Topologically associating domain (TAD) boundaries stable across diverse cell types are evolutionarily constrained and enriched for heritability

Authors:
Evonne McArthur[1] (evonne.mcarthur@vanderbilt.edu)
John A. Capra[1,2]* (tony.capra@vanderbilt.edu)

Affiliations:
[1]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, 37235, USA
[2]Department of Biological Sciences, Vanderbilt University, Nashville, TN, 37235, USA
*Corresponding author

ABSTRACT

Background: Topologically associating domains (TADs) are fundamental units of three-dimensional (3D) nuclear organization. The regions bordering TADs—TAD boundaries—regulate gene expression by restricting interactions of cis-regulatory sequences to their target genes. TAD and TAD boundary disruption have been implicated in rare disease pathogenesis; however, we have a limited framework for integrating cross-cell-type TAD maps into the study of genome evolution and interpretation of common trait-associated variants. Here, we investigate an unexplored attribute of 3D genome architecture—the stability of TAD boundaries across cell types—and demonstrate its relevance to understanding how genetic variation in TADs contributes to complex disease.

Results: By synthesizing TAD maps across 37 diverse cell types with multiple annotations and 41 genome-wide association studies (GWAS), we investigate the differences in function and evolutionary constraint between TADs versus TAD boundaries. We demonstrate that TAD boundaries are more evolutionarily constrained than TADs. We find that boundaries are enriched for disease heritability, especially for immunologic, hematologic, and metabolic traits. Next, we stratify boundaries by their stability across cell types and show that more than 30% of boundaries are unique to a specific tissue. Compared to these unique boundaries, cell-type-stable boundaries are enriched for CTCF binding, evolutionary constraint, housekeeping genes, and trait heritability, suggesting greater functional importance.

Conclusions: Cell-type-stable TAD boundaries have greater sequence-level constraint and complex trait heritability enrichment than cell-type-unique boundaries and TADs overall. Thus, considering TAD and boundary stability across cell types provides valuable context for understanding the genome's functional landscape and enabling 3D-structural aware variant interpretation.

Keywords: Topologically associating domain, Hi-C, heritability, evolutionary conservation

BACKGROUND

The three-dimensional (3D) conformation of the genome facilitates the regulation of gene expression.[1–4] Using chromosome conformation capture technologies (3C, 4C, 5C, Hi-C),[5–7] recent studies have demonstrated that modulation of gene expression via 3D chromatin structure is important for many physiologic and pathologic cellular functions, including cell type identity, cellular differentiation, and risk for multiple rare diseases and cancer.[8–13] Nonetheless, many fundamental questions remain about the functions of and evolutionary constraints on 3D genome architecture. For example, how does genetic variation in different 3D contexts contribute to the risk of common complex disease? Furthermore, disease-causing regulatory variation is well known to be tissue-specific; however, there has been little characterization of 3D structure variation across multiple cell types.[13–15] Understanding how different attributes of 3D genome architecture influence disease risk in a cell-type-specific manner is crucial for interpreting human variation and, ultimately, moving from disease associations to an understanding of disease mechanisms.[16]

3D genome organization can be characterized at different scales. Globally, chromosomes exist in discrete territories in the cell nucleus.[7,17,18] On a sub-chromosomal scale, chromatin physically compartmentalizes into topologically associating domains (TADs). TADs are megabase-long genomic regions that self-interact, but rarely contact regions outside the domain (Fig. 1A).[7,19–21] They are likely formed and maintained through interactions between CTCF zinc-finger transcription factors and cohesin ring-shaped complexes, among other proteins both known and unknown.[7,22] TADs are identified as regions of enriched contact density in Hi-C maps (Fig. 1A). TADs modulate gene regulation by limiting interactions of cis-regulatory sequences to target genes.[7] The extent to which chromatin 3D topology affects gene expression is still debated. In extensively rearranged *Drosophila* balancer chromosomes, few genes had expression changes.[23] In contrast, subtle chromatin interaction changes in induced pluripotent stem cells (iPSCs) from seven related individuals were associated with proportionally large differential gene expression.[24] Thus, further cell-type-specific investigation into properties of TAD organization and disruption is needed to clarify which parts of the genome are sensitive to changes in 3D structure.

TAD organization can be divided into two basic features: the TAD and the TAD boundary. TADs are the self-associating, loop-like domains which contain interacting cis-regulatory elements and target genes. TAD boundaries—regions in between TADs—are insulatory elements that restrict interactions of cis-regulatory sequences, like enhancers, to target genes.[7] Previous work suggests the functional importance of maintaining both the “self-associating” TADs and the “insulatory” boundaries. For example, in cross-species multiple sequence alignments, syntenic break enrichment at TAD boundaries suggests a long-term evolutionary preference for rearrangements that “shuffle” intact TADs, rather than “break” them.[25,26] Additionally, TADs often contain clusters of co-regulated genes—*e.g.*, cytochrome genes and olfactory receptors.[7,20,27] Intra-TAD structural variation that deletes or duplicates enhancers has been implicated in polydactyly, B cell lymphoma, and aniridia.[28] Together, these data suggest that the genome is under pressure to preserve TADs as functional units.

Other evidence suggests the greater importance of maintaining TAD boundaries. TAD boundaries are enriched for housekeeping genes and transcription start sites.[7,19] Removing insulatory TAD boundaries leads to ectopic gene expression in cultured cells and *in vivo*. For example, TAD structure disruption at the *EPHA4* locus leads to inappropriate rewiring of developmental genes implicated in limb formation defects.[7,28,29] In cancer, large structural alterations that disrupt TAD boundaries cause pathogenic gene expression in acute myeloid leukemia (AML) and medulloblastoma.[30,31] Structural variation (SV) that disrupts TAD boundaries causes gain-of-function, loss-of-function, and misexpression in many forms of rare neurodevelopmental disease.[28] Accordingly, TAD boundaries have evidence of purifying

selection on SV.[32] Finally, human haplotype breakpoints do not align with chromatin boundaries, which indicates that recombination may be deleterious at TAD boundaries.[33] Collectively, these suggest that TAD boundaries are functionally important and constrained, especially on the scale of human evolution.

In addition to the need for further characterization of TADs versus TAD boundaries, there is also a gap in our understanding of the variability in TAD organization across cell types. TADs and TAD boundaries have been characterized as largely invariant across cell types[19,20,34–36] and species[7,19,26,37,38]. However, previous pair-wise comparisons of five 3D maps suggest that 30-50% of TADs differ across cell types.[35,39] Boundaries shared across two cell types have evidence of stronger SV purifying selection than "unique" boundaries, suggesting that "stable" boundaries are more intolerant of disruption.[32] Although this provides some characterization of "stable" versus "unique" TAD boundaries across cell types, most previous attempts to categorize TAD boundaries have focused on insulatory strength or hierarchical level in a single cell type. For example, stratifying boundaries by their strength (in a single cell type) facilitated discovery that greater CTCF binding confers stronger insulation and that super-enhancers are preferentially insulated by the strongest boundaries.[40] Stratifying by hierarchical properties of TADs—TADs often have subTADs—demonstrated that boundaries flanking higher level structures are enriched for CTCF, active epigenetic states, and higher gene expression.[41]

Despite these preliminary indications that the stability of components of the 3D architecture may influence functional constraint, there has been no comprehensive analysis comparing genomic functionality and disease-associations between 3D structural elements stable across multiple cell types versus those that are unique to single cell types. Quantifying stability across cell types is important to for interpreting new variation within the context of the 3D genome given our knowledge that disease-associated regulatory variation is often tissue-specific.[13–15]

To investigate differences in TAD boundaries across cell types, we quantify boundary "stability" as the number of tissues that share a TAD boundary. If a TAD boundary is found in many tissues, it is "stable"; whereas, if it is found in few tissues, it is "unique" (Fig. 1B). Using this characterization, we address two main questions that aim to expand our framework for cell-type-aware interpretation of genetic variation and disease associations in the context of the 3D genome (Fig. 2B):

1. How do **TADs** versus **TAD boundaries** differ in terms of their functionality and evolutionary constraint?
2. Are there functional and evolutionary differences in TAD boundaries **stable** across multiple cell types versus TAD boundaries **unique** to specific tissues?

Synthesizing 3D genome maps across 37 diverse cell types with multiple functional annotations and genome-wide association studies (GWAS), we show that TAD boundaries are more evolutionary constrained and enriched for heritability of common complex traits compared to TADs. Furthermore, TAD boundaries stable across multiple cell types are more evolutionarily constrained and enriched for functional elements than unique boundaries. Genetic variation in stable TAD boundaries also accounts for more of the heritability of immunologic, hematologic, and metabolic traits than variation in TADs. Together, our work suggests that TAD boundary stability across cell types provides valuable context for understanding the genome's functional landscape and enabling 3D-structural aware variant interpretation.

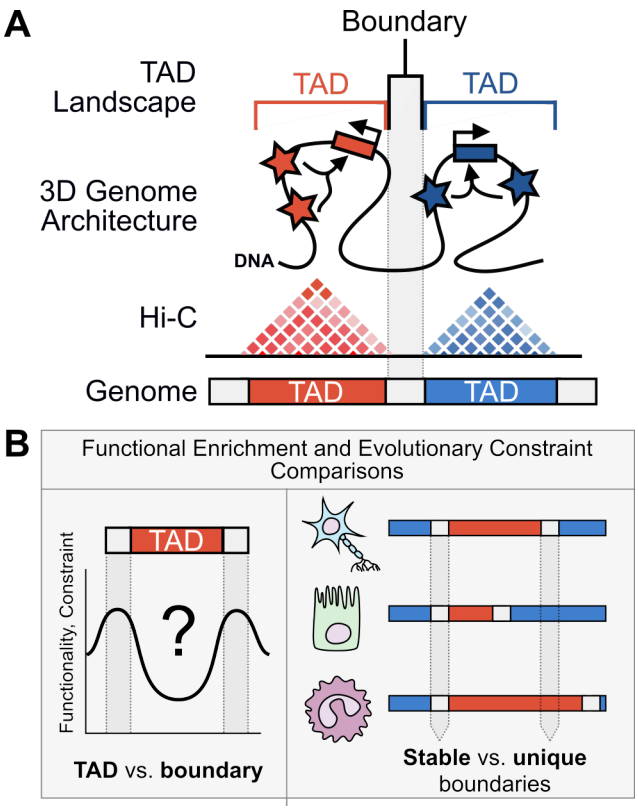


Figure 1. Schematic of our exploration into 3D chromatin TAD boundary stability and functionality. (A) Chromatin is organized in 3D space into TADs which are experimentally determined by Hi-C experiments. Regions bordering TADs are TAD boundaries. Regions within a TAD are much more likely to interact with one another than regions outside of the TAD. Boxes with right-angled arrows represent genes and stars represent gene regulatory elements, like enhancers. **(B)** This work addresses two main questions: (1) Are TADs or TAD boundaries more enriched for functional annotations, evolutionary conservation, and complex trait heritability? (2) Are stable TAD boundaries (*i.e.*, those observed across multiple tissues) more or less functionally enriched than TAD boundaries unique to specific tissues?

RESULTS

TAD maps across diverse cellular contexts vary substantially

To characterize the stability of TAD structure across diverse cellular contexts, we examined 37 TAD maps from the 3D Genome Browser (Supplemental Table 1).[42] The cellular contexts include primary tissues, stem cells, and cancer cell lines; for simplicity, we will refer these as "cell types".[34–36,43–46] All TAD maps were systematically predicted from Hi-C data using the hidden Markov model (HMM) pipeline from Dixon *et al*. (2012) (Fig. 2A).[19,42] The resulting maps are defined with respect to the same 40 kb windows across the genome for each cell type.

The median TAD length across all cell types is 1.15 Mb (IQR: 0.71 - 1.82 Mb) and the median number of TADs per cell type is 1844 (IQR: 1625 - 2277). We observed an inverse relationship between TAD length and number of TADs in a cell type: cells with longer TADs have fewer TADs (Supplemental Fig. 1). Primary tissues have longer TADs, whereas naïve cell types like stem cells and de-differentiated leukemia cell-lines have shorter TADs (Supplemental Fig. 1). This is consistent with previous examination of neuronal development which found that, during differentiation, TAD number decreases with a corresponding increase in size.[12]

To compare TAD boundaries across cell types, we defined boundaries as the 40 kb window surrounding (± 20 kb) TAD start and stop sites. This size is consistent with TAD boundaries defined by other studies and the 25-40 kb Hi-C resolution in these maps.[36] Since the maps for each cell type are defined using the same genomic windows, we identify overlapping boundaries based on these 40 kb windows. To demonstrate the robustness of our results, we also consider a more permissive definition of overlapping TAD boundaries by collapsing all adjacent 40 kb TAD boundaries across cell types up to a 200 kb resolution (Supplemental Fig. 2A). We will refer to this permissive definition as the "200 kb collapsed" boundaries and confirm that all results hold using this definition.

Using the cross-cell-type TAD boundary intersection, we find that boundaries vary substantially across cell types. Only 4.9% of TAD boundaries are shared in 20+ of the 37 cell types and 33.8% of TAD boundaries are unique to a single cell type (Fig. 2B). Even with "200 kb collapsed" permissive boundary overlap definition, 32.6% of boundaries are unique to a single cell type (Supplemental Fig. 2C). To quantify boundary stability for further analyses, we defined genome-wide boundary "stability maps" based on stability percentiles across TAD boundaries; for example, boundaries present in 25 of 37 contexts are at the 98th percentile of stability when compared to other TAD boundaries (CDF in Fig. 2B).

Although there is high variability in TAD boundary landscape across different cell types, we found that biologically similar cell types have more similar TAD boundary maps. For example, cell type classes (*e.g.* organ/tissue, stem cell, and cancer) generally cluster together. The two melanoma cell lines cluster together, as do left ventricle, right ventricle, and aorta. This trend held at both the 40 kb boundary resolution (Fig. 2C, Supplemental Fig. 3A) and with the 200 kb collapsed boundaries (Supplemental Fig. 3B).

Previous studies have found contrasting results about the level and patterns of similarity across cell types. Our finding of TAD map similarity among functionally similar cell types contrasts with previous work by Sauerwald *et al.* (2018) which found that most similar TAD map pairs have no biological connection; however, they investigate a different set of cells (predominantly cancer cell lines) and TAD maps generated at 100 kb resolution.[13] Similar to the Sauerwald comparisons, our results may be confounded by batch effects because the Hi-C are generated by different groups. Additionally, comparisons with highly mutated cancer cell lines that may not reflect natural boundary patterns. However, our similarity quantifications agree with previous estimates. Rao et al. (2014) reported Jaccard indices from 0.21 - 0.30 for comparisons of GM12878 to each of IMR90, HMEC, HUVEC, K562, KBM7 and NHEK.[13,35] The Jaccard similarity for our comparisons of these cell types is 0.24 - 0.37. The median Jaccard similarity for all 37 x 37 cell type comparisons is 0.18 (IQR: 0.15 - 0.23). Analyses that examined cell type TAD map similarity among larger windows have reported similarity coefficients between 0.4 - 0.5.[13] This is in line with our pairwise comparisons between the 200 kb collapsed boundaries (Median: 0.39, IQR: 0.34 - 0.44).

In summary, although TADs and TAD boundaries have been characterized as largely invariant across cell types, we demonstrate that there is substantial 3D variability between cell types.[19,20,34–36] We also find that biologically related cell types have more similar TAD maps, providing preliminary evidence for the cell type specificity of the 3D genome and providing further rationale for investigating TAD map cell type differences.

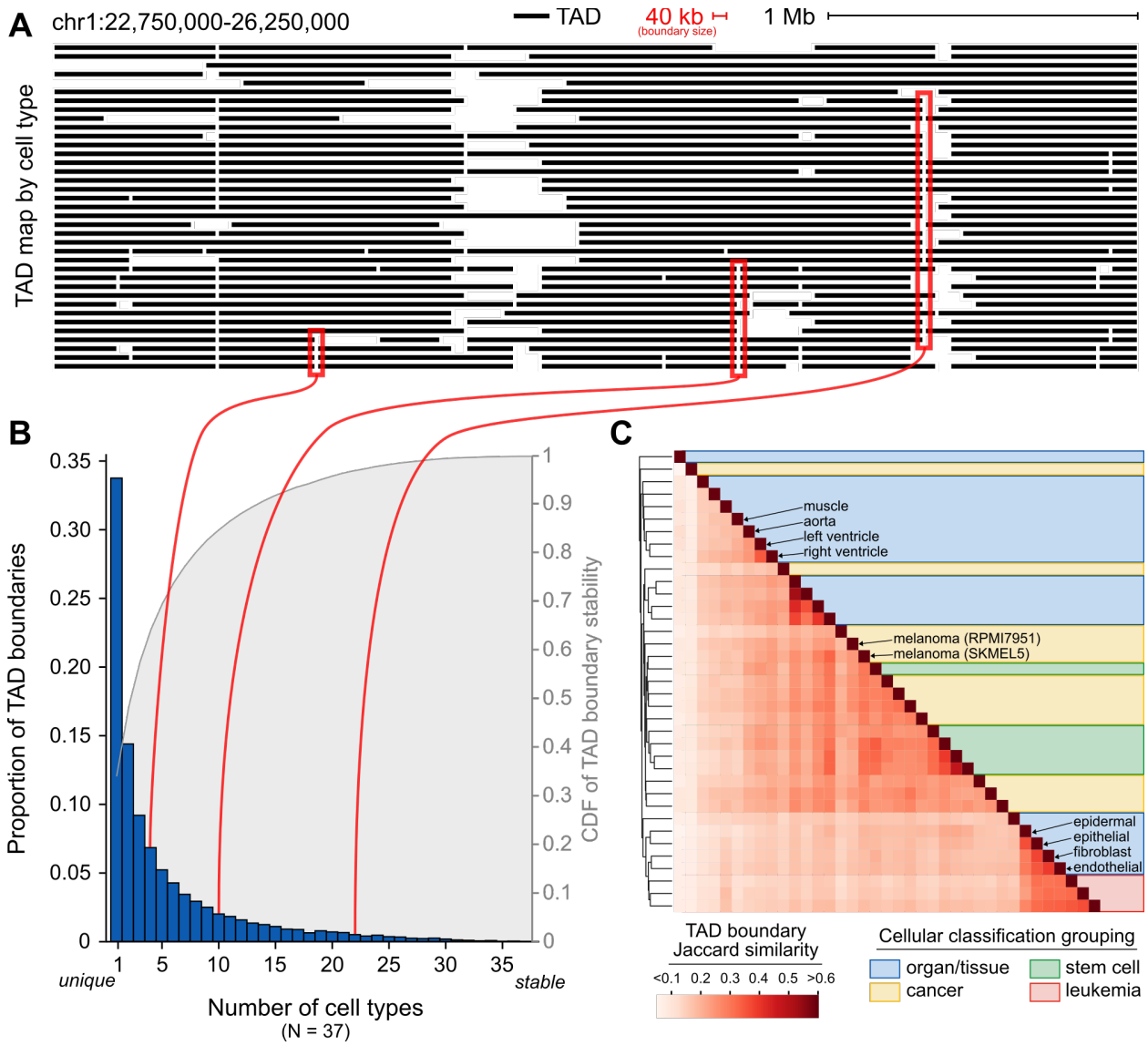


Figure 2. TAD boundaries vary across cell types. 37 cell type TAD maps (rows) for a 3.5 Mb window from human chr1. Each black line represents the extent of a TAD. Example boundaries of different stabilities are outlined in red. **(A)** Histogram of TAD boundaries by the number of cell types they are observed in (their "stability"). Most boundaries are present in only a few cell types. The right axis and gray distribution represent the empirical CDF of boundary stability shown in the histogram. **(B)** Clustering of TAD boundary maps for 37 cell types using the pairwise Jaccard similarity metric (lower triangle). The TAD maps cluster by similar cell types (upper triangle). Supplemental Fig. 3A shows a larger version with all cell type labels.

Stable TAD boundaries are enriched for CTCF binding, evolutionary constraint, and housekeeping genes

To characterize the genomic elements and evolutionary conservation across the 3D TAD landscape, we compared CTCF binding, PhastCons conserved elements, and gene annotations between the 40 kb TAD boundaries and matched 40 kb windows inside TADs. We then repeated the analysis with boundaries stratified by their stability across cell types (defined using stability percentiles depicted in Fig. 2B).

As expected, CTCF binding is enriched at TAD boundaries compared to TADs (2.3 vs. 1.4 per 40 kb, Welch's *t*-test $p \sim 0$, Fig. 3A) and the genomic background (2.3 vs. 1.8 per 40 kb,

$p \sim 0$). CTCF binding was determined through CTCF ChIP-seq analyses obtained from ENCODE.[45,46] Boundary insulatory strength is positively associated with CTCF binding; however, it is unknown whether CTCF binding plays a role in boundary stability across cell types.[19,40] We find that TAD boundary stability and CTCF binding are positively correlated; for example, boundaries stable across 10+ cell types have 1.5x more CTCF sites on average compared to TAD boundaries unique to one cell type (2.7 vs 1.8, $p \sim 0$, Fig. 3A).

TAD boundaries are more evolutionarily conserved than matched windows in TADs; on average, TAD boundaries overlap PhastCons elements at 216 more base pairs (bp) per 40 kb window (1920 vs. 2136 bp per 40 kb, $p \sim 0$, Fig. 3B). This aligns with the previous findings underscoring the importance of maintaining TAD boundaries: TAD boundaries are undergoing purifying selection against structural variation (SV) and are depleted of recombination events.[32,33] Moreover, we find that TAD boundaries stable across multiple cell types are even further constrained. TAD boundaries stable across 10+ cell types have an average of 202 more conserved bases pairs than TAD boundaries unique to a cell type (2235 vs. 2033 bp, $p = 1\text{x}10^{-9}$, Fig. 3B).

TAD boundaries are enriched for genes and housekeeping genes (1.1 vs. 0.74 genes per 40 kb, $p \sim 0$, Fig. 3C; 0.17 vs 0.096 housekeeping genes per 40 kb, $p \sim 0$, Supplemental Fig. 4C); this supports previous findings that transcription start sites (TSS) and housekeeping genes are enriched at TAD boundaries.[7,19] Stable TAD boundaries are 1.5 times enriched for genes compared to cell-type-unique TAD boundaries (1.3 vs 0.88 genes per 40 kb, $p \sim 0$, Fig 3C). Although stable TAD boundaries are 1.5x enriched for genes, stable TAD boundaries have further relative enrichment for overlap with housekeeping genes; housekeeping genes are 2.6x enriched in stable TAD boundaries (Fig. 3D). The relationship between stable TAD boundaries and housekeeping gene enrichment may result from many factors, including strong enhancer-promoter interactions, specific transcription factor binding, or because highly active sites of transcription may cause chromatin insulation.[12]

In summary, TAD boundaries are enriched for CTCF binding, evolutionary conservation, genes, and housekeeping genes compared to TADs. TAD boundaries stable across multiple cell types are further enriched for these attributes. These trends hold for other metrics of conservation, CTCF binding, and gene overlap (Supplemental Fig. 4) and the 200 kb collapsed boundaries (Supplemental Fig. 5).

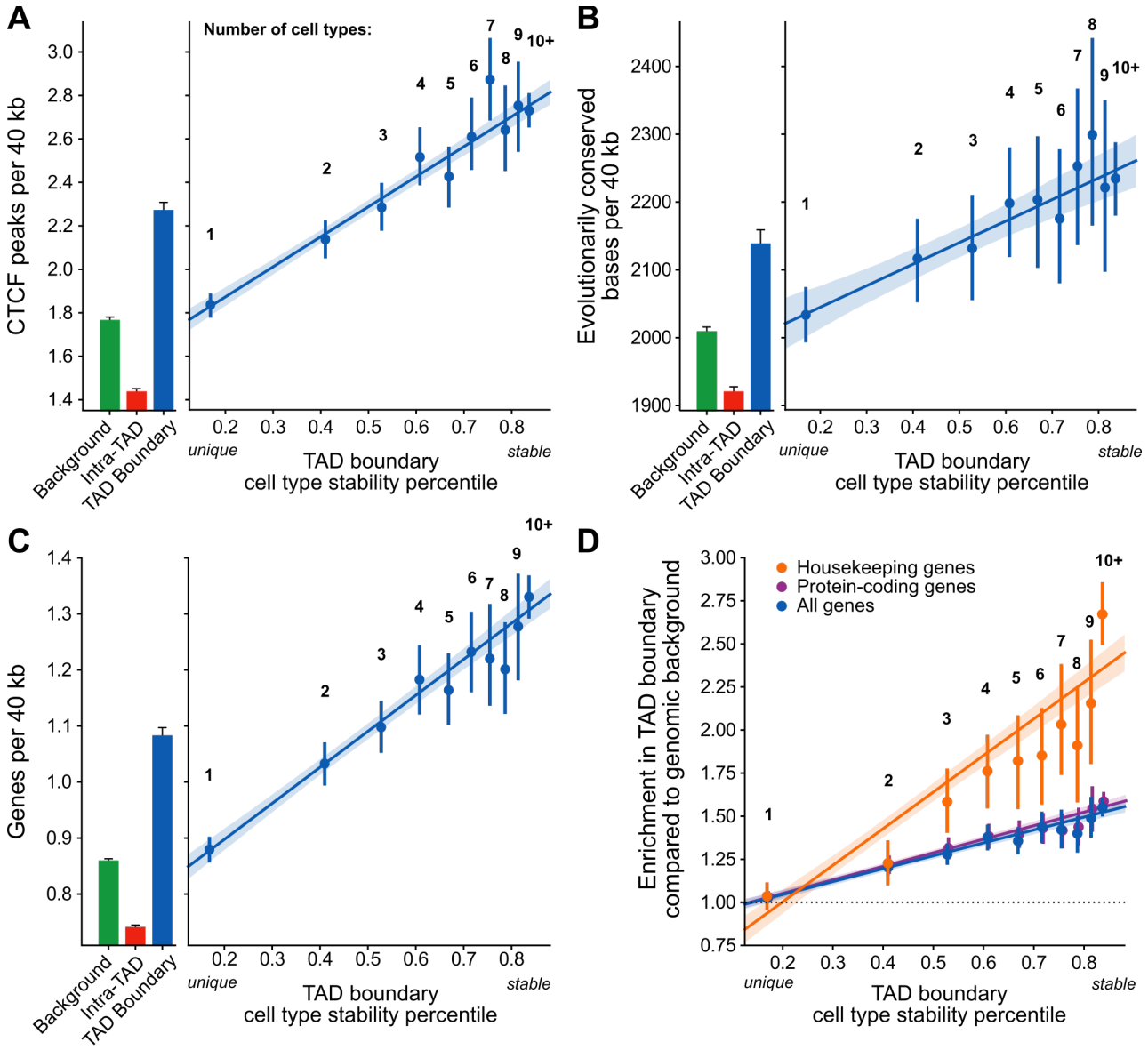


Figure 3. Stable TAD boundaries are enriched for CTCF binding, sequence-level conservation, and housekeeping genes. Genomic annotation overlap is compared across 40 kb windows representing the genomic background (green bar), TADs (red bar), and TAD boundaries (blue bar). TAD boundaries are also stratified by their stability across cell types (blue points). TAD boundary stability is defined by the number of cell types sharing that TAD boundary (black numbers) and its corresponding percentile of stability (x-axis, generated from the CDF in Fig. 2B). Per 40 kb window, TAD boundaries and, more specifically, stable TAD boundaries have increased: **(A)** CTCF binding (boundaries are 1.6 times enriched for CTCF peaks than TADs [$p \sim 0$] and the most stable boundaries have 0.9 more peaks than the least stable boundaries, [$p \sim 0$]), **(B)** Number of conserved bases (boundaries overlap vertebrate PhastCons elements at an average of 216 more bases than matched 40 kb windows in TADs [$p \sim 0$] and the most stable boundaries have 202 more conserved bases than the least stable boundaries [$p = 1\text{x}10^{-9}$]); and **(C)** Overlap with all RefSeq genes (boundaries are 1.5 times enriched for genes than TADs [$p \sim 0$] and the most stable boundaries have 0.45 more genes than the least stable boundaries [$p \sim 0$]). **(D)** Compared to genomic background, stable TAD boundaries have further relative enrichment for overlap with housekeeping genes (orange) than protein-coding genes (navy) or all genes (blue, Fig. 3C shows the absolute number of genes). All error bars/bands signify 95% confidence intervals.

TAD boundaries are enriched for complex trait heritability

Disruption of 3D genome architecture has recently been shown to play a role in rare disease and cancer; however, the relationship between 3D genome landscape and common disease has not been investigated. Small-scale variation (*e.g.* single nucleotide polymorphisms [SNPs]) may disrupt 3D genome structure, for example, by modifying CTCF-binding site motifs necessary for TAD formation.[7,22] However, the contribution of small-scale variation in different 3D contexts to common phenotypes is unknown. We use partitioned heritability analysis to quantify the relationship between different attributes of the 3D genome architecture and the genetic architecture of common phenotypes.

We estimated common trait heritability enrichment among SNPs within annotations of interest (*e.g.* TADs or TAD boundaries from different cell types) using stratified-LD score regression (S-LDSC).[47,48] We considered GWAS summary statistics from a previously-described representative set of 41 diseases and complex traits (average N = 329,378, M = 1,155,239, h^2_{SNP} = 0.19, Supplemental Table 2).[49–59] Motivated by the approach to partitioning TADs from Krefting *et al.*, we analyzed TADs plus 50% of their total length on each side and subdivided this region into 20 equal-sized partitions.[25] In cases where a TAD is in very close proximity to another TAD, the ±50% region flanking the TAD may extend into a neighboring TAD. However, generally, TADs represent the center 10 partitions and we define TAD boundaries as the partitions that flank the TAD. We then conducted S-LDSC for each of the 20 partitions across the 37 cell types for the 41 traits to estimate the enrichment (or depletion) of heritability for that trait across the TAD landscape.

TAD boundaries are enriched for complex trait heritability; whereas TADs are marginally depleted for heritability (1.07x enrichment in boundaries vs. 0.99x enrichment in TADs, $p \sim 0$) (Fig. 4A). However, there is a spike of heritability enrichment centrally in TADs; we explore this further in the next section. These results are consistent whether averaged across traits or meta-analyzed using a random-effects model[48,50,60] (R^2 = 0.85, $p = 7\text{x}10^{-9}$, Supplemental Fig. 6); therefore, further analyses will average across traits for simplicity and interpretability. The heritability enrichment values are highly significant but, as expected, relatively small in magnitude given the large genomic regions considered by this analysis—only a small fraction of the base pairs in a boundary are likely to be functionally relevant. We are most interested the comparison of heritability enrichment patterns between TADs and boundaries.

The complex trait heritability enrichment at TAD boundaries is consistent across cell types (Fig. 4B). We note that heritability enrichment occurs in a window “bookending” the TAD start and stop that is larger than the exact 40 kb TAD start/stop site (Fig. 4A). To capture this larger region flanking the TAD start/stop, we define a second set of TAD boundaries termed “100 kb bookend boundaries” which represent regions 100 kb upstream of the TAD start and 100 kb downstream of the TAD end. Using this 100 kb bookend boundary, we find a modest, but significant positive relationship between cell-type-stability and trait heritability enrichment (R^2 = 0.04, p = 0.009, Fig. 4C). The most stable boundaries (fourth quartile) have 1.08x enrichment of trait heritability compared to 0.97x enrichment in unique boundaries (first quartile). This positive relationship between heritability and boundary stability holds even with the more local 40 kb windows; however, it is significantly weakened (p = 0.8, Supplemental Fig. 10), suggesting that the 100 kb bookend boundaries better capture the trait heritability enrichment at TAD boundaries.

The heritability enrichment at TAD boundaries is likely due at least in part to their overlap with CTCF binding sites, conserved elements, and genes—each a functional category that is enriched for heritability. To assess whether the TAD boundary heritability enrichment is greater than expected given other functional elements overlapping TAD boundaries, we calculate standardized effect sizes (τ^*).[50,61] This metric quantifies heritability unique to the focal annotation by conditioning on a broad set of 86 functional regulatory, evolutionary conservation, coding, and LD-based annotations (baseline v2.1).[48,50,61,62] TAD boundaries did not show more

heritability than expected based on their enrichment for other annotations (Supplemental Fig. 7). Thus, existing annotations likely capture relevant functional elements (*e.g.*, CTCF and other regulatory element binding sites) that determine and maintain boundary function.

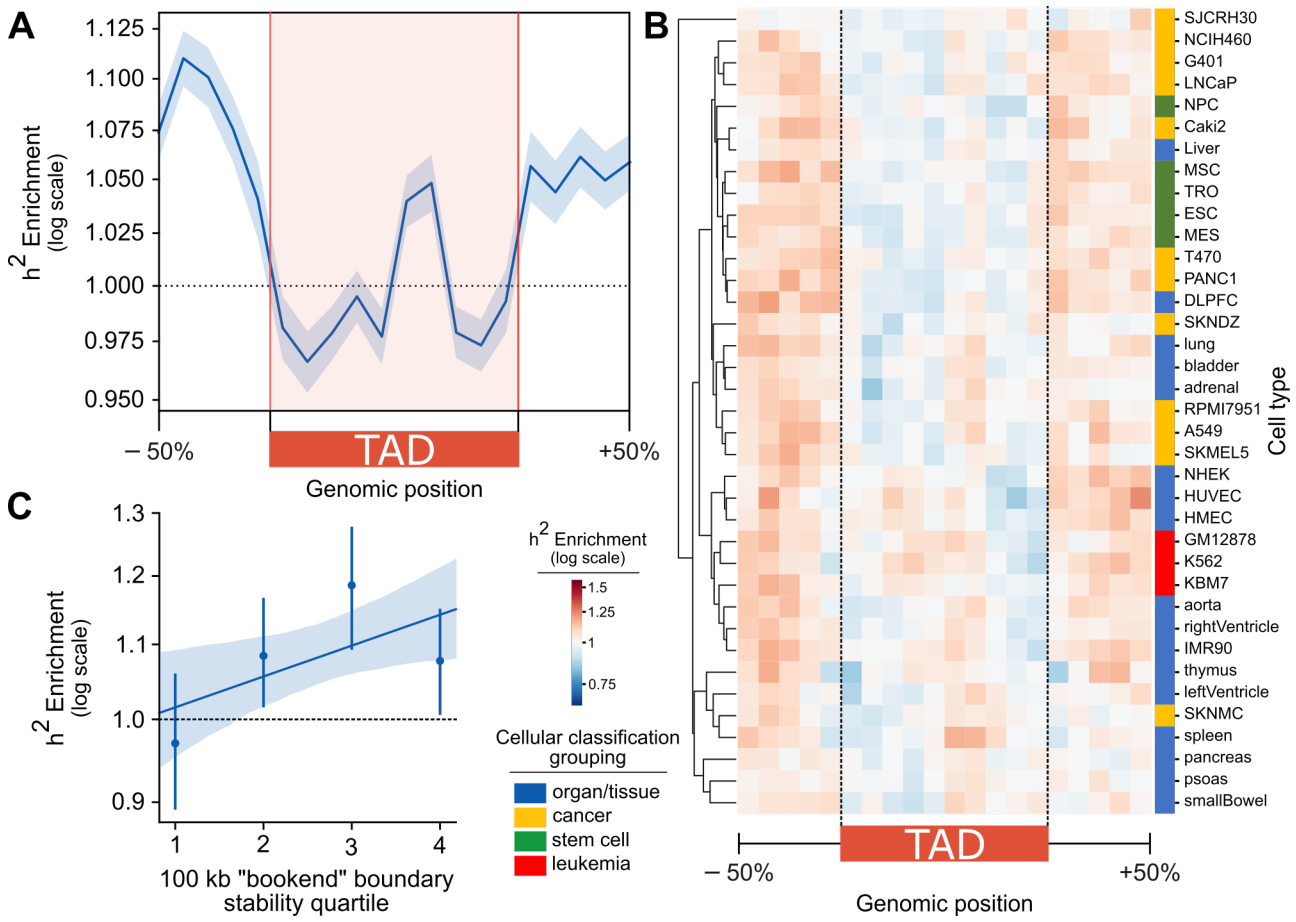


Figure 4. TAD boundaries are enriched for heritability of diverse common complex traits. (A) For TADs across 37 cell types, heritability is enriched near TAD boundaries and in the center of TADs when averaged across 41 common complex phenotypes. Error band signifies 99% confidence interval. **(B)** TAD boundary heritability patterns are consistent across the 3D genome landscape for 37 cell types. **(C)** Boundary stability is positively correlated with trait heritability enrichment (R^2 = 0.04, p = 0.009). Boundaries were defined using the 100 kb "bookend" strategy. Error bars/band signify 95% confidence intervals.

The heritability landscape across the 3D genome varies across phenotypes

Although complex trait heritability is generally enriched at TAD boundaries across multiple cell types and traits, we hypothesized that TAD boundaries may be more important for certain traits than others. Other studies have shown preliminary evidence that different traits have unique enrichment profiles among different functional annotations.[48] To investigate trait-specific heritability across the TAD landscape, we hierarchically clustered heritability enrichment profiles across the 3D genome partitions by trait (Fig. 5A). We observed two distinct trait clusters.

One cluster of traits ("boundary-enriched" cluster) is strongly enriched for complex trait heritability at TAD boundaries (Fig. 5B). Across TAD maps in 37 cell types, these traits have on average 1.15x heritability enrichment at TAD boundaries compared to genomic background and 1.18x enrichment compared to TADs ($p \sim 0$). The other cluster of traits ("boundary-depleted"

cluster) shows a weak inverted pattern compared to the boundary-enriched cluster, with marginal heritability depletion at TAD boundaries (0.96x enrichment, $p \sim 0$) and a spike of heritably enrichment within the TAD center (Fig. 5C).

The traits in the boundary-enriched cluster are predominantly hematologic (*e.g.* white blood cell and red blood cell counts), immunologic (*e.g.* rheumatoid arthritis, Crohn's disease), and metabolic traits (*e.g.* type 2 diabetes, lipid counts) (Fig. 5E). The traits in the boundary-depleted cluster are mostly neuropsychiatric (*e.g.* schizophrenia, years of education, Autism spectrum disorder) and dermatologic (*e.g.* skin color, balding) (Fig. 5E). Our stratification of complex diseases into phenotypic classes does not perfectly reflect the traits' pathophysiology. For example, some dermatologic traits fall into the boundary-enriched cluster. However, these dermatologic traits, like eczema, also have a significant immunologic and hematologic basis, which are hallmarks of other traits in the boundary-enriched cluster. Additionally, body mass index (BMI) clustered with the psychiatric-predominant boundary-depleted cluster instead of with other metabolic traits in the boundary-enriched cluster. This is interesting given previous findings that BMI heritability is enriched in central nervous system (CNS)-specific annotations rather than metabolic tissue (liver, adrenal, pancreas) annotations.[48] Skeletal, cardiopulmonary, and reproductive traits do not consistently segregate into one of the clusters (Fig. 5E). This is likely because of the small sample size and heterogeneity of traits in these phenotypic classes.

The relationship between heritability enrichment in TAD boundaries and the trait clusters is not confounded by GWAS trait sample size (N), number of SNPs (M), or the traits' SNP-based heritability (h^2_{SNP}) (Supplemental Fig. 8). Despite using a diverse set of cell types, we recognize that the heritability pattern differences between traits could be affected by the representation of cell types investigated. However, given that the pattern of heritability enrichment is consistent across all cell types (Fig. 4B), we are confident that no single cluster of cell types is driving the heritability pattern differences between traits. Furthermore, these patterns are maintained even when calling TADs using a variety of computational methods (Armatus, Arrowhead, DomainCaller, HiCseg, TADbit, TADtree, TopDom), suggesting that the finding of immunologic and hematologic heritability enrichment at TAD boundaries is robust to technical variation (Suppplemental Fig. 9).

Although analysis over all traits revealed a positive relationship between boundary cell-type-stability heritability enrichment (Fig. 4C), we found that this trend differs between the two trait clusters. Traits in the boundary-enriched cluster have further heritability enrichment in cell-type-stable boundaries ($R^2 = 0.26$, $p = 4\text{x}10^{-7}$, Fig. 5D). The most stable boundaries (fourth quartile) have 1.24x enrichment of trait heritability compared to 0.91x enrichment in unique boundaries (first quartile). In contrast, traits in the boundary-depleted cluster have further heritability depletion in stable boundaries ($R^2 = 0.042$, $p = 0.07$, Fig. 5F).

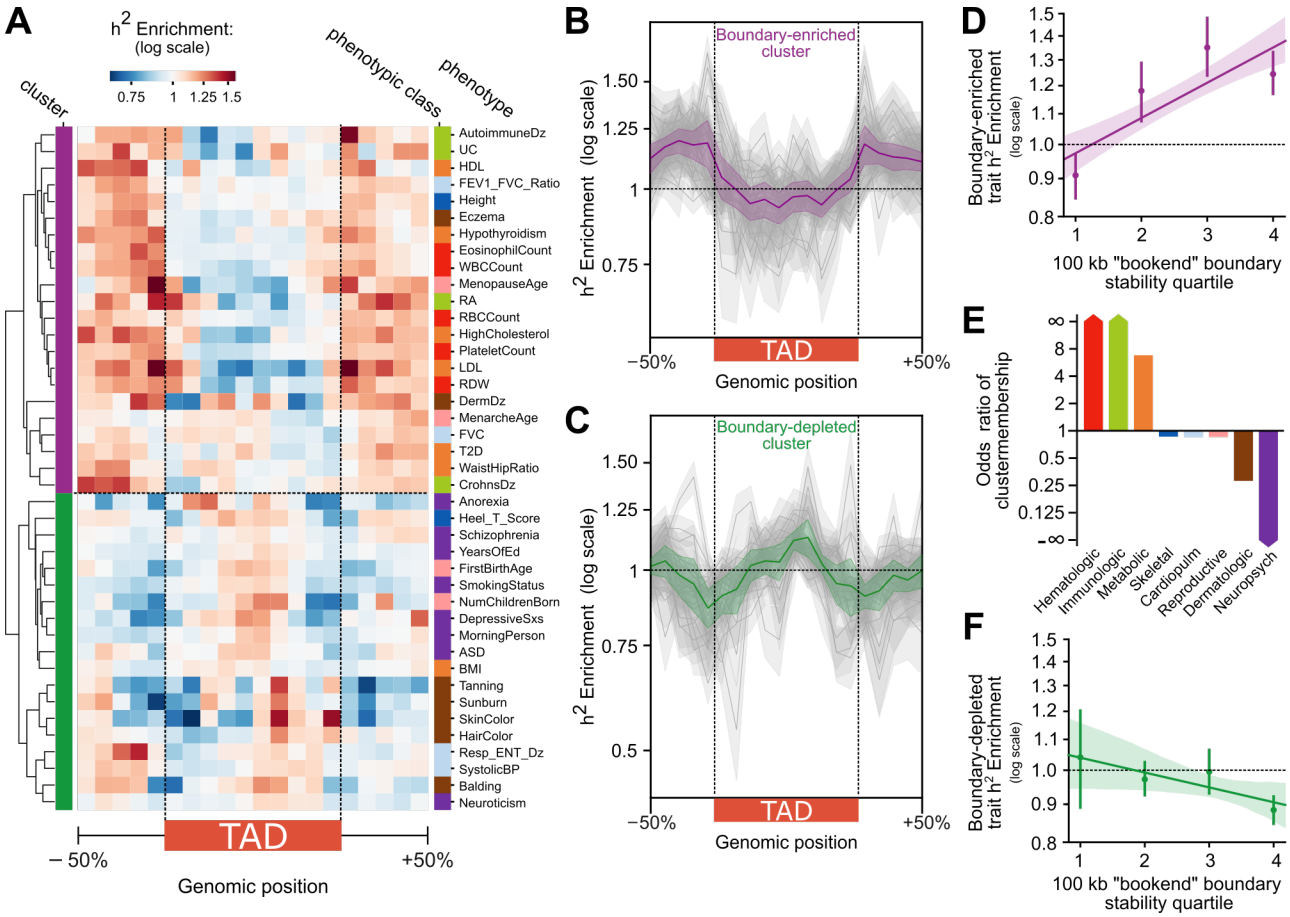


Figure 5. The heritability landscape across the 3D genome varies across phenotypes. TAD boundary trait heritability patterns across the 3D genome organize into two clusters. Some traits are strongly enriched for complex trait heritability at TAD boundaries ("boundary-enriched" cluster, purple), while others are weakly depleted at TAD boundaries and enriched centrally within the TAD ("boundary-depleted" cluster, green). **(A)** Heritability enrichment landscape over TADs for traits in the boundary-enriched cluster (N = 22). The gray lines represent the heritability pattern for each trait in the cluster; the purple line is the average over all the traits. **(B)** Heritability enrichment landscape over TADs for traits in the boundary-depleted cluster (N = 19). The gray lines represent the heritability pattern for each trait in the cluster; the green line is the average over all the traits. **(C)** The positive correlation between (100 kb "bookend") boundary stability and trait heritability (Fig. 4C) is even stronger for the subset of traits in the boundary-enriched cluster ($R^2 = 0.26$, $p = 4x10^{-7}$). **(D)** Odds of cluster membership across phenotype categories. The boundary-enriched cluster is predominantly hematologic, immunologic, and metabolic traits. The boundary-depleted cluster is predominantly neuropsychiatric traits. **(E)** There is a negative correlation between (100 kb "bookend") boundary stability and trait heritability for traits in the boundary-depleted cluster ($R^2 = 0.042$, $p = 0.07$). Error bars/bands signify 99% confidence intervals in B & C, 95% confidence interval in D & F.

DISCUSSION

Although we are beginning to understand the role of 3D genome disruption in rare disease and cancer, we have a limited framework for integrating maps of 3D genome structure into the study of genome evolution and the interpretation of common disease-associated variation. Here, we show that TAD boundaries are enriched for evolutionary constraint and common complex trait heritability compared to TADs. Furthermore, in exploring TAD boundaries stable across cell types, we find they are further enriched for CTCF binding, evolutionary constraint, and heritability of hematologic, immunologic, and metabolic traits. These findings demonstrate the relationship between 3D genome structure and the genetic architecture of common complex

disease and reveal differences in the evolutionary pressures acting on different components (*i.e.* TADs and TAD boundaries) of the 3D genome.

Previous work has predominantly characterized the importance and evolutionary constraint of different components of the 3D genome from the perspective of SV and rearrangement events. We address functionality at the level of common single nucleotide variants and evolutionary constraint across diverse vertebrate species (~13-450 mya) and within humans (~100,000 ya). Over vertebrate evolution, we show that TAD boundaries have more sequence-level constraint than TADs. This provides a complementary perspective to Krefting *et al.* (2018) which found that TAD boundaries are enriched for syntenic breaks between humans and 12 vertebrate species, thus, concluding that intact TADs are shuffled over evolutionary time.[25] While shuffling a TAD may "move" its genomic location, preserving the TAD unit also requires maintaining at least part of its boundary. Our work suggests that even though TADs are shuffled, the boundary-defining sequences are under more constraint than the sequences within the TAD. This is further supported by the high TAD boundary concordance between syntenic blocks in different species and depletion of SVs at TAD boundaries in humans and primates.[7,19,26,32,37,38]

At the scale of common human variation, we show that TAD boundaries are enriched for SNPs that account for the heritability of common complex traits. By demonstrating this relationship between 3D genome structure and common disease-associated variation, we suggest that TAD boundaries are more intolerant of variation than TADs. This aligns with the finding of Whalen et al. (2019) that human haplotype breakpoints—which are associated with increased variation due to the mutagenic properties of recombination—are depleted at chromatin boundaries.[33]

Slight variation in 3D structure can cause proportionally large changes in gene expression.[24] Thus, it is tempting to speculate that the trait-associated common variants in TAD boundaries modify or disrupt elements important for the formation, maintenance, or dynamics of TADs. For example, CTCF helps maintain and form TAD boundaries; consequently, altering CTCF binding often leads to functional gene expression changes, *e.g.*, oncogenic gene expression in gliomas.[27] We hypothesize that altering gene regulation though common variant-disruption of transcription-factor motifs important in 3D structure organization, like CTCF, contributes to the enrichment for complex disease heritability. However, variation at TAD boundaries may also disrupt regulatory elements like enhancers, that are known to be enriched at boundaries, without disrupting the TAD architecture. A deeper mechanistic understanding of TAD formation will be critical to further understanding how TAD boundary disruption contributes to both rare and common disease at potentially nucleotide-level and cell type resolution.

Our finding of divergent patterns of TAD boundary heritability enrichment for different traits (enrichment for hematologic, immunologic and metabolic traits versus depletion for psychiatric and dermatologic) suggests that the 3D genome architecture may play differing roles in the genetic architecture of different traits. As a preliminary test of this hypothesis, we evaluated the relationship between boundary stability and intra-TAD heritability enrichment. We find that, for traits with heritability depletion at boundaries (psychiatric, dermatologic traits), increasingly stable TAD boundaries insulate increased intra-TAD heritability (Supplemental Fig. 11). For these traits, we speculate that stable boundaries may function to insulate important intra-TAD elements (*e.g.* enhancers or genes); whereas, for the boundary-enriched traits, functional elements are located at the stable boundaries. This idea is consistent with previous work showing that super-enhancers are insulated by the strongest boundaries (in a single cell type).[40] Thus, it will be important to further explore how TAD boundaries (or other functional elements at TAD boundaries) may play different regulatory roles in different traits and diseases.

Finally, we identify substantial variation among 3D maps across cell types. While TAD stability across cell types is greater than expected by chance, our findings expand the number and diversity of cell types compared and identify a large proportion of boundaries unique to

single cell types. Depending on the TAD boundary definition used and the subset of cell types compared, median pairwise Jaccard similarities range from 0.18 to 0.39. This highlights the sensitivity of stability comparisons across cell types to the definition of TAD boundaries used. For example, the median pairwise Jaccard similarity between 40 kb boundaries across 21 tissues defined by Schmitt et al. (2016) is 0.106 (IQR: 0.086 - 0.123). However, they collapsed boundaries to 200 kb "boundary regions" to conclude that TAD boundaries are highly stable (stating that over 35% of TAD boundaries are present in 21 of 21 tissues). These previous studies often investigated more homogenous groups of cell types which could lead to higher estimates of stability; however, even when we remove cancer cell lines from our analysis, 36% of boundaries are unique to a single cell type and median pairwise Jaccard similarity is 0.18 (IQR: 0.15 – 0.22).

Our analysis is, however, limited by Hi-C data availability; despite analyzing TADs from the same pipeline, there could be batch or protocol-specific effects because the Hi-C are generated by different groups. This underscores the need for higher resolution Hi-C across replicates of diverse cell types and continued efforts to integrate data from multiple TAD-calling algorithms to more precisely define TAD boundaries, especially given their hierarchical nature.[41,63] Despite these complexities in identifying TAD boundaries, our findings replicate with all our boundary definitions. Even with our most inclusive definition, we estimate that more than 50% of TAD boundaries differ between any two cell types and only 5% of TAD boundaries are highly shared (in 20+ of the 37 cell types). Furthermore, using this metric of cell-type-stability to stratify TAD boundaries identifies meaningful biological differences: stable boundaries are enriched for CTCF binding, evolutionary constraint, housekeeping genes, and common trait heritability.

CONCLUSIONS

Here, we introduce the stability of a TAD boundary across cell types as a functionally relevant metric. At these stable TAD boundaries, we observe enrichment of trait heritability, CTCF binding, and sequence-level constraint. Our work suggests the utility of incorporating 3D structural data across multiple cell types to aid context-specific non-coding variant interpretation. Starting from this foundation, much further work is needed to elucidate the molecular mechanisms, evolutionary history, and cell-type-specificity of TAD structure disruption. Furthermore, while we have identified properties of TAD boundaries stable across cell types, it would also be valuable to identify differences in TAD boundary stability across species to find human-specific structures. Finally, as high-resolution Hi-C becomes more prevalent from diverse tissues and individuals, we anticipate that computational prediction of personalized cell-type-specific TAD structure will facilitate understanding of how any possible variant is likely to affect 3D genome structure, gene regulation, and disease risk.

METHODS

Defining TADs

TAD maps for 37 different cell types were obtained from the 3D genome browser in BED format (Supplemental Table 1).[42] All TAD map analyses were conducted using the hg19 genome build. All cell types were available in hg19 format, except the Liver data, which we downloaded in hg38 and used the UCSC liftOver tool to convert to hg19.[64,65] All TAD maps were systematically predicted from Hi-C data using the hidden Markov model (HMM) pipeline from Dixon *et al*. (2012).[19,42] All analyses on the TAD map BED files were performed using the pybedtools wrapper for BedTools.[66,67] Most maps were already defined using the same 40 kb windows.

However, 7 cell types (GM12878, HMEC, HUVEC, IMR90, K562, KBM7, NHEK) were defined using 25 kb windows. We converted each of these TAD start and stop positions to the corresponding 40 kb windows used for the other cell types. Therefore, all resulting maps are defined with respect to the same 40 kb windows across the genome for each cell type.

Quantifying boundary overlap and stability

40 kb boundaries

For each cell type we defined a set of 40 kb boundaries. These were the 40 kb windows surrounding (± 20 kb) TAD start and stop sites. For example, if a TAD was located at chr1: 2,000,000-3,000,000, we would define its TAD boundaries to be at chr1:1,980,000-2,020,000 (boundary around the start) and chr1:2,980,000-3,020,000 (boundary around the end). These 40 kb boundaries would be with respect to the same 40 kb windows across the genome for all cell types because of how the TADs were defined (see above). Therefore, to quantify stability, we examined each 40 kb window across the genome (*e.g.* chr1:20,000-60,000, chr1:60,000-100,000, chr1: 100,000-140,000, etc.). If there was a TAD boundary at that window for any of the cell types, we counted how many cell types (out of 37) shared that boundary. If only one cell type had a boundary at that location, it would be considered a “unique” boundary; whereas if it was observed in many cell types, it would be considered “stable. To quantify the stability, we calculated a percentile corresponding to each level of stability. For example, if 5% of boundaries are shared in 20 or more of the cell types, a boundary observed in 20 cell types would be at the 95th percentile of stability compared to other boundaries.

200 kb collapsed boundaries

To demonstrate the robustness of our results to a more permissive boundary and overlap definition, we defined a second set of boundaries used for analysis shown in the supplementary figures. For this we grouped all adjacent boundaries together regardless of their cell type (Supplemental Fig. 2A). For example, if liver cells had a boundary at chr1:60,000-100,000 and muscle cells had a boundary at chr1:100,000-140,000, we would consider this to represent the same larger boundary of chr1:60,000-140,000. We show the size distribution of the collapsed boundaries in Supplemental Fig. 2B. We did not consider the 930 (11%) boundaries that, when collapsed, were larger than 200 kb. Therefore, the collapsed boundaries are either 40, 80, 120, 160, or 200 kb long. We calculated the boundary stability and stability percentiles in the same way as the 40 kb boundaries (above).

Quantifying TAD boundary profile similarity

To quantify TAD boundary profile similarity between any two cell types, we calculate a Jaccard similarity coefficient. We do this by counting the number of shared boundaries (intersection) and dividing by the total boundaries over both tissues (union). For the TAD boundary similarity heatmaps in Fig. 2C and Supplemental Fig. 3, we clustered the cell types using complete-linkage (*i.e.* farthest neighbor) with the Jaccard distance metric (1-stability). The plots were made using Seaborn’s clustermap.[68]

TAD boundary stability and enrichment analyses

CTCF enrichment

CTCF binding sites were determined through ChIP-seq analyses from ENCODE.[45,46] We searched for all CTCF ChIP-seq data with the following criteria: experiment, released, ChIP-seq, human (hg19), all tissues, adult, BED NarrowPeak file format, exclude any experiments with biosample tratments. Across all files, the CTCF peaks were concatenated, sorted, and merged into a single bed file. Overlapping peaks were merged into a single larger peak. The CTCF bed file was intersected with the TAD boundary bed files (both 40 kb boundaries and 200 kb collapsed boundaries, described above) to quantify CTCF ChIP-seq peak’s overlap with each

TAD boundary, stratified by that boundary's stability. The two metrics of quantification were the number of CTCF ChIP-seq peaks per boundary (peaks per 40 kb) and the number of CTCF peak base pairs overlapping each boundary (base pairs per 40 kb).

Evolutionary constraint

Evolutionary constraint was quantified by PhastCons elements. These were determined through a maximum likelihood fitting of a phylo-HMM across a group of 46 vertebrate genomes to predict conserved elements.[69] The BED file for these conserved elements was downloaded from the UCSC table browser in hg19.[65,70] Each element has a score describing its level of conservation (a transformed log-odds score between 0 and 1000). This PhastCons elements file was then intersected with the TAD boundary bed files (both 40 kb boundaries and 200 kb collapsed boundaries). The two metrics of overlap quantification are number of PhastCons base pairs per boundary regardless of score (base pairs per 40 kb) and the average PhastCons element score per boundary (average score of elements in the 40 kb boundary).

Genes & Protein-coding genes

A list of all RefSeq genes and hg19 coordinates were downloaded from the UCSC table browser.[65,70,71] These were converted into a simplified list which included coordinates of only one transcript per gene (the longest) and included only autosomal and sex chromosome genes. From the simplified list of RefSeq genes, a subset of protein-coding genes were also created (identified using RefSeq accession numbers starting with NM). The simplified RefSeq gene list contains 27,090 genes. The simplified protein-coding RefSeq gene list contains 19,225 genes. These two gene sets were intersected with the TAD boundary bed files stratified by boundary stability (both 40 kb boundaries and 200 kb collapsed boundaries). The two metrics of overlap quantification are the number of genes or protein-coding genes per boundary (genes per 40 kb) and the number of gene or protein-coding gene base pairs per boundary (base pairs per 40 kb).

Housekeeping genes

3804 housekeeping genes were downloaded from Eisenberg & Levanon (2013).[72] We converted the gene names into hg19 coordinates using RefSeq genes from the UCSC table browser, resulting in a BED file list of coordinates for 3681 genes (coordinates for a small number of genes were not found in the RefSeq list).[65,70,71] For each gene, coordinates of the longest transcript were considered. These housekeeping genes were intersected with the TAD boundary bed files stratified by boundary stability (both 40 kb boundaries and 200 kb collapsed boundaries). Because the number of housekeeping genes is different than the total number of genes, it is hard to directly compare the relationship between TAD boundary stability and density of genes versus housekeeping genes. To compare these relationships on the same scale, we normalized each absolute result by calculating an enrichment from genomic background (see controls, below). For example, stable boundaries (observed in 10+ cell types) have 1.5 times enrichment of genes compared to genomic background ($\frac{1.33\ genes\ /\ 40\ kb\ boundary}{0.86\ genes\ /\ 40\ kb\ window\ in\ genomic\ background} = 1.5$) whereas they have 2.6 times enrichment of housekeeping genes ($\frac{0.29\ HKG\ /\ 40\ kb\ boundary}{0.11\ HKG\ /\ 40\ kb\ window\ in\ genomic\ background} = 2.6$).

Controls

To compare the density of annotations (CTCF, PhastCons, genes) in boundaries versus in TADs or genomic background, we defined matched sets of randomly-shuffled equivalently-sized (40 kb) windows within TADs or the background genome. To define windows within the TADs, we shuffled the 40 kb boundaries for each cell type and required them to fall outside TAD boundaries and inside the TADs (using TAD maps defined by the 3D genome browser). We

shuffled the TAD boundaries three times for each cell type (N = 111). Using this comparative intra-TAD set, we calculated density of these annotations (CTCF, PhastCons, genes) using the same method described above for TAD boundaries. To define a genomic background control set, we considered the region interrogated by Hi-C for each chromosome (using the first start coordinate and last stop coordinate for each chromosome observed in TAD maps). The Y chromosome was excluded from genomic background because our TAD maps do not cover the Y chromosome. We also excluded blacklisted regions[73] and genomic gaps (centromeric/telomeric repeats from UCSC table browser).[65,70] We performed ten shuffles of the 23,553 40 kb TAD boundaries (observed across all cell types) into this genomic background subset region. Using this control set, we calculated density of these annotations (CTCF, PhastCons, genes) using the same method described above for TAD boundaries.

Quantifying partitioned heritability with S-LDSC

We conducted partitioned heritability using stratified-LD Score Regression v1.0.1 (S-LDSC) to test whether an annotation of interest (e.g. TADs or TAD boundaries) is enriched for heritability of a trait.[47,48] We considered GWAS summary statistics from a previously-described representative set of 41 diseases and complex traits. [49–59] Previous studies using these traits had GWAS replicates (genetic correlation > 0.9) for six of these traits. For these six traits (BMI, Height, High Cholesterol, Type 2 Diabetes, Smoking status, Years of Education), we considered only the GWAS with the largest sample size so our combined analysis did not overrepresent these six. All GWAS are European-ancestry only. We use 1000 Genomes for the LD reference panel (SNPs with MAF > 0.05 in European samples)[74] and HapMap Project Phase 3 (HapMap 3)[75] excluding the MHC region for our regression SNPs to estimate heritability enrichment and standardized effect size metrics.

Heritability enrichment

S-LDSC estimates the heritability enrichment, defined as the proportion of heritability explained by SNPs in the annotation divided by the proportion of SNPs in the annotation. The enrichment of annotation c is estimated as

$$Enrichment_c = \frac{\%h^2{}_{(c)}}{\%SNP_{(c)}} = \frac{h^2{}_{(c)}/h^2}{|\,c\,|\,/\,M} \text{ (Equation 1),}$$

where $h^2{}_{(c)}$ is the heritability explained by common SNPs in annotation c, h^2 is the heritability explained by the common SNPs over the whole genome, $|c|$ is the number of common SNPs that lie in the annotation, and M is the number of common SNPs considered over the genome.[48,50] To investigate trends over all traits, we average the heritability enrichment and provide a confidence interval. When compared to meta-analysis using a random effects model conducted using Rmeta (function meta.summaries())[48,50,60,76] the trends are consistent (Supplemental Fig. 6); therefore, we use averaging for future analyses to improve interpretability and reduce over-representation of higher powered GWAS traits.

Standardized effect size

In contrast to heritability enrichment, standardized effect size ($\tau^*{}_c$) quantifies effects that are unique to the focal annotation compared to a set of other annotations. The standardized effect size for annotation c *is*

$$\tau^*{}_c = \frac{M * \sqrt{\%SNP_{(c)}(1-\%SNP_{(c)})}}{h^2}\,\tau_c \text{ (Equation 2),}$$

where τ_c, estimated by LDSC, is the per-SNP contribution of (a one standard deviation increase in) annotation c to heritability, jointly modeled with other annotations.[50,61] The estimate of τ_c is conditioned on 86 diverse annotations from the baseline v2.1 model including coding, UTR, promoter and intronic regions, histone marks (H3K4me1, H3K4me3, H3K9ac, H3K27ac), DNAse I hypersensitivity sites (DHSs), chromHMM and Segway predictions, super-enhancers, FANTOM5 enchancers, GERP annotations, MAF bins, LD-related, and conservation annotations.[48,61,62] Standard errors are computed by LDSC using a block-jackknife. To combine across traits, we meta-analyze using a random effects model using Rmeta.[76]

Heritability enrichment across the TAD landscape

We analyzed TADs plus 50% of their total length on each side and subdivided into 20 equal-sized partitions, motivated by Krefting *et al*.[25] Hence, the center 10 bins (6-15) are inside the TAD. Bins 1-5 are upstream of the TAD and 16-20 are downstream of the TAD. We ran S-LDSR on these 20 bins across TAD maps from 37 cell types to calculate heritability enrichment over 41 traits. We investigated the heritability enrichment (or depletion) trends averaged across all traits and cell types (Fig 4A), by cell type (Fig. 4B), and by trait (Fig. 5A). For the analyses by cell type and by trait, we clustered the heritability landscapes to determine if related cell types or related traits had similar patterns of heritability across the 3D genome. To do so, correlation distance was used as the distance metric and the clustering metric used average linkage. When clustering traits by their heritability landscape across the 3D genome, two agglomerative clusters were defined and termed “boundary-enriched” and “boundary-depleted”.

Using other TAD callers

To assess the influence of technical variation of TAD calling on these findings, we assessed the heritability patterns in human embryonic stem cells across TADs called by seven diverse methods (Armatus, Arrowhead, DomainCaller, HiCseg, TADbit, TADtree, TopDom). The TAD were called and published by Dali et al. 2017[77] using Hi-C from Dixon et al. 2015.[34] All other analysis was conducted the same as described using the 3DGenomeBrowser TADs.

Heritability enrichment by TAD boundary stability

Based on the heritability enrichment patterns observed “bookending” the TAD start and end rather than the exact TAD start/end site, we defined an additional “100 kb bookend” TAD boundary definition to quantify this effect. These are regions 100 kb upstream of the TAD start and 100 kb downstream of the TAD end. For example, if a TAD was at chr1: 2,000,000-3,000,000, we would define its TAD boundaries to be at chr1:1,900,000-2,000,000 (boundary around the start) and chr1:3,000,000-3,100,000 (boundary around the end). These boundaries were divided into quartiles of cell-type-stability. The cell type stability for quartile 1 included only boundaries observed in a single cell type, quartile 2 included boundaries in [2,4] cell types, quartile 3 included boundaries in [5,12], and quartile 4 included boundaries in 13 or more cell types. S-LDSC was conducted on each quartile of stability for all 41 traits. Enrichment values were log-scaled and a linear regression (log10(heritability enrichment) ~ quartile of stability) was conducted.

Defining GWAS phenotypic classes

To determine if similar phenotypes had similar heritability patterns across the 3D genome, we defined eight different phenotypic classes (Supplemental Table 2): cardiopulmonary (N = 4), dermatologic (N = 7), hematologic (N = 5), immunologic (N = 4), metabolic (N = 7), neuropsychiatric (N = 8), reproductive (N = 4), and skeletal (N = 2). Our clusters originated from GWAS atlas’ domains[78]; however, the categories were modified to place more emphasis on disease pathophysiology instead of involved organ system (*e.g.* Crohn’s Disease and Rheumatoid Arthritis were moved from the gastrointestinal and connective tissue categories

respectively to an immunologic category). Similar categories were also combined (*e.g.* metabolic and endocrine, cardiovascular and respiratory).

Data analysis and figure generation

Data and statistical analyses were conducted in Python 3.5.4 (Anaconda distribution) and R 3.6.1. Figure generation was aided by Matplotlib, Seaborn, and Inkscape.[68,79,80] This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, TN.

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The publicly available datasets and software used for analysis are available in the following repositories: TAD maps from 3D Genome Browser [https://promoter.bx.psu.edu/hi-c/publications.html][42], LDSC [https://github.com/bulik/ldsc][47,48], GWAS traits formatted for LDSC from the Alkes Price lab [https://data.broadinstitute.org/alkesgroup/LDSCORE/independent_sumstats/], housekeeping genes [https://www.tau.ac.il/~elieis/HKG/HK_genes.txt][72], PhastCons elements, RefSeq Genes, and genome gaps [https://genome.ucsc.edu/cgi-bin/hgTables][65,69–71], and CTCF ChIP-seq peaks [https://www.encodeproject.org/][45,46].

The datasets we generated are available in the TAD-stability-heritability GitHub repository [https://github.com/emcarthur/TAD-stability-heritability] with DOI available at Zenodo[81] and include all results of our boundary calling (40 kb, 100 kb bookend, and 200 kb collapsed) and all partitioned heritability analysis output (by cell type and trait). Code is available upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Institutes of Health grant R35GM127087 to JAC and T32GM007347. The funding bodies had no role in the design of the study and collection, analysis, or interpretation of data, or in writing the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' contributions

EM and JAC conceived and designed all the work presented here. EM conducted all the analysis. EM and JAC interpreted the results, drafted the work, and substantively revised the manuscript. EM and JAC have approved the submitted version and have agreed to be accountable for their contributions.

Acknowledgements

The authors would like to thank Katherine S. Pollard, Emily Hodges, Geoff Fudenberg, Sarah Fong, Mary Lauren Benton, and other members of the Capra Lab for helpful discussions and manuscript comments. They would like to thank Margaux L.A. Hujoel and Steven Gazal for their

help with LDSC implementation and heritability result interpretation. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

REFERENCES

1. Cavalli G, Misteli T. Functional implications of genome topology. *Nat Struct Mol Biol*. 2013. doi:10.1038/nsmb.2474
2. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2001. doi:10.1038/35066075
3. Duggal G, Wang H, Kingsford C. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res*. 2014. doi:10.1093/nar/gkt857
4. Le Dily FL, Baù D, Pohl A, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev*. 2014. doi:10.1101/gad.241422.114
5. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science (80-)*. 2002. doi:10.1126/science.1067799
6. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi:10.1126/science.1181369
7. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell*. 2016;62(5):668-680. doi:10.1016/J.MOLCEL.2016.05.018
8. Fudenberg G, Getz G, Meyerson M, Mirny LA. High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat Biotechnol*. 2011. doi:10.1038/nbt.2049
9. Hnisz D, Weintrau AS, Day DS, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science (80-)*. 2016. doi:10.1126/science.aad9024
10. Meaburn KJ, Gudla PR, Khan S, Lockett SJ, Misteli T. Disease-specific gene repositioning in breast cancer. *J Cell Biol*. 2009. doi:10.1083/jcb.200909127
11. Misteli T. Higher-order genome organization in human disease. *Cold Spring Harb Perspect Biol*. 2010. doi:10.1101/cshperspect.a000794
12. Bonev B, Mendelson Cohen N, Szabo Q, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 2017;171(3):557-572.e24. doi:10.1016/J.CELL.2017.09.043
13. Sauerwald N, Kingsford C. Quantifying the similarity of topological domains across normal and cancer human cell types. *Bioinformatics*. 2018;34(13):i475-i483. doi:10.1093/bioinformatics/bty265
14. Yu J, Hu M, Li C. Integrative analyses of multi-tissue Hi-C and eQTL data demonstrate close spatial proximity between eQTLs and their target genes. 2018. doi:10.1101/392266
15. Aguet F, Ardlie KG, Cummings BB, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017. doi:10.1038/nature24277
16. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: Illuminating the Dark Road from Association to Function. *Am J Hum Genet*. 2013;93(5):779-797. https://www.sciencedirect.com/science/article/pii/S0002929713004692?via%3Dihub. Accessed March 19, 2019.
17. Boveri T. Die blastomerenkerne von Ascaris meglocephala und die theorie der chromosomenindividualitaet. *Archiv fur Zellforschung.* 1909;3:81–268.
18. Rabl C. Uber Zellteilung. *Morphologisches Jahrbuch*. 1885;10:214–330.
19. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376-380. doi:10.1038/nature11082

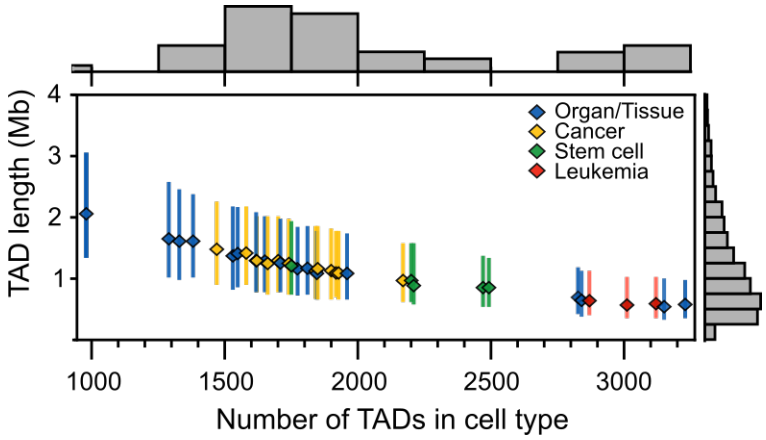
20. Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012. doi:10.1038/nature11049
21. Sexton T, Yaffe E, Kenigsberg E, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. 2012;148(3):458-472. doi:10.1016/j.cell.2012.01.010
22. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. *Nat Rev Genet*. 2018;19(12):789-800. doi:10.1038/s41576-018-0060-8
23. Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet*. July 2019:1. doi:10.1038/s41588-019-0462-3
24. Greenwald WW, Li H, Benaglio P, et al. Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat Commun*. 2019;10(1):1054. doi:10.1038/s41467-019-08940-5
25. Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol*. 2018;16(1):87. doi:10.1186/s12915-018-0556-x
26. Vietri Rudan M, Barrington C, Henderson S, et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep*. 2015;10(8):1297-1309. doi:10.1016/J.CELREP.2015.02.004
27. Flavahan WA, Drier Y, Liau BB, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016. doi:10.1038/nature16490
28. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. *Nat Rev Genet*. 2018;19(7):453-467. doi:10.1038/s41576-018-0007-0
29. Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*. 2015;161(5):1012-1025. doi:10.1016/J.CELL.2015.04.004
30. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell*. 2014. doi:10.1016/j.cell.2014.02.019
31. Northcott PA, Lee C, Zichner T, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*. 2014. doi:10.1038/nature13379
32. Fudenberg G, Pollard KS. Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A*. 2019;116(6):2175-2180. doi:10.1073/pnas.1808631116
33. Whalen S, Pollard KS. Most chromatin interactions are not in linkage disequilibrium. *Genome Res*. January 2019:gr.238022.118. doi:10.1101/gr.238022.118
34. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331-336. doi:10.1038/nature14222
35. Rao SSP, Huntley MH, Durand NC, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. 2014;159(7):1665-1680. doi:10.1016/J.CELL.2014.11.021
36. Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep*. 2016;17(8):2042-2059. https://www.sciencedirect.com/science/article/pii/S2211124716314814?via%3Dihub. Accessed January 10, 2019.
37. Dekker J. Two ways to fold the genome during the cell cycle: Insights obtained with chromosome conformation capture. *Epigenetics and Chromatin*. 2014. doi:10.1186/1756-8935-7-25
38. Dekker J, Heard E. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*. 2015. doi:10.1016/j.febslet.2015.08.044
39. Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D

genome. *Nat Rev Mol Cell Biol*. 2016;17(12):771-782. doi:10.1038/nrm.2016.138
40. Gong Y, Lazaris C, Sakellaropoulos T, et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun*. 2018;9(1):542. doi:10.1038/s41467-018-03017-1
41. An L, Yang T, Yang J, et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol*. 2019;20(1):282. doi:10.1186/s13059-019-1893-y
42. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol*. 2018;19(1):151. doi:10.1186/s13059-018-1519-9
43. Leung D, Jung I, Rajagopal N, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*. 2015;518(7539):350-354. doi:10.1038/nature14217
44. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods*. 2015;72(C):65-75. doi:10.1016/j.ymeth.2014.10.031
45. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247
46. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794-D801. doi:10.1093/nar/gkx1081
47. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211
48. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404
49. Hormozdiari F, Gazal S, Van De Geijn B, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat Genet*. 2018. doi:10.1038/s41588-018-0148-2
50. Hujoel MLA, Gazal S, Hormozdiari F, van de Geijn B, Price AL. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. *Am J Hum Genet*. 2019;104(4):611-624. doi:10.1016/j.ajhg.2019.02.008
51. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014. doi:10.1038/nature13595
52. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
53. Boraska V, Franklin CS, Floyd JAB, et al. A genome-wide association study of anorexia nervosa. *Mol Psychiatry*. 2014. doi:10.1038/mp.2013.187
54. Smoller JW, Kendler K, Craddock N, et al. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet*. 2013. doi:10.1016/S0140-6736(12)62129-1
55. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012. doi:10.1038/nature11582
56. Okbay A, Baselmans BML, De Neve JE, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016. doi:10.1038/ng.3552
57. Barban N, Jansen R, De Vlaming R, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet*. 2016. doi:10.1038/ng.3698
58. Teslovich TM, Musunuru K, Smith A V., et al. Biological, clinical and population relevance

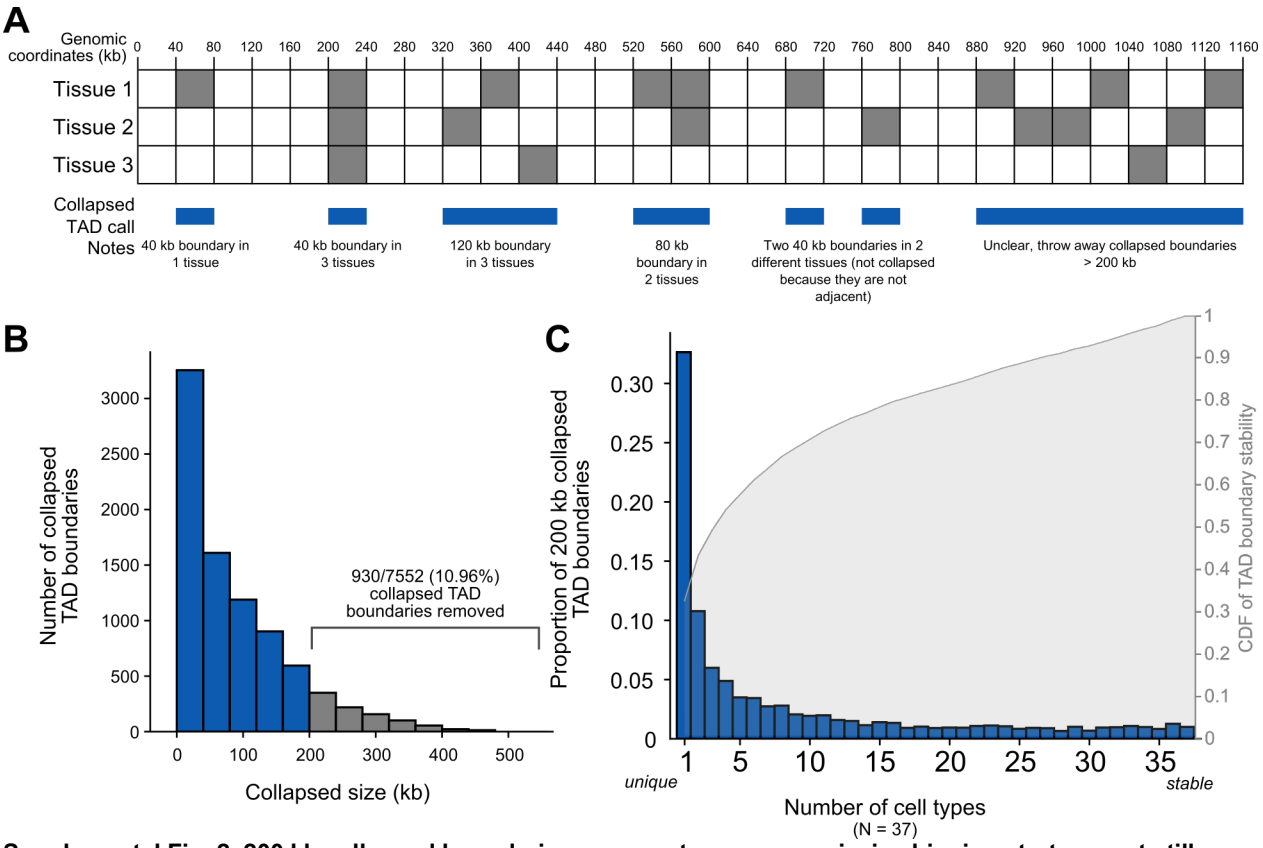
of 95 loci for blood lipids. *Nature*. 2010. doi:10.1038/nature09270
59. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014. doi:10.1038/nature12873
60. Hormozdiari F, van de Geijn B, Nasser J, et al. Functional disease architectures reveal unique biological role of transposable elements. 2018. doi:10.1101/482281
61. Gazal S, Finucane HK, Furlotte NA, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet*. 2017. doi:10.1038/ng.3954
62. Gazal S, Loh PR, Finucane HK, et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat Genet*. 2018. doi:10.1038/s41588-018-0231-8
63. Zufferey M, Tavernari D, Oricchio E, Ciriello G. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol*. 2018;19(1):217. doi:10.1186/s13059-018-1596-9
64. Hinrichs AS, Karolchik D, Baertsch R, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34(90001):D590-D598. doi:10.1093/nar/gkj144
65. Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gky1095
66. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010. doi:10.1093/bioinformatics/btq033
67. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*. 2011. doi:10.1093/bioinformatics/btr539
68. Waskom M, Botvinnik O, O'Kane D, et al. mwaskom/seaborn: v0.9.0 (July 2018). July 2018. doi:10.5281/ZENODO.1313201
69. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15(8):1034-1050. doi:10.1101/gr.3715005
70. Karolchik D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004. doi:10.1093/nar/gkh103
71. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016. doi:10.1093/nar/gkv1189
72. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569-574. doi:10.1016/J.TIG.2013.05.010
73. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*. 2019. doi:10.1038/s41598-019-45839-z
74. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015. doi:10.1038/nature15393
75. International HapMap 3 Consortium TIH 3, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-58. doi:10.1038/nature09298
76. Lumley T. rmeta: Meta-Analysis. R package version 3.0. 2018. https://cran.r-project.org/package=rmeta.
77. Dali R, Blanchette M. A critical assessment of topologically associating domain prediction tools. *Nucleic Acids Res*. 2017. doi:10.1093/nar/gkx145
78. Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019. doi:10.1038/s41588-019-0481-0
79. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007. doi:10.1109/MCSE.2007.55
80. InkscapeProject. Inkscape. 2018. https://inkscape.org.

81. McArthur E, Capra JA. emcarthur/TAD-stability-heritability. https://doi.org/10.5281/zenodo.3601560.

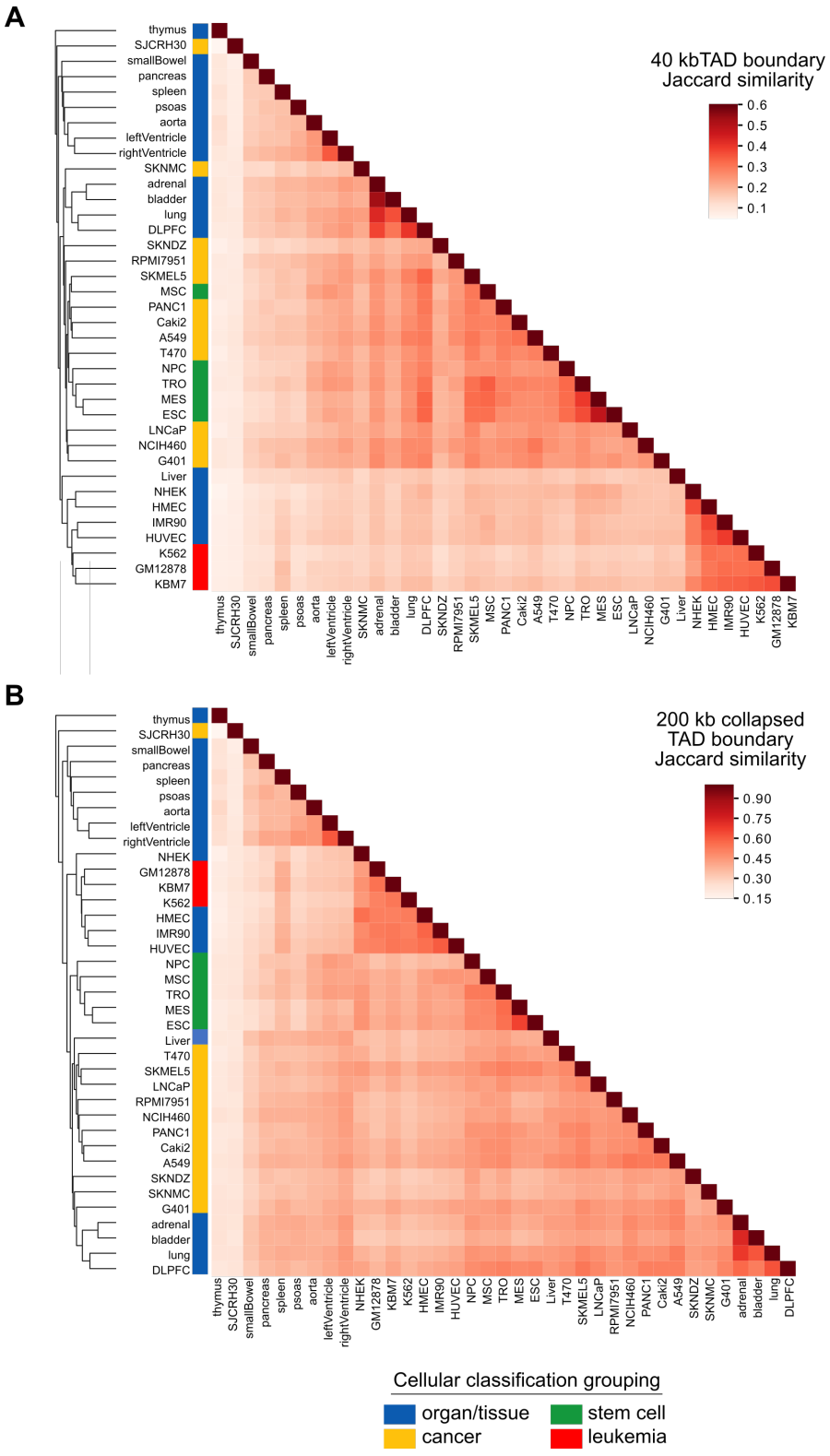
SUPPLEMENTAL FIGURES



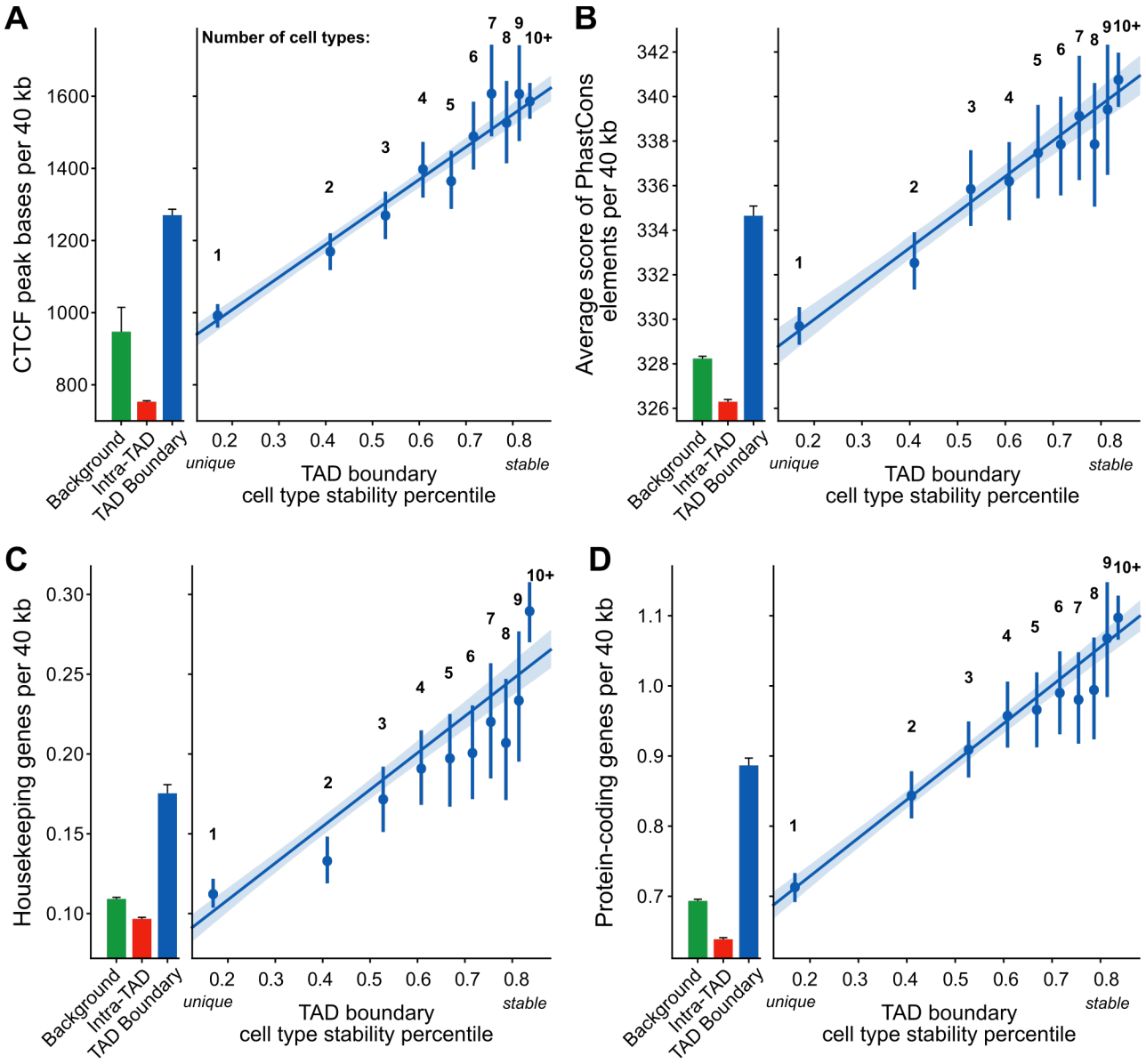
Supplemental Fig. 1. Across 37 cell types, there is an inverse relationship between TAD length and number of TADs. Organ/tissue cell types generally have the longest (and fewest) TADs. Leukemia and stem cells have the shortest (and most) TADs. Error bands signify the IQR.



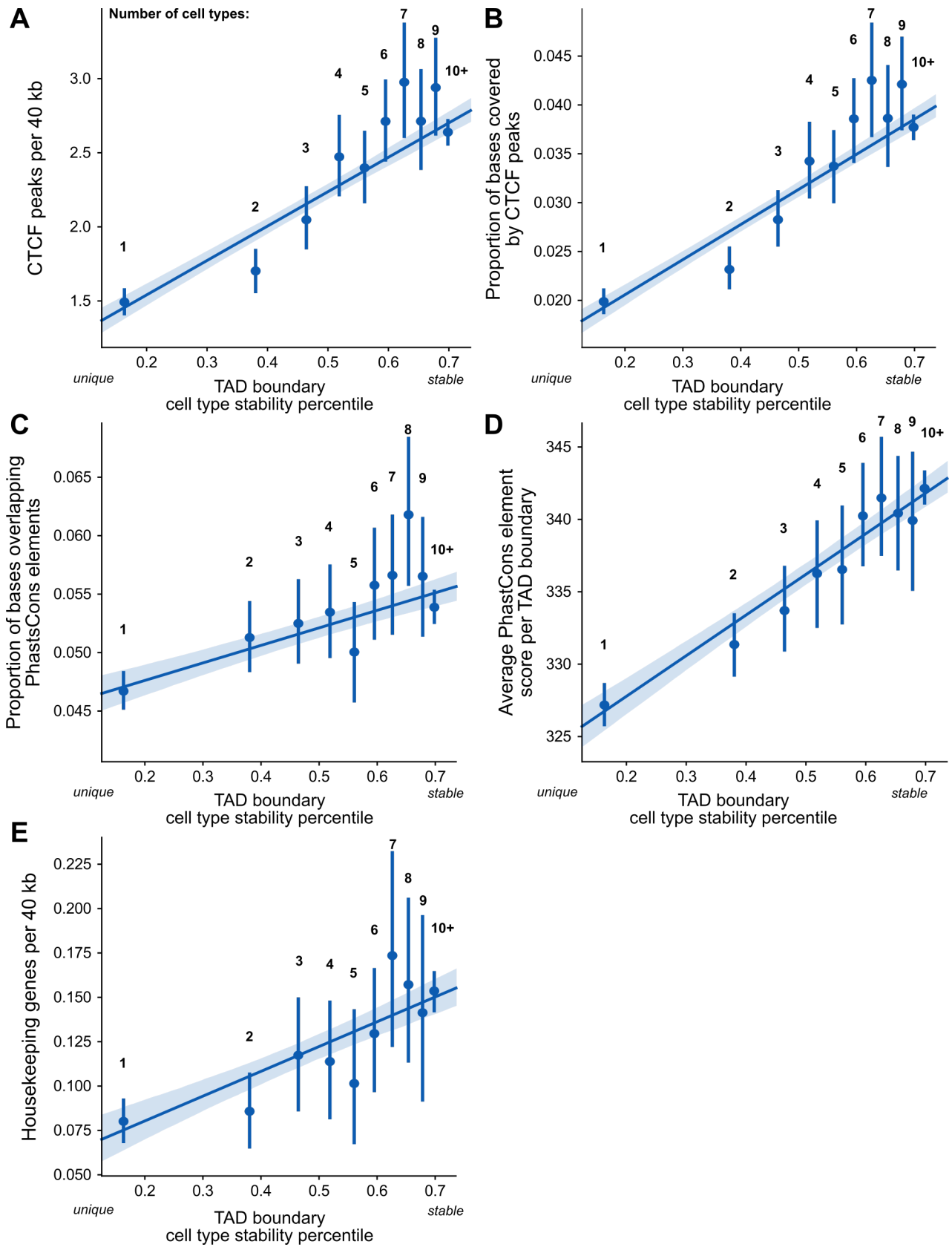
Supplemental Fig. 2. 200 kb collapsed boundaries represent a more permissive binning strategy, yet still have a large proportion of TAD boundaries unique to a single tissue. (A) Example cases demonstrating the 200 kb collapsed TAD boundary binning strategy. Boundaries are all called using the same 40 kb windows across the genome. If a boundary has no adjacent boundaries, it is a singular 40 kb boundary. If a boundary has adjacent boundaries, it is collapsed to an 80 kb (2 adjacent), 120 kb (3 adjacent), 160 kb (4 adjacent), or 200 kb (5 adjacent) boundary. If adjacent boundaries collapse to a size larger than 200 kb, these boundaries are removed from analysis. **(B)** Size distribution showing the collapsed boundary size. Boundaries that collapse to a size larger than 200 kb (10.96%) are removed from analysis. **(C)** Histogram of TAD boundaries (using this 200 kb collapsed boundary strategy) by the number of cell types they are observed in (their "stability"). Most boundaries (32.6%) are unique to a single cell type. The right axis and gray distribution represent the empirical CDF of boundary stability shown in the histogram.



Supplemental Fig. 3. Biologically similar cell types cluster by TAD map similarity. (A) Using the 40 kb boundaries, clustering for 37 cell types using the pairwise Jaccard similarity metric with colors labelling cellular classification groups. This is the same data shown in the bottom triangle of Fig. 2C but with added labels. **(B)** Using the 200 kb collapsed boundaries, clustering for 37 cell types.

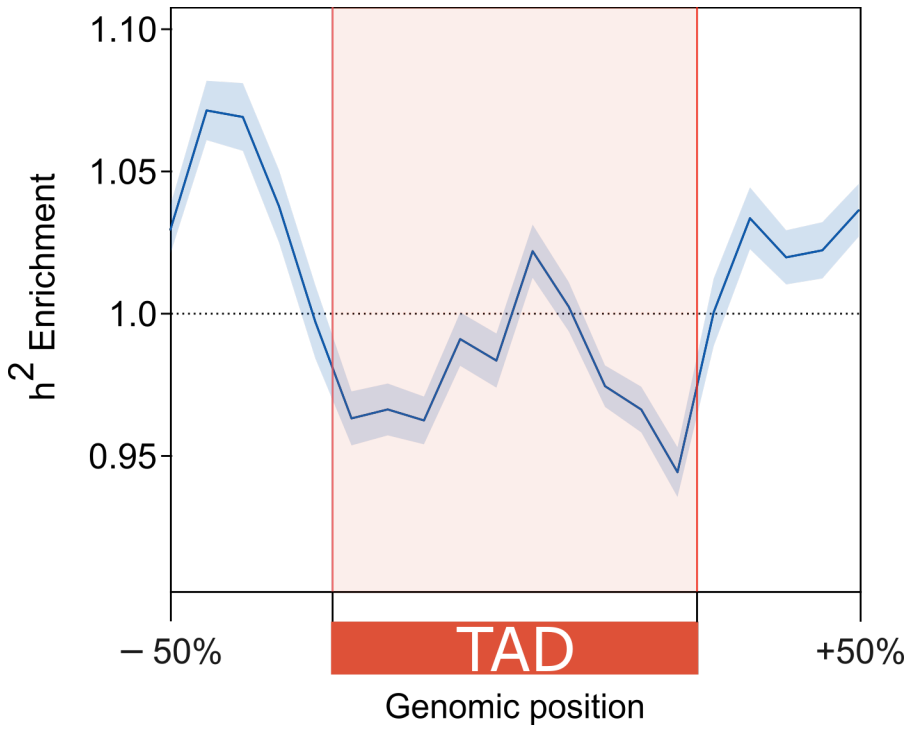


Supplemental Fig. 4. Stable TAD boundaries are enriched for CTCF binding, sequence-level conservation, and housekeeping genes (using metrics expanded on those shown in Fig. 3). Genomic annotation overlap is compared across 40 kb windows for genomic background (green bar), TADs (red bar), and TAD boundaries (blue bar). TAD boundaries are also stratified by their stability across cell types (blue points). TAD boundary stability is defined by the number of cell types sharing that TAD boundary (black numbers) and its corresponding percentile of stability (x-axis, generated through the CDF in Fig. 2B). Per 40 kb window, TAD boundaries and, more specifically, stable TAD boundaries have increased **(A)** bases overlapping CTCF ChIP-seq peaks, **(B)** average (46-way Vertebrate) PhastCons element score, **(B)** absolute overlap with housekeeping genes (relative enrichment from background is shown in Fig. 3D), and **(D)** absolute overlap with protein-coding genes (relative enrichment from background is shown in Fig. 3D). Error bars/bands signify 95% confidence intervals.


A
Number of cell types:
CTCF peaks per 40 kb
B
Proportion of bases covered by CTCF peaks
C
Proportion of bases overlapping PhastsCons elements
D
Average PhastCons element score per TAD boundary
E
Housekeeping genes per 40 kb
1
2
3
4
5
6
7
8
9
10+
unique
stable
TAD boundary
cell type stability percentile

1018

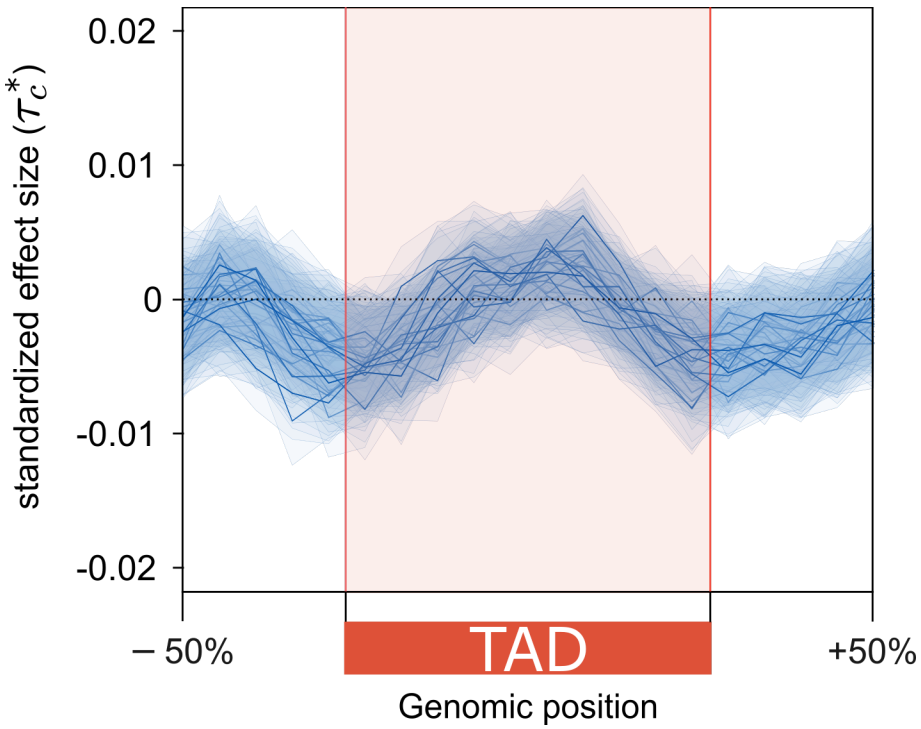
1019 **Supplemental Fig. 5. Using the 200kb collapsed definition of TAD boundaries, stable boundaries are enriched**
1020 **for CTCF binding, sequence-level conservation, and housekeeping genes.** TAD boundary stability is defined by
1021 the number of cell types sharing that TAD boundary (black numbers) and its corresponding percentile of stability (x-
1022 axis, generated through the CDF in Supplemental Fig. 2C). The stable 200 kb collapsed TAD boundaries have
1023 increased **(A)** density of CTCF peaks (scaled to peaks per 40 kb), **(B)** proportion of bases covered by CTCF ChIP-
1024 seq peaks, **(C)** proportion of bases overlapping (46-way Vertebrate) PhastCons elements, **(D)** average PhastCons
1025 element score, and **(E)** number of housekeeping genes (scaled to genes per 40 kb). Error bars/bands signify 95%
1026 confidence intervals.



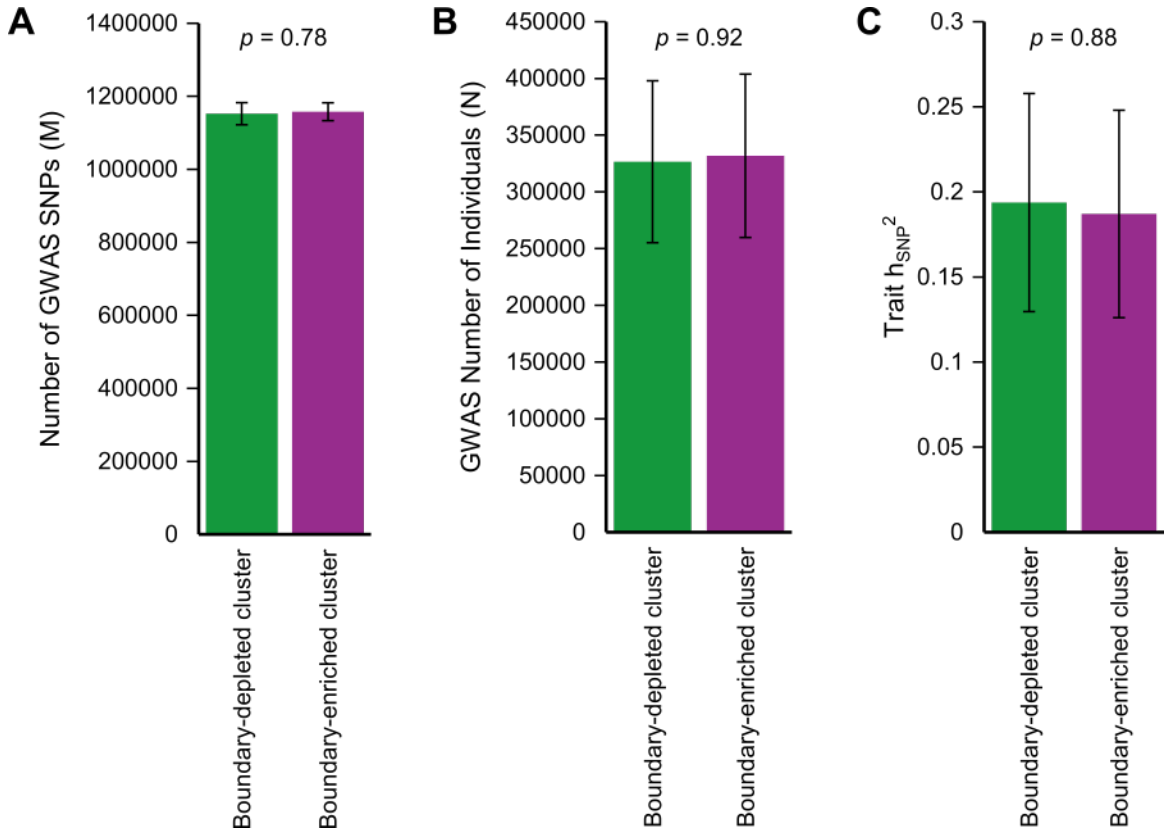
1027

1028 **Supplemental Fig. 6.** For TADs across 37 cell types, heritability is enriched near TAD boundaries when meta-
1029 analyzed across 41 common complex phenotypes. When combining data across traits, The heritability enrichment
1030 results are consistent using random-effects meta-analysis model (here) versus averaging (R^2 = 0.85, p = $7x10^{-9}$,Fig.
1031 4A). The error band signifies a 99% confidence interval.

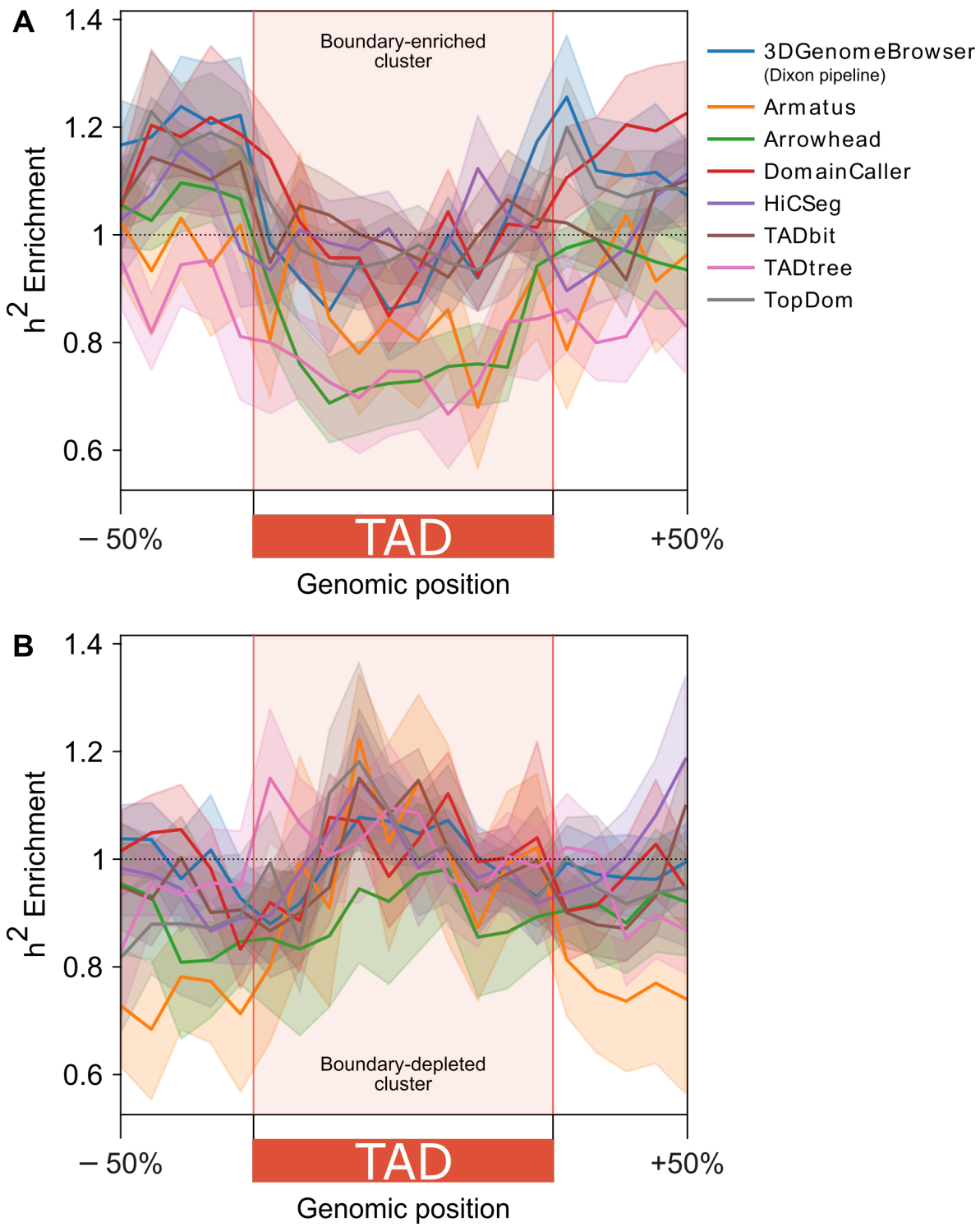
1032



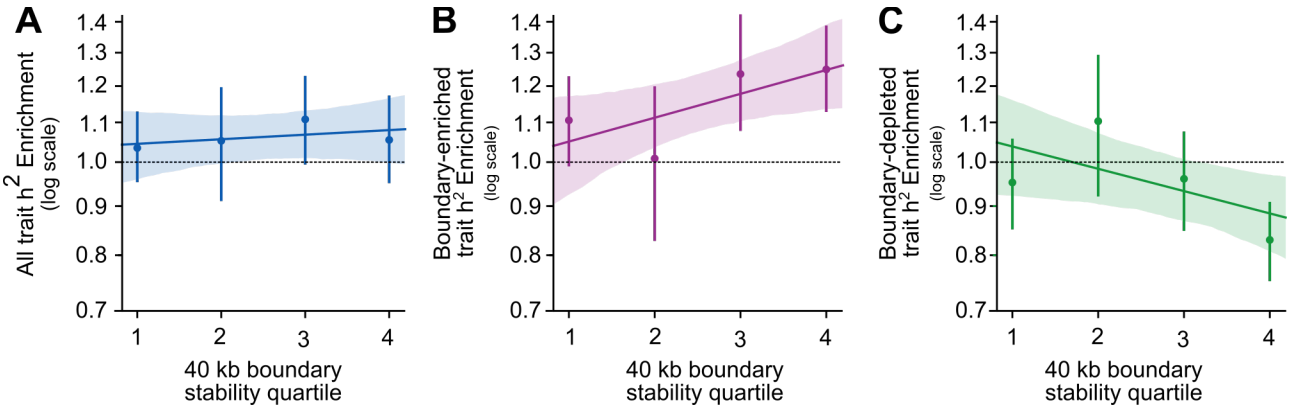
Supplemental Fig. 7. When meta-analyzed across all traits, there is no single part of the TAD landscape that is uniquely informative for trait heritability across all cell types when conditioned on a broad set of 86 functional (*e.g.* regulatory, conservation, coding, LD-related) annotations. Each line represents the standardized effect size meta-analyzed across all traits for that cell type (n = 37). The error bands signify 99% confidence intervals.



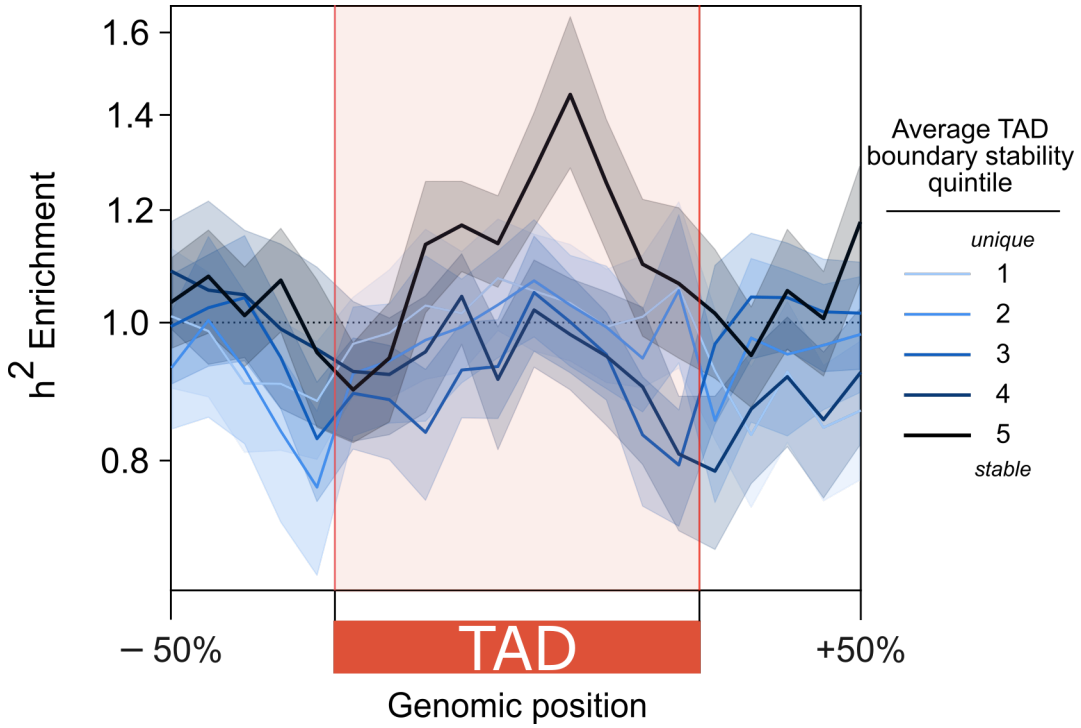
Supplemental Fig. 8. Traits in the boundary-depleted cluster and boundary-enriched cluster do not differ for GWAS parameters including: (A) Number of GWAS SNPs (p = 0.78, t-test with equal variances), **(B)** Number of individuals in the GWAS (p = 0.92), or **(C)** The traits SNP-based heritability (p = 0.88). Error bars signify 95% confidence intervals.



Supplemental Fig. 9. Patterns of heritability enrichment across the 3D genome in human embryonic stem cells (ESC) are robust to the TAD definition algorithm used. (A) Heritability enrichment landscape over TADs in ESCs called by eight different algorithms for traits in the boundary-enriched cluster. Like findings shown in Fig. 5B (which use TADs from the Dixon pipeline), TAD boundaries are enriched for heritability compared to TADs. **(B)** Heritability enrichment landscape over TADs in ESCs for traits in the boundary-depleted cluster. Like findings shown in Fig. 5C (which use TADs from the Dixon pipeline), TADs are centrally enriched for heritability compared to TAD boundaries. Error bands signify 95% confidence intervals.



Supplemental Fig. 10. The relationship between 40 kb boundary cell-type-stability and trait heritability is positive for the boundary-enriched trait cluster and negative for the boundary-depleted trait cluster. (A) There is a very weak positive relationship between 40 kb boundary stability and heritability enrichment for all traits. This positive correlation (R^2 = 0.0003, p = 0.8) is consistent with, but much weaker than, the positive correlation using the 100 kb "bookend" boundary definition (Fig. 3C). **(B)** 40 kb TAD boundaries in the boundary-enriched cluster traits (Fig. 5B) are further enriched for heritability at the highest quartile of cell-type-stability. This positive correlation (R^2 = 0.04, p = 0.07) is consistent with, but weaker than, the positive correlation using the 100 kb "bookend" boundary definition (Fig. 5D). **(C)** 40 kb TAD boundaries in the boundary-depleted cluster traits (Fig. 5C) are further depleted for heritability at the highest quartile of cell-type-stability. This negative correlation (R^2 = 0.03, p = 0.1) is consistent with, but weaker than, the negative correlation using the 100 kb "bookend" boundary definition (Fig. 5F). Error bars/bands signify 95% confidence intervals.



Supplemental Fig. 11. For the boundary-depleted cluster traits, TADs flanked by the most stable boundaries (measured by taking the average stability of its two boundaries and binning into quintiles) have increased heritability centrally. This analysis was performed in a random subset of 7 cell types (aorta, H1_ESC, leftVentricle, Liver, psoasMuscle, SKNDZ, T470). Error bands signify 95% confidence intervals.

SUPPLEMENTAL TABLES

Supplemental Table 1. Cell types used for all analyses from the 3DGenomeBrowser

FileNameFrom3DGenomeBrowser	CellTypeDescription	Abbreviation	BiologicalCluster	Citation
A549_raw-rep1_TADs.txt	A549_lungAdenocarcinoma_dekker	A549	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
AdrenalGland_Donor-AD2-raw_TADs.txt	adrenal_schmitt2016	adrenal	organ/tissue	Schmitt et al. (2016)[36]
Aorta_STL002_Leung2015-raw_TADs.txt	aorta_leung2015	aorta	organ/tissue	Leung et al. (2015)[43]
Bladder_Donor-BL1-raw_TADs.txt	bladder_schmitt2016	bladder	organ/tissue	Schmitt et al. (2016)[36]
Bowel_Small_Donor-SB2-raw_TADs.txt	smallBowel_schmitt2016	smallBowel	organ/tissue	Schmitt et al. (2016)[36]
Caki2_raw-rep1_TADs.txt	Caki2_clearCellRenalCellCarcinoma_dekker	Caki2	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
Cortex_DLPFC_Donor-CO-raw_TADs.txt	cortex_DLPFC_schmitt2016	DLPFC	organ/tissue	Schmitt et al. (2016)[36]
G401_raw-rep1_TADs.txt	G401_Wilms_tumor_dekker	G401	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
GM12878_Lieberman-raw_TADs.txt	GM12878_lymphoblastoid_Lieberman	GM12878	leukemia	Rao et al. (2014)[35]
H1-ESC_Dixon2015-raw_TADs.txt	H1_ESC_Dixon2015	ESC	stem cell	Dixon et al. (2015)[34]
H1-MES_Dixon2015-raw_TADs.txt	H1_mesendoderm_Dixon2015	MES	stem cell	Dixon et al. (2015)[34]
H1-MSC_Dixon2015-raw_TADs.txt	H1_mesenchymalSC_Dixon2015	MSC	stem cell	Dixon et al. (2015)[34]
H1-NPC_Dixon2015-raw_TADs.txt	H1_neuralSC_Dixon2015	NPC	stem cell	Dixon et al. (2015)[34]
H1-TRO_Dixon2015-raw_TADs.txt	H1_trophoblastLike_Dixon2015	TRO	stem cell	Dixon et al. (2015)[34]
HMEC_Lieberman-raw_TADs.txt	HMEC_humanMammaryEpithelial_Lieberman	HMEC	organ/tissue	Rao et al. (2014)[35]
HUVEC_Lieberman-raw_TADs.txt	HUVEC_Lieberman	HUVEC	organ/tissue	Rao et al. (2014)[35]
IMR90_Lieberman-raw_TADs.txt	IMR90_fetalLungFibroblast_Lieberman	IMR90	organ/tissue	Rao et al. (2014)[35]
K562_Lieberman-raw_TADs.txt	K562_CML_Lieberman	K562	leukemia	Rao et al. (2014)[35]
KBM7_Lieberman-raw_TADs.txt	KBM7_CML_Lieberman	KBM7	leukemia	Rao et al. (2014)[35]
Liver_STL011_Leung_2015-raw_TADs_hg19From38.txt	Liver_leung2015	Liver	organ/tissue	Leung et al. (2015)[43]
LNCaP_raw-rep1_TADs.txt	LNCaP_prostateAdenocarcinoma_dekker	LNCaP	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
Lung_Donor-LG1-raw_TADs.txt	lung_schmitt2016	lung	organ/tissue	Schmitt et al. (2016)[36]
Muscle_Psoas_Donor-PO1-raw_TADs.txt	psoasMuscle_schmitt2016	psoas	organ/tissue	Schmitt et al. (2016)[36]
NCIH460_raw-rep1_TADs.txt	NCIH460_NSCLC_dekker	NCIH460	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
NHEK_Lieberman-raw_TADs.txt	NHEK_epidermalKeratinocytes_Lieberman	NHEK	organ/tissue	Rao et al. (2014)[35]
PANC1_raw-rep1_TADs.txt	PANC1_pancreaticCarcinoma_dekker	PANC1	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
Pancreas_Donor-PA2-raw_TADs.txt	pancreas_schmitt2016	pancreas	organ/tissue	Schmitt et al. (2016)[36]
RPMI7951_raw-rep1_TADs.txt	RPMI7951_melanoma_dekker	RPMI7951	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
SJCRH30_raw-rep1_TADs.txt	SJCRH30_BMrhabdomyosarcoma_dekker	SJCRH30	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
SKMEL5_raw-rep1_TADs.txt	SKMEL5_melanoma_dekker	SKMEL5	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
SKNDZ_raw-rep1_TADs.txt	SKNDZ_neurblastoma_dekker	SKNDZ	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
SKNMC_raw-rep1_TADs.txt	SKNMC_neuroblastoma_dekker	SKNMC	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
Spleen_Donor-PX1-raw_TADs.txt	spleen_schmitt2016	spleen	organ/tissue	Schmitt et al. (2016)[36]
T470_raw-rep1_TADs.txt	T470_breastCancer_dekker	T470	cancer	Lajoie, Dekker et al. (2015)[44], ENCODE[45,46]
Thymus_STL001_Leung2015-raw_TADs.txt	thymus_leung2015	thymus	organ/tissue	Leung et al. (2015)[43]
VentricleLeft_STL003_Leung2015-raw_TADs.txt	leftVentricle_leung2015	leftVentricle	organ/tissue	Leung et al. (2015)[43]
VentricleRight_Donor-RV3-raw_TADs.txt	rightVentricle_schmitt2016	rightVentricle	organ/tissue	Schmitt et al. (2016)[36]

Supplemental Table 2. Genome-wide association study (GWAS) traits used for heritability analyses

Nickname	Trait	M	h2	h2_SE	N	Phenotypic class	actual cluster	Source
Anorexia	Anorexia	931184	0.2153	0.0169	32143	Neuropsych	Boundary-depleted	Boraska et al. 2014 Mol Psych[53]
ASD	Autism_Spectrum	1173307	0.4607	0.0517	10263	Neuropsych	Boundary-depleted	PGC Cross-Disorder Group, 2013 Lancet[54]
AutoimmuneDz	Auto_Immune_Traits_(Sure)	1187056	0.0068	0.0013	459324	Immunologic	Boundary-enriched	UKBiobank[52]
Balding	Balding_Type_I	1187056	0.2154	0.019	208336	Dermatologic	Boundary-depleted	UKBiobank[52]
BMI	BMI	1187056	0.252	0.0071	457824	Metabolic	Boundary-depleted	UKBiobank[52]
CrohnsDz	Crohn's_Disease	1051514	0.4723	0.0575	20883	Immunologic	Boundary-enriched	Jostins et al., 2012 Nature[55]
DepressiveSxs	Depressive_symptoms	1115393	0.0473	0.0037	161460	Neuropsych	Boundary-depleted	Okbay et al., 2016 Nat Genet[56]
DermDz	Dermatologic_Diseases	1187056	0.0094	0.0014	459324	Dermatologic	Boundary-enriched	UKBiobank[52]
Eczema	Eczema	1187056	0.0675	0.0038	458699	Dermatologic	Boundary-enriched	UKBiobank[52]
EosinophilCount	Eosinophil_Count	1187056	0.1977	0.0143	439938	Hematologic	Boundary-enriched	UKBiobank[52]
FEV1_FVC_Ratio	FEV1-FVC_Ratio	1187056	0.2336	0.0113	371949	Cardiopulmonary	Boundary-enriched	UKBiobank[52]
FirstBirthAge	Age_first_birth	1079424	0.0617	0.0033	222037	Reproductive	Boundary-depleted	Barban et al., 2016 Nat Genet[57]
FVC	Forced_Vital_Capacity_(FVC)	1187056	0.2068	0.0065	371949	Cardiopulmonary	Boundary-enriched	UKBiobank[52]
HairColor	Hair_Color	1187056	0.4523	0.1497	452720	Dermatologic	Boundary-depleted	UKBiobank[52]
HDL	HDL	1019272	0.1362	0.0166	99900	Metabolic	Boundary-enriched	Teslovich et al., 2010 Nature[58]
Heel_T_Score	Heel_T_Score	1187056	0.3628	0.0307	445921	Skeletal	Boundary-depleted	UKBiobank[52]
Height	Height	1187056	0.6034	0.027	458303	Skeletal	Boundary-enriched	UKBiobank[52]
HighCholesterol	High_Cholesterol	1187056	0.0468	0.0039	459324	Metabolic	Boundary-enriched	UKBiobank[52]
Hypothyroidism	Hypothyroidism	1187056	0.0459	0.0037	459324	Metabolic	Boundary-enriched	UKBiobank[52]
LDL	LDL	1017973	0.121	0.0166	95454	Metabolic	Boundary-enriched	Teslovich et al., 2010 Nature[58]
MenarcheAge	Age_at_Menarche	1187056	0.2457	0.0102	242278	Reproductive	Boundary-enriched	UKBiobank[52]
MenopauseAge	Age_at_Menopause	1187056	0.1215	0.0086	143025	Reproductive	Boundary-enriched	UKBiobank[52]
MorningPerson	Morning_Person	1187056	0.1002	0.0035	410520	Neuropsych	Boundary-depleted	UKBiobank[52]
Neuroticism	Neuroticism	1187056	0.1113	0.0037	372066	Neuropsych	Boundary-depleted	UKBiobank[52]
NumChildrenBorn	Number_children_ever_born	1080059	0.0256	0.0018	318863	Reproductive	Boundary-depleted	Barban et al., 2016 Nat Genet[57]
PlateletCount	Platelet_Count	1187056	0.349	0.0294	444382	Hematologic	Boundary-enriched	UKBiobank[52]
RA	Rheumatoid_Arthritis	1125155	0.1694	0.023	38242	Immunologic	Boundary-enriched	Okada et al., 2014 Nature[59]
RBCCount	Red_Blood_Cell_Count	1187056	0.2434	0.0191	445174	Hematologic	Boundary-enriched	UKBiobank[52]
RDW	Red_Blood_Cell_Distribution_Width	1187056	0.2234	0.0198	442700	Hematologic	Boundary-enriched	UKBiobank[52]
Resp_ENT_Dz	Respiratory_and_Ear-nose-throat_Diseases	1187056	0.0483	0.0034	459324	Cardiopulmonary	Boundary-depleted	UKBiobank[52]
Schizophrenia	Schizophrenia	1083014	0.4512	0.0189	70100	Neuropsych	Boundary-depleted	SCZ Working Group of the PGC, 2014 Nature[51]
SkinColor	Skin_Color	1187056	0.1896	0.0539	453609	Dermatologic	Boundary-depleted	UKBiobank[52]
SmokingStatus	Smoking_Status	1187056	0.0972	0.0032	457683	Neuropsych	Boundary-depleted	UKBiobank[52]
Sunburn	Sunburn_Occasion	1187056	0.0915	0.0162	344229	Dermatologic	Boundary-depleted	UKBiobank[52]
SystolicBP	Systolic_Blood_Pressure	1187056	0.1966	0.007	422771	Cardiopulmonary	Boundary-depleted	UKBiobank[52]
T2D	Type_2_Diabetes	1187056	0.043	0.0025	459324	Metabolic	Boundary-enriched	UKBiobank[52]
Tanning	Tanning	1187056	0.172	0.0609	449984	Dermatologic	Boundary-depleted	UKBiobank[52]
UC	Ulcerative_Colitis	1076834	0.2424	0.032	27432	Immunologic	Boundary-enriched	Jostins et al., 2012 Nature[55]
WaistHipRatio	Waist-hip_Ratio	1187056	0.1423	0.0067	458417	Metabolic	Boundary-enriched	UKBiobank[52]
WBCCount	White_Blood_Cell_Count	1187056	0.1873	0.0105	444502	Hematologic	Boundary-enriched	UKBiobank[52]
YearsOfEd	College_Education	1187056	0.1299	0.0037	454813	Neuropsych	Boundary-depleted	UKBiobank[52]