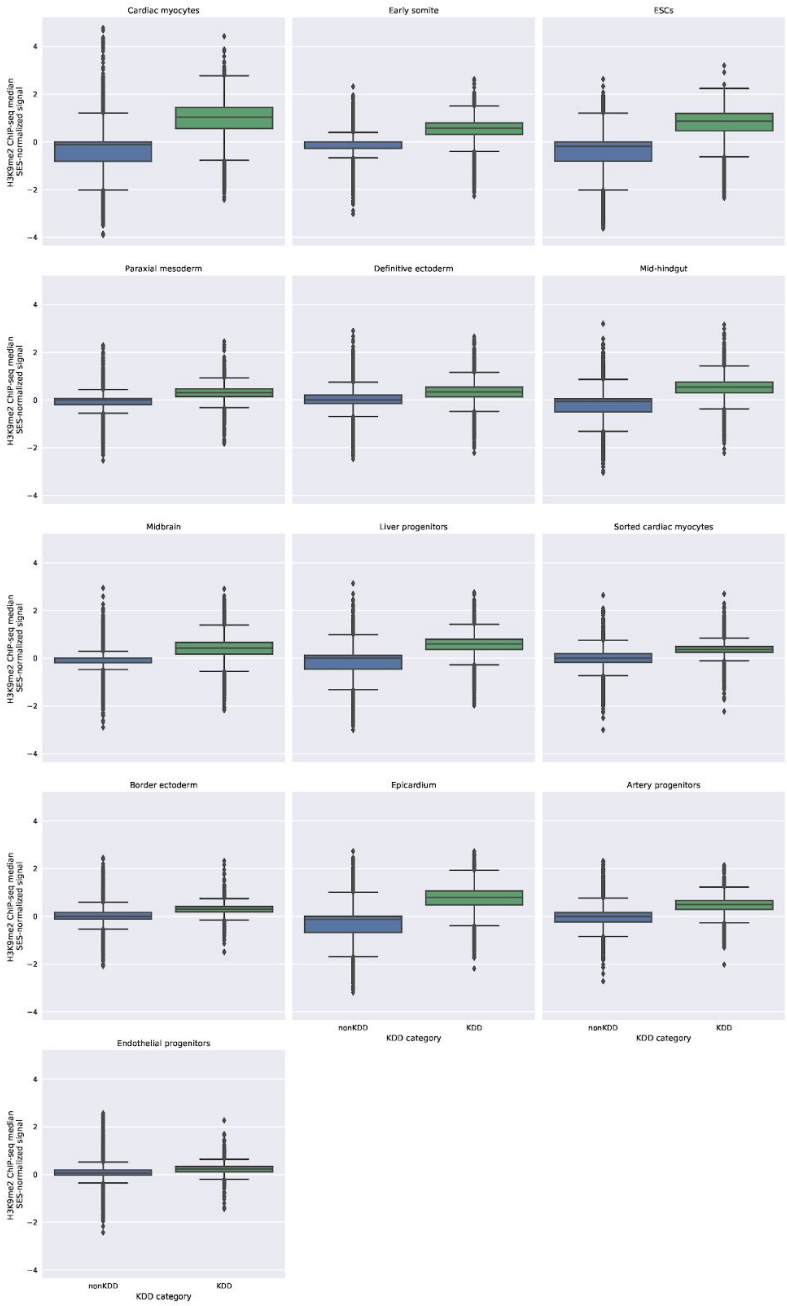
SFig. 8: Enrichment of TE groups in T1-, T2- and nonLADs (100 permutations) across all cell types assayed in this study.

SFig. 9: Enrichment of TE families in T1- and T2-LADs across cell types



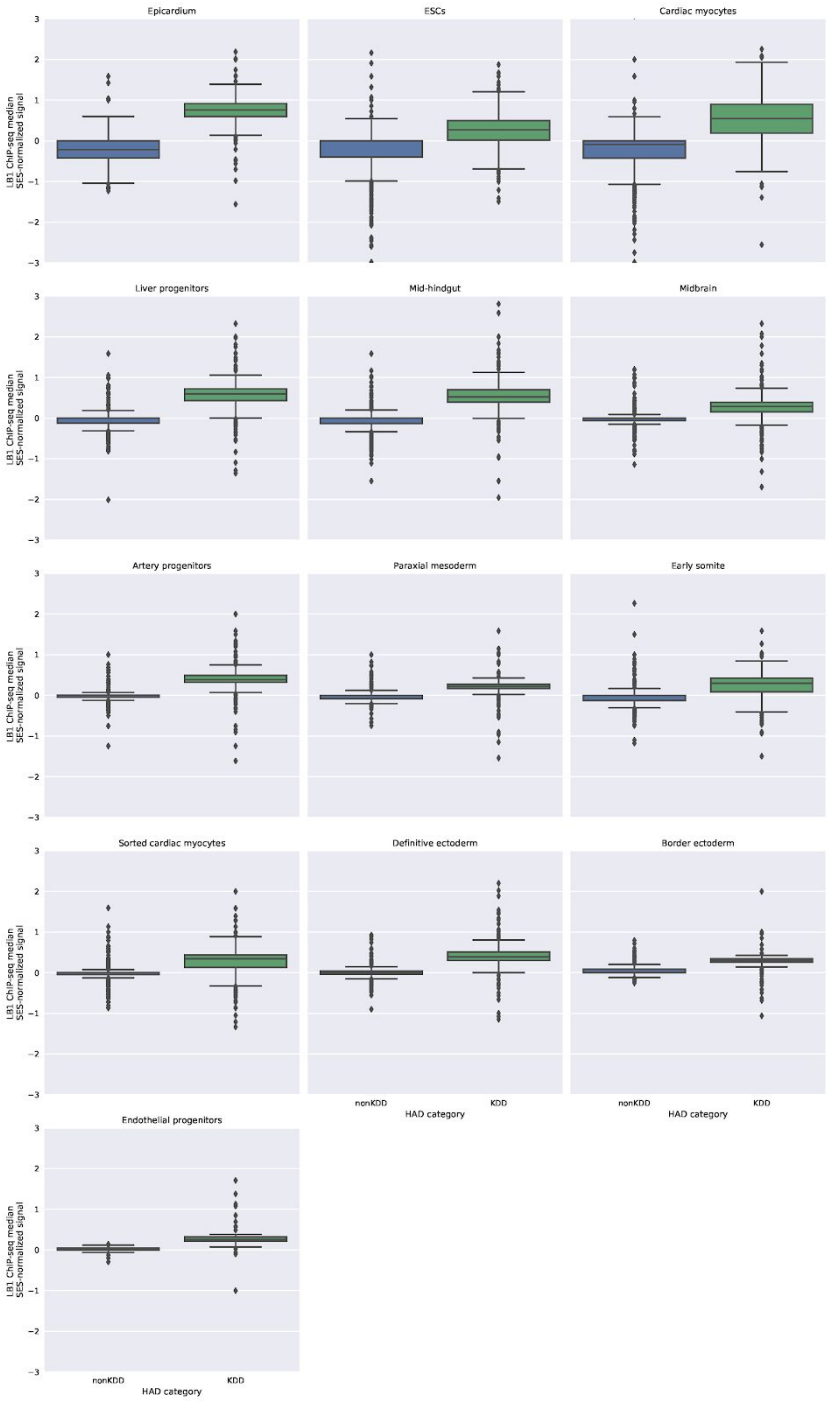
SFig. 9: Enrichment of TE families in T1-, T2- and nonLADs (100 permutations) across all cell types assayed in this study. For clarity, only significant results ($p < 0.01$) are shown.

SFig 10: H3K9me2 occupancy in KDDs in all cell types



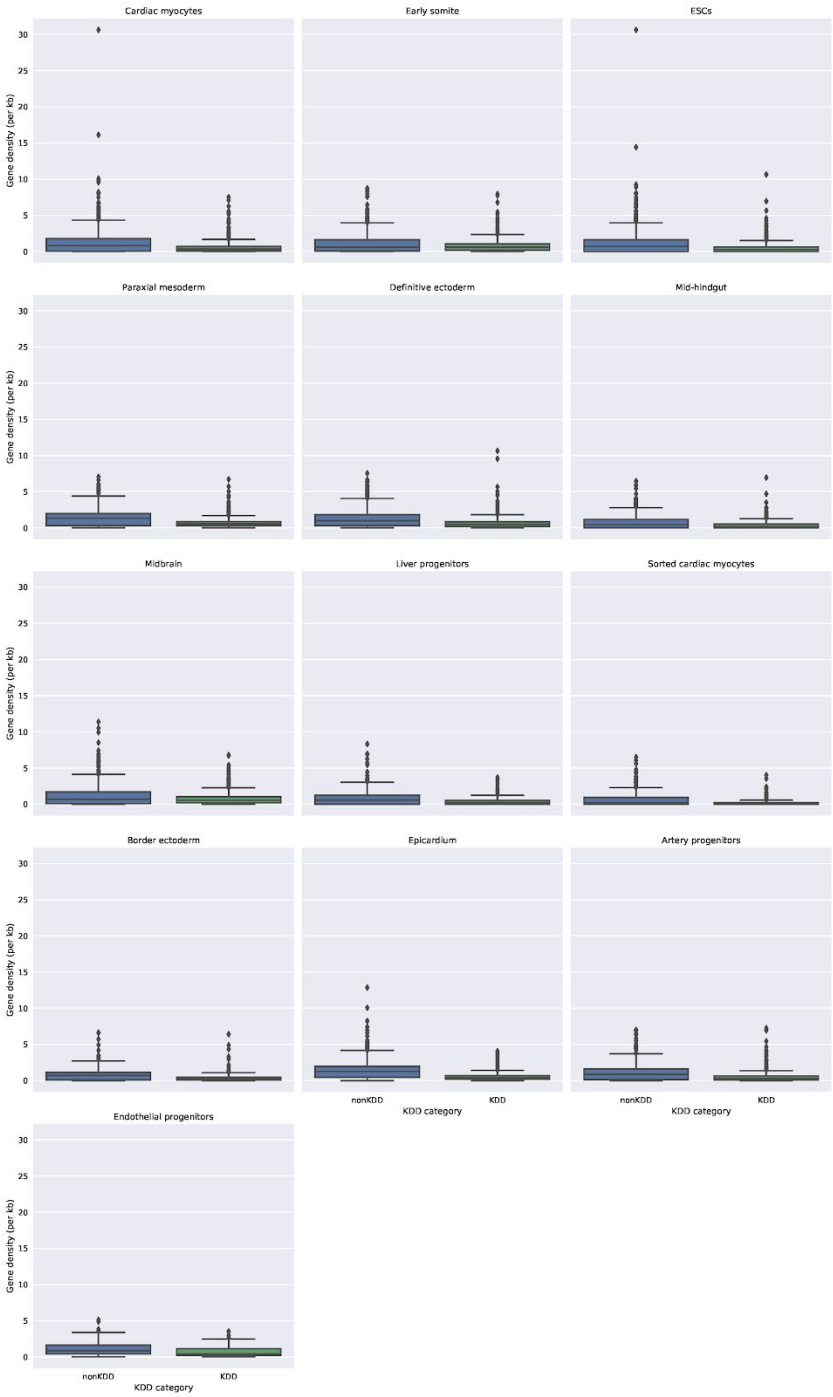
SFig. 10: H3K9me2 occupancy in KDDs and nonKDDs across all cell types assayed in this study.

SFig 11: LB1 occupancy in KDDs in all cell types



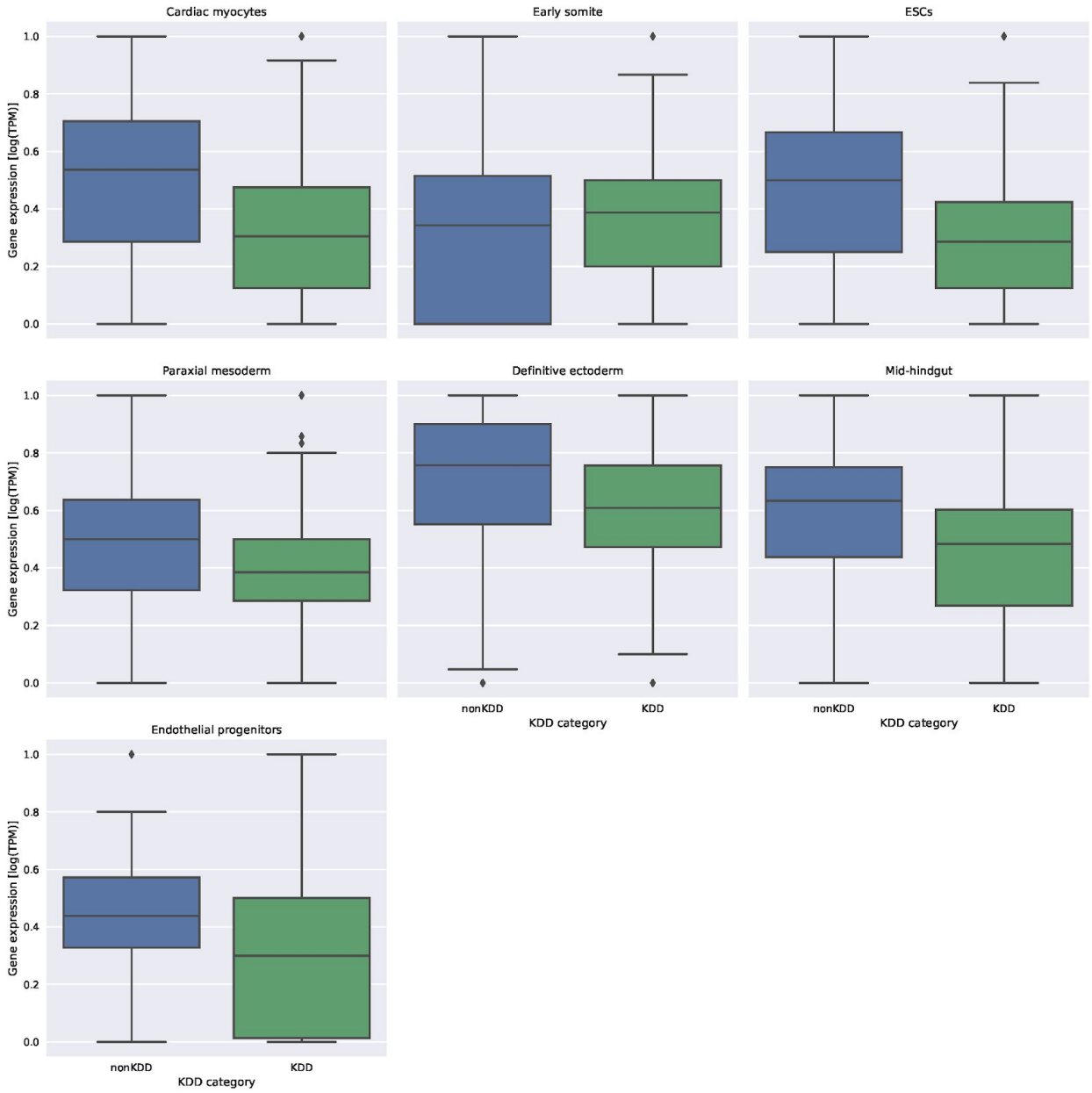
SFig. 11: LB1 occupancy in KDDs and nonKDDs across all cell types assayed in this study.

SFig 12: Gene density in KDDs in all cell types



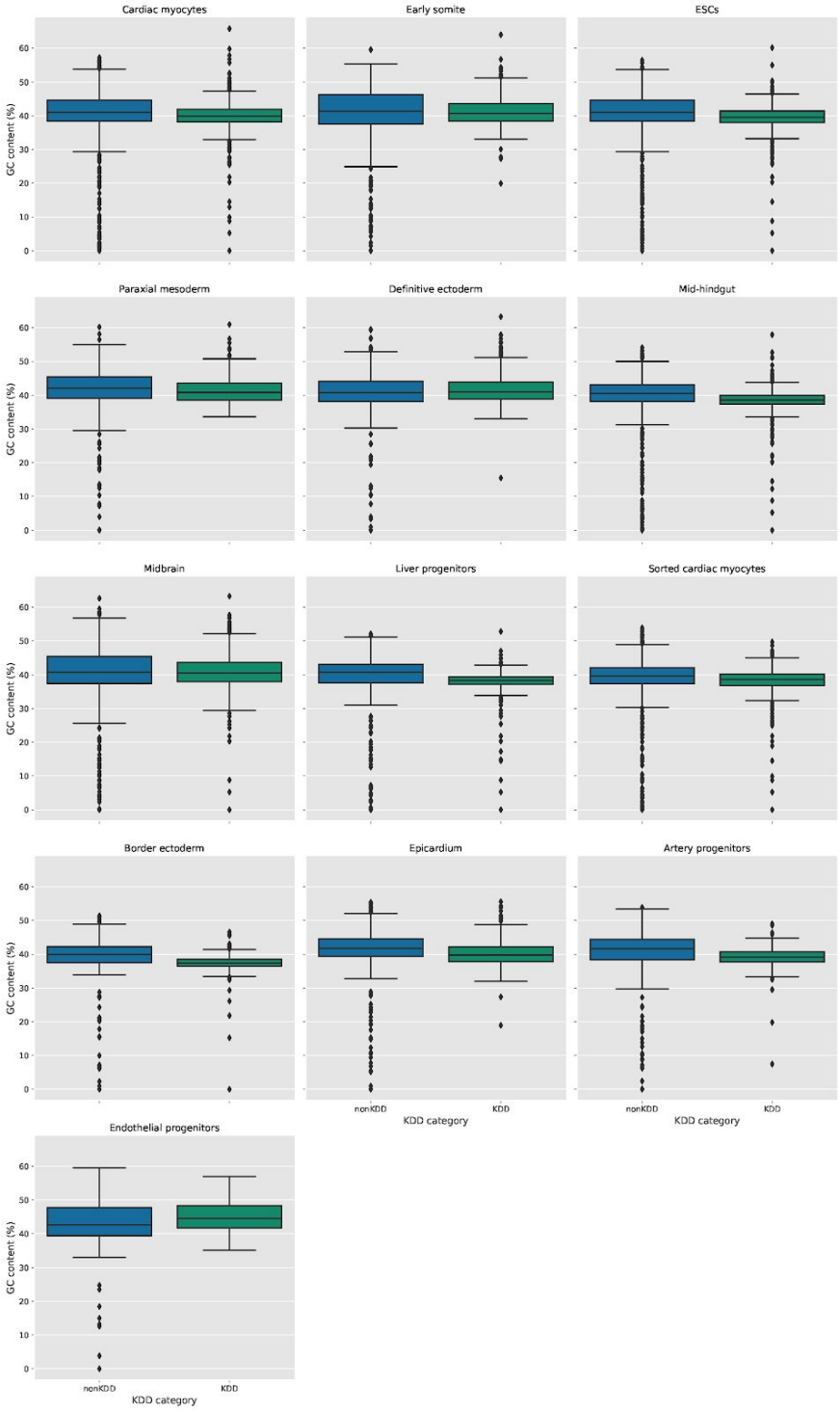
SFig. 12: Gene density in KDDs and nonKDDs across all cell types assayed in this study.

SFig 13: Gene expression in KDDs in all cell types



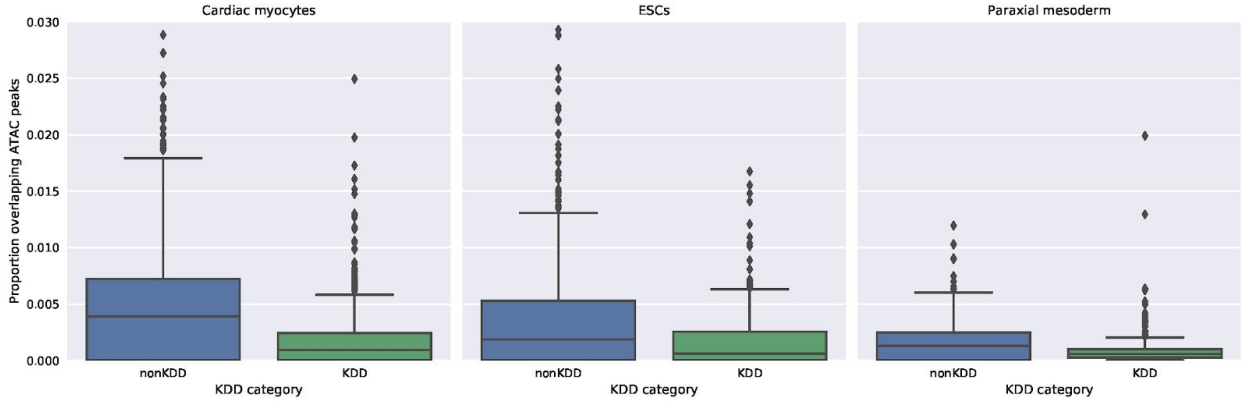
SFig. 13: Gene expression in KDDs and nonKDDs across all cell types assayed in this study with matched RNA-seq data available.

SFig 14: GC content in KDDs in all cell types



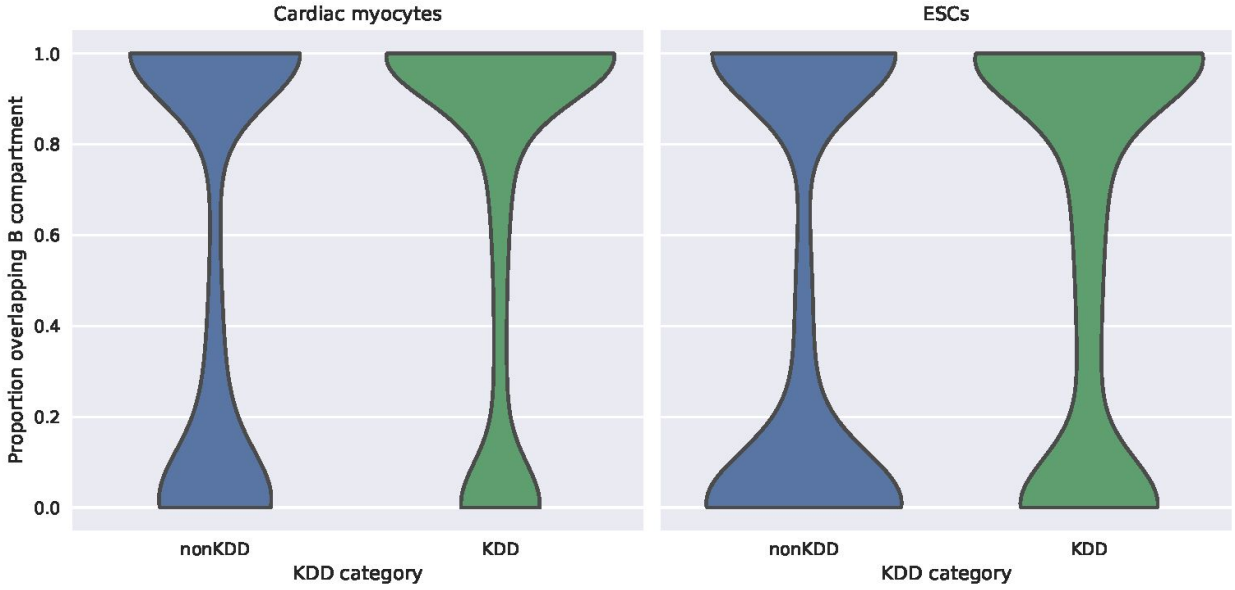
SFig. 14: GC content in KDDs and nonKDDs across all cell types assayed in this study.

SFig 15: ATAC-seq peak overlap in KDDs in all cell types



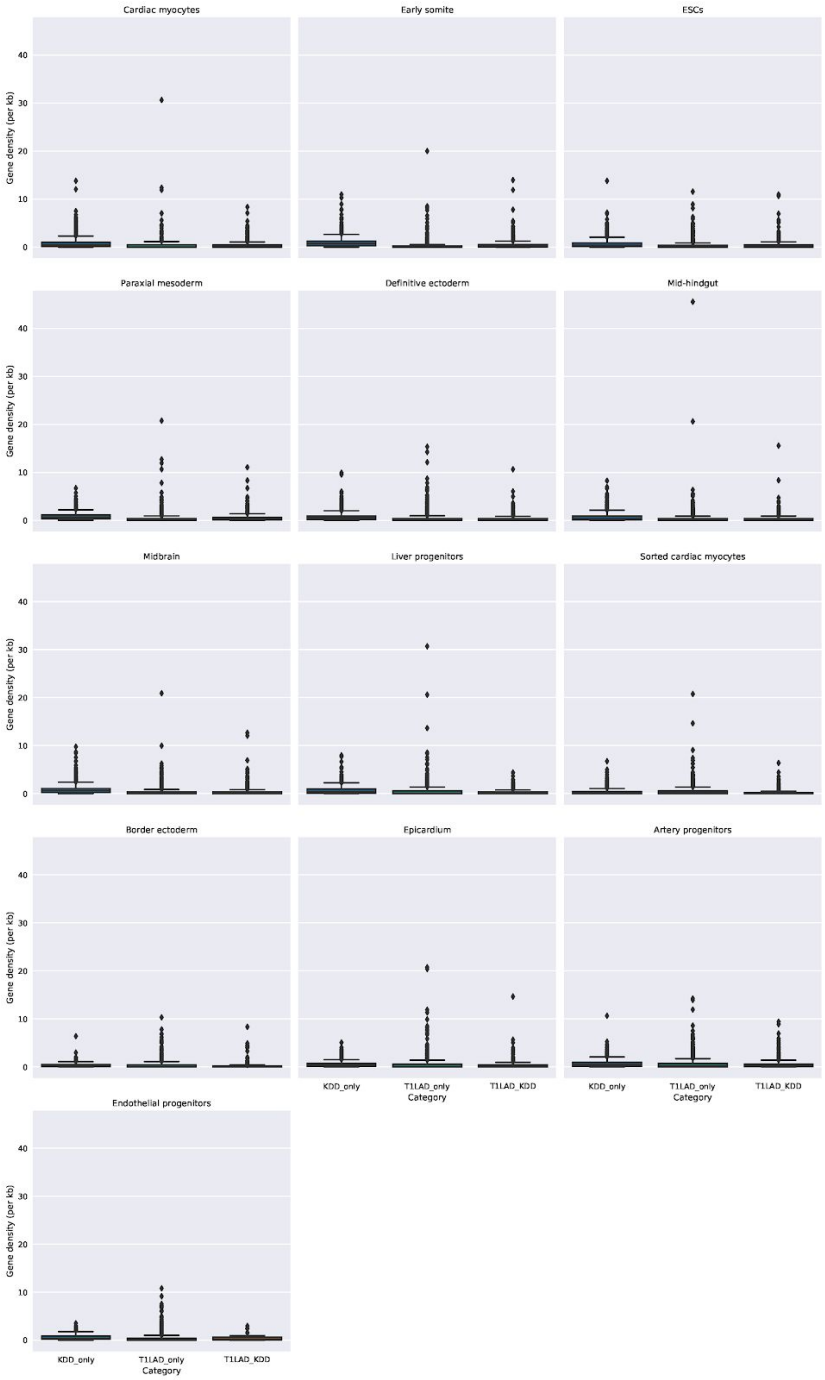
SFig. 15: ATAC-seq peak overlap from previous work (Zhang et al., 2019a) in KDDs and nonKDDs across matched cell types from this study.

SFig 16: B compartment overlap in KDDs in all cell types



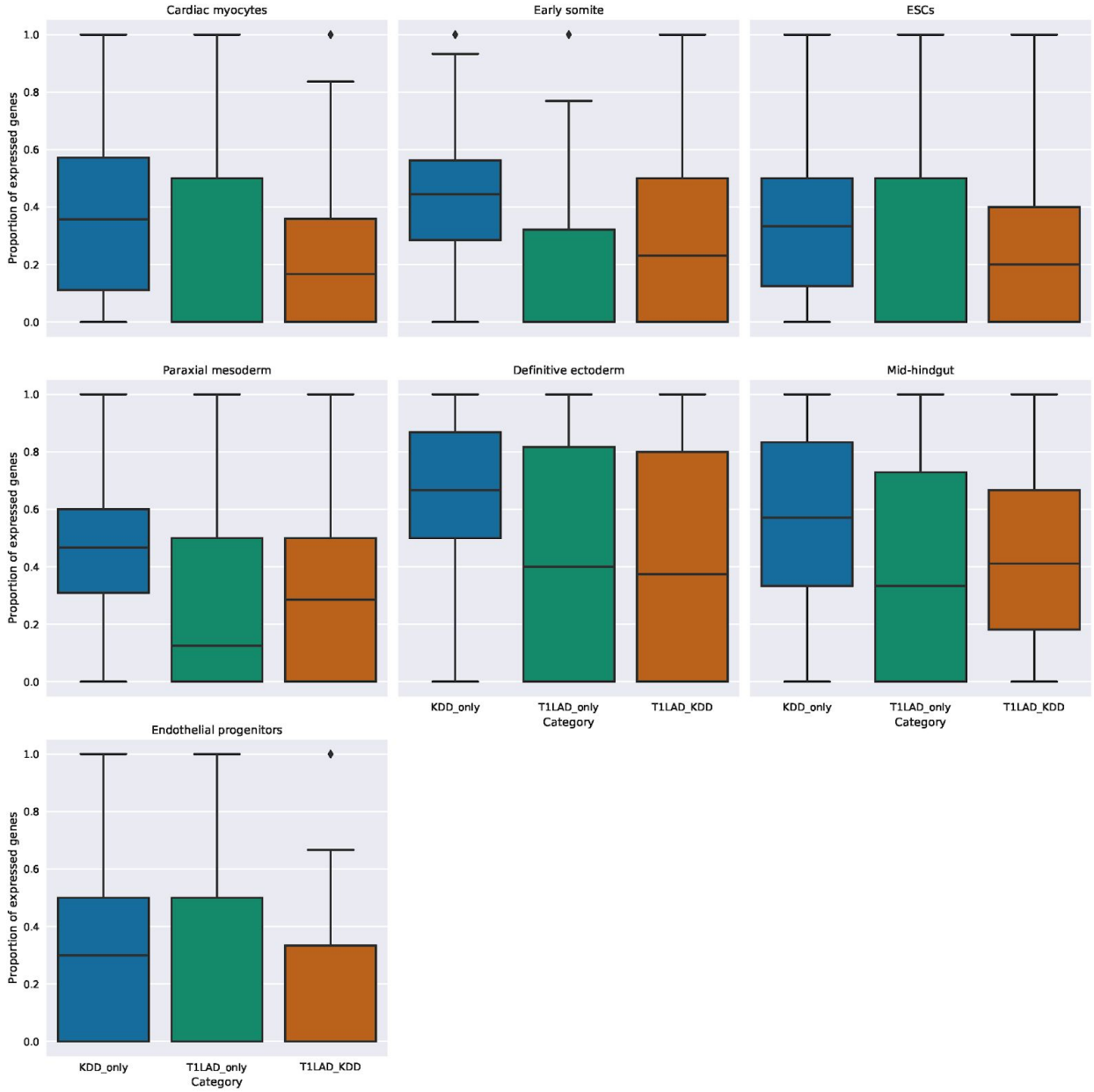
SFig. 16: B compartment overlap from Hi-C data generated in previous work (Zhang et al., 2019a) in KDDs and nonKDDs across matched cell types from this study.

SFig 17: Gene density in in T1-LADs compared to KDDs in all cell types



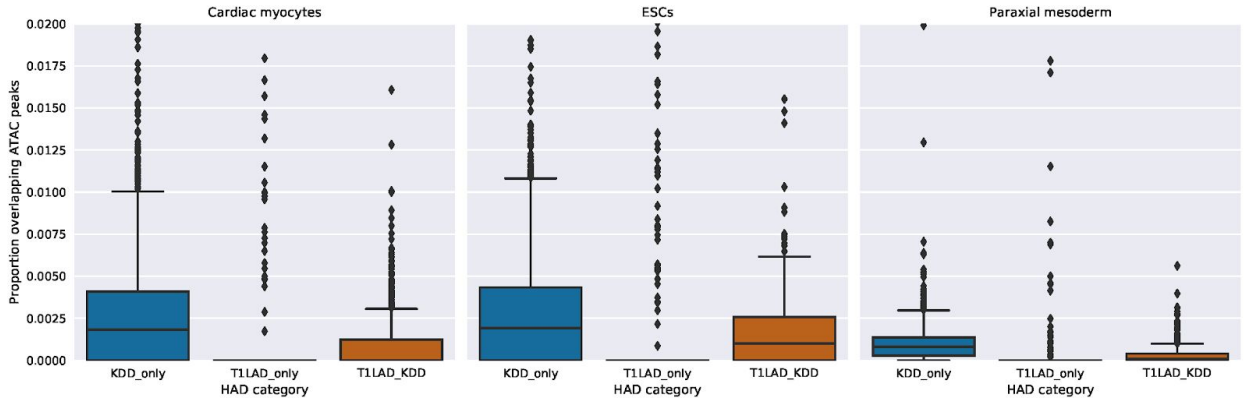
SFig. 17: Gene density in T1-LADs compared to KDDs across cell types.

SFig 18: Proportion genes expressed in T1-LADs compared to KDDs in all cell types



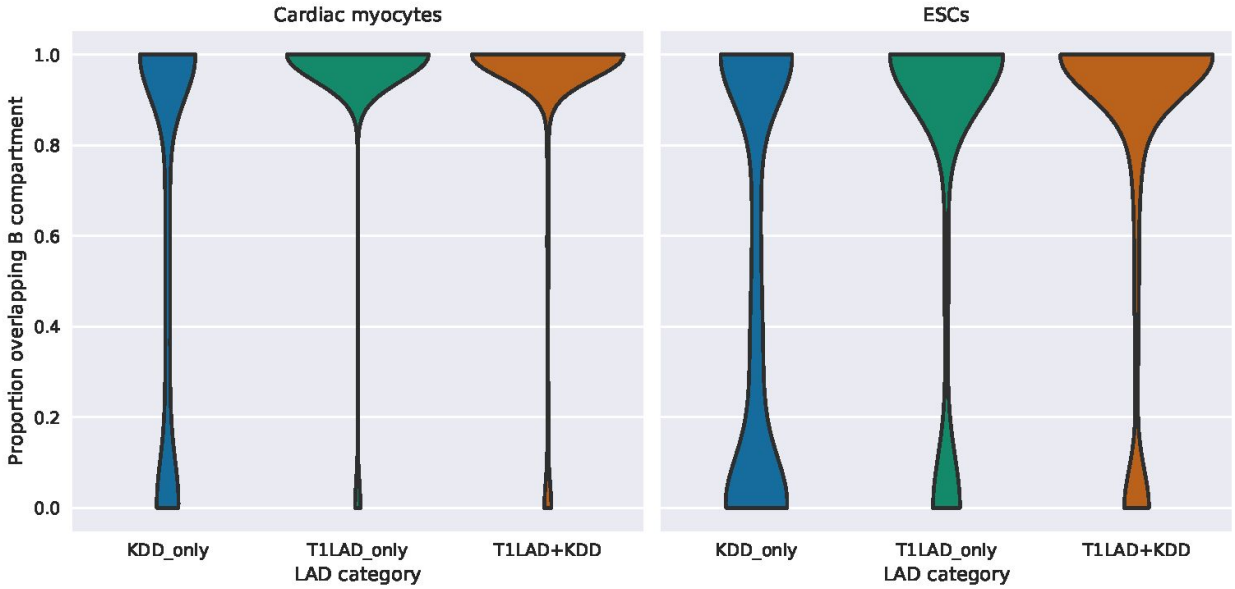
SFig. 18: Proportion genes expressed in T1-LADs compared to KDDs across cell types.

SFig 19: ATAC-seq peak overlap in T1-LADs compared to KDDs in matched cell types



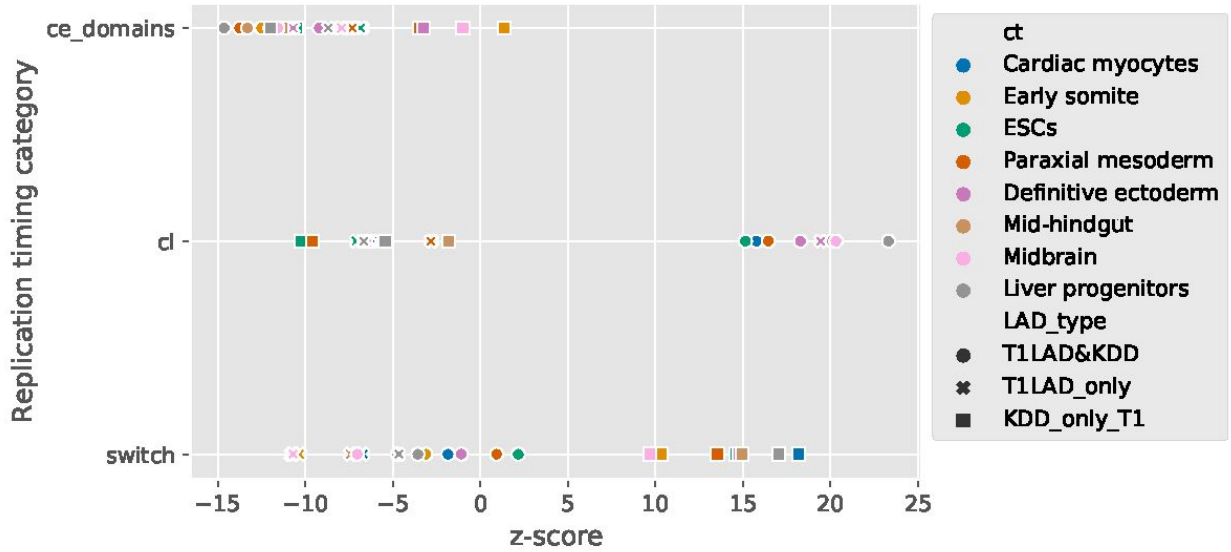
SFig. 19: ATAC-seq peak overlap from previous work (Zhang et al., 2019a) in T1-LADs compared to KDDs across cell types.

SFig 20: B compartment overlap in T1-LADs compared to KDDs in matched cell types



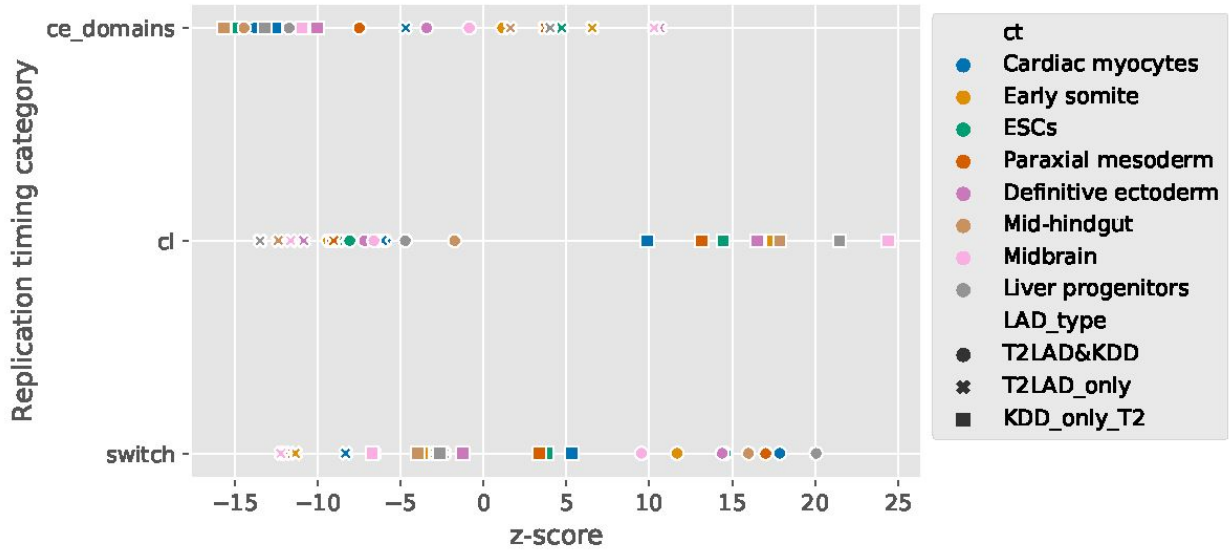
SFig. 20: B compartment overlap from Hi-C data generated in previous work (Zhang et al., 2019a) in T1-LADs compared to KDDs across cell types.

SFig 21: RT domain enrichment in T1-LADs compared to KDDs in all cell types



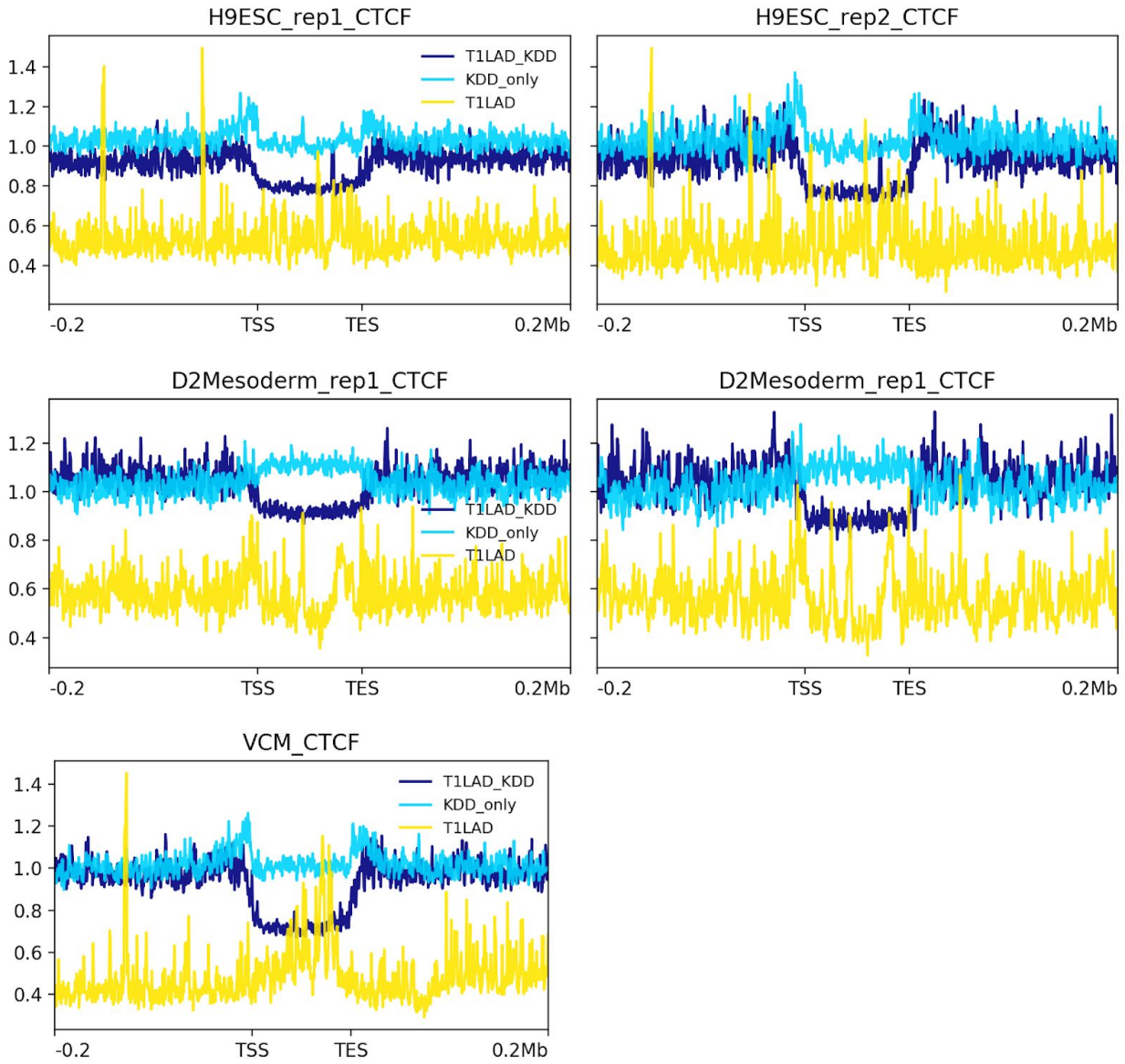
SFig. 21: Enrichment of RT domains (100 permutations) in T1-LADs compared to KDDs across cell types.

SFig 22: RT domain enrichment in T2-LADs compared to KDDs in all cell types



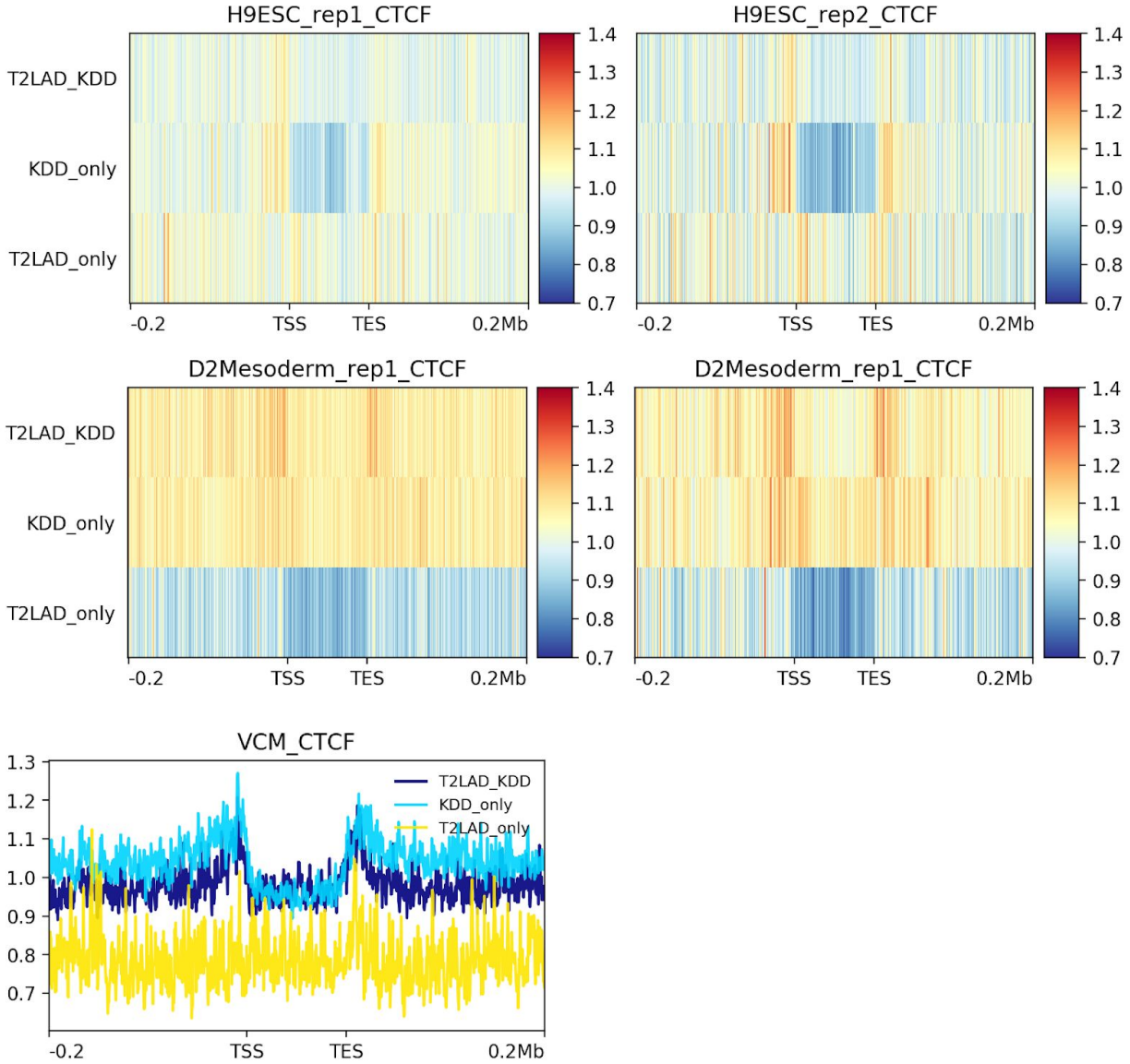
SFig. 22: Enrichment of RT domains (100 permutations) in T2-LADs compared to KDDs across cell types.

SFig 23: CTCF occupancy in T1-LADs compared to KDDs in all cell types



SFig. 23: CTCF occupancy with CTCF ChIP-seq data from previous work (Zhang et al., 2019b) in T1-LADs compared to KDDs across matched cell types. Cell type names are from the CTCF data, where H9ESC was matched with ESC data from our study, D2Mesoderm with Paraxial Mesoderm, and VCM (Ventricular Cardiac Myocytes) with CM data.

SFig 24: CTCF occupancy in T2-LADs compared to KDDs in all cell types



SFig. 24: CTCF occupancy with CTCF ChIP-seq data from previous work (Zhang et al., 2019b) in T2-LADs compared to KDDs across matched cell types.