Deeply hidden genome organization directly mediated by SATB1

Yoshinori Kohwi[1*], Mari Grange[1], Hunter W. Richards[1], Ya-Chen Liang[2], Cheng-Ming Chuong[2]
Yohko Kitagawa[3†], Shimon Sakaguchi[3], Vladimir A. Botchkarev[4], Ichiro Taniguchi[5], and Terumi Kohwi-Shigematsu[1*]

[1] Department of Orofacial Sciences, University of California San Francisco, San Francisco, CA, 94143, USA
[2] Department of Pathology, University of Southern California, Los Angeles, CA, 90033, USA
[3] Laboratory of Experimental Immunology, Immunology Frontier Research Center, Osaka University, Osaka, 565-0871, Japan
[4] Department of Dermatology, Boston University School of Medicine, Boston, Massachusetts, 02118, USA
[5] Laboratory for Transcriptional Regulation, RIKEN Center for Integrative Medical Sciences, Yokohama, 230-0045, Japan
† present address: Center of iPS Research and Applications, Kyoto University, Kyoto, 606-8507, Japan

* Correspondence Email: TKohwi-Shigematsu@lbl.gov, YKohwi@lbl.gov

Abstract:

Mammalian genomes are organized by multi-layered chromatin folding. Whether and how three-dimensional genome organization contributes to cell-type specific transcription remains unclear. Here we uncover genome architecture formed by specialized sequences, base-unpairing regions (BURs), bound to a nuclear architectural protein, SATB1. SATB1 regulates cell-type specific transcription that underlies changes in cellular phenotypes. We developed a modified ChIP-seq protocol that stringently purifies genomic DNA only with its directly-associated proteins and unmasked previously-hidden BURs as direct SATB1 targets genome-wide. These SATB1-bound BURs are mutually exclusive from CTCF binding sites, and SATB1 is dispensable for CTCF/cohesion-mediated topologically associated domains (TADs). Instead, BURs largely overlap with lamina associated domains (LADs), and the fraction of BURs tethered to the SATB1 protein network in the nuclear interior is cell type-dependent. Our results reveal TAD-independent chromatin folding mediated by BUR sequences which serve as genome architecture landmarks for direct targeting by cell type-specific gene regulator, SATB1.

One-Sentence Summary: Genome-wide chromatin folding by direct tethering of base-unpairing regions to SATB1 nuclear architecture is unveiled.

Introduction

Three-dimensional (3D) chromatin architecture, formed by chromatin folding at multiple hierarchical levels, brings together far-distal genomic loci and is thought to promote accessibility of regulatory factors to their target gene loci [1-9]. However, whether and how complex 3D chromatin organization contributes to regulation of genome function still remain unclear[10, 11].

High resolution chromatin interaction maps, generated by chromosome conformation capture (Hi-C), revealed chromatin to be partitioned largely into two types of megabase-sized compartments, A and B [12]. The euchromatic A compartment is gene-rich and actively transcribed, conferring high chromatin accessibility, whereas the B compartment is transcriptionally repressive and has low gene density. The largely transcriptionally repressive lamina-associated domains (LADs) are among the B compartment chromatin [13, 14]. LADs are tethered to nuclear lamina, an intermediate filament meshwork adjacent to the inner nuclear membrane, and represent discrete chromatin domains of 40kb to 30Mb in size that consist of internal loops [13]. Approximately 1000-1500 LADs have been identified in mouse and human cells and collectively cover ~30% of the genome [6]. By sequestering chromatin to the repressive nuclear compartment, nuclear lamina plays an important role in genome organization.

At a sub-megabase level, chromosomes are partitioned into topologically associated domains (TADs) preferentially forming intra-domain contacts via looping and are relatively insulated from neighboring TADs [15, 16]. TADs are well conserved amongst different cell types and across species, but reorganization of some TADs has been reported during lineage specification, and disruption of TAD boundaries can cause gene mis-expression and disease [17-19]. The CCCTC-binding factor (CTCF) colocalizes with cohesin genome-wide, and these factors are enriched at the boundaries of TADs[15, 20] and LADs[13]. CTCF and cohesin are the two major architectural proteins required for folding chromatin into loops, as acute depletion of CTCF, cohesin or its loading factor NIPBL results in a major reduction in loops and TADs [21-24]. In contrast to their role in TAD formation, however, these factors are not required for the genome segregation into the A and B compartments [21-24]. TAD formation has been proposed to facilitate enhancer-promoter contact within individual TADs by providing spatial restriction and proximity[25,26]. Therefore, genome-wide changes in gene expression was predicted upon loss of TADs. However, when TADs and loops are dramatically weakened by acute depletion of CTCF, cohesin or NIPBL, immediate transcriptional change remains unexpectedly modest, affecting a relatively low number of protein-coding genes [7, 21-23]. Furthermore, two recent reports studying chromatin structure (loops, domains and compartments) and specific gene expression during early embryonic development in *Drosophila* have shown compelling evidence that TADs and the compartments do not function as main determinants for gene activity[27, 28]. Thus, additional chromatin-binding proteins may exist which contribute to the formation of distinct types of 3D chromatin organization for gene regulation.

Besides CTCF/cohesin-mediated loops and the nuclear lamina, only few genome organizing proteins are known to date. Unlike nuclear lamina, ubiquitous expressed CTCF, cohesin, and YY1 that dynamically mediates smallest scale loops for enhancer-promoter contacts [7, 29], SATB1 is a cell-type specific nuclear architectural protein, which has also been reported to form chromatin looping at specific gene loci [30-32]. For cells in which SATB1 is expressed or induced, expression of hundreds to a thousand of protein-coding genes are regulated depending on SATB1 to either maintain specific cell properties or to acquire new phenotypes [2]. While CTCF consensus sequence is divergent, SATB1 binds specifically to genomic sequences strongly potentiated for unwinding under negative superhelical stress, called Base Unpairing Regions (BURs) of ~200bp in length [33, 34]. BURs are at least 60% AT-rich, but importantly, not all AT-rich sequences are BURs. BURs contain a cluster of short sequence segments (25-50 bp), displaying an ATC (or ATG) sequence context, where one strand of DNA of each segment contains exclusively either G or C [33, 35]. A nuclear protein, SATB1, was identified by virtue of its specific binding to double-stranded BURs, recognizing their altered phosphate backbone structure [35]. SATB1 expression is mostly restricted to specific adult progenitor cells, including immature thymocytes [36], ameloblasts [37], and basal layers of epidermis [38]. It is also expressed in aggressive epithelial cancers [2] as well as in postnatal neurons in cortex and amygdala [39, 40]. SATB1 confers unique nuclear architectural distribution, such as an interconnected cage-like distribution in thymocytes[41] and a fine spider web-like distribution in neurons[39]. We refer to such SATB1 distribution as SATB1 nuclear architecture. An early work on SATB1 function reported that upon T helper 2 cell activation, multiple BURs are tethered to the SATB1 nuclear architecture to form dense chromatin looping over the 200kb-cytokine gene cluster, bringing multiple distal genes into close proximity. SATB1 then recruits chromatin remodeling complexes, epigenetic factors and transcription factors to this region to co-induce interleukin genes [30]. Through such activities, SATB1 enables cells to change their phenotypes, including T cell activation [30], cell development of multiple lineages [32, 36-39, 42-47], and cancer metastasis [48, 49]. Once expressed in breast cancer, SATB1 regulates ~1000 genes known to promote tumor progression [48, 50-52] and its expression has high prognostic significance [48, 50]. Such roles of SATB1 in cancer are now validated and extended to at least 20 different epithelial cancer types [2, 49, 53-57].

Although BURs are known as specific binding targets of SATB1, mysteriously the genome-wide SATB1 binding profiles obtained by chromatin immunoprecipitation-deep sequencing (ChIP-seq) did not detect SATB1 interaction with BURs. Instead, SATB1 mapped mainly to enhancers [32, 44] and to a region that would become an active super-enhancer upon differentiation [45]. A recent study also describes SATB1 interaction with enhancers in T cells (Zelenka et al, doi: https://doi.org/10.1101/2021.07.09.451769). However, given that SATB1 functions as a protein platform to recruit multiple gene regulatory proteins, standard ChIP-seq approaches cannot differentiate between genomic sites directly bound by SATB1 from secondary interactions mediated through its interacting proteins. Thus, whether a SATB1-BUR-mediated genome organization exists remained elusive.

Here we developed a modified ChIP-seq platform, urea-ChIP-seq, to stringently purify genomic DNA with only its directly bound proteins to identify true SATB1-bound sites. Using a BUR reference map containing all potential BURs in the entire mouse genome, we found that SATB1 binds BURs with high specificity, and that SATB1-bound BURs are mutually exclusive with CTCF-bound sites. Consistently, SATB1 does not impact TAD formation. We found that BURs largely map to LADs, and the fraction of BURs associated with the SATB1 protein network in the nuclear interior is dependent on cell type. Our results reveal functional nuclear architecture organization that is mediated by the interaction between BUR genomic landmarks and the cell type specific gene regulator, SATB1. These interactions are independent of TADs, and the degree of interaction between the genome-wide distributed BURs and SATB1 is cell type dependent.

Results

Development of urea ChIP-seq method to detect genome-wide SATB1-binding profile

We hypothesized that genome-wide interaction of SATB1 with BURs could not be detected to date by widely used ChIP-seq (standard ChIP-seq) because unlike many nuclear proteins (e.g. transcription and chromatin modifying factors), SATB1 and its direct interacting BURs are embedded in the "insoluble nuclear substructures". In fact, SATB1 resides in nuclear substructures exhibiting resistance to high-salt extraction [58]. Consistent with this, after extraction of nuclei with buffer containing 2M NaCl that removed most of the proteins from DNA, individually cloned SATB1-bound BURs remained anchored to the residual SATB1 nuclear architecture found in the interior of nuclei [41]. By contrast, in high-salt extracted *Satb1*$^{-/-}$ (SATB1-KO) thymocyte nuclei, these BURs lost their anchored sites and were exclusively found in the distended DNA halos that spread around the residual nuclei, which is illustrated (Fig. 1A) [41].

To unveil BURs as SATB1' primary targets at the genome-wide scale, we devised a modified urea ChIP-seq method and found that the vast majority of SATB1-binding sites are BURs in the genome. The urea ChIP-seq protocols first purifies the crosslinked chromatin under a highly-stringent condition and then use the purified chromatin fragments for ChIP-seq. For this purpose, we subjected lysed crosslinked cells to the 8M urea ultracentrifugation to remove all loosely-bound and free proteins from crosslinked genomic DNA to isolate the pure genomic DNA fiber retaining only directly-bound crosslinked proteins. Thus, these genomic regions can be studied solely based on their direct protein binding status. The chromatin purification method by urea centrifugation alone [59] was not sufficient to obtain reproducible genome-wide DNA binding profiles by ChIP-seq, however. It was essential to further optimize the method to quantitatively solubilize the unsheared crosslinked genomic DNA fiber that sedimented at the bottom of the tube as well as to maintain intact chromatin-protein interactions during fragmentation. We optimized this step and established the modified urea ChIP-seq method as illustrated (Fig. 1B). The critical steps that are different between urea and standard ChIP-seq are highlighted.

Urea ChIP-seq detects BURs as direct binding targets of SATB1 *in vivo.*

To investigate the chromatin organization by SATB1 on a genome-wide scale, it was essential to validate SATB1-bound sites as BURs. We used a BUR reference map which contains all potential BURs in the entire mouse genome [59]. BURs do not have a primary sequence consensus for an easy identification and no short motif within AT-rich sequences can serve as reliable markers. BURs used to be empirically determined by placing them in supercoiled DNA to test their reactivity to unpaired DNA-specific chloroacetaldehyde [33, 60]. Because SATB1 binds BURs specifically, we used SATB1 as a biological detecting tool for identifying BURs to map all potential BURs by capturing purified genomic DNA fragments containing BURs *in vitro*, followed by deep sequencing [59]. Approximately 240,000 BURs have been detected reproducibly ($q<0.01$) as the BUR references for both mouse and human genomes.

Using the BUR maps as a validating tool, urea ChIP-seq protocol was extensively optimized to detect direct SATB1 target regions quantitatively and reproducibly in mammalian cells. Among SATB1 urea ChIP-seq peaks, an average of 77% of two replicas coincided with BURs (Fig. 1C). Strikingly, urea ChIP-seq produced the SATB1 binding profiles that are entirely different from those produced by the standard ChIP-seq method (Fig. 1D). In mouse thymocytes, only 0.9% of peaks in urea SATB1 ChIP-seq were co-mapped with those in standard SATB1 ChIP-seq (Table S1). In a wide view covering a 4.48Mb region in chromosome 17 (Fig. 1D, top), a gene-rich region was found in the valley of BUR-enriched domains (track 3) and this region was heavily bound by SATB1 according to standard ChIP-seq (Fig. 1D, track 2). By contrast, by urea ChIP-seq the domains surrounding the gene rich region, heavily enriched in BUR reference sites (Fig. 1D, track 3) were specifically bound to SATB1 (Fig. 1D, track 5). In a zoom-in view, covering a 430kb region, SATB1 sites identified from urea ChIP-seq precisely coincided with a subset of BUR peaks and excluded the standard SATB1 ChIP-seq peaks. SATB1-binding sites from urea ChIP-seq avoided CTCF sites, H3K27ac-marked sites and DNase 1 hypersensitive sites (DNase 1 HS), while those from standard ChIP-seq peaks largely co-mapped to these sites (Fig. 1D, bottom). These results show that SATB1-binding profiles obtained by standard and urea ChIP-seq methods are mutually exclusive and display complementary distributions. Our results show that indeed BURs are direct targets of SATB1 *in vivo*, serving as important genomic marks for SATB1-mediated chromatin looping at the genome-wide scale. The unbiased feature of urea ChIP-seq, which solely captures DNA interaction with its direct binding proteins, without the influence of protein-DNA complexes or the original chromatin state, was able to unmask the normally hidden BUR-mediated chromatin organization directly regulated by SATB1.

CTCF-binding profiles remain unchanged by urea ChIP-seq

To investigate the potential cause for the discrepancy in the SATB1-binding profiles, we examined whether the urea ChIP-seq method could be reliably applied to other factors. The CTCF-binding profile generated by urea ChIP-seq (Fig. 2, track 2) was very similar to its profile obtained by standard ChIP-seq from Encode [61] (Fig. 2, tracks 4 and 5). This shows that the urea ChIP-seq method is not only uniquely designed to detect the direct binding sites of SATB1, but it can be applied for detecting the binding sites for CTCF identified by Encode

obtained by standard ChIP-seq. This was further confirmed by other nuclear factors (e.g. JARID2, SUZ12 and modified histone H3) (data not shown). Among standard SATB1 ChIP-seq peaks, 37% of these peaks (average of two replicas) intersected with CTCF peaks, whereas only 0.8% of urea SATB1 ChiP-seq peaks conincided with CTCF peaks (average of two replicas) (Fig. 2B), confirming the ucsc browser views (Fig. 1D and 2A). These results indicate that urea ChIP-seq produces the contrasting DNA-binding profile uniquely for SATB1 and not for other non-BUR-binding proteins, compared to their profiles generated by standard ChIP-seq.

SATB1 ablation has no effects on CTCF-binding, TADs or compartments

We addressed if CTCF-binding to genomic DNA depends on SATB1. The CTCF-binding profiles were generated by urea ChIP-seq in *Satb1*$^{-/-}$ (SATB1-KO) and *Satb1*$^{+/+}$ (SATB1-WT) thymocytes. These profiles showed no major changes (Fig. 2A, tracks 2 and 3), indicating that CTCF-binding to genomic DNA is SATB1 independent. Therefore, besides their mutually exclusive genomic distribution, this result provides further evidence supporting that CTCF and SATB1 are independent factors. The loop extrusion model postulates that cohesin functions in loop extrusion until it encounters CTCF bound to its target sites in a convergent orientation [62, 63]. To confirm that SATB1 does not use cohesin for chromatin looping, we examined their interactions. Cohesin, a ring-shaped multiple-protein complex, physically interact with CTCF as a stable protein complex *in vitro* [64]. We used urea ChIP-Western assay to study SATB1 interaction with either CTCF or cohesin on chromatin. Urea-purified chromatin fragments immunoprecipitated with anti-SATB1 antibody (SATB1-bound urea ChIP fragments) contain SATB1, detected by Western-blot assay of the formaldehyde decrosslinked urea ChIP sample. However, neither CTCF nor RAD21 (a cohesin complex component) was detected in SATB1 urea-ChIP samples. Reciprocal experiments using CTCF and RAD21-bound urea ChIP samples confirmed lack of SATB1 co-binding to these chromatin fragments (Fig. 2C, fig S1A). Therefore, these results strongly suggest that SATB1 mediated chromatin organization is independent from cohesin or CTCF.

Given that CTCF deletion causes most of TADs to disappear, we next examined whether SATB1 depletion has any effects on TADs formation. We compared Hi-C results at 100kb resolution between thymocytes isolated from wild-type (Satb1$^{+/+}$:*Cd4-cre*) and SATB1-deficient Satb1$^{F/F}$:*Cd4-cre*) mice. This study revealed indistinguishable TAD profiles between the two thymocytes (Fig. 2D). Neither TAD number nor TAD sizes were changed by SATB1 depletion from thymocytes (fig. S1). TAD interactivity (the sum of Hi-C chromatin contacts inside the TAD) also remained mostly unchanged (Fig. 2E). These results indicate that CTCF and SATB1 have distinct functions in genome organization. Similar to the results from the CTCF depletion study [21], SATB1 depletion also did not alter segregation of active and inactive chromatin compartmentalization (Fig. 2F). This result may be unexpected considering changes in gene expression by SATB1 depletion. However, TADs and compartments detected by Hi-C appear to be quite stable as these domains remain largely unaltered across tissues, despite clear changes in tissue-specific chromatin state and gene expression during *Drosophila*

development [28]. Therefore, the present example with SATB1 depletion is also in agreement with the notion that domains detected by Hi-C appear are not associated with changes in gene activity.

Non-specific peaks appear in *Satb1*$^{-/-}$ thymocytes by standard SATB1 ChIP-seq

An important question is whether SATB1-binding profiles obtained by standard ChIP-seq is strictly SATB1 dependent, thus reflect real SATB1 association. To address this question, we performed standard ChIP-seq on *Satb1*$^{-/-}$ (KO) thymocytes using two different anti-SATB1 antibodies (1583 and ab109112, both of which are validated for immunoprecipitation). Both antibodies generated similar peaks in KO thymocytes (KO peaks), and a strong tendency was detected for standard ChIP-seq to produce peaks in open chromatin regions marked by DNase 1 HS in KO thymocytes, coinciding with peaks originally detected in WT thymocytes (fig. 2A and 2B). This is in full agreement with an earlier report that ChIP-seq generate false-positive peaks (or phantom peaks) in the region of active promoters [65]. Some of these peaks were also detected in input chromatin by standard ChIP-seq, with varying peak intensity depending on regions. By contrast, input chromatin was found to be generally flat for urea ChIP-seq (fig. S2A and S2B, top, tracks 3, 4, 8 and 9). These results are highlighted in two randomly chosen regions, a 3.1 Mb region in chromosome 17 and a 3.6 Mb in X chromosome each containing a gene-rich region surrounded by gene-poor regions (fig. S2A and S2B). Thus, validation using knockout cells is critical to identify real SATB1-binding sites by the standard ChIP-seq method.

Many SATB1-binding sites identified by standard ChIP-seq were detected only in WT thymocytes and absent in KO thymocytes. These sites are associated with SATB1, most likely through indirect binding, and might play significant biological roles. We focused on three individual gene loci, i. e. *Rag1* and *Rag2* genes [32], *Foxp3* [45], and *Runx3* [44], for which specific SATB1-binding sites by standard ChIP-seq have been reported (Fig. 3). These genes are regulated by SATB1 in thymocytes. SATB1 functions as a regulator of a *Rag* locus through binding to the anti-silence element (ASE), which counteracts the activity of an intergenic silencer and promotes expression of Rag 1 and Rag 2 [32]. During development of Foxp3^{+}-regulatory T cells (Treg), SATB1 expressed in Treg precursor cells plays an essential role in establishing a Treg cell–specific super-enhancer at the *Foxp3* locus in Treg cells to activate *Foxp3* expression [45]. Thus, SATB1 deficiency in precursor cells (*Satb1*^{FF}:*Cd4*-cre mice), but not *Satb1*^{FlF}:*Foxp3*-cre mice, impairs Treg super-enhancer activation and expression of *Foxp3* [45]. This causes impaired Treg cell development, leading to autoimmune disease [44-46, 66, 67]. SATB1 is also essential for generation of CD4^{+} and CD8^{+} T cells and NKT cells [44]. In response to T cell receptor signaling by self-peptide/MHC I and II, SATB1 regulates enhancers function in gene loci encoding lineage-specifying factors (e.g. *Runx3*) and directs lineage-specific transcriptional programs in the thymus [44]. In all three loci, we found SATB1-bound sites to be strictly SATB1 dependent and detected only in WT (Fig. 3, track 3) but not in SATB1 KO-thymocytes (Fig. 3, tracks 7 and 8 for *Rag/Rag2* and *Runx3*, and tracks 9 and 10 for *Foxp3*). In these regions, SATB1-bound BURs, detected by urea ChIP-seq, was found infrequently with a single major site appeared within 30-50kb regions. Such SATB1-bound BUR sites did not co-map to sites identified by standard ChIP-seq containing enhancers

marked by H3K27ac and H3Kme1 modifications. Of note, SATB1-bound BUR sites were found to reside within 3 kb from the nearest SATB1-bound enhancers for the *Rag1/Rag2* and *Runx3* loci and from the super-enhancer once it is fully established in *Foxp3* loci in Treg cells. These results show that at least for these three loci, the SATB1-binding sites by standard ChIP-seq were clearly SATB1 dependent.

The SATB1 binding to BURs is cell-type dependent

We next examined whether SATB1 binds BURs in a cell-type dependent manner. Three different types of primary mouse cells, thymocytes, frontal cortical neurons (brain) and embryonic skin (basal layer cells), were studied addressing whether different subsets of BURs are targeted by SATB1, depending on cell type. UCSC browser images are compared for two selected regions, ~2.4 Mb region in chromosome 6 containing the *Cd4* locus expressed in thymocytes (Fig. 4A), and a 406kb region in chromosome 5 containing *Gabra2* and *Gabrg1* loci expressed in the mammalian brain (Fig. 4B). Regardless of the difference in cell-type specific gene expression, BURs in these gene loci appeared to be most frequently bound by SATB1 in brain compared to thymocytes, and they are least frequently bound by SATB1 in skin. Whereas SATB1-bound BURs show cell-type dependency, CTCF binding sites that largely coinciding with DNase 1 HS remain essentially unchanged among these three types of cells (Fig. 4A).

We analyzed whether BURs are targeted by SATB1 depending on cell type by peak intersection studies. Consistent with the cell-type specific function of SATB1, we observed cell-type dependent difference in the frequency of BURs targeted by SATB1. For all three types of cells, high percentages of their urea ChIP-seq peaks (close to 80%, Table S1) represent BURs, revealed by intersection studies comparing against the BUR reference map, using Deeptools [68] (Fig. 4C, Table S1). Despite similar total aligned sequence reads for urea ChIP-seq, the number of SATB1-bound BUR peaks (referring to the BUR reference map) greatly differed depending on cell type (Table S2). Among 238,380 BURs mapped ($q<0.01$), percent of these BURs intersected with urea SATB1 ChIP-seq peaks in skin, thymocytes, and brain are 3.6%, 13.5% and 35.0% (based on average of two replicas/cell type), respectively (Fig. 4D and Table S1). These results indicate that frequency of BURs targeted by SATB1 *in vivo* can vary more than 10-fold depending on cell type. Interestingly, the majority of BURs targeted by SATB1 in skin (83.9%) are bound to SATB1 in thymocytes, and SATB1-bound BURs in thymocytes (91.6%) were bound to SATB1 in brain. Therefore, rather than completely different sets of BURs are bound by SATB1 in a cell-type dependent manner, different number of BURs in a common group of BURs appear to be targeted by SATB1 depending on cell type. The exceptionally high frequency of BURs targeted by SATB1 in cortical neurons may reflect specialized function of neurons that might require highly dynamic transcriptional regulation to enable rapid response to a wide variety of environmental stimuli. There is growing evidence supporting the critical roles of chromatin topology and nuclear architecture in neuronal response to external stimuli to establish precise expression program associated with cognition [69]. In fact, SATB1 has roles in neuronal

connectivity[39, 70]. How differential BUR utility (or BUR targeting frequency) is functionally linked to the ultimate cell type-specific gene regulation is an exciting question to be addressed in future research.

BURs are key sequence elements of LADs

To gain an insight into the biological significance of SATB1-BUR interactions in chromatin architecture, we hypothesized that the widely-spread BURs in the genome might be associated with some known features of nuclear structure. We superimposed the BUR reference map with LADs mapped by Lamin B1-adenine methyltransferase identification(Dam ID) [71] for a zoom-out region (47 Mb) in chromosome 5 covering the *Gabrg1* and *Gabra2* loci in Fig. 4B. Strikingly, we found that the main cluster of BUR distribution in this region largely overlapped with LADs (fig. S3). We confirmed this overlap examining a larger view covering 47.4 Mb in chrosome 3. Indeed, BUR clusters were found to correspond to both Lamin B1-ChIP-seq [72] and Lamin B1-DamID sites [71] (Fig. 5A). LADs are known to be enriched in AT-rich sequences [73]. In fact, BURs are a discrete subgroup of AT-rich sequences, with the ATC sequence context [2]. Besides being generally AT-rich, whether LADs have any additional features remained unknown. BURs are revealed as an important sequence component to LADs, which are specifically targeted by BUR-binding proteins. Because the binding sites of CTCF are enriched at the border of LADs [6], and not overlapping with LADs, this explains why SATB1-bound BURs exclude CTCF-binding sites. The 47.4 Mb region shows that CTCF-binding sites, DNase 1 HS, and standard SATB1 ChIP-seq peaks (Fig. 5, tracks 10-12) all converge in the gene-rich valleys of DamID and Lamin B1 ChIP-seq (Fig. 5, tracks 1 and 4) clusters. The cell-type dependent targeting frequency of BURs is verified in this large region: SATB1-bound BURs appear to be most frequent in the brain, followed by thymocytes, and the least in skin (Fig. 5, tracks 7-9).

Although LADs form heterochromatin lining the nuclear envelope, some LADs dynamically contact the nuclear lamina[6]. Gene movement in and out of the nuclear periphery in a circadian rhythm has been reported [74]. Also, during differentiation of mouse cells, hundreds of genes display dynamic intranuclear spatial repositioning relative to the nuclear lamina [75]. A dramatic relocation of all LADs occurs in the retinal rod cells of nocturnal animals [76]. Recently, relocation of the *hunchback* gene in *Drosophila* neuroblasts to nuclear lamina requires polycomb proteins binding to a specific intronic element of the gene locus. This process is important for heritable gene silencing to regulate neuroblast competence [77]. Little information is available concerning mechanisms specifying nuclear interior destination of dynamic LADs. By immunostaining combined with DNA fluorescent in situ hybridization (immune-FISH), individually cloned SATB1-bound BURs in thymocytes are found to exclusively localized within the inner nuclear location, colocalized with the SATB1 nuclear architecture [41, 58]. Taken together with the current results, we propose a model (Fig. 5B) which illustrates that upon SATB1 expression, SATB1 provides counterforces by recruiting specific BURs from LADs onto the nuclear interior SATB1 architecture to setup the BUR-mediated genome organization regulated by SATB1 for cell-type dependent gene expression.

Discussion

To better understand whether and how 3D chromatin organization contributes in genome activity, it is important to investigate additional mechanisms of chromatin folding and functions of other key chromatin-binding proteins that might also play roles in chromatin architecture. Here we describe genome organization in which chromatin is folded by direct binding of BUR genomic landmarks to the SATB1 nuclear architecture. This chromatin folding mechanism is distinct from CTCF and cohesin-mediated chromatin folding that forms TADs. At least in our experimental system employed in this study, SATB1 is independent from CTCF and cohesin, based on several evidence: SATB1 and CTCF exhibits mutually exclusive binding profiles, neither CTCF nor cohesin associate with SATB1 on chromatin, and unlike CTCF and cohesin, SATB1 has no role in TAD formation. Unlike many chromatin-binding proteins described, SATB1 is a cell-type specific nuclear architectural protein that is essential for global reprogramming of transcription profiles required to induce changes in cell activities[2]. We show that SATB1 binding to BURs in the genome also depends on cell-type, suggesting that the 3D genome organization mediated by SATB1-BUR interactions is closely linked to gene expression.

SATB1 is equipped with BUR-binding domains [78-81] as well as domains responsible for protein:protein interactions, enabling SATB1 to bind BURs in the genome and also associate with chromatin remodeling factors, histone modifying and transcription factors [30, 41, 43, 48, 82-86]. This dual binding activity of SATB1 could provide a clue in explaining the contrasting DNA-binding profiles generated by distinct ChIP-seq protocols. SATB1-binding sites at enhancers identified by standard ChIP-seq are probably generated by SATB1 binding to these regulatory regions through indirect association with the large protein complexes that are assembled at the loci. Based on a previous locus-specific study, SATB1-bound BURs also make physical contact with T_H2-specific regulatory regions through chromatin looping within the 200kb-cytokine gene cluster in T_H2 cells [30]. Thus, it can be conceived that SATB1 directly binds BURs to provide a chromatin platform onto which a SATB1-regulatory network is assembled by recruiting distant loci through indirect association of SATB1 with regulatory factors bound to such loci. Other chromatin-binding factors such as YY1 could be a component of such regulatory network to promote enhancer-promoter interactions. Besides enhancers associated with specific gene loci critical for T cell development [32, 44, 45], SATB1 is also reported to make a protein complex with POU-homeodomain transcription factor (Pit1) and β-catenin to tether Pit 1-target enhancers to a matrin-3-rich nuclear substructure to activate Pit1 regulated transcriptional program in pituitary glands [43]. Thus, mutli-layered of interaction is expected to take place for SATB1, including many BURs in the genome, association with regulatory nuclear factors, and regulatory loci, to constitute the SATB1-regulatory network. Such “3D interactomes” may connect a multitude of BURs to enhancers and promoters from many chromosomes to “poise” regulation of a large number of genes. Our ongoing research using urea 4C-seq also shows that strictly depending on SATB1 a single BUR site in breast cancer cells interacts, both *cis* and *trans*, with many loci, including DNase 1 HS and enhancers

(manuscript in preparation). To accurately and quantitatively analyze SATB1-mediated interactions genome-wide involving BURs, urea Hi-C might be useful in the future.

SATB1 is not the only BUR-binding protein. Although SATB1 confers the highest affinity to BURs, several nuclear factors (e.g. nucleolin, HMGI/Y, PARP1, Ku70/86 heterodimer) have also been verified to directly and specifically bind BURs *in vitro* [87-89]. Recently, we have identified an additional strong BUR-direct binding protein from mouse embryonic stem cells (mESCs) that is essential for pluripotency (manuscript submitted). Similar to SATB1, this nuclear factor also binds specifically to BURs in mESCs, and identification of BURs as its direct targets in mESCs requires urea ChIP-seq as well. Despite the identical BUR binding specificity of SATB1 and this stem cell factor, these two proteins confer entirely different functions in distinct cell types. Therefore, we propose that BURs in the genome are the core genomic marks that can be selectively targeted by distinct BUR-binding proteins to form unique 3D chromatin regulatory network depending on the BUR-binding protein for specific cell functions. Future research on BUR-mediated genome organization regulated by distinct BUR-binding proteins will provide the opportunity to learn precise mechanisms as to how 3D chromatin architecture underlies major changes in gene expression and cell phenotypes.

Materials and Methods.

Primary cells

All animal studies were performed under Institutional Animal Care & Use Program at University of California San Francisco for collecting brain and thymus and the Boston University for collecting skin. Thymi were isolated from 2 weeks-old *Satb1*$^{+/+}$ and *Satb1*$^{-/-}$ mice, as well as from *Satb1*$^{+/+}$:*Cd4*-cre and *Satb1* $^{F/F}$:*Cd4*-cre mice. Brain was isolated from 6-10 weeks-old *Satb1*$^{+/+}$ mice. Thymocytes were prepared from thymi and cortex was excised from brain for this study. Skin was isolated from postnatal day 0.5 of C57Bl/6 mice.

Antibodies

Antibodies used and for their purposes are the followings. CTCF for ChIP-seq: Active MOTIF (61311. lot# 34614003), CTCF for Western: Millipore (07-729; lot#-3429114), RAD21 for ChIP-seq and Western: ABclonal (A18850, lot# 3521809001) and SATB1 for ChIP-seq and Western: rabbit polyclonal 1583 prepared by Kohwi-Shigematsu's laboratory and abcam (ab109122, lot#GR137720-13).

Urea-purified chromatin immunoprecipitation

Thymocytes and single cell suspension prepared for cortex and skin were cross-linked with 1% formaldehyde for 10 min at room temperature, followed by addition of glycine at 0.125 M for 5 min to quench cross-linking reaction. These cells were washed with 1X PBS, suspended in 1% BSA-PBS and cells were precipitated in a tube. These

cells were re-suspended in lysis buffer (4% SDS, 50 mM Tris-HCl [pH 8.0], 10 mM EDTA, 100 mM NaCl) with Protease inhibitor (Roche) and RNase inhibitor. The resulting cell lysate was incubated for 30 min at room temperature and loaded on top of 8 M urea followed by ultracentrifugation in a Beckman TL-100 Ultracentrifuge for 4-5 hrs at 45K rpm (181,000g, g=relative centrifugation force or RCF; TLS55 rotor). Recovered chromatin pellet was sonicated to generate fragment size ranging 300–1,1,200bp using Branson sonifier (15%-20% power, 10 sec on and 30 sec off for 4-6 cycles) in sonication buffer (0.5% Na-laurylsarcosine, 0.1% DOC, 1mM EDTA, 0.5mM EGTA, 10mM TrisHCl, 100mM NaCl) with Protease inhibitor and RNase inhibitor. Note: urea-purified crosslinked chromatin is highly vulnerable to sonication. The sonicated chromatin with 1% Triton X-100 were dialyzed against dialysis buffer (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA, 5% glycerol), and dialyzed for 16 hrs at 4°C (this dialysis is optional). The chromatin sample was centrifuged at 15,000 g for 5 min at 4°C to remove a trace amount of residual insoluble components. After reverse crosslinking an aliquot of sonicated chromatin, shearing efficiency was examined by Bioanalyzer 2100 (Agilent Technologies Inc.). The reverse cross-linking of chromatin aliquot was done at 65°C overnight in buffer containing 1% SDS, 100 mM NaCl, 1 mM EDTA, 10mMTris. Samples were further treated by 100 µg/ml RNase A at 37°C for 1 hr, 200 µg/ml proteinase K at 60°C for more than 3 hrs, followed by DNA purification using QIAquick PCR purification kit (Qiagen). For chromatin-immunoprecipitation, 1 µg of an antibody of choice were added to 10 ~ 20 µg of sonicated solubilized chromatin in 350 µl in dilution buffer (1% Triton X-100, 0.01% SDS, 1 mM EDTA, 15 mM Tris-HCl, 100 mM NaCl) and incubated overnight at 4°C. At least three chromatin samples were prepared for ChIP-seq from independent experiments to confirm the results. The chromatin fragments bound to antibodies (after incubation overnight) were captured using 40 µl of protein-A or protein-G Dynabeads (Invitrogen; pre-washed with 0.1% BSA [Fraction V, Sigma), by 4 hrs incubation time at 4°C. Dynabeads bound to chromatin fragments were washed twice with Binding buffer (10 mM Tris, pH8.0, 1 mM EDTA, 1 mM EGTA, 0.5% Triton X-100, 0.1% DOC, 100 mM NaCl), twice with LiCl buffer (100 mM Tris [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, and 0.5% DOC). The tubes were changed and washed twice with 0.1M NaCl-TE buffer. The immunoprecipitated samples were eluted from Dynabeads by elution buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 100 mM NaCl and 1% SDS) at 65°C for 20 min. After elution, ChIP DNAs was reverse-crosslinked at 65°C overnight, followed by RNase-A treatment at 37°C for 1 hr and proteinase K treatments for 3 hrs at 60°C. The recovered DNA was further purified with a QIAquick PCR purification kit (Qiagen), with elution in 15µl of Qiagen Elution Buffer (alternatively, Phenol-chloroform extraction and ethanol precipitation). Briefly for library construction, ChIP DNA fragments were end-repaired and ligated to indexed sequencing adapters and amplified with 8-11 rounds of PCR, depending on the DNA amount. Using SPRI beads (Beckmancoulter), small fragments (<250bp) were removed from the amplified library. The size-selected library was sequenced to 50bp single-end read lengths on an Illumina HiSeq4000 (Vincent J. Coates Genomic Sequencing Laboratory, UC Berkeley). Analysis is described under Bioinformatic Analysis.

Urea Chromatin immunoprecipitation (urea ChIP)-Western analysis

Urea-purified chromatin prepared as described above (~10μg of chromatin-DNA) was Immunoprecipitated with indicated antibody to produce urea ChIP samples, and the chromatin fragments bound by the antibody was trapped by Protein-A beads. The anitibody-bound chromatin fragment-trapped beads were washed with binding buffer/washing buffer/NaCL-TE (total 6 times), and the chromatin fragments were released and isolated from beads (65°C for 20 min) in 100 μl of 1% SDS, 10 mM Tris, pH7.6, 1mM EDTA, 100mM NaCl solution. The isolated chromatin fragments were subjected to reverse-crosslinking by incubation at 65°C for additional 6 hrs. The released proteins were dissolved in the Laemmli buffer, boiled and subjected to Western blot assay.

Hi-C

Hi-C was performed as described [20]. Primary thymocyte samples from *Satb1*$^{+/+}$:*Cd4*-cre and *Satb1* $^{F/F}$:*Cd4*-cre mice were formaldehyde crosslinked at 2% at room temperature for 10 min. Two biological replica per sample were prepared for Hi-C. Hi-C libraries was prepared using the Arima HiC^{+} kit (Arima Genomics) exactly following the associated protocol. Per sample, 200-300 million pair-end reads were obtained by Illumina NovaSeq 6000.

Bioinformatics analysis:

Hi-C analysis

In-situ Hi-C datasets were analyzed using the HiC-bench platform [90]. The application of this method is described in detail [19]. Briefly, datatypes were aligned against the mouse reference genome (mm10) by bowtie2 (version 2.3.1) [91] with mostly default parameters (specific settings: --very-sensitive-local --local). For Hi-C, aligned reads were filtered by the GenomicTools[92] tools-hic filter command (integrated in HiC-bench). The absolute number of accepted intra-chromosomal read-pairs were >75 million reads for all samples, which were sufficient to call topologically associated domains (TADs). Interaction matrices for each chromosome separately were created by the HiC-bench platform at a 40kb resolution. Filtered read counts were normalized by “iterative correction and eigenvector decomposition” (ICE)[93]. To account for variances of read counts of more distant loci, distance normalization for each matrix was performed as recently described[94].

TADs were called using the algorithm developed within HiC-bench [90] setting the insulating window to 500kb. Principal Component Analysis (PCA) on Hi-C datasets was performed in R (prcomp, with scale=TRUE and center=TRUE) using the genome-wide Hi-C “ratio” insulation scores for 500kb windows, as defined in Lazaris et al. [90]. For visualization of Hi-C data, we generated HiC files using the Juicer [95] ‘pre’ tool with default parameters and visualized them using Juicebox. HiC-bench [90] was used to determine the A and B compartments. HiC-bench uses the housekeeping genes to identify compartments A.

urea ChIP-Seq analysis

Sequenced reads were quality-checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) prior to downstream analysis. Unaligned reads were mapped to the mouse reference genome (NCBI37/mm9) using Bowtie (v1.0.0) [96] allowing for three mismatches in the seed region and only unique matches. Candidate binding sites were called using MACS2 (v2.2.7) [97] using a q-value threshold of 0.01. Coverage tracks are presented in RPM (read pileups per million reads) and utilities from UCSC were used for creating browser tracks [98]. To determine intersection of peaks across ChIP-seq samples, deepTools [68] was used.

Acknowledgement

We thank Dr. Aristotelis Tsirigos, Ziyan Li, and Javier Rodriguez Hernaez at New York University School of Medicine for Hi-C and intersection analyses, and Dr. Anthony Schmitt from Arima Genomics for help with Hi-C experiments and Ms. Bao Ho for providing technical assistance. This project was supported by NIHR01ES023854 to TKS, NIH R01AR071727 to VAB and TKS, and NIH R01 AR047364 to C-MC and TKS.

Author contribution

TKS and YK conceptualized and designed this study. Data were generated by YK and MG, and checked by Y-CS, and TKS wrote the manuscript. Data were analyzed by TKS and YK, and bioinformatic analysis was done by TKS, HWR and Y-CS. *Satb1*$^{F/F}$:*Cd4*-cre thymocytes were prepared by IT for Hi-C, and VAB provided skin samples for ChIP-seq. IT, YK, Y-CS and SS critically reviewed and edited the manuscript. TKS, VAB and Y-CS contributed in funding acquisition.

Competing interests: The authors declare no competing interests

Data availability: All sequencing data for ChIP-seq and Hi-C in this study was uploaded to NCBI Gene Expression Omnibus (GEO) and is available under the accession GSE191146.

Figure legends

Figure 1. Identification of BURs as direct targets of SATB1 *in vivo* by urea ChIP-seq.

A) An illustration for a BUR anchored to the high-salt extraction resistant SATB1 nuclear architecture [41]. DNA-FISH signals (red dots) for a cloned SATB1-bound DNA are indicated in *Satb1*$^{+/+}$(WT) and *Satb1*$^{-/-}$(KO) thymocytes treated with 2M NaCl solution on slides to produce DNA halos around the nuclei (blue loops). B) A comparative overview of the urea ChIP-seq and standard ChIP-seq protocols. Steps with critical differences are marked by blue boxes. C) The percentage of urea and standard (std) SATB1 ChIP-seq peaks that intersect with the BUR reference map (see Table S1). D) Contrasting distributions of SATB1-binding profiles determined by urea ChIP-seq (Urea) compared to standard (Std) in mouse thymocytes (top). A zoom-in view of the blue-highlighted region is shown (bottom) The track 3 (top and bottom) represents all potential BUR sites serving as a reference for BUR sites. Two independent urea and standard ChIP-seq experiments show high reproducibility (tracks 1 and 2 and tracks 4 and 5). Tracks 6 and 7 [ENCODE from Bring Ren's Laboratory at the Ludwig Institute for Cancer Research (LICR) and tracks 8 and 9 (from the UW ENCODE group).

Figure 2. Identical CTCF-binding profiles are produced by urea and standard ChIP-seq, independent of SATB1.

A) Urea ChIP-seq (Urea) profiles are shown for CTCF in *Satb1*$^{+/+}$ (WT) thymocytes and *Satb1*$^{-/-}$ (KO) thymocytes (track 2 and 3), and for SATB1 in WT thymocytes (track 1). These binding patterns were compared against CTCF-binding profiles generated by Standard ChIP-seq (Std) from Encode (tracks 4-6) for testis and thymocytes (from Bing Ren's laboratory at LICR), along with the DNase 1 HS profile generated by the UW ENCODE group B) The percentage of either urea or standard (std) SATB1 ChIP-seq peaks that intersect with CTCF ChIP-seq peaks. C) Urea-ChIP-Western studies for CTCF, RAD21, and SATB1, show no association of SATB1 with CTCF nor RAD21 on chromatin. D) Hi-C interaction heatmaps for *Satb1*$^{+/+}$:*Cd4*-cre (WT) and *Satb1* $^{F/F}$:*Cd4*-cre (KO) thymocytes show essentially identical patterns. E) Volcano plot of pair-wise comparisons of TAD activities (the sum of Hi-C chromatin contacts inside the TAD), analyzed by hic-ratio TAD caller, between WT and KO thymocyte samples shows well-conserved intra-TAD activity in KO thymocytes. F) The A and B compartments at 100kb windows analyzed by the HiC-bench platform show similar compartment distributions between WT and KO thymocytes, each shown in duplicates.

Figure 3. SATB1-bound sites identified by standard ChIP-seq at regulatory regions are SATB1 dependent.

SATB1-binding profiles in immature thymocytes determined by urea ChIP-seq (Urea) and standard ChIP-seq (Std) covering three loci at *Rag1/Rag2* (top) *Foxp3* (middle), and *Runx3* (bottom) are shown. For each of the

three loci, the standard SATB1 ChIP-seq generates peaks in *Satb1*$^{+/+}$ (WT) thymocytes (track 3), but not in *Satb1*$^{-/-}$ (KO) thymocytes (tracks 8 for *Rag 1/2* and *Runx 3*, and tracks 10 for *Foxp3*). These SATB1 peaks in *Rag1/Rag2* and *Runx3* loci largely overlap with enhancers (marked by H3K4me1 and H3K27ac). At the *Foxp3* locus in immature thymocytes SATB1 peaks (track 3) appear in a region marked by low levels of H3K27ac and H3Kme1 (tracks 4 and 5) where superenhancer is established upon differentiation into Treg cells (tracks 6 and 7)[45]. SATB1-bound peaks produced by urea ChIP-seq (track 2) are consistently located in distinct positions from those peaks obtained by standard ChIP-seq in these gene loci. Urea ChIP-seq did not generate peaks in KO thymocyes with anti-SATB1 antibody. H3K27ac and H3Kme1 data (tracks 4 and 5) are from Encode, generated by Bing Ren's laboratory at LICR, and data for Treg shown in (tracks 6 and 9) are from Kitagawa [45].

Figure 4. SATB1 binds to BURs in a cell-type dependent manner.

A) SATB1-binding profiles in a ~2.4 Mb region in chromosome 6 containing *Cd4* in a gene-rich region show mutually exclusive patterns between urea ChIP-seq (Urea) (track 2) and standard ChIP-seq (Std) (tracks 3-8). Urea ChIP-seq-derived SATB1 profiles in thymocytes (track 3), brain (track 4) and skin (track 5) show difference in peak numbers depending on cell type. This is in contrast to mostly identical CTCF patterns in these cell types (tracks 11-13). B) SATB1-binding profiles in a 406kb region in chromosome 5 containing *Gabra2* and *Gabrg1*, highly expressed in brain, also shows much increased BUR binding in brain, compared to thymocytes and skin similar to A). In this region, there is no SATB1-binding site detected by standard ChIP-seq (track 2), and CTCF binding sites are mostly absent (track 9). C) Percentages of SATB1 urea-ChIP-seq peaks in skin, thymocytes and brain intersecting with the BUR reference map. D) The percentage of all BURs in the BUR reference map intersected with urea SATB1 ChIP-seq peaks identified in skin, thymocytes and brain. DNase 1 HS data (from Encode by the UW group), CTCF (track 11 in Fig. 4A, and H3K27ac track 9 in Fig. 4B) (from Bing Ren's laboratory at LICR).

Figure 5. BURs largely co-map with LADs

A) The BUR distribution assessed by the BUR reference map (tracks 2 and 3 with different display modes) was compared against Lamin B1-DamID (track 1) [71] and Lamin B1 std ChIP-seq (track 4) [72] over a 47.4Mbp region in chromosome 3. The BUR distribution, shown in both styles (maximum and smoothened display), shows very similar overall distribution of Lamin B1-DamID and Lamin B1 ChIP-seq profiles. The cell-type specificity for targeting BURs by SATB1 is shown by the lowest (skin, track 9), medium (thymocytes, track 7) to highest (brain, track 8) BUR binding frequency. SATB1-binding regions determined by std ChIP-seq (Std) (track 10) largely coincide with DNase H1 (track 11) and CTCF-binding regions (track 12), but these regions avoid LADs and SATB1-binding sites derived from urea ChIP-seq (Urea) (tracks 7-9). DNase 1 HS profile (track 11) is from Encode,generated by UW group, Lamin B1 ChIP-seq (track 3) and H3K4me3 data (track 6) are from Handoko

et al. [72], and Lamin B1-DamID data are from [71] B) A model showing that some of the BURs (red dots), within LADs, move to SATB1 nuclear architecture in the interior of nuclei, once SATB1 is expressed in cells.

References

1. Politz, J.C., Scalzo, D. & Groudine, M. Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu Rev Cell Dev Biol* **29**, 241-270 (2013).
2. Kohwi-Shigematsu, T. et al. Genome organizing function of SATB1 in tumor progression. *Semin Cancer Biol* (2013).
3. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245-257 (2015).
4. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 772 (2016).
5. Cruz-Molina, S. et al. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell stem cell* **20**, 689-705 e689 (2017).
6. van Steensel, B. & Belmont, A.S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791 (2017).
7. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).
8. Bashkirova, E. & Lomvardas, S. Olfactory receptor genes make the case for inter-chromosomal interactions. *Curr Opin Genet Dev* **55**, 106-113 (2019).
9. Loubiere, V., Papadopoulos, G.L., Szabo, Q., Martinez, A.M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
10. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28-45 (2020).
11. Furlong, E.E.M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345 (2018).
12. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
13. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951 (2008).
14. Kind, J. et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-147 (2015).
15. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
16. Nora, E.P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
17. Lupianez, D.G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237 (2016).

18. Gonzalez-Sandoval, A. & Gasser, S.M. On TADs and LADs: Spatial Control Over Gene Expression. *Trends Genet* **32**, 485-495 (2016).
19. Kloetgen, A. et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet* **52**, 388-400 (2020).
20. Rao, S.S. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665-1680 (2014).
21. Nora, E.P. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).
22. Rao, S.S.P. et al. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017).
23. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
24. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* **36**, 3573-3599 (2017).
25. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).
26. Beagrie, R.A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524 (2017).
27. Espinola, S.M. et al. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nat Genet* **53**, 477-486 (2021).
28. Ing-Simmons, E. et al. Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nat Genet* **53**, 487-499 (2021).
29. Weintraub, A.S. et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e1528 (2017).
30. Cai, S., Lee, C.C. & Kohwi-Shigematsu, T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* **38**, 1278-1288 (2006).
31. Wang, L. et al. Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster. *PLoS One* **4**, e4629 (2009).
32. Hao, B. et al. An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J Exp Med* **212**, 809-824 (2015).
33. Kohwi-Shigematsu, T. & Kohwi, Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29**, 9551-9560. (1990).
34. Bode, J. et al. Biological significance of unwinding capability of nuclear matrix- associating DNAs. *Science* **255**, 195-197. (1992).
35. Dickinson, L.A., Joh, T., Kohwi, Y. & Kohwi-Shigematsu, T. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**, 631-645. (1992).

36. Alvarez, J.D. et al. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev* **14**, 521-535. (2000).
37. Zhang, Y. et al. SATB1 establishes ameloblast cell polarity and regulates directional amelogenin secretion for enamel formation. *BMC Biol* **17**, 104 (2019).
38. Fessing, M.Y. et al. p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* **194**, 825-839 (2011).
39. Balamotis, M.A. et al. Satb1 ablation alters temporal expression of immediate early genes and reduces dendritic spine density during postnatal brain development. *Mol Cell Biol* **32**, 333-347 (2012).
40. Huang, Y. et al. Distribution of Satb1 in the central nervous system of adult mice. *Neurosci Res* **71**, 12-21 (2011).
41. Cai, S., Han, H.J. & Kohwi-Shigematsu, T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* **34**, 42-51 (2003).
42. Satoh, Y. et al. The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. *Immunity* **38**, 1105-1115 (2013).
43. Skowronska-Krawczyk, D. et al. Required enhancer-matrin-3 network interactions for a homeodomain transcription program. *Nature* **514**, 257-261 (2014).
44. Kakugawa, K. et al. Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. *Cell Rep* **19**, 1176-1188 (2017).
45. Kitagawa, Y. et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat Immunol* **18**, 173-183 (2017).
46. Yasuda, K. et al. Satb1 regulates the effector program of encephalitogenic tissue Th17 cells in chronic inflammation. *Nat Commun* **10**, 549 (2019).
47. Goolam, M. & Zernicka-Goetz, M. The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. *Development* **144**, 1450-1461 (2017).
48. Han, H.J., Russo, J., Kohwi, Y. & Kohwi-Shigematsu, T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* **452**, 187-193 (2008).
49. Fromberg, A., Engeland, K. & Aigner, A. The Special AT-rich Sequence Binding Protein 1 (SATB1) and its role in solid tumors. *Cancer Lett* **417**, 96-111 (2018).
50. Ordinario, E. et al. ATM suppresses SATB1-induced malignant progression in breast epithelial cells. *PLoS One* **7**, e51786 (2012).
51. Liu, X. et al. Expression of SATB1 and HER2 in breast cancer and the correlations with clinicopathologic characteristics. *Diagn Pathol* **10**, 50 (2015).
52. Li, Q.Q. et al. Overexpression and involvement of special AT-rich sequence binding protein 1 in multidrug resistance in human breast carcinoma cells. *Cancer Sci* **101**, 80-86 (2010).

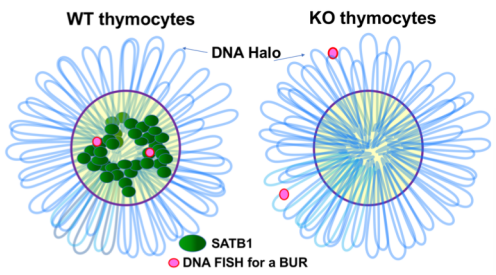
53. Hedner, C. et al. SATB1 is an independent prognostic factor in radically resected upper gastrointestinal tract adenocarcinoma. *Virchows Arch* **465**, 649-659 (2014).
54. Tu, W. et al. Upregulation of SATB1 promotes tumor growth and metastasis in liver cancer. *Liver Int* (2012).
55. Nodin, B., Hedner, C., Uhlen, M. & Jirstrom, K. Expression of the global regulator SATB1 is an independent factor of poor prognosis in high grade epithelial ovarian cancer. *J Ovarian Res* **5**, 24 (2012).
56. Mao, L. et al. SATB1 is overexpressed in metastatic prostate cancer and promotes prostate cancer cell growth and invasion. *J Transl Med* **11**, 111 (2013).
57. Mir, R., Pradhan, S.J., Patil, P., Mulherkar, R. & Galande, S. Wnt/beta-catenin signaling regulated SATB1 promotes colorectal cancer tumorigenesis and progression. *Oncogene* **35**, 1679-1691 (2016).
58. de Belle, I., Cai, S. & Kohwi-Shigematsu, T. The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J Cell Biol* **141**, 335-348. (1998).
59. Kohwi-Shigematsu, T. et al. SATB1-mediated functional packaging of chromatin into loops. *Methods* (2012).
60. Kohwi-Shigematsu, T. & Kohwi, Y. Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *Methods Enzymol* **212**, 155-180 (1992).
61. Davis, C.A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-D801 (2018).
62. Sanborn, A.L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456-6465 (2015).
63. Fudenberg, G. et al. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-2049 (2016).
64. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6** (2017).
65. Jain, D., Baldi, S., Zabel, A., Straub, T. & Becker, P.B. Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Res* **43**, 6959-6968 (2015).
66. Tanaka, Y. et al. SATB1 Conditional Knockout Results in Sjogren's Syndrome in Mice. *J Immunol* **199**, 4016-4022 (2017).
67. Kondo, M. et al. SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J Immunol* **196**, 563-572 (2016).
68. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

69. Watson, L.A. & Tsai, L.H. In the loop: how chromatin topology links genome structure to function in mechanisms underlying learning and memory. *Curr Opin Neurobiol* **43**, 48-55 (2017).
70. Close, J. et al. Satb1 is an activity-modulated transcription factor required for the terminal differentiation and connectivity of medial ganglionic eminence-derived cortical interneurons. *J Neurosci* **32**, 17690-17705 (2012).
71. Chen, S. et al. A Lamina-Associated Domain Border Governs Nuclear Lamina Interactions, Transcription, and Recombination of the Tcrb Locus. *Cell Rep* **25**, 1729-1740 e1726 (2018).
72. Handoko, L. et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**, 630-638 (2011).
73. Meuleman, W. et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**, 270-280 (2013).
74. Zhao, H. et al. PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Mol Cell* **59**, 984-997 (2015).
75. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603-613 (2010).
76. Solovei, I. et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356-368 (2009).
77. Lucas, T., Hafer, T.L., Zhang, H.G., Molotkova, N. & Kohwi, M. Discrete cis-acting element regulates developmentally timed gene-lamina relocation and neural progenitor competence in vivo. *Dev Cell* **56**, 2649-2663 e2646 (2021).
78. Galande, S., Dickinson, L.A., Mian, I.S., Sikorska, M. & Kohwi-Shigematsu, T. SATB1 cleavage by caspase 6 disrupts PDZ domain-mediated dimerization, causing detachment from chromatin early in T-cell apoptosis. *Mol Cell Biol* **21**, 5591-5604. (2001).
79. Dickinson, L.A., Dickinson, C.D. & Kohwi-Shigematsu, T. An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. *J Biol Chem* **272**, 11463-11470. (1997).
80. Nakagomi, K., Kohwi, Y., Dickinson, L.A. & Kohwi-Shigematsu, T. A novel DNA-binding motif in the nuclear matrix attachment DNA-binding protein SATB1. *Mol Cell Biol* **14**, 1852-1860. (1994).
81. Wang, Z. et al. Crystal structure of the ubiquitin-like domain-CUT repeat-like tandem of special AT-rich sequence binding protein 1 (SATB1) reveals a coordinating DNA-binding mechanism. *J Biol Chem* **289**, 27376-27385 (2014).
82. Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P. & Kohwi-Shigematsu, T. SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419**, 641-645 (2002).
83. Pavan Kumar, P. et al. Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. *Mol Cell* **22**, 231-243 (2006).

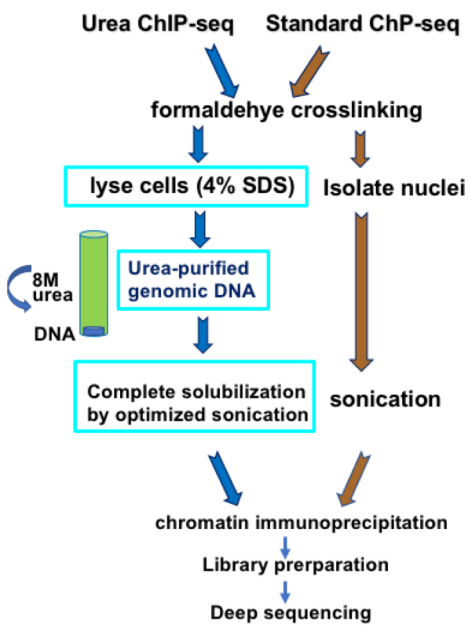
84. Notani, D. et al. Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. *PLoS Biol* **8**, e1000296 (2010).
85. Notani, D. et al. N-terminal PDZ-like domain of chromatin organizer SATB1 contributes towards its function as transcription regulator. *J Biosci* **36**, 461-469 (2011).
86. Stephen, T.L. et al. SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. *Immunity* **46**, 51-64 (2017).
87. Dickinson, L.A. & Kohwi-Shigematsu, T. Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Mol Cell Biol* **15**, 456-465. (1995).
88. Galande, S. & Kohwi-Shigematsu, T. Poly(ADP-ribose) polymerase and Ku autoantigen form a complex and synergistically bind to matrix attachment sequences. *J Biol Chem* **274**, 20521-20528. (1999).
89. Liu, W.M., Guerra-Vladusic, F.K., Kurakata, S., Lupu, R. & Kohwi-Shigematsu, T. HMG-I(Y) recognizes base-unpairing regions of matrix attachment sequences and its increased expression is directly linked to metastatic breast cancer phenotype. *Cancer Res* **59**, 5695-5703. (1999).
90. Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC Genomics* **18**, 22 (2017).
91. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
92. Tsirigos, A., Haiminen, N., Bilal, E. & Utro, F. GenomicTools: a computational platform for developing high-throughput analytics in genomics. *Bioinformatics* **28**, 282-283 (2012).
93. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003 (2012).
94. Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun* **9**, 542 (2018).
95. Durand, N.C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
96. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
97. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
98. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).

FIGURE 1

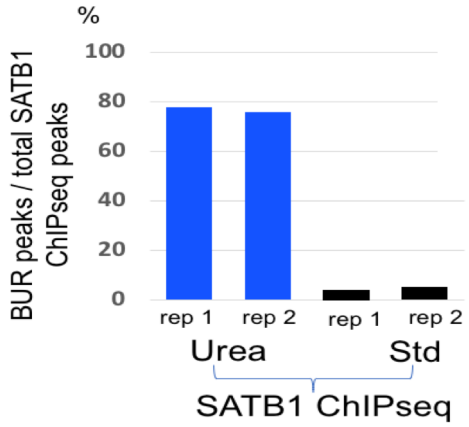
A



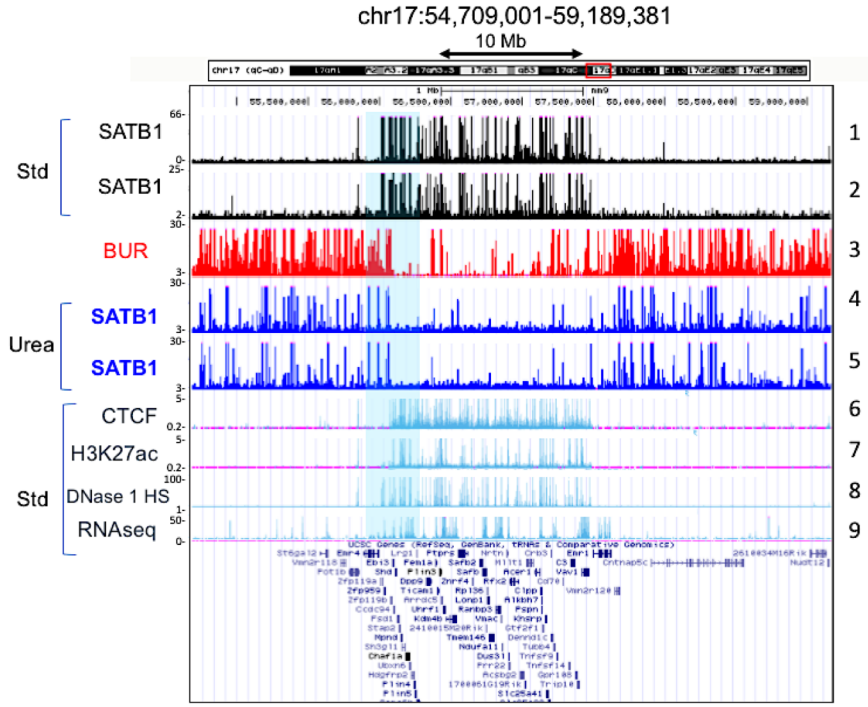
B



C



D



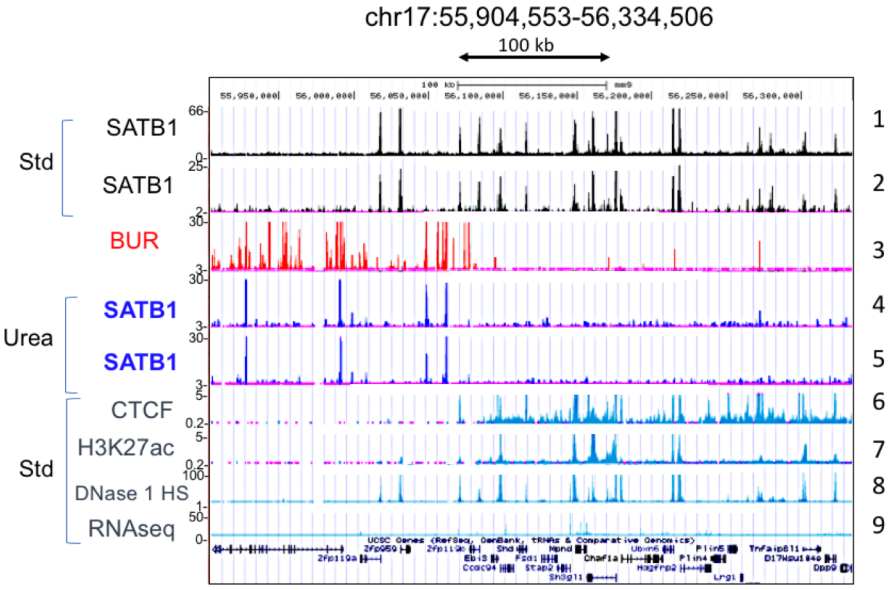


FIGURE 2

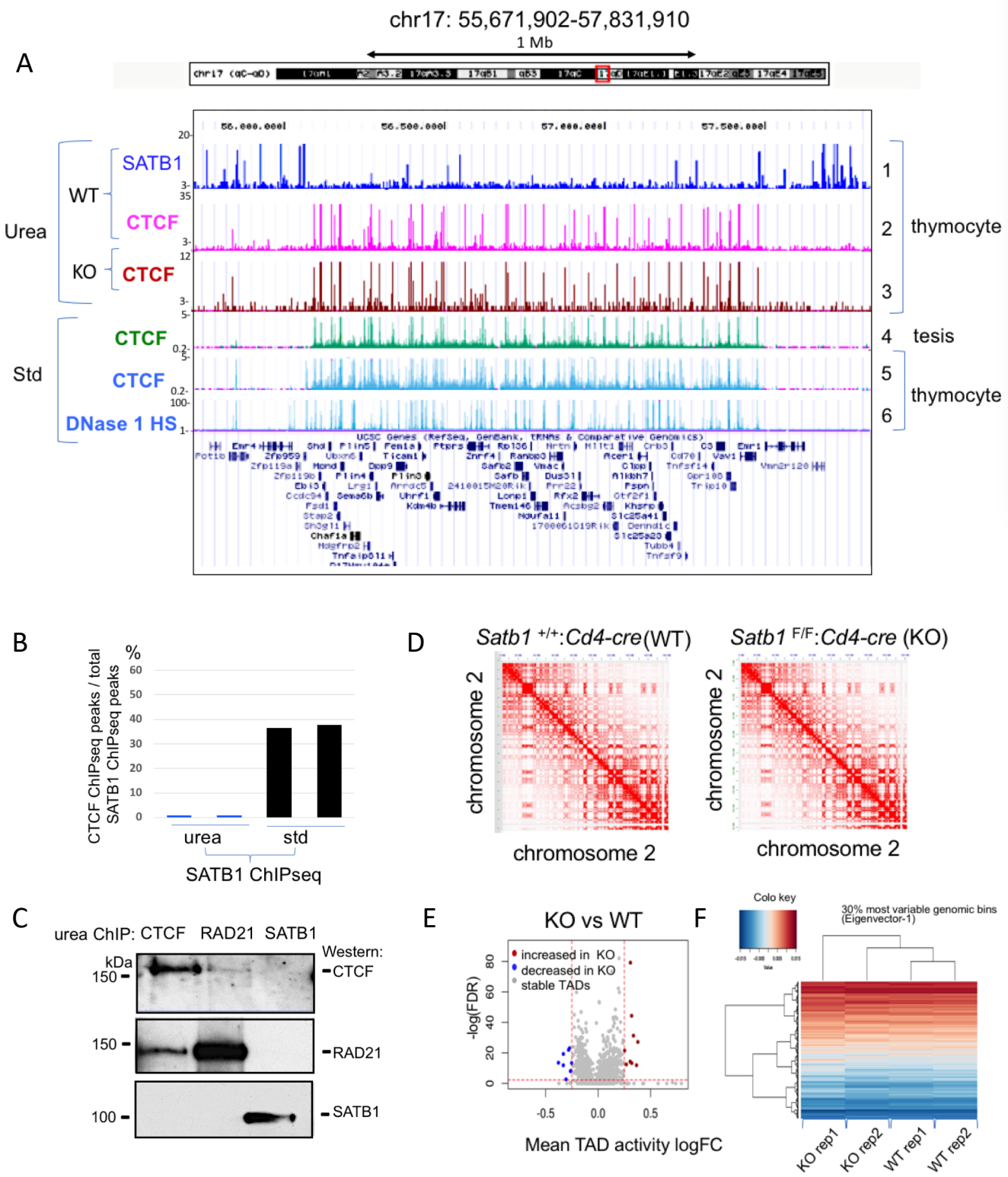
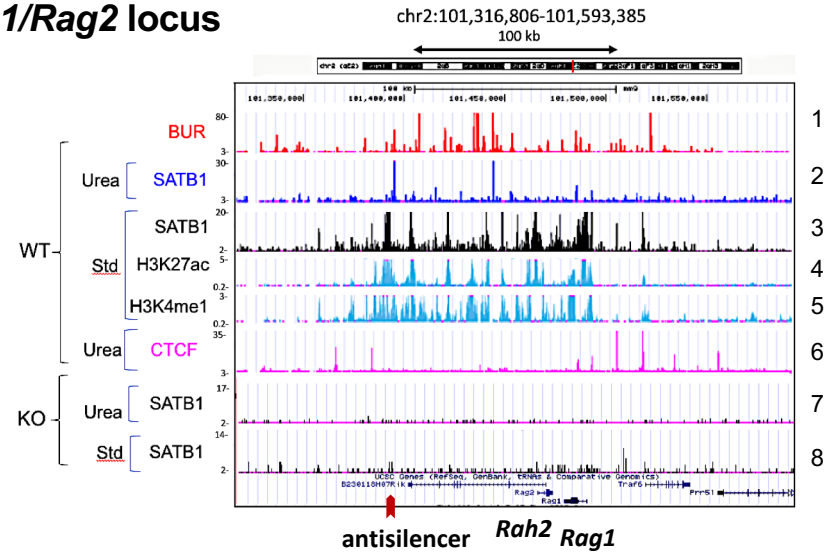
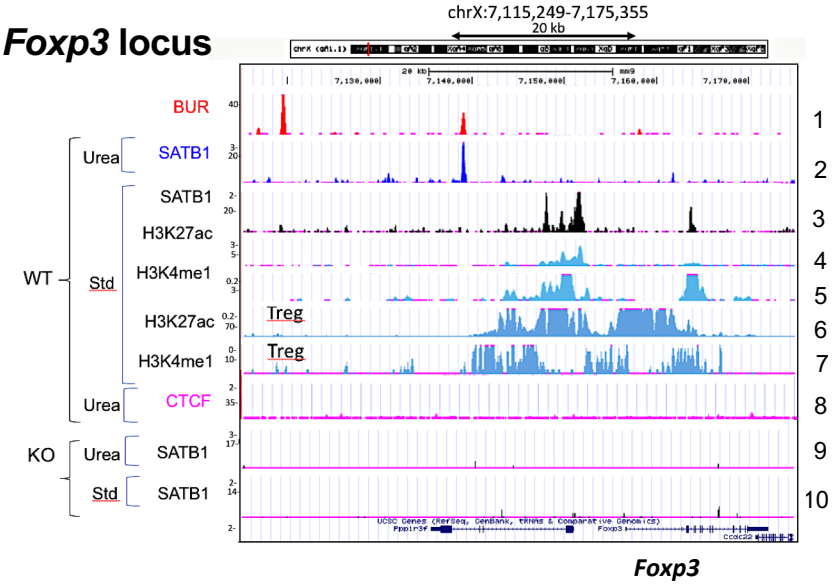

A
chr17: 55,671,902-57,831,910
1 Mb
Urea
WT
SATB1
CTCF
KO
CTCF
Std
CTCF
CTCF
DNase 1 HS
1
2
3
thymocyte
4
tesis
5
6
thymocyte
B
CTCF ChIPseq peaks / total SATB1 ChIPseq peaks
%
urea
std
SATB1 ChIPseq
C
urea ChIP: CTCF RAD21 SATB1
kDa
150
150
100
Western:
CTCF
RAD21
SATB1
D
Satb1 +/+:Cd4-cre(WT)
Satb1 F/F:Cd4-cre (KO)
chromosome 2
chromosome 2
chromosome 2
chromosome 2
E
KO vs WT
increased in KO
decreased in KO
stable TADs
-log(FDR)
Mean TAD activity logFC
F
Colo key
30% most variable genomic bins (Eigenvector-1)
KO rep1
KO rep2
WT rep1
WT rep2

FIGURE 3

Rag1/Rag2 locus



Foxp3 locus



Runx3 locus

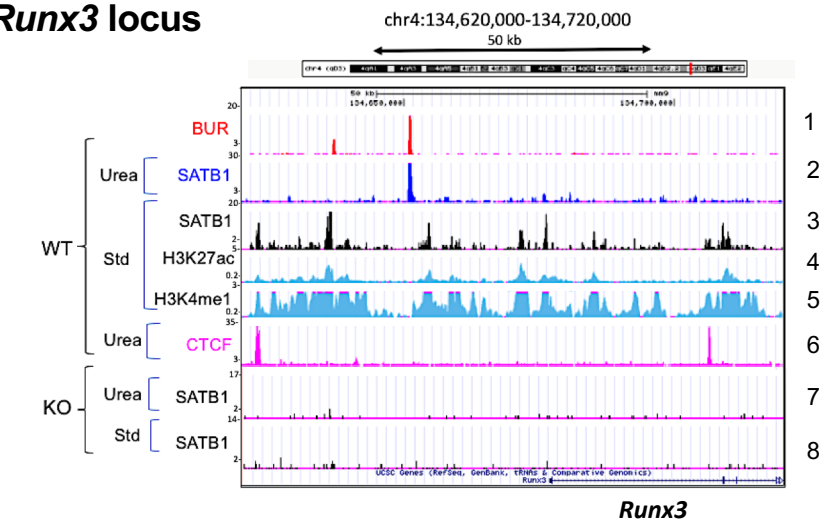


FIGURE 4

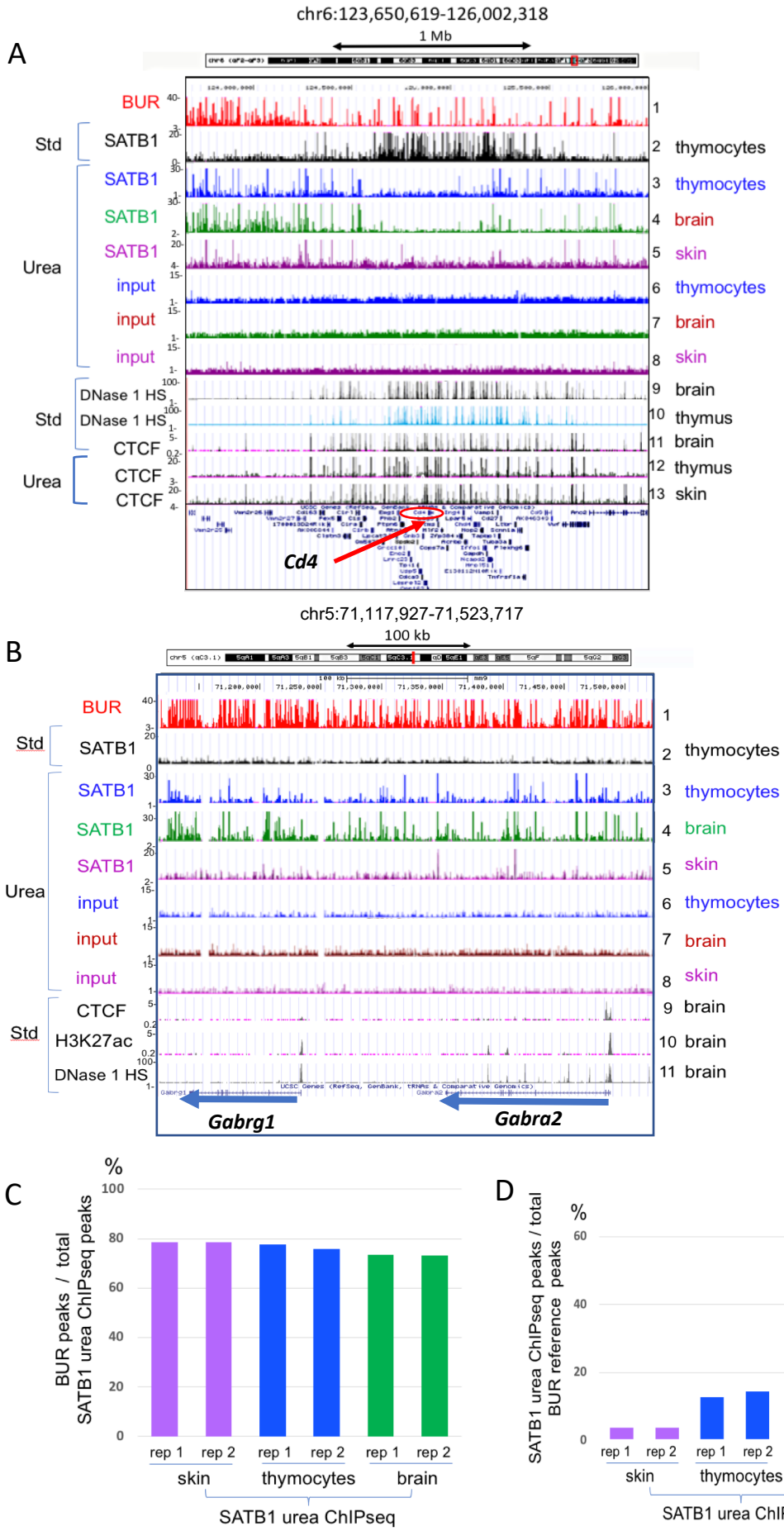
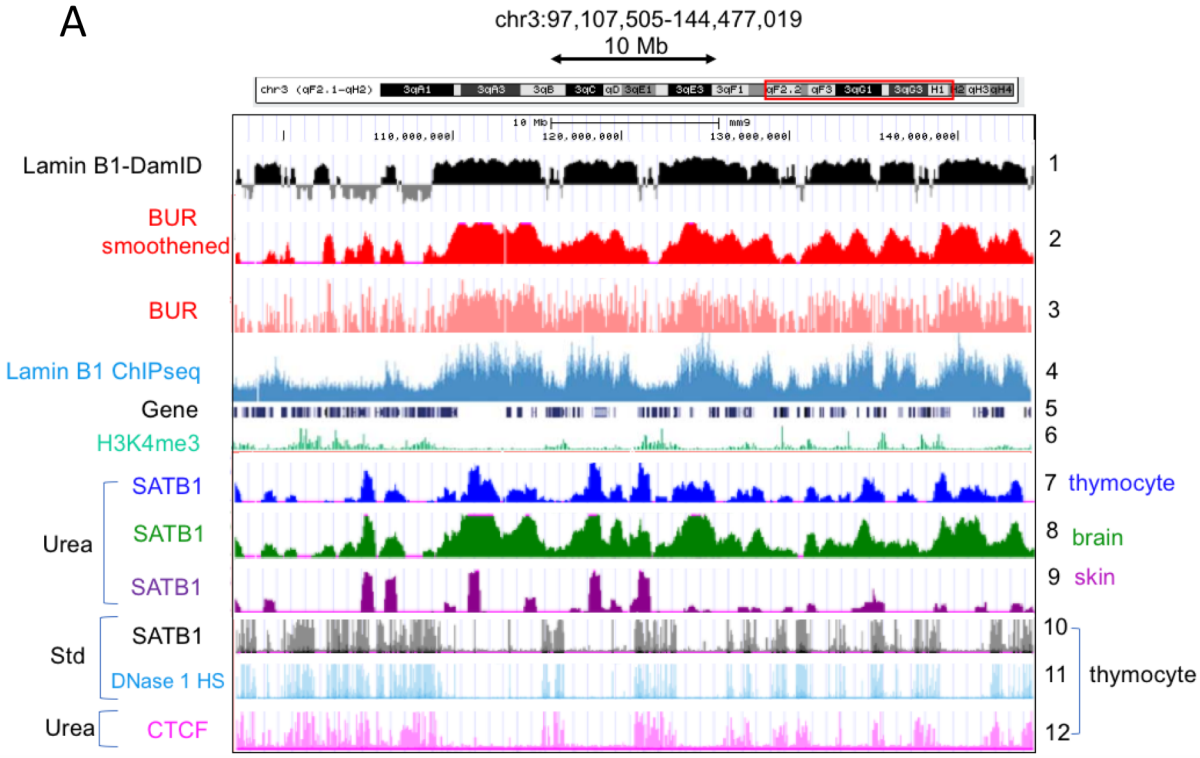
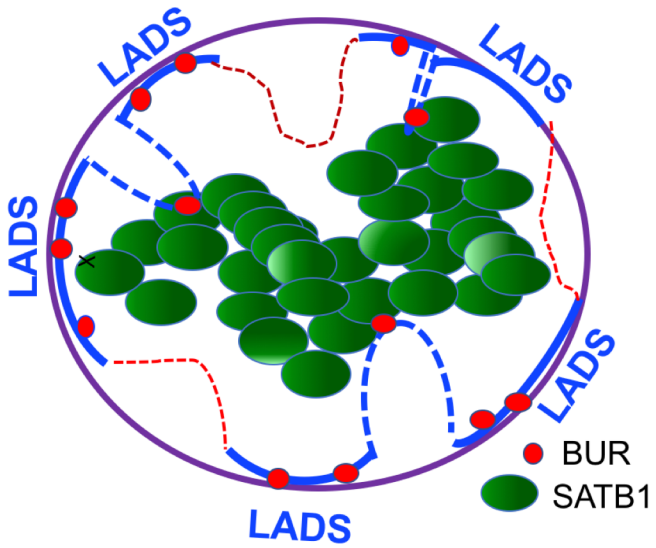

A
chr6:123,650,619-126,002,318
1 Mb
BUR
Std
SATB1
Urea
SATB1
SATB1
SATB1
input
input
input
Std
DNase 1 HS
DNase 1 HS
CTCF
Urea
CTCF
CTCF
1
2 thymocytes
3 thymocytes
4 brain
5 skin
6 thymocytes
7 brain
8 skin
9 brain
10 thymus
11 brain
12 thymus
13 skin
Cd4
B
chr5:71,117,927-71,523,717
100 kb
BUR
Std
SATB1
Urea
SATB1
SATB1
SATB1
input
input
input
Std
CTCF
H3K27ac
DNase 1 HS
1
2 thymocytes
3 thymocytes
4 brain
5 skin
6 thymocytes
7 brain
8 skin
9 brain
10 brain
11 brain
Gabrg1
Gabra2
C
%
BUR peaks / total
SATB1 urea ChIPseq peaks
100
80
60
40
20
0
rep 1
rep 2
rep 1
rep 2
rep 1
rep 2
skin
thymocytes
brain
SATB1 urea ChIPseq
D
%
SATB1 urea ChIPseq peaks / total
BUR reference peaks
60
40
20
0
rep 1
rep 2
rep 1
rep 2
rep 1
rep 2
skin
thymocytes
brain
SATB1 urea ChIPseq

FIGURE 5

A


chr3:97,107,505-144,477,019
10 Mb
Lamin B1-DamID
BUR smoothened
BUR
Lamin B1 ChIPseq
Gene
H3K4me3
Urea
SATB1
SATB1
SATB1
Std
SATB1
DNase 1 HS
Urea
CTCF
1
2
3
4
5
6
7 thymocyte
8 brain
9 skin
10
11
12
thymocyte

B


LADS
LADS
LADS
LADS
LADS
BUR
SATB1