Genome organization by SATB1 binding to base-unpairing regions (BURs) provides scaffold for SATB1-regulated gene expression

Short title: SATB1 binds BUR genomic elements to organize chromatin

Yoshinori Kohwi[1*], Xianrong Wong[2^], Mari Grange[1], Thomas Sexton[3], Hunter W. Richards[1], Yohko Kitagawa[4†], Shimon Sakaguchi[4], Ya-Chen Liang[5], Cheng-Ming Chuong[5], Vladimir A. Botchkarev[6], Ichiro Taniguchi[7], Karen L. Reddy[2] and Terumi Kohwi-Shigematsu[1*]

[1] Department of Orofacial Sciences, University of California San Francisco, San Francisco, CA, 94143, USA

[2] Department of Biological Chemistry, School of Medicine, Johns Hopkins University, Baltimore, MD, 21205, USA

[3] Institute of Genetics and Molecular and Cellular biology (IGBMC), INSERM U1258, Illkirch, France.

[4] Laboratory of Experimental Immunology, Immunology Frontier Research Center, Osaka University, Osaka, 565-0871, Japan

[5] Department of Pathology, University of Southern California, Los Angeles, CA, 90033, USA

[6] Department of Dermatology, Boston University School of Medicine, Boston, Massachusetts, 02118, USA

[7] 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

^ Agency for Science, Technology, and Research (A*STAR) Skin Research Labs (A*SRL), Singapore 138632

* co-corresponding authors: TKohwi-Shigematsu@lbl.gov, YKohwi@lbl.gov

Abstract

Mammalian genomes are organized by multi-level folding, yet how this organization contributes to cell type-specific transcription remain unclear. We uncovered that the nuclear protein SATB1 establishes two-tiered chromatin organization, one through indirect binding and another by direct binding of base-unpairing regions (BURs), which are genomic elements with high unwinding propensities. Published ChIP-seq datasets show SATB1 binding to highly accessible chromatin at enhancers and CTCF sites, but not to BURs. By employing urea ChIP-seq, which retains only directly bound protein:DNA complexes, we found that BURs, but not CTCF sites, are direct SATB1 binding targets. SATB1-bound BUR interactions with accessible chromatin can cross multiple topologically associated domains (TADs) and SATB1 is required for these megabase-scale interactions linked to cell type-specific gene expression. BURs are mainly found within lamina associated domains (LADs) sequestered at the nuclear lamina, but also in inter-LADs, and SATB1 binds a subset of BURs depending on cell type. Notably, despite the mutually exclusive SATB1-binding profiles uncovered by the two ChIP-seq methods, we found most peaks in both profiles are real and require SATB1. Together, we propose that SATB1 has functionally distinct modes of chromatin interaction by directly binding BURs to form a chromatin scaffold to which it indirectly tethers open chromatin. Such chromatin organization may provide a gene-regulatory network underlying cell type-specific gene expression.

Introduction

Three-dimensional (3D) chromatin architecture, formed by chromatin folding at multiple hierarchical levels, is thought to play a fundamental role in genome activity such as gene expression(*1-7*). Whether and how genome folding contributes to gene regulation, however, remains largely unknown. High resolution chromatin interaction maps, generated by chromosome conformation capture methods such as Hi-C (all-to-all contacts), have revealed chromatin to be partitioned mainly into two types of megabase (Mb)-sized compartments, A and B (*8*). The euchromatic A compartment is gene-rich, transcriptionally active and enriched in histone marks associated with active chromatin, conferring high chromatin accessibility. On the other hand, the B compartment is gene-poor, transcriptionally repressive, and enriched in repressive histone marks. The largely transcriptionally repressive lamina-associated domains (LADs) are among the B compartment chromatin displaying the typical features of heterochromatin (*9, 10*). LADs have been identified by the DamID method (*11*) as genomic regions that are spatially proximal to the nuclear lamina, an intermediate filament meshwork adjacent to the inner nuclear membrane. LADs represent discrete chromatin domains of 40kb to 30 Mb in size (*9*). Approximately 1000-1500 LADs have been identified in mouse and human cells and, in a given cell type, typically cover ~30% of the genome (*2*). By sequestering chromatin to the repressive nuclear compartment, the nuclear lamina is thought to play an important role in genome organization.

At sub-megabase scales, using Hi-C, chromatin domains called "topological-associated domains" (TADs) were previously detected that are relatively insulated from adjacent TADs by preferentially making intra-domain contacts via looping (*12-14*). CTCF and cohesin are the two major architectural proteins required for folding chromatin into loops. The loop extrusion model postulates that cohesin functions in loop extrusion until it encounters the CCCTC-binding factor (CTCF) bound to its target sites in a convergent orientation (*15, 16*). CTCF- and cohesion-binding sites are enriched at the boundaries of TADs (*12, 17*) and LADs (*9*). Acute depletion of CTCF (*18*), cohesin (*19, 20*) or induced removal of cohesin loading factor NIPBL (*21*) results in a major loss of TADs. Unexpectedly, the majority of protein-coding genes remained largely unchanged in their expression levels in the absence of TADs. These results challenge the view that TADs provide a unit of gene expression control by insulating enhancer-promoter interactions from neighboring TADs. In the last few years, using super-resolution live-cell imaging, genome interactions were studied at the single cell levels. These studies revealed that chromatin looping formed by CTCF and cohesin are highly dynamic and variable among individual single cells, suggesting that TADs are an emergent property of the cell population average of heterogeneous chromatin folding (*22-26*). Therefore, TADs, enriched in contacts, may not represent a stable structural unit of chromatin flanked by two boundary loci stably bound by CTCF and cohesin [reviewed in (*25, 27, 28*)]. Super-resolution microscopy detected not only single cohesin-mediated loops but also radially stacked loops forming "rosette" like structures in individual single cells. These loops were also transient and a single cell spends a minority of its time in the configuration, and depletion of

95 cohesion led to increased variability in gene expression at single cell levels (*29*). As such, although
96 research on CTCF and cohesin has greatly advanced our understanding on chromatin loop formation,
97 there is still much to learn how 3D chromatin structure underlies gene activity, especially cell-type
98 specific gene expression. It is increasingly clear that it is necessary to explore additional nuclear factors
99 that potentially drive chromatin folding. In this paper, we studied function of a tissue-specific
100 architectural protein SATB1 in genome organization.

101 SATB1 was originally identified by virtue of its specific binding to genomic sequences called
102 Base Unpairing Regions (BURs) (*30-32*). BURs have a strong unwinding property under negative
103 superhelical stress, and BURs are empirically identified using unpaired DNA-specific chemical probes
104 (*30, 33*) because of the lack of primary sequence consensus. However, BURs have a unique sequence
105 context. A typical BUR (~200-300bp) contains a cluster of short sequence segments (25-50 bp/segment)
106 containing exclusively As, Ts, and Cs on one strand (ATC context) (*30, 32*). BURs are at least 65% AT-
107 rich, but importantly, not all AT-rich sequences are BURs. SATB1 binds to BURs in the double-stranded
108 DNA form, recognizing their altered phosphate backbone structure (*32*). Disruption of the ATC context
109 by mutagenesis, without affecting the AT content, results in loss of unwinding property of a BUR and
110 SATB1 binding. SATB1 is most abundantly expressed in thymocytes among multiple adult progenitor
111 cells (*34-36*), as well as terminal differentiated postnatal neurons in the cortex and amygdala (*37, 38*).
112 Located in nuclear interior, SATB1 has an interconnected cage-like distribution in thymocytes(*39*) and a
113 fine spider web-like distribution in neurons (*37*) forming unique SATB1 nuclear architectures that are
114 resistant to high-salt extraction. Some BURs have been identified and characterized as in vivo SATB1
115 binding sites and when bound to SATB1, are also resistant to salt extraction (*39, 40*). SATB1 has been
116 shown to be essential for loop formation connecting a subset of these BURs and recruiting chromatin
117 remodeling complexes, epigenetic factors and transcription factors to specific gene loci (*41*). SATB1
118 regulates expression of hundreds of protein-coding genes in a cell-type specific manner to enable cells
119 to acquire new phenotypes, such as during T cell activation (*42*), adult progenitor cell development (*34-*
120 *37, 43-49*), and cancer metastasis (*41, 50-52*). By directing lineage-specific transcriptional programs,
121 SATB1 plays an essential role in development of $CD4^+$ T cells, $CD8^+$ T cells, NKT cells, and $Foxp3^+$
122 regulatory T cells (Treg) in the thymus(*46, 47*) as well as in promoting the pathogenic effector program of
123 tissue Th17 cells in autoimmune disease in mice(*48*). These activities have been shown to mediate co-
124 induction of interleukin genes in T helper cells (*42*) and regulate ~1000 genes in cancer to promote
125 metastasis(*50*). Thus, SATB1 has long been thought to play a role in regulating genome architecture
126 and in gene expression.

127 Although BURs have been identified to be specific binding target sequences of SATB1,
128 mysteriously the genome-wide SATB1 binding profiles obtained by the widely used, standard chromatin
129 immunoprecipitation-deep sequencing (ChIP-seq) did not detect SATB1 interaction with BURs. Instead,
130 SATB1 has been mapped mainly to enhancers and to a region that would become an active super-
131 enhancer upon differentiation (*45-47*). In line with this, recent studies on chromatin looping with Hi-C also

132 shows that SATB1 mediates interactions involving a subset of enhancers and promoters for cell identity
133 gene expression and that SATB1 binding sites coincide with many CTCF-bound loci (*53-55*). These
134 results are consistent with the SATB1 function in assembling transcription complexes at regulatory
135 regions, but at the same time, are puzzling in light of the SATB1-BUR relationship. We hypothesized
136 that SATB1-BUR interactions are somehow undetectable by the standard ChIP-seq approach as wells
137 as the CUT&Tag method, and if identified, would provide additional mechanisms for chromatin looping
138 events *in vivo* that link to gene expression.

139 ChIP-seq has been a gold standard for profiling chromatin-protein interactions. However,
140 standard ChIP-seq approaches have several drawbacks. First, chromatin structure can introduce biases
141 into ChIP-seq analysis. Fragmentation of crosslinked chromatin in whole cells/nuclei can result in biased
142 capture of transcriptionally active, highly accessible "open" chromatin and generation of false-positive
143 "phantom" peaks (*56-60*). Second, it fails to distinguish between direct and indirect (piggybacking)
144 chromatin binding of a protein of interest. Here, we optimized urea-ChIP-seq, a modified ChIP-seq that
145 first isolate stringently purified whole intact genomic DNA that retains only its directly bound proteins
146 from crosslinked cells prior to fragmentation and immunoprecipitation. This results in an unbiased
147 capture of the entire genome, regardless of its original chromatin accessibility status. This approach
148 has allowed to identify genomic sites directly bound by SATB1 and revealed BURs to be the primary
149 *direct* SATB1 target sites, previously hidden by standard ChIP-seq. The SATB1-bound BURs are
150 mutually exclusive with enhancer-enriched SATB1-bound sites mapped by standard ChIP-seq. Instead,
151 a majority of BURs are mapped within LADs, and depending on cell type, SATB1 binds *in vivo* to
152 selective subsets of these BURs. Using the urea 4C-seq (one-to-all contacts), select SATB1-bound
153 BURs were found to interact extensively over the >5.7Mb gene-rich region within DNase 1
154 hypersensitive (DNase 1 HS) regions, but not over adjacent gene-poor regions. SATB1 is essential for
155 both extensive megabase-level chromatin interactions and proper expression of multiple genes in the
156 gene-rich region. Collectively, our results suggest that SATB1-mediated genome organization exhibits
157 two distinct features, a "skeletal framework" where SATB1 binds BURs directly, and a network of gene-
158 rich open chromatin that is tethered to BURs indirectly through SATB1-mediated chromatin looping,
159 which is correlated to gene expression.

160 Results

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

162 Unlike many nuclear proteins (e.g. transcription and chromatin modifying factors), SATB1 resides in
163 nuclear substructures exhibiting resistance to extraction with high salt or lithium 3,5-diidosalicylate (LIS)
164 (*40, 44*). Consistent with this, after extraction of nuclei with buffer containing 2M NaCl that removed
165 most of the proteins from DNA, individually cloned SATB1-bound BURs remained anchored to the
166 residual SATB1 nuclear architecture found in the interior of nuclei (*39*). By contrast, in high-salt

167 extracted *Satb1*$^{-/-}$ (SATB1-KO) thymocyte nuclei, these BURs lost their anchored sites and were
168 exclusively found in the distended DNA halos that spread around the residual nuclei, which is illustrated
169 (*Figure 1A*) (*39*). Thus, we hypothesized that SATB1-BUR interactions are likely embedded in the "highly
170 insoluble nuclear substructures" that could not be detected to date by the widely used standard ChIP-
171 seq methods. This is likely because ChIP-seq experiments have inherent biases arising from chromatin
172 structure such as heterochromatin or transcriptionally active regions.

173 To detect BURs as the primary targets of SATB1 genome wide, we devised a urea ChIP-seq
174 method to purify genomic DNA with only its directly bound proteins. Through this apporach, we found
175 that a vast majority of *in vivo* SATB1-binding sites are BURs in the genome. In this method, cells are
176 first crosslinked with formaldehyde and then crosslinked chromatin in lysed cells is subjected to 8M urea
177 ultracentrifugation. In this manner, crosslinked chromatin is stringently purified by removing all free-
178 floating proteins and proteins bound to DNA indirectly. Therefore, urea ChIP-seq is performed with the
179 pure genomic DNA fiber that retains only directly bound proteins. Without prior formaldehyde crosslinking,
180 urea centrifugation removes 99.9% of chromatin-associated protein from chromatin and sediments
181 unsheared pure genomic DNA at the bottom of the tube (*61*). Critically, the urea ChIP-seq method (*62*)
182 allows detection of genomic regions solely based on their direct protein binding and is not biased by their
183 original chromatin status, such as accessible or inaccessible chromatin. To obtain reproducible direct
184 binding profiles of SATB1 at the genome-wide scale, it was necessary to make additional optimizations.
185 The modified urea ChIP-seq protocol allows quantitative solubilization of the unsheared crosslinked
186 genomic DNA fiber after urea ultracentrifugation, while still maintaining intact chromatin-protein
187 interactions during fragmentation. The critical steps that are different between urea and standard ChIP-
188 seq are highlighted (*Figure 1B*). A putative scenario as to how DNA-binding sites of chromatin-
189 associated factors are captured by the two different ChIP-seq method is illustrated (*Figure 1C*). Urea
190 ChIP-seq allows detection of chromatin-bound proteins regardless of the chromatin's accessibility
191 because the whole genome is purified intact. In contrast, in standard ChIP-seq the insoluble fraction is
192 discarded and only the soluble fraction is retained for analysis, resulting in a bias toward detection of
193 accessible "open" chromatin. Moreover, standard ChIP-seq captures both directly- and indirectly-bound
194 proteins, and these cannot be distinguished. Thus, urea ChIP-seq is uniquely suited for detecting direct
195 DNA binding profiles of SATB1 that resides in the high-salt extraction resistant nuclear substructure.
196

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

198 To determine whether SATB1 binds to BURs *in vivo*, it is essential to validate *in vivo* SATB1-bound
199 sites as BURs. We previously generated a genome-wide BUR reference map by deep sequencing
200 DNA fragments generated from purified genomic DNA bound by recombinant SATB1 protein *in vitro*
201 *(62)*. Since SATB1 binds specifically to BURs *in vitro*, we used SATB1 as a biological probe to identify
202 all potential BURs in the genome. Using this approach, we reproducibly identified approximately
203 240,000 BURs ($q<0.01$), establishing a BUR reference map for both mouse and human genomes.

204 Comparing ChIP-seq peaks for SATB1 with the BUR reference map allows us to assess whether SATB1
205 binding sites *in vivo* correspond to BURs.
206 We performed urea ChIP-seq for SATB1 in thymocytes and compared the results with standard
207 ChIP-seq for this protein and the BUR reference map. Strikingly, urea ChIP-seq produced SATB1
208 binding profiles that are entirely different from those produced by the standard ChIP-seq method (*Figure*
209 *2A*). The genome-wide analysis in mouse thymocytes using deepTools revealed that only 0.92% (an
210 average of two replicates per ChIP-seq method) of peaks in urea SATB1 ChIP-seq coincided with those
211 in standard SATB1 ChIP-seq (*supplement Table 1 and Table 2*). In a large genomic region covering a
212 4.48Mb region in chromosome 17 (*Figure 2A, top*), a gene-rich region was found in the valley between
213 BUR-enriched domains (*Figure 2A, track 3*), and standard ChIP-seq detects clusters of SATB1 binding
214 sites in this gene-rich region (*Figure 2A, tracks 1 and 2*). In contrast, urea ChIP-seq detected SATB-
215 binding to the domains surrounding the gene rich region, corresponding to regions heavily enriched in
216 BUR sites (*Figure 2A, tracks 3- 5*). In a zoomed-in view, covering a 430 kb region, SATB1 sites
217 identified from urea ChIP-seq precisely coincided with BUR peaks and are excluded in the standard
218 SATB1 ChIP-seq peaks (*Figure 2B, tracks 1-3*). SATB1 bound sites uncovered from urea ChIP-seq
219 avoided CTCF sites (*Figure 2B, tracks 4 and 5 versus track 6)*, H3K27ac-marked sites and DNase 1
220 HS (*Figure 2B tracks 7 and 8)*, whereas SATB1 sites from standard ChIP-seq primarily co-mapped to
221 these sites (*Figure 2B, tracks 1,2,6-9*). We confirmed the relationship of SATB1-binding sites by urea
222 ChIP-seq and BURs at the genome-wide level using deepTools. Among SATB1 urea ChIP-seq peaks,
223 an average of 77% of two replicates coincided with BURs identified in our reference maps (*Figure 2C*).
224 The statistical significance of enrichment of BURs compared to a random overlap in either urea ChIP-
225 seq or standard ChIP-seq peaks is $p<0.001$ or $p=1$, respectively, using 1000 bootstrapped genomic
226 intervals sampled using the *bootRanges* approach (*63*). Consistent with this, genome-wide distribution
227 profile analyses show that while SATB1 binding profiles by standard ChIP-seq specifically exclude
228 BURs, SATB1-binding profiles by urea ChIP-seq coincide with BUR distribution, showing a sharp
229 indentation at the peak of BUR distribution (*Figure 2D*). These results indicate that SATB1
230 predominantly and directly targets BURs genome-wide *in vivo* and that the SATB1-binding profiles
231 obtained by standard and urea ChIP-seq methods are mutually exclusive. We conclude that the
232 unbiased feature of urea ChIP-seq, which solely captures DNA with its directly bound proteins unmasks
233 the SATB1 direct binding profiles.

234

235 **CTCF-binding profiles remain unchanged between urea and standard ChIP-seq**

236 To investigate the potential cause for the discrepancy in the SATB1-binding profiles between the two
237 ChIP-seq approaches, we applied the urea ChIP-seq method to other factors and asked whether
238 binding profiles differed between urea and standard ChIP-seq. We investigated the DNA-binding profile
239 for CTCF by urea ChIP-seq and found its profile to be very similar to that obtained by standard ChIP-

240 seq from ENCODE, which largely overlapped with DNase 1 HS (*64*) (*Figure 3A, tracks 2, 4, 5 and 6*).
241 Similarly, we also found that urea ChIP-seq generated essentially identical DNA-binding profiles for
242 polycomb repressive complex 2 (PRC2) core subunits, SUZ12, JARID2 and EZH2 as generated by
243 others using standard ChIP-seq (*65*) (*Figure 3—supplement figure 1*). In line with recent studies (*53-*
244 *55*), 37% of SATB1 peaks from standard ChIP-seq (ENCODE) (average of two replicates, *supplement*
245 *Table 1*) intersected with CTCF peaks genome-wide. In contrast, only 0.8% of SATB1 peaks from urea
246 ChIP-seq coincided with CTCF peaks (average of two replicas, *supplement Table 1*), confirming the
247 UCSC browser track views (*Figure 2A and 3A*). These results verify that urea ChIP-seq can accurately
248 identify the DNA-binding sites of chromatin associated proteins. Thus, urea ChIP-seq, compared to
249 standard ChIP-seq, produces contrasting DNA-binding profiles for SATB1, but this is not necessarily
250 the case for other nuclear proteins. Thus, we conclude that for SATB1, urea and standard ChIP-seq
251 produces contrasting SATB1 binding profiles, and importantly, only urea ChIP-seq can faithfully detect
252 direct and specific SATB1 interactions with BURs. This highlights a different mode of SATB1 chromatin
253 interaction compared to that detected by standard ChIP-seq. It is important to note that these urea ChIP-
254 seq results support a large body of cytological data for SATB1, indicating that SATB1 forms salt-
255 extraction resistant nuclear architecture network, and that the genomic sites directly bound to SATB1
256 reside in the "inaccessible" chromatin.
257
258 **CTCF and SATB1 have distinct and independent binding profiles**
259 Our results above indicate that CTCF and SATB1 direct binding sites do not overlap. However, because
260 both CTCF and SATB1 are known to affect chromatin organization, there might be a functional link
261 between the two proteins. Recently, CTCF-binding sites mapped by CUT&Tag were found to be SATB1
262 independent in specific gene loci (*55*). We validated this result by generating the CTCF-binding profiles
263 by urea ChIP-seq in *Satb1*$^{-/-}$ (SATB1-KO) and *Satb1*$^{+/+}$ (SATB1-WT) thymocytes. These CTCF-binding
264 profiles showed no major changes depending on SATB1 (*Figure 3A, compare tracks 2 and 3*), as shown
265 in a randomly chosen region in chromosome 17 in wild-type (WT) and SATB1-KO thymocytes. This was
266 confirmed by genome-wide analyses which showed that CTCF urea ChIP-seq peak distributions from
267 SATB1-WT and SATB1-KO thymocytes coincided (*Figure 3B*). This indicates that CTCF binding to
268 genomic DNA is SATB1-independent genome wide.
269 To corroborate the independence of CTCF/cohesin and SATB1, we next examined whether
270 SATB1 interacts with CTCF and/or cohesin or both, components known to be important for loop
271 formation. Instead of protein co-immunoprecipitation (co-IP) using whole cell or nuclear extracts, we
272 examined the direct co-binding status of SATB1 and CTCF or cohesion on chromatin *in vivo* by urea
273 ChIP-Western. Urea-purified crosslinked-chromatin fragments immunoprecipitated with anti-SATB1
274 antibody (SATB1-bound urea-ChIP fragments) contained SATB1 as expected, detected by Western-
275 blot of the de-crosslinked urea-ChIP sample. However, neither CTCF nor RAD21 (a cohesin complex
276 component) was detected in the SATB1 urea-ChIP samples. Reciprocal experiments using either

277 CTCF or RAD21-bound urea ChIP samples confirmed lack of SATB1 co-binding to these chromatin
278 fragments (*Figure 3C*, *Figure 3—supplement figure 2A*). We confirmed that a similar amount of urea-
279 purified crosslinking chromatin was used for each ChIP-Western sample with an anti-H3 antibody.
280 Therefore, these results from urea ChIP-Western indicate that at least for the direct DNA co-binding
281 status of SATB1 with either CTCF or RAD21, SATB1 does not co-bind chromatin with either of these
282 proteins (*Figure 3C*). These results are consistent with mutually exclusive direct binding sites of CTCF
283 and SATB1 (*Figure 3A* and *3B*).
284 If cohesin and CTCF bind directly to chromatin for loop formation according to the loop extrusion
285 model, then we expect that the urea ChIP-Western results will show that cohesin-bound DNA fragments
286 are co-bound by CTCF at the base of loops. We found that chromatin fragments co-bound by RAD21
287 and CTCF represent only a small fraction of either total RAD21-bound or CTCF-bound chromatin
288 fragments (*Figure 3C*). These results show that CTCF and RAD21 direct co-binding to chromatin occurs
289 much less frequently than their individual binding to chromatin. This result agrees with a recent report
290 which used super-resolution imaging of individual cells to show that the stable CTCF-mediated loops
291 are rare, and >95% of the time, the loops are either only partially extruded, or no loops are formed (*25*).
292 Given that CTCF deletion causes most TADs to disappear, we next examined whether SATB1
293 depletion has any effects on TAD formation. We compared bulk Hi-C results between thymocytes
294 isolated from wild-type (*Satb1*$^{+/+}$: *Cd4-cre*) and SATB1-deficient (*Satb1*$^{F/F}$:*Cd4-cre*) mice in duplicates
295 (*Figure 3—supplement figure 2B*). Our Hi-C results revealed largely indistinguishable TAD profiles
296 between the two thymocytes (*Figure 3D*) examined at 10kb resolution**.** Neither TAD number nor TAD
297 sizes were changed by SATB1 depletion from thymocytes (*Figure 3—supplement figure 2C).* Intra-TAD
298 interactivity (the sum of Hi-C chromatin contacts inside the TAD) also remained mostly unchanged
299 (*Figure 3—supplement figure* 2D). These bulk Hi-C results upon SATB1 deletion are in good agreement
300 with recently published studies (*53, 54*). Similar to the results from the CTCF depletion study (*18*),
301 SATB1 depletion also did not alter segregation of active (A) and inactive chromatin
302 compartmentalization (B) (*Figure 3E*). Furthermore, our scaling plot displaying how the Hi-C contacts
303 decrease as a function of genomic distance shows that SATB1 depletion does not affect chromosome
304 compaction (*Figure 3—supplement figure* 2F). These Hi-C results taken together with SATB1-
305 independent CTCF-binding to DNA and non-overlapping direct genomic DNA-binding of SATB1 and
306 CTCF/cohesion, strongly suggest that SATB1 and CTCF are two independent proteins that have distinct
307 roles in chromatin organization.
308
309 **Non-specific peaks appear in *Satb1*$^{-/-}$ thymocytes by standard SATB1 ChIP-seq**
310
311 An important question is whether SATB1-binding profiles obtained by standard ChIP-seq are strictly
312 SATB1 dependent. If so, these binding sites likely reflect real but indirect SATB1 association with
313 chromatin. To address this question, we performed standard ChIP-seq on *Satb1*$^{-/-}$ (KO) thymocytes
314 using two different anti-SATB1 antibodies (1583 and ab109112, both of which are validated for

315 immunoprecipitation). The results are highlighted in two randomly chosen ~3Mb regions in chromosome
316 17 and in X chromosome, each containing a gene-rich region surrounded by gene-poor regions. Both
317 antibodies generated similar peaks in KO thymocytes (KO peaks) by standard ChIP-seq. These KO
318 peaks (false-positive) were predominantly detected in gene-rich, open chromatin regions in KO
319 thymocytes, and many of them coincided with peaks originally detected in WT thymocytes (WT peaks)
320 (*Figure 4—supplement figure 1 A and 1B, top and bottom, tracks 9 and 10*). This is in agreement with
321 an earlier report that ChIP-seq generate false-positive peaks in highly accessible, transcriptionally
322 active chromatin region (*56-60*). Intersection analyses with deepTools show that 4.1% of SATB1
323 standard ChIP-seq peaks obtained from our experiments are false-positive peaks (based on average
324 of two standard KO ChIP-seq samples) (Supplement Table 3 and 4). Besides those KO peaks that
325 overlapped with WT peaks, there were several thousands of KO-specific peaks generated by standard
326 ChIP-seq in our study. In contrast, SATB1 urea ChIP-seq had no false-positive peaks (0.0%). Also,
327 only ~100 KO-specific peaks (not overlapping with WT peaks) were detected (average of two samples)
328 compared to ~40,000 total WT peaks from SATB1 urea ChIP-seq (*supplement Table 3 and 4*).
329 Therefore, validation using knockout cells is critical to distinguish real peaks that depend on SATB1
330 from SATB1-independent phantom peaks derived from the standard ChIP-seq method. The
331 percentages of phantom peaks in standard ChIP-seq likely depend on the specificity of antibodies used
332 and the condition of sample preparation.

333 After validation with SATB1-KO cells, we conclude that many SATB1 sites detected from
334 standard ChIP-seq are bound by SATB1, but most likely through indirect binding, as these peaks are
335 not detected by urea ChIP-seq, which removes any indirect binding proteins by urea ultracentrifugation.
336 SATB1 proteins are found in high salt-resistant fraction as well as salt-extracted fraction(*40*). Thus, it is
337 possible that soluble SATB1 may associate with "open" chromatin. We focused on three individual
338 gene loci, i. e. *Rag1* and *Rag2* genes (*45*), *Foxp3* (*47*), and *Runx3* (*46*), for which specific SATB1-
339 binding sites by standard ChIP-seq have been reported (*Figure 4*). In addition, these genes are known
340 to be regulated by SATB1 in thymocytes and are thus excellent candidate regions to investigate how
341 SATB1 is regulating chromatin and gene expression. By either direct urea-ChIP or standard ChIP, we
342 found SATB1-bound sites within all three loci (*Rag1/2*, *Foxp3*, and *Runx3*) to be SATB1-dependent.
343 These sites were detected only in WT (*Figure 4, track 3*) but not in SATB1 KO-thymocytes (*Figure 4,*
344 *tracks 7 and 9 for Rag/Rag2* and *Runx3*, *and tracks 9 and 11 for Foxp3*). In these regions, input sample
345 (without antibodies) had no peaks (*Figure 4*, *tracks 8 and 10 for Rag1/Rag2 and Runx 3, and tracks 10*
346 *and 12 for Foxp3*). For the *Rag1* and *Rag2* loci, SATB1 has been shown by standard ChIP-seq to bind
347 the anti-silencer element (ASE) and this binding is thought to counteract the activity of an intergenic
348 silencer, thus promoting expression of *Rag 1* and *Rag 2* (*45*). During development of Foxp3^{+}-regulatory
349 T cells (Treg) in the thymus, SATB1 expression in Treg precursor cells is essential for the future
350 establishment of a Treg cell–specific super-enhancers at the *Foxp3* locus in Treg cells (*47*). Thus,
351 thymocyte precursor-specific SATB1 deficiency, using *Satb1*$^{fl/fl}$ CD4-cre mice but not *Satb1*$^{fl/fl}$ *Foxp3*-

352 cre mice, impairs Treg super-enhancer activation and expression of *Foxp3*(*47*). Impaired Treg cell
353 development results in autoimmune disease (*46, 47, 66, 67*). Finally, SATB1 is also essential for
354 generation of $CD4^+$ and $CD8^+$ T cells and NKT cells (*46*). In response to T cell receptor signaling by
355 MHC class I and II, SATB1 regulates enhancer function in gene loci encoding lineage-specifying factors
356 (e.g. *Runx3*) and directs lineage-specific transcriptional programs in the thymus (*46*). Strikingly, in these
357 three regions, SATB1-bound BURs, as detected by urea ChIP-seq, are found infrequently, and appear
358 as isolated single peaks within 30-50kb regions. Such SATB1-bound BUR sites are located close to,
359 but not overlapping with the cluster of enhancers marked by H3K27ac and H3Kme1 modifications.
360 These SATB1-bound BUR sites reside within 3kb of the nearest putative SATB1-bound enhancers
361 (detected by standard ChIP-seq) for the *Rag1/Rag2* and *Runx3* loci as well as from the super-enhancer
362 that becomes established at the *Foxp3* loci in matured Treg cells. Results from these three loci show
363 that, while the SATB1-bound peaks generated by urea ChIP and standard ChIP-seq did not overlap,
364 they were nonetheless proximal to each other, and both were SATB1 dependent. The two distinct ChIP-
365 seq protocols are likely detecting two different modes of DNA binding for SATB1 i.e., direct binding to
366 BURs and indirect binding to open chromatin regions.
367

368 **Long-distance chromatin interactions involving a BUR depend on SATB1**

369 We explored whether there is any relationship between SATB1-bound BURs detected by urea-
370 ChIP-seq and SATB1-bound sites in the open chromatin region detected by standard ChIP-seq. We
371 hypothesized that chromatin architecture formed by direct binding of SATB1 to BURs may recruit open
372 chromatin regions by indirect association with their chromatin-bound protein complexes by SATB1. If
373 so, such interactions would necessarily be dependent on SATB1, and they might be correlated to gene
374 expression. To address these questions, we studied a gene-rich region in chromosome 2 that contains
375 the *Rag1* and *Rag2* genes whose expression is SATB1 dependent (*45*). We first examined genome-
376 wide interactions from a SATB1-bound BUR that appeared as a single, strong peak (*Figure 4 and 5,*
377 *top)* as the viewpoint or bait (BUR-1). The BUR-1 primer sequence is located near the border between
378 the 3.9 Mb gene-desert region and the 5.8Mb gene rich region on chromosome 2 where the major BUR
379 bound to SATB1 is located. BUR-1 resides 1.4 kb away from the ASE (*Figure 4, top*), which is essential
380 for proper *Rag1* and *Rag2* expression in double-positive thymocytes and for differentiation of single-
381 positive thymocytes(*68*).
382 To study interactions from the SATB1-bound BUR, we performed 4C-seq using urea purified
383 chromatin (urea 4C-seq, one-to-all) to better access and capture BUR-mediated interactions, which are
384 expected to emanate from the highly inaccessible nuclear substructure. In our urea 4C-seq, we used
385 *Hind*III for initial digestion. The 6-base cutter, *Hind*III, was used instead of a 4-base cutter like *Dpn*II
386 (frequently used for 4C-seq), because multiple *Dpn*II sites are often found within individual SATB1-
387 bound BURs, thus producing small fragments, likely disrupting SATB1-binding to BURs. The "BUR-1"

388 bait region at a SATB1-bound BUR is designed at the end of the *Hind*III fragment, which is adjacent to
389 the *Hind*III fragment containing ASE. We examined interaction events from the fragment containing
390 BUR-1 and its interaction sites were mapped on the genome for visualization.

391 While urea purification removes indirectly associated protein and thus may reduce some
392 chromatin interactions, we speculated that the overall 3D chromatin structure of fixed chromatin will still
393 retain its proximity events, especially in the environment of chromatin which is strongly anchored to the
394 nuclear substructure (as we predicted for SATB1 scaffold anchoring at BURs) to enable detection in
395 urea 4C-seq. In the ChIP-seq protocol (in general), we expect that in the absence of a ligation step as
396 is routinely done in 4C protocols (as described below), subsequent steps for ChIP-seq (sonication,
397 stringent washes of immunoprecipitated fragments, etc) results in dissociation of interacting fragments.
398 We carried out urea 4C-seq experiments to test if we could indeed detect such strong SATB1-mediated
399 3D interactions. Briefly, for urea 4C-seq (outlined in *Figure 5—supplement figure 1*), crosslinked
400 chromatin purified by urea ultracentrifugation is first digested with a restriction enzyme (*Hind*III).
401 Digestion is followed by ligation of chromatin fragments whose physically proximity status in 3D space
402 was fixed at the time of formaldehyde crosslinking. After these steps, DNA was purified. Beyond this
403 point urea-4C-seq differs from the commonly used 4C-seq called circular chromosome conformation
404 capture (referred to here as circular 4C-seq)(*69-74*). In circular 4C-seq, ligated DNA fragments are
405 circularized after second restriction enzyme digestion before PCR amplification. In contrast, in urea 4C-
406 seq a single strand DNA capture strategy was used to isolate a specific pool of single strand DNA, each
407 containing a ligation-captured DNA fragment surrounded by the bait fragment and the DNA linker (see
408 details in *Figure 5—supplement figure 1* legend and Materials and Methods). The captured DNA
409 sequences in the single strand DNA were PCR amplified using a primer from the bait sequence and
410 another primer from the DNA linker, thus avoiding amplification of any non-specific DNA fragments.
411 This was confirmed by the lack of PCR products using either primer alone (data not shown). This critical
412 step results in highly specific signals with very low background signals. The high signal-to-noise ratio in
413 urea 4C-seq contrasts with uniform and strong genome-wide background signals typically seen with
414 circular 4C-seq, which are thought to reflect random chromatin interaction events. Some non-interacting
415 sequences could also be trapped at the circulization step, and even if such incidence is rare, would
416 result in background signals upon PCR amplification. Because the background signals for urea 4C-
417 seq are extremely low in contrast to circular 4C-seq, the visual appearances of the interacting profiles
418 from the two methods are greatly different. Nevertheless, urea 4C-seq detects large-scale interactions
419 involving BURs that are strictly dependent on SATB1 as described below.

420 In wild-type thymocytes, we detected BUR-1 making extensive long-distance interactions
421 covering most of the ~5.8Mb adjacent gene-rich region (*Figure 5A*). Interestingly, BUR-1 interactions
422 are found to be scarce over the gene-desert region upstream of BUR-1, despite the presence of many
423 SATB1-bound BURs in this region (*Figure 5A, top track for urea SATB1 ChIP-seq, Figure 5—*
424 *supplement figure 2*). Though 4C-seq can detect both up- and downstream interactions from a chosen

425 single bait site, we observed heavily direction-biased interactions of BUR-1 with the gene-rich region,
426 perhaps indicative of its involvement in gene regulation. BUR-1 interactions were found particularly
427 concentrated over the ~3 megabase (Mb) gene-rich region, containing SATB1-dependent *Rag1* and
428 *Rag2* genes. These dense and strong interaction signals must not be due to their spatial proximity from
429 the bait because similar dense interactions were not detected with the SATB1-KO cells. In fact, most of
430 the BUR-1 interactions were absent in SATB1-KO cells (*Figure 5A, track KO-BUR1*), indicating that
431 SATB1 is required for the numerous long-distance interactions over the gene-rich region. We also
432 validated these interactions using a second bait (BUR-2) as an experimental control as BUR-1 interacts
433 with the *Hind*III fragment containing BUR-2, located in the second intron of *B230118H07Rik* gene (see
434 *Figure 5B* for specific location). This fragment contains two strong BUR peaks. Although somewhat
435 weakly, SATB1 binds to one of these BURs. Hence, urea 4C-seq using BUR-2 serves as a reciprocal
436 experiment of BUR-1 to validate reproducibility of the BUR-1 interaction profile. By comparing with the
437 Hi-C heatmap in this 5.8Mb region we found that BUR-1 interactions traverse multiple TADs and are
438 not constrained by any specific TADs (*Figure 5—supplement figure 2*), further supporting the CTCF
439 independent nature of these SATB1-mediated interactions. Within this gene rich region, numerous
440 interactions from BUR-1 were detected within DNase 1 HS containing enhancers marked by H3K27ac
441 and H3Kme1 modifications. Therefore, dense chromatin looping events involving these SATB1-bound
442 BURs indicate a highly interactive chromatin architecture for the gene-rich region, bringing together
443 many gene regulatory sequences into proximity. We further examined a wider region covering >44Mb
444 that includes distal neighboring gene-rich regions *(Figure 5 and Figure 5—supplement figure 3A and*
445 *3B).* Although interaction frequencies were lower than that within the first ~5.8Mb gene-rich region, we
446 detected BUR-1 interactions with further distal gene-rich region coinciding with enhancer-rich DNase 1
447 HS regions, where clusters of SATB1 binding sites were detected by standard ChIP-seq, skipping over
448 gene-poor regions and a gene-rich region containing transcriptionally-silent olfactory receptor genes.
449 A zoomed-in view (chr2:101,308,846-101,599,827; 291Kb) of BUR-1 interactions confirmed a
450 marked difference in BUR-1 interactions (~200Kb) over the *Rag1* and *Rag 2* loci between wild-type and
451 *Satb1*$^{-/-}$thymocytes, confirming BUR-1 interactions require SATB1 (*Figure 5B*). In addition to many
452 BUR-1 interacting chromatin fragments close to or containing enhancers (marked by H3K27ac and
453 H3Kme1), some of these fragments contain additional BUR sites, including BUR-2 adjacent to
454 enhancer(s) *(Figure 5B,* gray shaded with red stars*).* BUR-2 interacted heavily over the *Rag1* and *Rag2*
455 loci in wild-type thymocytes and well reproduced the interactions observed with BUR-1, consistent with
456 the larger region (*Figure 5A*), suggesting that both BUR-1 and BUR-2 are part of the common interaction
457 network that requires SATB1. Concerning the correlation of chromatin interactions and gene
458 expression, among 43 genes identified with official gene symbols mapped in the 5.8Mb gene rich region,
459 8 genes display various levels of SATB1-dependent expression ($p<0.05$) (*54*). These genes were either
460 up- or downregulated upon SATB1 deletion in thymocytes based on RNA-seq data (*Figure 5-*
461 *supplement figure 4*), and *Rag1* and *Rag2* were most strongly downregulated by SATB1 depletion in

462 this region. Collectively, these data suggest a strong correlation between SATB1-dependent gene
463 expression and extensive looping events over long distance linking SATB1-bound BURs and regulatory
464 regions in the open chromatin regions. Therefore, SATB1 function on chromatin is not limited to merely
465 connecting a specific target gene locus with nearby enhancers and promoters within TADs. Instead,
466 SATB1 appears to form a large interaction network covering the entire gene-rich regions.
467
468
469 **SATB1 binds BURs in a cell-type dependent manner**
470 We next examined whether SATB1 binds different subsets of BURs depending on cell type, by studying
471 SATB1-binding profiles in three different types of mouse tissues– thymus, frontal cortex and skin – that
472 contain SATB1-expressing cells. SATB1 expression is primarily restricted to adult progenitor cells, such
473 as thymocytes and in basal layer of epidermis. SATB1 is also expressed in differentiated neurons in
474 specific brain subregions (e.g. frontal cortex and amygdala). We compared UCSC browser images of
475 two selected regions, a ~2.4 Mb region in chromosome 6 containing the *Cd4* locus expressed in
476 thymocytes (*Figure 6A*) and a 406kb region in chromosome 5 containing *Gabra2* and *Gabrg1* loci
477 expressed in the mammalian brain (*Figure 6B*). BURs were most frequently bound by SATB1 in brain
478 compared to thymocytes, and the least frequently bound in skin. Whereas SATB1-bound BURs show
479 cell-type dependency, CTCF binding sites that largely coincide with DNase1 HS were largely similar
480 between these three types of cells (*Figure 6A*).
481 We analyzed whether BURs are targeted by SATB1 depending on cell type. Consistent with the
482 cell-type specific function of SATB1, we observed cell-type dependent differences in BUR occupancy
483 by SATB1. For all three types of cells, high percentages of urea ChIP-seq peaks (close to 80%,
484 *supplement Table 1*) correspond to BURs, as revealed by peak intersection studies using deepTools
485 (*75*) and comparing to the BUR reference map (*Figure. 6C*, *supplement Table 1*). Despite similar total
486 aligned sequence reads in urea ChIP-seq, the number of SATB1-bound BUR peaks greatly differed
487 between cell types (*supplement Table 2*). Among 238,380 mapped BURs (q<0.01), the percentage of
488 SATB1-bound BURs in skin, thymocytes, and brain were 3.7%, 13.5% and 35.0% (based on average
489 of two replicas/cell type), respectively (*Figure 6D and supplement Table 1*). These results indicate that
490 frequency of BURs targeted by SATB1 *in vivo* can vary more than 10-fold depending on cell type.
491 Interestingly, while the number of BURs bound by SATB1 greatly differs between cell types, the majority
492 (83.9%) of BURs targeted by SATB1 in skin were bound to SATB1 in thymocytes, and nearly all (91.6%)
493 SATB1-bound BURs in thymocytes were bound to SATB1 in brain. Therefore, each cell type has an
494 overlapping set of SATB1-bound BURs in common, with increasing/decreasing numbers of sites
495 depending on cell type, rather than distinct subsets. Despite the highest BUR occupancy by SATB1
496 was observed in the brain, this is not due to the brain expressing the highest levels of SATB1 protein.
497 In fact, the thymus expresses by far the highest levels of SATB1, which comprises ~0.3% of the total
498 protein (*76*).

499
500 **BURs are enriched in LADs**
501
502 LADs are large mostly heterochromatic domains in close contact with the nuclear lamina and are known
503 to be enriched in AT-rich sequences (*77*). Although AT-rich sequences are not necessarily BURs,
504 BURs have a minimum of 65% AT content (many of them have a higher AT content) with the ATC
505 sequence context with an exceptionally strong unwinding propensity (*41*). We therefore asked whether
506 and to what extent any of the ~240,000 BURs identified in the genome overlap with LADs. We first
507 visually compared the distribution of all potential BURs with LADs mapped by lamin B1-mediated
508 adenine methyltransferase identification (Dam ID) and noted some correlation of BURs and LADs. As
509 an example, the 9 Mb region in chromosome 5 containing the neuronal gene cluster locus that encodes
510 subunits of $GABA_A$ receptor complex (*Gabrg1,* and *Gabra1, Gaba6, Gabrb2)* has a BUR-enriched
511 region coinciding closely with LAD organization (*Figure 7A*). We checked if this LAD-BUR overlap is
512 not restricted to this specific region by examining a larger view of a randomly chosen region covering
513 47.4 Mb in chromosome 3 (*Figure 7—supplement figure 1*). In this region CTCF-binding sites, DNase1
514 HS, and standard SATB1 ChIP-seq peaks all converge in the gene-rich inter-LAD regions, whereas
515 LADs largely overlap with SATB1 urea ChIP-seq peaks and BUR distributions.
516 We then examined the relationship between BUR and LAD distribution at the genome-wide
517 scale. We compared the genomic proximity to lamin B1 (as measured by Dam ID), SATB1-bound
518 genomic sites (from standard ChIP-seq), SATB1-bound BURs (from urea ChIP-seq) over LADs and
519 flanking regions (+/- 10% outside of the LADs). We found that the genome-wide distribution of LADs
520 coincides with the majority of SATB1-bound BURs, but it is mutually exclusive with SATB1-bound
521 genomic sites that are mapped by standard ChIP-seq, which avoid LADs (*Figure 7B*). In agreement
522 with the result shown in *Figure2A*, CTCF-bound sites mapped by urea-ChIP-seq coincide with those
523 mapped by standard ChIP-seq and are found at LAD borders and inter-LAD regions, as previously
524 described(*9*)(*Figure 7C*). Collectively, these analyses confirmed that the majority of potential BURs
525 from the BUR reference map as well as directly SATB1-bound BURs overlap with LADs (Figure 7B and
526 C), whereas CTCF-binding sites avoid LAD-BUR regions and are confined to inter-LADs and LAD
527 boundaries (*Figure 7C*). We conclude that, at the genome-wide scale, the majority of BURs overlap
528 with LADs, suggesting that BURs represent as an important sequence component of LADs, some of
529 which are specifically targeted by SATB1. This agrees with earlier findings that LADs are enriched in
530 AT sequences as many BURs are AT-rich (>65% AT) and further identifies LADS as specifically
531 enriched in BURs.
532 Within larger LAD domains identified, there are smaller regions (<10kb) that display localized
533 looping away from the lamina (*78-80*). These DiPS (Depleted in Peripheral Signal) display all the
534 hallmarks of poised or active promoters or enhancers and are flanked by CTCF and cohesin. Thus,
535 these small regions embedded in larger LADs, are akin to small inter-LADs that are constrained at the
536 lamina due to their short lengths. Given that SATB1 has been shown to have both repressive and

537 activator functions in genome regulation, we next asked if SATB1 binds BURs in these active regions
538 within LADs. We examined the distribution of average signals of SATB1-bound BURs relative to
539 enhancers within LADs marked by H3K4me1 and H3K27ac modifications over a +/- 3kb window
540 centered at H3K4me1 peaks. The peaks of enhancer histone marks (H3K27ac and H3K4me1)
541 corresponded to minimum occupancies of SATB1 within LADs and also excluded BURs, indicating that
542 enhancers are largely excluded from BURs and SATB1-directly bound BURs within LADs (*Figure 7D*).
543 We also examined SATB1-bound BURs outside LADs (inter-LADs) as well. Although less-
544 densely distributed compared to LADs, SATB1-bound BURs are detected in inter-LAD regions enriched
545 in actively transcribed genes. SATB1-bound BUR distribution in representative inter-LADs containing
546 *Rag1/Rag2, Foxp3,* and *Runx3* relative to nearby LADs are shown (*Figure 7—supplement figure*
547 *2*). Genome-wide analyses showed that in inter-LAD regions, SATB1-bound BURs are generally
548 excluding the peaks of enhancers marked by H3K4me1 and H3K27ac (Figure 7E). Although the
549 positions are not precisely coincident, ~10% of SATB1-bound BURs in inter-LADs are located close
550 (within +/- 5kb) to the peaks of H3K4me1 or H3K27ac. We also examined the distribution of BUR-1 or
551 BUR-2 interactions relative to LADs. These SATB1-mediated interactions were found to be well
552 constrained within inter-LADs enriched in histone marks for enhancers *(Figure 7—supplement figure*
553 *3).*
554

555 # Discussion

556 Here we used a systematic approach to compare and interrogate the chromatin binding profiles of a
557 genome organizing protein SATB1 detected by two alternative ChIP-seq approaches. Critically, by
558 establishing a urea purification-based ChIP-seq method that detects direct DNA-binding sites of nuclear
559 proteins, we found that SATB1 directly binds BURs genome-wide in stark contrast to previously
560 identified enhancers and promoters in the open chromatin by standard ChIP-seq. Our results revealed
561 that the two distinct DNA-binding profiles for SATB1, produced by either direct or indirect binding, are
562 both real and that SATB1 is required for megabase-level interactions, connecting BURs to open
563 chromatin associated with cell-type specific gene expression regulated by SATB1 in thymocytes.
564 By urea ChIP-seq that uses stringently purified crosslinked chromatin to map DNA binding profiles
565 of SATB1, we found a new type of chromatin organization formed by direct binding of SATB1 with BURs,
566 genomic elements distributed across the mouse genome. This SATB1-mediated chromatin organization
567 contrasts with the previous held notion generated by the standard ChIP-seq that SATB1 mainly interacts
568 with enhancers and promoters located within the "open chromatin" regions. SATB1-binding sites
569 generated by standard ChIP-seq or CUT&Tag co-map with CTCF-binding sites and regulatory
570 sequences (enhancers marked by H3K27ac and H3Kme1) within DNase1 HS (*53-55*). However,
571 SATB1's direct binding sites identified by urea ChIP-seq consistently avoids these regions in "open
572 chromatin". These SATB1 binding sites were mutually exclusive and less than 1% of SATB1 peaks from

573 standard ChIP-seq were found in urea ChIP-seq (*supplement Table 1*). This striking difference in
574 SATB1-binding profiles is due to the differences in the ChIP-seq protocols used. Standard ChIP-seq
575 protocols use crosslinked whole cells or nuclei without purification of crosslinked chromatin, and
576 immunoprecipitation is typically performed with the spun-down fraction of sonicated crosslinked
577 chromatin, removing most of the insoluble debris. We envision that SATB1 bound to inaccessible
578 nuclear regions may be lost in the insoluble fraction, while SATB1 bound to open chromatin regions,
579 albeit indirectly, may become the predominantly captured fraction that is amplified during PCR-based
580 library preparations. Thus, SATB1 directly bound to BURs in the high-salt extraction resistant nuclear
581 substructure is likely missed by standard ChIP-seq (*53, 54*) and by CUT&Tag (*55*), which is also known
582 to preferentially detect protein interactions in open chromatin regions. In contrast, in urea ChIP-seq,
583 whole, intact crosslinked chromatin is purified through urea ultracentrifugation, first removing all
584 indirectly bound proteins from chromatin. Subsequently, the entire urea-purified crosslinked chromatin
585 is quantitatively solubilized. The two main advantages of urea ChIP-seq are that (1) only direct-binding
586 profiles will be obtained for a protein of interest, and (2) chromatin tightly associated with nuclear
587 substructures, which are typically discarded in the insoluble fractions in standard ChIP-seq, are retained
588 as whole intact crosslinked genomic DNA in urea ChIP-seq. Therefore, all chromatin binding profiles
589 can be obtained, regardless of the original chromatin state (i.e. no bias toward open chromatin). Thus,
590 urea ChIP-seq is uniquely suited to detect SATB1's direct binding sites genome wide. This is not the
591 case for non-BUR-binding proteins (e.g. CTCF and PRC2 core subunits) and urea ChIP-seq generate
592 identical DNA binding profiles as standard ChIP-seq.

593 CTCF is a widely studied chromatin organizer. Results obtained using urea-purified crosslinked
594 chromatin show no relationship between CTCF and SATB1. Several pieces of evidence support this
595 conclusion. First, genome-wide CTCF-binding profiles and Hi-C profiles remained largely unaltered by
596 SATB1 deletion in thymocyte (*Figure 3*), a finding supported by other studies (*53, 54*). Second, by urea
597 ChIP-seq, SATB1:BUR binding sites precisely exclude CTCF-binding sites, which contrasts with recent
598 findings that SATB1 and CTCF-binding profiles overlap by standard ChIP-seq(*54, 55*) or by CUT&Tag
599 (*55*).Third, chromatin interactions (>5.7Mb) from a SATB1-bound BUR are not constrained to individual
600 TADs but span multiple TADs. Fourth, by urea ChIP-Western, SATB1 does not interact with CTCF or
601 cohesin on chromatin. Therefore, chromatin organization mediated by SATB1 through direct binding to
602 BURs is distinct from that mediated by CTCF and cohesin.

603 We also identify BURs as genomic elements enriched in LADs located proximal to the nuclear
604 lamina. LADs and LAD borders have critical roles in gene regulation (e.g. the *Tcrb* locus) (*2, 81*). Except
605 for being generally AT-rich, genomic features of LADs have thus far been poorly characterized. BURs
606 are not simply any AT-rich sequence, but rather, are characterized by a specific ATC sequence context
607 that is specifically targeted by SATB1 having a role in cell-type specific gene expression(*41*). Thus,
608 discovering BUR enrichment in LADs uncovers mechanistic insights into the biological function of LADs
609 in gene regulation. Although LADs form heterochromatin proximal to the nuclear envelope, some LAD

610 organization is dynamic. For instance, during differentiation of mouse cells, hundreds of genes display
611 dynamic intranuclear spatial repositioning relative to the nuclear lamina (*82*). Some LAD regions
612 dynamically contact the nuclear lamina, and gene movement in and out of the nuclear periphery in a
613 circadian rhythm has been reported (*83*). Recently, relocation of the *hunchback* gene in *Drosophila*
614 neuroblasts to nuclear lamina has been shown to require polycomb proteins binding to a specific intronic
615 element of the gene locus. This process is important for heritable gene silencing to regulate neuroblast
616 competence (*84*). The organization of genomic DNA at the lamina has been linked to both chromatin
617 state and the nuclear lamins, along with other nuclear membrane proteins (*2, 85-90*). While chromatin
618 state is likely important for both positioning at and de-localization from the lamina, little information is
619 available concerning *in vivo* mechanisms specifying which regions within LADs could be targeted for
620 de-localization and subsequent cell-type specific gene activation. The SATB1 protein network is found
621 mostly in the nuclear interior of thymocytes where many individually cloned BURs have been detected
622 by DNA-FISH. Among such interiorly localized BURs, a BUR 5' of *Myc* was confirmed to co-localize
623 with SATB1 by immunoDNA-FISH in thymocytes (*39, 40*). Interaction of select BURs with SATB1 may,
624 thus, provide a mechanism by which a subset of genes (e.g. *Gabrg1 and Gabra2* in neurons, Figure
625 7A) within LADs can be re-localized to the nuclear interior. A closer look at immunostaining profiles of
626 SATB1 distribution in thymocyte nuclei, however, shows that a small reproducible fraction of SATB1
627 resides in the peripheral zone of nuclei and in some cells the tip of the SATB1 network extends to and
628 touches this region of the nucleus(*39*). This supports our data showing that SATB1 binds selected BURs
629 within LADs at the nuclear periphery. In addition to LADs, we note that inter-LAD regions also contain
630 BURs, albeit at a considerably lower density compared to LADs. Some of these BURs are directly
631 SATB1 bound and located near enhancers. As we show for an inter-LAD containing the *Rag1/Rag2*
632 locus, for instance, we speculate that these BURs in inter-LADs, enriched in actively transcribing genes,
633 may be important for cell-type specific gene regulation. Given the unique network-like distribution of
634 SATB1, we envision that some transcriptionally active gene loci in the inter-LADs are pinned down to
635 the SATB1 nuclear architecture via SATB1 binding to BURs. How the subset of BURs that interact with
636 SATB1 are determined, and how SATB1 regulates gene expression in the nuclear interior or the
637 periphery, and ultimately, how these interactions are regulated in distinct SATB1-expressing cell types
638 are important topics for future research.

639 It is of interest that the number of BURs targeted by SATB1 greatly varies depending on cell type,
640 where neurons have by far the greatest number of BURs captured by SATB1. This may reflect the level
641 of plasticity for gene transcription or number of genes that must be dynamically regulated in each cell
642 type in which SATB1 is expressed. There is growing evidence supporting the critical roles of chromatin
643 topology and nuclear architecture in neuronal response to external stimuli to establish precise
644 expression program associated with cognition (*91*). In fact, SATB1 has roles in early postnatal cortical
645 neuron development and neuronal connectivity (*37, 92*). The exceptionally large number of BURs
646 directly interacting with SATB1 in cortical neurons may reflect specialized function of neurons that might

647 require highly dynamic transcriptional regulation to enable rapid response to a wide variety of
648 environmental stimuli.

649 Besides BURs, SATB1 may also bind to other genomic loci as suggested by standard ChIP-seq.
650 Standard ChIP-seq is known to generate "phantom" or false-positive peaks (*56-60*). Although we
651 detected such peaks using SATB1-KO cells by standard ChIP-seq (Table 3 and 4), under our
652 experimental condition, we confirmed that the majority of SATB1-binding peaks are SATB1-dependent.
653 Importantly, these SATB1 peaks from standard ChIP-seq were essentially undetected by urea ChIP-
654 seq. Therefore, it is likely that SATB1 indirectly binds to these sites by associating with nuclear proteins
655 bound to their target sites in open chromatin. SATB1 is equipped with domains responsible for specific
656 binding to BURs (*93-96*) as well as for associating with and recruiting chromatin binding proteins to
657 specific genomic sites (e.g. CHRAC and ACF1 nucleosome mobilizing complex subunits, Brg1, p300,
658 Hdac1, Gata3, Stat6, c-Maf, Pit1, etc) (*39, 42, 44, 50, 97-101*). Therefore, we anticipate that SATB1
659 binds indirectly and dynamically to regulatory regions in open chromatin by associating with many
660 chromatin-modifying and transcription factors to form protein complexes to support tissue-specific gene
661 expression.

662 Considering that the two contrasting SATB1-binding profiles are both real, an important question
663 emerges as to whether there is a functional or structural link between the two chromatin regions. We
664 conceive that SATB1 might have a role in bringing together chromatin structure formed by BURs tightly
665 bound to nuclear substructure and a highly accessible "open" chromatin enriched in transcription related
666 proteins. Our urea 4C-seq results suggest that these two regions are connected via SATB1-mediated
667 interactions involving BURs. The ligation step in urea 4C-seq enabled us to capture chromatin sites that
668 were in close spatial proximity to the bait site (e.g. a BUR) at the time of crosslinking. Apparently, 3D
669 structure fixed by formaldehyde retained sufficient chromatin proximity after urea purification to enable
670 capture of BUR-interacting fragments. The urea 4C-seq revealed that a SATB1-bound BUR site (BUR-
671 1 as a bait), located at the border of a gene desert and a gene-rich region containing SATB1-regulated
672 genes (*Rag 1* and *Rag2*), interacts extensively over a 5.8Mb gene-rich region with many gene loci, such
673 as enhancers, promoters, BURs and intergenic regions. Such ultra-long-distance interactions from
674 BUR-1 predominantly cover the entire 5.8Mb gene-rich region, mostly avoiding the gene desert region.

675 The SATB1-mediated chromatin interactions from BUR-1 and BUR-2 (validation control) as
676 baits, covering the 5.8Mb *Rag1/Rag2*-containing gene-rich region, are tightly linked to SATB1-
677 dependent expression of eight genes including *Rag1* and *Rag2* ($P<0.05$) out of 43 genes in this region
678 in thymocytes. BUR-1 further interacts with far distal gene-rich regions *in cis* in chromosome 2 (>44Mb),
679 skipping interstitial gene desert regions in a SATB1-dependent manner, albeit through weaker
680 interactions (*Figure 5—supplement figure 3*). These chromatin interactions are at a much greater scale
681 than those previously observed, such as T helper 2 cell activation induced dense chromatin looping in
682 the 200kb cytokine gene cluster, including several BURs and regulatory regions, tightly linked to
683 activation of the cytokine genes (*42*) and interactions at the subTAD levels (or <~500Kb) between

684 specific enhancers or enhancers to promoters in thymocytes (*53-55*). Combining results from SATB1-
685 binding profiles derived from urea ChIP-seq and standard ChIP-seq as well as urea 4C-seq, a model is
686 illustrated how the two chromatin regions might interact depending on SATB1 (*Figure 8*). In this model,
687 the stable SATB1 binding to BURs in the inaccessible nuclear substructure forms a foundational
688 chromatin scaffold and SATB1 tethers highly accessible "open" chromatin regions to this scaffold by
689 binding to regulatory protein complexes. We anticipate that the latter interactions to be dynamic and
690 form the highly interacting chromatin architecture over gene-rich inter-LADs, bringing regulatory
691 sequences in close vicinities. Both components of SATB1-mediated chromatin organization likely
692 underlie regulation of gene expression. SATB1 proteins may be heterogeneous with regards to protein
693 mobility. Consistent with the results that some SATB1 proteins can be extracted by high salt (*40*), we
694 observed at least two groups of SATB1 proteins based on mobility, one essentially immobile and the
695 other with high mobility by fluorescence recovery after photobleaching (FRAP) assay in thymocytes (our
696 unpublished results). Future studies are required to understand how SATB1-mediated 3D chromatin
697 structure is formed. We need to examine the potential role of the soluble (highly mobile) SATB1 proteins
698 act as a bridge connecting the SATB1-bound BURs on the nuclear substructure to "open" chromatin.
699 Although the current study focused on SATB1 targeting select BURs to form 3D chromatin structures
700 in the inter-LADs, it is of great interest to study how SATB1 binding to BURs within LAD affects 3D
701 chromatin structure. We observed specific RNA to remain associated with crosslinked chromatin after
702 urea purification (unpublished). Since RNA is known to play an important role in the maintenance of
703 3D chromatin structure by acting as a scaffold to facilitate chromatin loop formation(*102, 103*),it would
704 be important to address the role of RNA in SATB1-mediated chromatin organization.

705 Importantly, SATB1 is not the only BUR-binding protein, indicating the broader biological
706 significance of BURs as genomic elements. Although SATB1 confers by far the highest affinity to BURs,
707 several other nuclear factors (e.g. nucleolin, HMGI/Y, PARP1, Ku70/86 heterodimer) have also been
708 verified to directly and specifically bind BURs *in vitro* (*76, 104, 105*). Recently, we identified Tip5/Baz2a
709 as an additional strong BUR-direct binding protein from mouse embryonic stem cells (mESCs) that is
710 essential for pluripotency (submitted for publication). Similar to SATB1, this nuclear protein also binds
711 specifically to BURs in mESCs, and identification of BURs as its direct targets in mESCs requires urea
712 ChIP-seq as well. Despite the identical BUR binding specificity of SATB1 and this stem cell factor, these
713 two proteins confer entirely different functions in distinct cell types. Therefore, BURs may serve as core
714 genomic marks that can be selectively targeted by distinct BUR-binding proteins to form unique 3D
715 chromatin regulatory network depending on the BUR-binding protein for specific cell functions. BURs
716 in the mammalian genome may have similarities in function with the "tethering elements" in *Drosophila*
717 (*106, 107*).These elements can be bound by regulators such as pioneering factors and polycomb
718 proteins and were shown promote interactions of regulatory regions up to ~250kb within TADs to
719 regulate transcription of genes. Most recently, meta-loops constituting a meta-domain spanning
720 numerous TADs in CNS of *Drosophila* were reported (*108, 109*). Although meta-loops are similar in size

721 with SATB1-mediated chromatin interactions, the CNS meta-loops have different features. They are
722 formed by transcription factors (e.g. CTCF and GAF), the half of them is anchored at the TAD borders,
723 and the metadomain boundaries are CTCF-bound. It appears that clusters of very long chromatin
724 looping events linked to gene expression may form by multiple mechanisms across species. Future
725 research on BUR-mediated genome organization regulated by distinct BUR-binding proteins will likely
726 provide more in-depth understanding about cell-type specific mechanisms driving 3D chromatin
727 architecture that contributes to major changes in gene expression and cell phenotypes.
728
729
730

731 Materials and Methods

732

733 Experimental design

734

735 Primary cells

736 All animal studies were performed under Institutional Animal Care & Use Program at University of
737 California San Francisco for collecting brain and thymus and the Boston University for collecting skin.
738 Thymi were isolated from 2 weeks-old *Satb1*$^{+/+}$ and *Satb1*$^{-/-}$ mice, as well as from *Satb1*$^{+/+}$:*Cd4*-cre and
739 *Satb1* $^{F/F}$:*Cd4*-cre mice. Brain was isolated from 6-10 weeks-old *Satb1*$^{+/+}$ mice. Thymocytes were
740 prepared from thymi and cortex was excised from brain for this study. Skin was isolated from postnatal
741 day 0.5 of C57Bl/6 mice.
742

743 Antibodies

744 Antibodies used and for their purposes are the followings. CTCF for ChIP-seq: Active MOTIF (61311.
745 lot# 34614003), CTCF for Western: Millipore (07-729; lot#-3429114), RAD21 for ChIP-seq and Western:
746 ABclonal (A18850, lot# 3521809001), SATB1 for ChIP-seq and Western: rabbit polyclonal 1583
747 prepared by Kohwi-Shigematsu's laboratory and abcam (ab109122, lot#GR137720-13) and Histone H3
748 for Western: abcam (ab1791, lotGR3236305-2).
749

750 Urea-purified chromatin immunoprecipitation and urea ChIP-seq

751 Thymocytes and single cell suspension prepared for cortex and skin were cross-linked with 1%
752 formaldehyde for 10 min at room temperature, followed by addition of glycine at 0.125 M for 5 min to
753 quench cross-linking reaction. These cells were washed with 1X PBS, suspended in 1% BSA-PBS and
754 cells were precipitated in a tube. These cells were re-suspended in lysis buffer (4% SDS, 50 mM Tris-
755 HCl [pH 8.0], 10 mM EDTA, 100 mM NaCl) with Protease inhibitor (Roche) and RNase inhibitor. The
756 resulting cell lysate was incubated for 30 min at room temperature and loaded on top of 8 M urea followed

757 by ultracentrifugation in a Beckman TL-100 Ultracentrifuge for 4-5 hrs at 45K rpm (181,000g, g=relative
758 centrifugation force or RCF; TLS55 rotor). Recovered chromatin pellet was sonicated to generate
759 fragment size ranging 300–1,1,200bp using Branson sonifier (20% power, 10 sec on and 30 sec off for
760 4-6 cycles) in 300μl of sonication buffer [0.5% Na-laurylsarcosine, 0.1% deoxycholate (DOC), 1mM
761 EDTA, 0.5mM EGTA, 10mM TrisHCl, 100mM NaCl] with Protease inhibitor and RNase inhibitor. Note:
762 urea-purified crosslinked chromatin is highly vulnerable to sonication, therefore, it needs to adjust the
763 sonication condition depending on cell type and/or the size of the precipitate DNA. The sonicated
764 chromatin with 1% Triton X-100 were dialyzed against dialysis buffer (10 mM Tris-HCl [pH 8.0], 100 mM
765 NaCl, 1 mM EDTA, 5% glycerol), and dialyzed for 16 hrs at 4°C (this dialysis is optional). The chromatin
766 sample was centrifuged at 15,000g for 5 min at 4°C to remove a trace amount of residual insoluble
767 components. After reverse crosslinking an aliquot of sonicated chromatin, shearing efficiency was
768 examined by Bioanalyzer 2100 (Agilent Technologies Inc.). The test-reverse cross-linking of chromatin
769 aliquot (5-10μl of the final volume) was done at 65°C for 2-3 hrs in buffer containing 1% SDS, 100 mM
770 NaCl, 1 mM EDTA, 10mMTris. Samples were further treated by 100 μg/ml RNase A at 37°C for 1 hr,
771 200 μg/ml proteinase K at 60°C for more than 3 hrs, followed by DNA purification using QIAquick PCR
772 purification kit (Qiagen). For chromatin-immunoprecipitation, 1 μg of an antibody of choice were added
773 to 10 ~ 20 μg of sonicated solubilized chromatin in 350 μl in dilution buffer (1% Triton X-100, 0.01% SDS,
774 1 mM EDTA, 15 mM Tris-HCl, 100 mM NaCl) and incubated overnight with rotation at 4°C. At least three
775 chromatin samples were prepared for ChIP-seq from independent experiments to confirm the results.
776 The chromatin fragments bound to antibodies (after incubation overnight) were captured using 20 μl of
777 protein-A, protein-G or protein AG Dynabeads (Invitrogen; pre-washed with 0.1% BSA [Fraction V,
778 Sigma), by 2 hrs incubation time at 4°C. Dynabeads bound to chromatin fragments were washed twice
779 with Binding buffer (10 mM Tris, pH8.0, 1 mM EDTA, 1 mM EGTA, 0.5% Triton X-100, 0.1% DOC, 100
780 mM NaCl), twice with LiCl buffer (100 mM Tris [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, and
781 0.5% DOC). The tubes were changed and washed twice with 0.1M NaCl-TE buffer. The
782 immunoprecipitated samples were eluted from Dynabeads by elution buffer (10 mM Tris-HCl [pH 8.0],
783 1 mM EDTA, 100 mM NaCl and 1% SDS) at 65ºC for 20 min. After elution, ChIP DNAs was reverse-
784 crosslinked at 65ºC for 4hrs, followed by RNase-A treatment at 37°C for 1 hr and proteinase K treatments
785 for 3 hrs at 60°C. The recovered DNA was further purified with a QIAquick PCR purification kit (Qiagen),
786 with elution in 15μl of Qiagen Elution Buffer (alternatively, Phenol-chloroform extraction and ethanol
787 precipitation). Briefly for library construction, ChIP DNA fragments were end-repaired and ligated to
788 indexed sequencing adapters (NEB or Rubicon kit) and amplified with 8-11 rounds of PCR, depending
789 on the DNA amount. Using SPRI beads (Beckmancoulter), small fragments (<200bp) were removed
790 from the amplified library. The size-selected library was sequenced to 50bp single-end read lengths on
791 an Illumina HiSeq4000 (Vincent J. Coates Genomic Sequencing Laboratory, UC Berkeley).
792
793 **Urea Chromatin immunoprecipitation (urea ChIP)-Western analysis**

794 Urea-purified chromatin prepared as described above (~10μg of chromatin-DNA) was
795 Immunoprecipitated with indicated antibody to produce urea ChIP samples, and the chromatin fragments
796 bound by the antibody was trapped by Protein-A beads. The antibody-bound chromatin fragment-trapped
797 beads were washed with binding buffer/washing buffer/NaCL-TE (total 6 times), and the chromatin
798 fragments were released and isolated from beads (65°C for 20 min) in 100 μl of 1% SDS, 10 mM Tris,
799 pH7.6, 1mM EDTA, 100mM NaCl solution. The isolated chromatin fragments were subjected to reverse-
800 crosslinking by incubation at 65°C for additional 4 hrs. The released proteins were dissolved in the
801 Laemmli buffer, boiled and subjected to Western blot assay. To compare multiple samples, we confirmed
802 that we used a similar amount of original urea-purified crosslinked chromatin before ChIP by subjecting
803 each chromatin sample to Western blot using anti-H3 antibody. The similar amount of urea-purified
804 crosslinked chromatin was then used for ChIP-Western studies.
805
806 **Urea 4C-seq**
807 Urea-purified crosslinked chromatin (described above for its preparation) was processed for urea-4C-
808 seq basically as described(*62*) with some modifications. For the current experiment, *Hind*III was used
809 to digest chromatin (10-20 units/μg DNA) after urea centrifugation overnight at 37°C. DNA
810 concentration was determined by reverse-crosslinking. Briefly, after *Hind*III digestion, *Hind*III was heat
811 inactivated at 80°C for 20 min, ligation was performed in 1.0 ml with 1X ligation buffer with ligase (600
812 units) at 16°C for overnight, followed by reverse-crosslinking at 65°C for overnight. The chromatin DNAs
813 were purified by Phenol-CHCl3 extraction (or spin column, Quagen), and the purified DNAs were 1-4
814 μg and were further digested with *Hae*III (100-125 units) overnight. The *Hind*III-*Hae*III digested DNA
815 fragments were treated with Klenow. The resulting blunt-end *Hind*III-*Hae*III DNA fragments were ligated
816 with our custom double-stranded (ds) Linker DNA with one strand-Biotin modified,
817 Linker-top sequence with Biotin modified (marked by *):
818 5'-AATTCGGTACCTCTAGAGA**T***A**T***CCGATCGCTCGAG(XhoI)AAGCTT(*Hind*III)-3';
819 Linker-bottom sequence: 5'-AAGCTTCTCGAGCGATCGGATATCTCTAGAGGTACCGAATT-3'
820 as described (*62*). The Linker-ligated dsDNA fragments were purified by spin column to remove free
821 Linker. The Linker-ligated dsDNA fragments were further digested with *Xho*I at the linker site (next site
822 from *Hind*III). Single stranded DNAs were isolated from the *Xho*I digested DNA fragments by coupling
823 to StreptAvidin Dynabeads (Invitrogen, Dynabeads M-280 streptavidin) followed by incubation in 40μl of
824 0.15M NaOH for 10 min at 25°C, followed by neutralization. Only single-stranded DNA, containing biotin
825 coupled with streptavidin, were retained on the Dynabeads, releasing the non-biotin modified single-
826 stranded DNA (single strand DNA capture) (*62*). The released non-biotin modified single strand DNAs
827 were amplified by PCR (AmpliTaq Gold system, AppliedBiosystems); 10 cycles at 94C for 30", 55°C for
828 60" and 72°C for 60", followed by 15-20 cycles at 94°C for 30", 65°C for 60", 72°C for 60", using the
829 custom Illumina-bait oligomers and Illumina-linker oligomer for library preparations as described below.
830 The PCR amplified DNAs were purified by high and low size selection by beads (SPRIselect, Beckman

831 Coulter). The purified DNAs were analyzed by Bioanalyzer to confirm their average size of ~300-500 bp
832 with mostly uniform distribution, followed by the Illumina sequencing using custom bait-specific sequence
833 oligomers and Linker specific barcode oligomer. We used two bait sequences: 5'
834 gaagggggtaggaaagaaagctt 3' (BUR-1 bait: adjacent to a SATB1-bound BUR, located1.4 kb from ASE)
835 and 5' cgactcattcctactgaagctt 3' (BUR-2-bait: enhancer located in the second intron of B230118H07Rik
836 close to a SATB1-bound BUR), both including a *Hind*III site.
837 The library was constructed by amplification using either Illumina-AS-bait or Illumina-IU bait primer with
838 Illumina-linker oligomer. (Capital letters represent Illumina sequence, lower case letters represent bait
839 sequences or the linker sequence)
840 Illumina-BUR-1-bait oligomer:
841 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC-gaagggggtaggaaagaaagctt
842 Illumina-BUR-2-bait oligomer:
843 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC-cgactcattcctactgaagctt
844 Illumina-linker oligomer: 5'-CAAGCAGAAGACGGCATACGAGAT (barcode; e.g. CGTGAT)
845 GTGACTGGAGTTC-gcgatcggatatctctagaggtaccgaattc
846 To validate successful experiments, we confirmed that no PCR amplified product was generated by
847 using either the bait oligomer or linker oligomer alone.
848 For the library construction, Taq-Gold polymerase (NEB) was used for PCR by the following condition:
849 94°C for 5-10min (one cycle), 94C for 30 seconds, 55C for 30 seconds, 72C for 60 seconds (for 5-10
850 cycles), 94°C for 30 second, 65°C for 30 seconds, 72°C for 60 seconds (for 20 cycles), followed by 72°C
851 for 10 min. PCR products were purified by Phenol-CHCl3 extraction twice, CHCl3 extraction once,
852 followed by EtOH precipitation (Optional; PCR purification by spin column, Qiagen).
853 The 4C libraries were sequenced by Illumina-HiSeq4000, using BUR-1-bait sequence and BUR-2-bait
854 sequence as custom sequence oligomers as follows:
855 BUR-1-bait sequence oligomer: 5'-CTTTCCCTACACGAC-gaagggggtaggaaagaaagctt
856 BUR-2-bait sequence oligomer: 5'-ACTCTTTCCCTACACGAC-cgactcattcctactgaagctt
857 Each library sequence was identified with Linker specific barcode oligomer;
858 5'-gaattcggtacctctagagatatccgat-GAACTCCAGTCAC. (Capital letters represent Illumina sequence and
859 lower-case letters represent linker sequence). For sequencing, the above custom barcode sequence
860 oligomer was required instead of the common Illumina barcode sequencing oligomer.
861

862 **Hi-C**

863 Hi-C was performed as described (*17*). Primary thymocyte samples from *Satb1*$^{+/+}$:*Cd4*-cre and *Satb1*
864 $^{F/F}$:*Cd4*-cre mice were formaldehyde crosslinked at 2% at room temperature for 10 min. Two biological
865 replica per sample were prepared for Hi-C. Hi-C libraries was prepared using the Arima HiC^{+} kit (Arima
866 Genomics) exactly following the associated protocol. Paired-end sequencing was performed by
867 Illumina NovaSeq 6000.

868

869 ## Statistical analysis

870 ### Significance of overlap of two sets of genomic features

871 We used the *bootRanges* function that is part of the *nullranges* bioconductor package (*63*). It was used
872 to compute the significance of overlap of two sets of genomic regions (e.g., BUR versus the called
873 peaks for each of the ChIP-seq data sets). This method is based on subsampling genomic regions of a
874 given size from within genomic segments of the same state. Here the mouse genome mm9 was
875 segmented into regions capturing 3 states based on density of known genes using the segmentDensity
876 function using the CBS method 4 (*110*) and assuming a segment length of 2Mb. Regions represented
877 blacklisted regions (queried using the terms "excluderanges" and "mm9") in the AnnotationHub package
878 were excluded from the analysis. For each comparison [reference regions (known BUR peak calls)
879 versus query regions (SATB1 Urea or Standard Chip-seq peak calls), the observed number of
880 overlapping regions is first computed, 1000 sets of random query region peak calls were sampled given
881 the estimated genome segmentation using blocks of length 300kb, the overlaps for each of these 1000
882 random query regions with the reference regions were then computed and finally, the estimated p-value
883 is based on the fraction of 1000 random overlaps greater than or equal to the observed number of
884 overlaps.

885

886 #### Hi-C data processing

887 Hi-C datasets were analyzed using the HiC-bench platform (*111*). The application of this method is
888 described in detail(*112*). Briefly, data were aligned against the mouse reference genome (mm10) by
889 bowtie2 (version 2.3.1)(*113*) with mostly default parameters (specific settings: --very-sensitive-local --
890 local). For Hi-C, aligned reads were filtered by the GenomicTools(*114*) tools-hic filter command
891 integrated in HiC-bench. Interaction matrices for each chromosome separately were created by the
892 HiC-bench platform at a 10kb resolution. The Hi-C filtered contact matrices were normalized using
893 "iterative correction and eigenvector decomposition" (ICE) algorithm (*115*) build into HiC-bench. To
894 account for variances of read counts of more distant loci, distance normalization for each matrix was
895 performed as recently described(*116*). TADs were called using the algorithm developed within HiC-
896 bench(*111*) at 10kb bin resolution with the insulating window of 100kb. For visualization of Hi-C data, we
897 generated HiC files using the Juicer(*117*).

898

899 ### Quality Control

900 The total numbers of reads per biological replicate for each genotype ranged from 200 to 230 million
901 reads. The percentage of reads aligned was >98% in all samples. The percentage of accepted reads
902 (ds-accepted-intra" and "ds-accepted-inter") was in the range of 50% to 55.5%. The absolute number

903 of accepted intra-chromosomal read-pairs were ~90-100 million reads for each sample. We analyzed
904 data from individual samples as well as combined data from two replicates (~200 million reads) for each
905 genotype.
906 Scaling plot was performed by aggregating the counts per 10kb distance bins. Values were
907 cpm-normalsized by dividing by the total number of valid intra-chromosomal pairs.
908
909 **Compartments**
910 Compartment analysis was performed using the Homer pipeline (v4.6)(*118*). Homer performs a principal
911 component analysis of the normalized interaction matrices and uses the PCA1 component (cis-
912 Eigenvector 1) to predict regions of transcriptionally active (A compartments) and transcriptionally
913 inactive chromatin (B compartments). Homer works under the assumption that gene-rich regions with
914 active chromatin marks have similar PC1 values, while gene deserts show the opposite
915 (http://homer.ucsd.edu/homer/interactions/HiCpca.html). HiC filtered matrices were given as input to
916 Homer together with a list of house-keeping genes for compartment prediction. The house-keeping
917 genes coordinates are used by Homer as prior information of active regions. We generated density
918 plots to compare the cis-eigenvector 1 values of the knock-out and wild-type samples with 10kb genomic
919 bins. Pearson correlation coefficients were calculated ('Cor' function in R) between the counts for
920 every pair of 10kb bins on chromosome 2. Heatmaps representing the Pearson correlation of interactions
921 of chromosome 2 for WT and KO thymocytes are displayed by juice box (500kb resolution).

922 **urea ChIP-seq analysis**

923 Sequenced reads were quality-checked using FastQC prior to downstream analysis. Unaligned reads
924 were mapped to the mouse reference genome (NCBI37/mm9) using bowtie2 (version 2.3.1)(*113*)
925 allowing for three mismatches in the seed region and only unique matches. Candidate binding sites were
926 called using MACS2 (v2.2.7)(*119*) using a q-value threshold of 0.01. Coverage tracks are presented in
927 RPM (read pileups per million reads) and utilities from UCSC were used for creating browser tracks (*120*).
928 To determine intersection of peaks across ChIP-seq samples, deepTools(*75*) was used.

929 **urea 4C-seq analysis**

930 Reads were mapped with Bowtie, then assigned to *HindIII* sites with the same utility tools used in (*69*).
931 These same tools also normalized the counts to per million intrachromosomal reads and removed reads
932 to restriction fragments comprising more than 2% of the total reads (namely the contiguous adjacent
933 restriction fragment corresponding to undigested chromatin during the urea 4C, and a rare (<5) numbers
934 of fragments with extreme PCR duplications). The generated bedGraph files and these were converted
935 to bigWig format using the bedGraphToBigWig program for direct visualization of the peaks on the
936 UCSC browser.

937

938 **Analysis of SATB1 bound sites and genomic features**

939 LADs were identified as genomic intervals using the previously described LADetector algorithm
940 (https://github.com/thereddylab/pyLAD) or sequencing data
941 (https://github.com/thereddylab/LADetector) (*89, 121*).

942 SATB1 urea-ChIP data from thymocytes were tested for statistically significant overlap of LADs using
943 the GenometriCorr package(*122*). The GenometriCorr package applies multiple spatial tests of
944 independence including a relative distance test (i.e., if two elements such as LADs and SatB1 sites are
945 spaced more often than expected at a certain relative distance from each other), projection test (i.e.,
946 testing for significant overlap between these two positions assuming they represent single points
947 points), and the Jaccard test (testing for correlation of these two positions assuming that they are not
948 points but instead occupy some interval of the genome).

949 Genome wide averages of intensities (relative score) of DamID and ChIP-seq peak data were generated
950 over interLAD (10% of the size of the adjacent LAD)/border flanked LADs scaled to the same relative
951 size using the Genomation R package(*123*). The average intensities were then normalized to fit a scale
952 of -1 to 1 and plotted over the LADs (normalized relative score).

953 To plot genomic features in the vicinity of enhancers, H3K4me1 peaks outside of LADs were
954 filtered by subtracting away peaks that intersected LADs using the BEDTools (*124*). To plot genomic
955 features in the vicinity DiPS in LADs, regions within LADs and DiPS were identified by LADetector. DiPS
956 were then mapped. Genome wide averages of the respective ChIPseq/DamID signal data were then
957 generated within a +/- 3kb window centered at H3K4me1 peaks, using Genomation R package and
958 plotted on a vertical scale ranging from -1 to 1.

959

960 **Figure legends**

961 **Figure 1 Scheme of urea ChIP-seq and difference from standard ChIP-seq**

962 A) An illustration for a BUR anchored to the high-salt extraction resistant SATB1 nuclear architecture
963 (*39*). DNA-FISH signals (red dots) for a cloned SATB1-bound DNA in both alleles are shown in
964 *Satb1*$^{+/+}$(WT) and *Satb1*$^{-/-}$(KO) thymocytes on slides treated with 2M NaCl solution to produce DNA
965 halos around the nuclei (blue loops). SATB1 and the SATB1-bound DNA remain in the residual nucleus
966 in WT thymocytes after salt treatment. However, in KO thymocytes the DNA regions normally bound
967 by SATB1 are located in the DNA halos after the treatment. B) A comparative overview of the urea
968 ChIP-seq and standard ChIP-seq protocols. In urea ChIP-seq, entire crosslinked chromatin purified by
969 8M urea ultracentrifugation is solubilized (see Materials and Methods) before chromatin
970 immunoprecipitation. By contrast, in standard ChIP-seq, crosslinked chromatin in whole nuclei is
971 sonicated, centrifuged and the soluble fraction is isolated for chromatin immunoprecipitation. These
972 critical steps that are different between urea and standard ChIP-seq are highlighted in boxed texts. C)

973 Differences in the profiles of the peaks generated from urea ChIP-seq versus standard ChIP-seq. A
974 diagram is shown for ChIP-seq peaks using antibodies (α) against a DNA direct-binding protein (pink),
975 a DNA indirect-binding protein (blue), and DNA direct-binding protein that resides in the inaccessible
976 chromatin region (green) Note: This diagram illustrate a hypothetical scenario where some of the
977 proteins (green) in the inaccessible chromatin region may also reside in the accessible chromatin
978 binding indirectly to chromatin.

979 **Figure 2. Identification of BURs as direct targets of SATB1 *in vivo* by urea ChIP-seq.**

980 A) Contrasting distributions of SATB1-binding profiles determined by urea ChIP-seq (Urea) compared
981 to standard ChIP-seq (Std) in mouse thymocytes. B) A zoomed-in view of the gray-highlighted region
982 shown in (panel A). The track 3 (panels A and B) represents all potential BUR sites serving as a
983 reference for BUR sites. For both panels A and B, two independent urea and standard ChIP-seq
984 experiments show high reproducibility (tracks 1 and 2 and tracks 4 and 5). Tracks 6 and 7 [ENCODE,
985 Bring Ren's Laboratory at the Ludwig Institute for Cancer Research (LICR)] and tracks 8 and 9
986 (ENCODE, the University of Washington (UW) group]. C) The percentages of urea (Urea) and standard
987 (Std) SATB1 ChIP-seq peaks that intersect with the BUR reference map (see *supplement Table 1*). The
988 significance (p values) of the overlap in each of the four cases (urea and standard ChIP-seq samples
989 with two replicates each) with BURs is estimated as the fraction of the overlap counts greater than or
990 equal to the observed count over 1000 random bootstrapped genomic features. $p<0.001$: none of 1000
991 random fragment showed greater overlap with BURs than observed in ChIP-seq samples, $p=1$ means
992 all 1000 random fragments showed greater overlap with BURs than observed in ChIP-seq, indicating
993 avoidance of BURs (see Methods). D) Normalized genome-wide average intensity distribution profiles
994 (normalized relative score) of SATB1 binding sites determined by standard ChIP-seq, urea ChIP-seq
995 as well as BURs, over a +/- 3kb window centered at BUR peaks (top). Relative distance, projection,
996 and Jaccard tests all indicate strong correlation ($p< 0.05$) between urea ChIP-seq peaks for SATB1 and
997 BURs, but no overlap ($p< 0.05$) between standard ChIP-seq peaks for SATB1 and BUR peaks.

998 **Figure 3. Identical CTCF-binding profiles are produced by urea and standard ChIP-seq,**
999 **independent of SATB1.**

1000 A) Urea ChIP-seq (Urea) profiles for CTCF in *Satb1*$^{+/+}$ (WT) thymocytes and *Satb1*$^{-/-}$ (KO) thymocytes
1001 (track 2 and 3), and for SATB1 in WT thymocytes (track 1). These binding patterns were compared
1002 against CTCF-binding profiles generated by standard ChIP-seq (Std) from ENCODE (tracks 4-6) for
1003 testis and thymocytes (ENCODE, LICR), along with the DNase1 HS profile (ENCODE, UW) as well as
1004 the SATB1-binding profile from standard ChIP-seq (track 7). B) Genome-wide average intensity
1005 distribution (relative score) of peaks from CTCF and SATB1 urea ChIP-seq for *Satb1*$^{+/+}$ thymocytes
1006 (WT), CTCF urea ChIP-seq for *Satb1*$^{-/-}$ thymocytes (KO), and SATB1 standard ChIP-seq (WT) over a
1007 +/- 3kb window centered at CTCF peaks (WT). Relative distance, projection, and Jaccard tests all

1008 indicate strong correlation ($p < 0.05$) between CTCF urea ChIP-seq peaks for WT and KO thymocytes,
1009 between these peaks and standard ChIP-seq peaks for SATB1, but no overlap ($p < 0.05$) between
1010 SATB1 urea ChIP-seq peaks. C) Urea-ChIP-Western results for CTCF, RAD21, and SATB1, showing
1011 that SATB1 does not associate with CTCF nor RAD21 on chromatin. Also, CTCF and RAD21 co-
1012 binding to chromatin appears to be infrequent, suggested by the weak ChIP-Western signals. D) Hi-C
1013 interaction heatmaps for *Satb1*$^{+/+}$:*Cd4*-cre (WT) (top right triangle) and *Satb1* $^{F/F}$:*Cd4*-cre (KO)
1014 thymocytes (bottom left triangle) show essentially identical patterns. E) Genome-wide comparison of
1015 distributions of cis-Eigenvector 1 values shows no difference between WT and KO thymocytes (high
1016 correlation at r=0.998), indicating that genomic compartmentalization is unaffected in *Satb1* KO
1017 thymocytes. Experiments were done in two replicates each for WT and KO samples. The data shown
1018 in D and E were generated by combining data from the replicates for each condition.
1019

1020 **Figure 4. SATB1-bound sites identified by standard ChIP-seq at regulatory regions are SATB1**
1021 **dependent.**

1022 SATB1-binding profiles in immature thymocytes determined by urea ChIP-seq (Urea) and standard
1023 ChIP-seq (Std) covering three loci at *Rag1/Rag2* (top) *Foxp3* (middle), and *Runx3* (bottom) are shown.
1024 For each of the three loci, the standard SATB1 ChIP-seq generates peaks in *Satb1*$^{+/+}$ (WT) thymocytes
1025 (track 3), but not in *Satb1*$^{-/-}$ (KO) thymocytes (tracks 9 for *Rag 1/2* and *Runx3*, and tracks 11 for *Foxp3*).
1026 Input chromatin for standard ChIP-seq in these regions did not produce any major peaks (track 10 in
1027 *Rag1/Rag2* and *Runx3*, and track 12 in *Foxp3*). The SATB1 peaks from standard ChIP-seq in
1028 *Rag1/Rag2* and *Runx3* loci largely overlap with enhancers (marked by H3K4me1 and H3K27ac). At
1029 the *Foxp3* locus in immature thymocytes SATB1 peaks (track 3) appear in a region marked by low levels
1030 of H3K27ac and H3Kme1 (tracks 4 and 5) where super-enhancer is established upon differentiation
1031 into Treg cells (tracks 6 and 7)(*47*). SATB1-bound peaks produced by urea ChIP-seq (track 2) are
1032 consistently located in distinct positions from those peaks obtained by standard ChIP-seq in these gene
1033 loci. Urea ChIP-seq did not generate peaks in KO thymocytes with anti-SATB1 antibody. Input
1034 chromatin also did not produce any peaks. H3K27ac and H3Kme1 data (tracks 4 and 5) are from
1035 Encode, generated by Bing Ren's laboratory at LICR, and data for Treg shown in (tracks 6 and 7) are
1036 from Kitagawa(*47*).
1037

1038 **Figure 5. SATB1 mediates dense chromatin interactions over *Rag1/2*-containing gene-rich**
1039 **region.**
1040 A) Chromatin interactions detected by urea 4C-seq using a SATB1-bound BUR (BUR-1) and another
1041 site (BUR-2) as baits. These baits are located near the border of a gene-poor region and a gene-rich
1042 region containing SATB1-regulated *Rag1* and *Rag2* genes (pink vertical bar with arrowhead). BUR-2
1043 was found to be one of the interacting sites of BUR-1. Therefore, it was used as a reciprocal control for

1044 BUR-1 to check the reproducibility of BUR-1 interactions. We used *Satb1*$^{+/+}$ (WT) and *Satb1*$^{-/-}$ (KO)
1045 thymocytes from 2 weeks old mice for urea 4C-seq. BUR-1 and BUR-2 interacts with many sites within
1046 the ~5.8Mb gene-rich region, containing *Rag1* and *Rag2* (tracks WT-BUR1 and WT-BUR2). These
1047 interactions were greatly reduced upon SATB1 deletion (track KO-BUR1). B) A zoom-out view, focusing
1048 on the *Rag1* and *Rag2* region, shows that BUR-1 interacts extensively over this region in WT
1049 thymocytes. Such interactions were virtually diminished in the absence of SATB1. The pink vertical
1050 bars with arrowheads indicate the positions of the BUR-1 and BUR-2 bait primers. Each of these primer
1051 sequences is located at the end of the *Hind*III fragment containing the primer. These *Hind*III fragments
1052 are indicated by pink bars adjacent to pink arrowheads pointing primer positions. The map of *Hind*III
1053 fragments is shown for interacting regions (track *Hind*III sites). BURs bound by SATB1 (confirmed by
1054 SATB1 urea ChIP-seq peaks and BUR reference) found within those *Hind*III fragments that interact with
1055 BUR-1 or BUR-2 are marked by small red stars and these *Hind*III fragments are light gray shaded. Note:
1056 Reads to the *Hind*III fragment immediately adjacent to each of the bait primers (BUR-1 and BURs) were
1057 removed as they comprised more than 2% of the total reads, typically derived from re-ligation of
1058 digested chromatin fragments to their original neighboring sequences and/or from a small undigested
1059 chromatin fraction.
1060

1061 **Figure 6. SATB1 binds to BURs in a cell-type dependent manner.**

1062 A) SATB1-binding profiles in a ~2.4 Mb region in chromosome 6 containing *Cd4* in a gene-rich region
1063 from standard ChIP-seq (Std) (track 2) and urea ChIP-seq (Urea) (tracks 3-8). Urea ChIP-seq-derived
1064 SATB1 profiles in thymocytes (track 3), brain (track 4) and skin (track 5) show difference in peak
1065 numbers depending on cell type (see supplement Table 1 and 2). This contrasts with mostly identical
1066 CTCF patterns in these cell types (tracks 11-13). B) SATB1-binding profiles in a 406kb region in
1067 chromosome 5 containing *Gabra2* and *Gabrg1*, highly expressed in brain, also shows much increased
1068 BUR binding in brain, compared to thymocytes and skin similar to A). In this region, there is no SATB1-
1069 binding site detected by standard ChIP-seq (track 2), and CTCF binding sites are mostly absent (track
1070 9). C) Percentages of SATB1 urea-ChIP-seq peaks in skin, thymocytes and brain intersecting with the
1071 BUR reference map, indicating that the majority of the peaks are BURs (supplement Table 1). D) The
1072 percentage of BURs among all BURs identified in the BUR reference map that intersected with urea
1073 SATB1 ChIP-seq peaks identified in skin, thymocytes and brain. DNase1 HS data are from ENCODE
1074 (UW), the CTCF-binding profile and H3K27ac data from standard ChIP-seq are from ENCODE (LICR).

1075

1076 **Figure 7. BURs largely co-map with LADs**

1077 A) The BUR distribution (BUR reference map, track 2) was compared against Lamin B1-DAM-ID
1078 sites(*81*) (track 1) and SATB1-binding sites derived from urea-ChIP-seq (Urea) (track 3 and 4) over a
1079 8.97 Mb region in chromosome 5 containing neuronal genes, *Gabrg1*, *Gabra2, Gabra4,* and *Gabrb1*.

1080 In this region, co-mapping of BURs and Lamin B1-DAM-ID regions is confirmed. Among BURs, SATB1-
1081 bound BURs derived from urea ChIP-seq (Urea) show difference in the number of peaks depending on
1082 cell type (see supplement Table 1 and 2). SATB1-binding sites derived from urea ChIP-seq avoid
1083 CTCF-binding sites, H3Kme1, H3K27ac sites (from Encode LICR). B) Average Lamin B1 DamID signal
1084 (black dotted), standard ChIP signal for SATB1 (black) and urea ChIP for SATB1 (blue) signal over
1085 LADs and flanking regions (interLAD sized at10% of adjacent LAD). C) Average Lamin B1 DamID signal
1086 (dotted black), BUR signal (red), and urea ChIP for CTCF (orange) signal over LADs (gray shaded) and
1087 flanking regions (interLAD sized at 10% of adjacent LAD). D) Average BUR signal (red), urea ChIP for
1088 SATB1 (blue) signal, Lamin B1 DamID (black, dotted), H3K4me1 (light green) and H3K27ac (light
1089 orange) signals over a +/- 3kb window centered at H3K4me1 peaks within LADs (LADs are gray shaded
1090 and DiPC that reside inside large LADs are unshaded, see text). E) Average BUR signal (red), urea
1091 ChIP for SATB1 (blue) signal, H3K4me1 (light green) and H3K27ac (light orange) signals over a +/- 3kb
1092 window centered at H3K4me1 peaks outside LADs. Relative distance, projection, and Jaccard tests all
1093 indicate strong correlation ($p < 0.05$) between urea ChIP-seq peaks for SATB1 and LADs, between
1094 BURs and LADs, but no overlap ($p<0.05$) between urea ChIP-seq SATB1 and H3K27ac and H3K4me3
1095 peaks. The results in B to E represent normalized average intensities of signals.
1096
1097 **Figure 8. A model for SATB1-mediated chromatin organization with direct and indirect binding**
1098 **of SATB1 with chromatin.**
1099
1100 The model shown is based on the results from urea ChIP-seq showing direct binding of SATB1 to BURs
1101 and standard ChIP-seq showing presumably indirect binding of SATB1 to gene-rich regions. In the
1102 inaccessible chromatin regions of nuclei, selected BURs are tightly anchored to the SATB1 nuclear
1103 architecture that resists high salt extraction. Therefore, we speculate that SATB1-BUR interactions are
1104 presumably stable, forming a chromatin scaffold. Such SATB1 chromatin scaffold further interacts
1105 dynamically with active gene-rich regions through indirect binding to form functional chromatin
1106 complexes (or hubs) to regulate gene expression. Future studies are required to determine the potential
1107 role of soluble SATB1 proteins in these interaction events.
1108

1109 # Acknowledgements

1110 We thank Dr. Aristotelis Tsirigos, Ziyan Lin, and Javier Rodriguez Hernaez at New York University
1111 School of Medicine for Hi-C and intersection analyses, Dr. Anthony Schmitt from Arima Genomics for
1112 help with Hi-C experiments. We also thank Dr. Elphege Nora for reviewing Hi-C results and providing
1113 critical comments on the manuscript and Ms. Bao Ho for providing technical assistance. We thank Dr.
1114 Minoree Kohwi for critical reviewing of the manuscript and Dr. Peter Krijger for kindly testing our samples
1115 and Dr. Elzo de Wit for discussion on the 4C-seq analysis.

Funding:

National Institute of Health grant R31CA39681 (TKS)

National Institute of Health grant R01ES023854 (TKS)

National Institute of Health grant R01GM132427 (KLR)

National Institute of Health grant R01AR071727 (VAB and TKS)

National Institute of Health grant R01 AR047364 (C-MC and TKS)

Author contribution:

Conceptualization: TKS, YKohwi

Methodology: TKS, YKohwi

Data production and bioinformatic analyses: TKS, YKohwi, XW, KLR, TS, MG, HWR

Supervision: TKS, YKohwi, KLR

Materials and/or Discussion: IT, SS, YKitagawa, C-MC, VAB, Y-CL

Writing—original draft: TKS, KLR, YKohwi

Writing—review & editing: TKS, KLR, YKohwi, IT, TS, SS

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.

References

1. J. C. Politz, D. Scalzo, M. Groudine, Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu Rev Cell Dev Biol* **29**, 241-270 (2013).
2. B. van Steensel, A. S. Belmont, Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791 (2017).
3. S. Schoenfelder, P. Fraser, Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).
4. E. Bashkirova, S. Lomvardas, Olfactory receptor genes make the case for inter-chromosomal interactions. *Curr Opin Genet Dev* **55**, 106-113 (2019).
5. A. Gondor, R. Ohlsson, Enhancer functions in three dimensions: beyond the flat world perspective. *F1000Res* **7**, (2018).
6. V. Loubiere, G. L. Papadopoulos, Q. Szabo, A. M. Martinez, G. Cavalli, Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
7. T. Misteli, The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28-45 (2020).
8. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
9. L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, B. van Steensel, Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951 (2008).

10. J. Kind, L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey, M. Bienko, Y. Zhan, B. Lajoie, C. A. de Graaf, M. Amendola, G. Fudenberg, M. Imakaev, L. A. Mirny, K. Jalink, J. Dekker, A. van Oudenaarden, B. van Steensel, Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-147 (2015).
11. B. van Steensel, J. Delrow, S. Henikoff, Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* **27**, 304-308 (2001).
12. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
13. E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
14. T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, G. Cavalli, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472 (2012).
15. A. L. Sanborn, S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden, 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).
16. G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L. A. Mirny, Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-2049 (2016).
17. S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665-1680 (2014).
18. E. P. Nora, A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).
19. S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017).
20. G. Wutz, C. Varnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S. Schoenfelder, G. Jessberger, M. Muhar, M. J. Hossain, N. Walther, B. Koch, M. Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, J. M. Peters, Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* **36**, 3573-3599 (2017).
21. W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, F. Spitz, Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
22. D. I. Cattoni, A. M. Cardozo Gizzi, M. Georgieva, M. Di Stefano, A. Valeri, D. Chamousset, C. Houbron, S. Dejardin, J. B. Fiche, I. Gonzalez, J. M. Chang, T. Sexton, M. A. Marti-Renom, F. Bantignies, G. Cavalli, M. Nollmann, Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat Commun* **8**, 1753 (2017).
23. B. Bintu, L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).
24. E. H. Finn, G. Pegoraro, H. B. Brandao, A. L. Valton, M. E. Oomen, J. Dekker, L. Mirny, T. Misteli, Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515 e1510 (2019).

1214 25. M. Gabriele, H. B. Brandao, S. Grosse-Holz, A. Jha, G. M. Dailey, C. Cattoglio, T. S. Hsieh, L.
1215 Mirny, C. Zechner, A. S. Hansen, Dynamics of CTCF- and cohesin-mediated chromatin
1216 looping revealed by live-cell imaging. *Science* **376**, 496-501 (2022).
1217 26. P. Mach, P. I. Kos, Y. Zhan, J. Cramard, S. Gaudin, J. Tunnermann, E. Marchi, J. Eglinger, J.
1218 Zuin, M. Kryzhanovska, S. Smallwood, L. Gelman, G. Roth, E. P. Nora, G. Tiana, L. Giorgetti,
1219 Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907-1918
1220 (2022).
1221 27. Y. Chi, J. Shi, D. Xing, L. Tan, Every gene everywhere all at once: High-precision
1222 measurement of 3D chromosome architecture with single-cell Hi-C. *Front Mol Biosci* **9**, 959688
1223 (2022).
1224 28. L. F. Chen, J. Lee, A. Boettiger, Recent progress and challenges in single-cell imaging of
1225 enhancer-promoter interaction. *Curr Opin Genet Dev* **79**, 102023 (2023).
1226 29. A. Hafner, M. Park, S. E. Berger, S. E. Murphy, E. P. Nora, A. N. Boettiger, Loop stacking
1227 organizes genome folding from TADs to chromosomes. *Mol Cell* **83**, 1377-1392 e1376 (2023).
1228 30. T. Kohwi-Shigematsu, Y. Kohwi, Torsional stress stabilizes extended base unpairing in
1229 suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29**, 9551-9560.
1230 (1990).
1231 31. J. Bode, Y. Kohwi, L. Dickinson, T. Joh, D. Klehr, C. Mielke, T. Kohwi-Shigematsu, Biological
1232 significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**, 195-197
1233 (1992).
1234 32. L. A. Dickinson, T. Joh, Y. Kohwi, T. Kohwi-Shigematsu, A tissue-specific MAR/SAR DNA-
1235 binding protein with unusual binding site recognition. *Cell* **70**, 631-645. (1992).
1236 33. T. Kohwi-Shigematsu, Y. Kohwi, Detection of non-B-DNA structures at specific sites in
1237 supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate.
1238 *Methods Enzymol* **212**, 155-180 (1992).
1239 34. J. D. Alvarez, D. H. Yasui, H. Niida, T. Joh, D. Y. Loh, T. Kohwi-Shigematsu, The MAR-binding
1240 protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell
1241 development. *Genes Dev* **14**, 521-535 (2000).
1242 35. Y. Zhang, L. Zheng, M. Le, Y. Nakano, B. Chan, Y. Huang, P. M. Torbaty, Y. Kohwi, R.
1243 Marcucio, S. Habelitz, P. K. Den Besten, T. Kohwi-Shigematsu, SATB1 establishes ameloblast
1244 cell polarity and regulates directional amelogenin secretion for enamel formation. *BMC Biol* **17**,
1245 104 (2019).
1246 36. M. Y. Fessing, A. N. Mardaryev, M. R. Gdula, A. A. Sharov, T. Y. Sharova, V. Rapisarda, K. B.
1247 Gordon, A. D. Smorodchenko, K. Poterlowicz, G. Ferone, Y. Kohwi, C. Missero, T. Kohwi-
1248 Shigematsu, V. A. Botchkarev, p63 regulates Satb1 to control tissue-specific chromatin
1249 remodeling during development of the epidermis. *J Cell Biol* **194**, 825-839 (2011).
1250 37. M. A. Balamotis, N. Tamberg, Y. J. Woo, J. Li, B. Davy, T. Kohwi-Shigematsu, Y. Kohwi,
1251 Satb1 ablation alters temporal expression of immediate early genes and reduces dendritic
1252 spine density during postnatal brain development. *Mol Cell Biol* **32**, 333-347 (2012).
1253 38. Y. Huang, L. Zhang, N. N. Song, Z. L. Hu, J. Y. Chen, Y. Q. Ding, Distribution of Satb1 in the
1254 central nervous system of adult mice. *Neurosci Res* **71**, 12-21 (2011).
1255 39. S. Cai, H. J. Han, T. Kohwi-Shigematsu, Tissue-specific nuclear architecture and gene
1256 expression regulated by SATB1. *Nat Genet* **34**, 42-51 (2003).
1257 40. I. de Belle, S. Cai, T. Kohwi-Shigematsu, The genomic sequences bound to special AT-rich
1258 sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the
1259 nuclear matrix at the bases of the chromatin loops. *J Cell Biol* **141**, 335-348 (1998).
1260 41. T. Kohwi-Shigematsu, K. Poterlowicz, E. Ordinario, H. J. Han, V. A. Botchkarev, Y. Kohwi,
1261 Genome organizing function of SATB1 in tumor progression. *Semin Cancer Biol*, (2013).
1262 42. S. Cai, C. C. Lee, T. Kohwi-Shigematsu, SATB1 packages densely looped, transcriptionally
1263 active chromatin for coordinated expression of cytokine genes. *Nat Genet* **38**, 1278-1288
1264 (2006).
1265 43. Y. Satoh, T. Yokota, T. Sudo, M. Kondo, A. Lai, P. W. Kincade, T. Kouro, R. Iida, K. Kokame,
1266 T. Miyata, Y. Habuchi, K. Matsui, H. Tanaka, I. Matsumura, K. Oritani, T. Kohwi-Shigematsu,
1267 Y. Kanakura, The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid
1268 lineages. *Immunity* **38**, 1105-1115 (2013).

44. D. Skowronska-Krawczyk, Q. Ma, M. Schwartz, K. Scully, W. Li, Z. Liu, H. Taylor, J. Tollkuhn, K. A. Ohgi, D. Notani, Y. Kohwi, T. Kohwi-Shigematsu, M. G. Rosenfeld, Required enhancer-matrin-3 network interactions for a homeodomain transcription program. *Nature* **514**, 257-261 (2014).
45. B. Hao, A. K. Naik, A. Watanabe, H. Tanaka, L. Chen, H. W. Richards, M. Kondo, I. Taniuchi, Y. Kohwi, T. Kohwi-Shigematsu, M. S. Krangel, An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J Exp Med* **212**, 809-824 (2015).
46. K. Kakugawa, S. Kojo, H. Tanaka, W. Seo, T. A. Endo, Y. Kitagawa, S. Muroi, M. Tenno, N. Yasmin, Y. Kohwi, S. Sakaguchi, T. Kowhi-Shigematsu, I. Taniuchi, Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. *Cell Rep* **19**, 1176-1188 (2017).
47. Y. Kitagawa, N. Ohkura, Y. Kidani, A. Vandenbon, K. Hirota, R. Kawakami, K. Yasuda, D. Motooka, S. Nakamura, M. Kondo, I. Taniuchi, T. Kohwi-Shigematsu, S. Sakaguchi, Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat Immunol* **18**, 173-183 (2017).
48. K. Yasuda, Y. Kitagawa, R. Kawakami, Y. Isaka, H. Watanabe, G. Kondoh, T. Kohwi-Shigematsu, S. Sakaguchi, K. Hirota, Satb1 regulates the effector program of encephalitogenic tissue Th17 cells in chronic inflammation. *Nat Commun* **10**, 549 (2019).
49. M. Goolam, M. Zernicka-Goetz, The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. *Development* **144**, 1450-1461 (2017).
50. H. J. Han, J. Russo, Y. Kohwi, T. Kohwi-Shigematsu, SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* **452**, 187-193 (2008).
51. A. Fromberg, K. Engeland, A. Aigner, The Special AT-rich Sequence Binding Protein 1 (SATB1) and its role in solid tumors. *Cancer Lett* **417**, 96-111 (2018).
52. E. Ordinario, H. J. Han, S. Furuta, L. M. Heiser, L. R. Jakkula, F. Rodier, P. T. Spellman, J. Campisi, J. W. Gray, M. J. Bissell, Y. Kohwi, T. Kohwi-Shigematsu, ATM suppresses SATB1-induced malignant progression in breast epithelial cells. *PLoS One* **7**, e51786 (2012).
53. D. Feng, Y. Chen, R. Dai, S. Bian, W. Xue, Y. Zhu, Z. Li, Y. Yang, Y. Zhang, J. Zhang, J. Bai, L. Qin, Y. Kohwi, W. Shi, T. Kohwi-Shigematsu, J. Ma, S. Liao, B. Hao, Chromatin organizer SATB1 controls the cell identity of CD4(+) CD8(+) double-positive thymocytes by regulating the activity of super-enhancers. *Nat Commun* **13**, 5554 (2022).
54. T. Zelenka, A. Klonizakis, D. Tsoukatou, D. A. Papamatheakis, S. Franzenburg, P. Tzerpos, I. R. Tzonevrakis, G. Papadogkonas, M. Kapsetaki, C. Nikolaou, D. Plewczynski, C. Spilianakis, The 3D enhancer network of the developing T cell genome is shaped by SATB1. *Nat Commun* **13**, 6954 (2022).
55. B. Wang, L. Ji, Q. Bian, SATB1 regulates 3D genome architecture in T cells by constraining chromatin interactions surrounding CTCF-binding sites. *Cell Rep* **42**, 112323 (2023).
56. D. Jain, S. Baldi, A. Zabel, T. Straub, P. B. Becker, Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Res* **43**, 6959-6968 (2015).
57. D. Park, Y. Lee, G. Bhupindersingh, V. R. Iyer, Widespread misinterpretable ChIP-seq bias in yeast. *PLoS One* **8**, e83506 (2013).
58. L. Teytelman, D. M. Thurtle, J. Rine, A. van Oudenaarden, Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A* **110**, 18602-18607 (2013).
59. C. A. Meyer, X. S. Liu, Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet* **15**, 709-721 (2014).
60. S. Baumgarten, J. Bryant, Chromatin structure can introduce systematic biases in genome-wide analyses of Plasmodium falciparum. *Open Res Eur* **2**, 75 (2022).
61. T. Kohwi-Shigematsu, I. deBelle, L. A. Dickinson, S. Galande, Y. Kohwi, Identification of base-unpairing region-binding proteins and characterization of their in vivo binding sequences. *Methods Cell Biol* **53**, 323-354 (1998).
62. T. Kohwi-Shigematsu, Y. Kohwi, K. Takahashi, H. W. Richards, S. D. Ayers, H. J. Han, S. Cai, SATB1-mediated functional packaging of chromatin into loops. *Methods* **58**, 243-254 (2012).
63. W. Mu, E. S. Davis, S. Lee, M. G. Dozmorov, D. H. Phanstiel, M. I. Love, bootRanges: flexible generation of null sets of genomic ranges for hypothesis testing. *Bioinformatics* **39**, (2023).

64. C. A. Davis, B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka, J. M. Cherry, The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-D801 (2018).
65. J. C. Peng, A. Valouev, T. Swigut, J. Zhang, Y. Zhao, A. Sidow, J. Wysocka, Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290-1302 (2009).
66. Y. Tanaka, T. Sotome, A. Inoue, T. Mukozu, T. Kuwabara, T. Mikami, T. Kowhi-Shigematsu, M. Kondo, SATB1 Conditional Knockout Results in Sjogren's Syndrome in Mice. *J Immunol* **199**, 4016-4022 (2017).
67. M. Kondo, Y. Tanaka, T. Kuwabara, T. Naito, T. Kohwi-Shigematsu, A. Watanabe, SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J Immunol* **196**, 563-572 (2016).
68. N. Yannoutsos, V. Barreto, Z. Misulovin, A. Gazumyan, W. Yu, N. Rajewsky, B. R. Peixoto, T. Eisenreich, M. C. Nussenzweig, A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nat Immunol* **5**, 443-450 (2004).
69. Y. Ben Zouari, A. Platania, A. M. Molitor, T. Sexton, 4See: A Flexible Browser to Explore 4C Data. *Front Genet* **10**, 1372 (2019).
70. E. de Wit, B. A. Bouwman, Y. Zhu, P. Klous, E. Splinter, M. J. Verstegen, P. H. Krijger, N. Festuccia, E. P. Nora, M. Welling, E. Heard, N. Geijsen, R. A. Poot, I. Chambers, W. de Laat, The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**, 227-231 (2013).
71. G. Geeven, H. Teunissen, W. de Laat, E. de Wit, peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res* **46**, e91 (2018).
72. P. H. L. Krijger, G. Geeven, V. Bianchi, C. R. E. Hilvering, W. de Laat, 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17-32 (2020).
73. D. Noordermeer, M. Leleu, E. Splinter, J. Rougemont, W. De Laat, D. Duboule, The dynamic architecture of Hox gene clusters. *Science* **334**, 222-225 (2011).
74. P. J. Wijchers, P. H. L. Krijger, G. Geeven, Y. Zhu, A. Denker, M. Verstegen, C. Valdes-Quezada, C. Vermeulen, M. Janssen, H. Teunissen, L. C. M. Anink-Groenen, P. J. Verschure, W. de Laat, Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol Cell* **61**, 461-473 (2016).
75. F. Ramirez, D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dundar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
76. L. A. Dickinson, T. Kohwi-Shigematsu, 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).
77. W. Meuleman, D. Peric-Hupkes, J. Kind, J. B. Beaudry, L. Pagie, M. Kellis, M. Reinders, L. Wessels, B. van Steensel, Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**, 270-280 (2013).
78. B. Wen, H. Wu, Y. H. Loh, E. Briem, G. Q. Daley, A. P. Feinberg, Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions. *BMC Genomics* **13**, 566 (2012).
79. C. Leemans, M. C. H. van der Zwalm, L. Brueckner, F. Comoglio, T. van Schaik, L. Pagie, J. van Arensbergen, B. van Steensel, Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell* **177**, 852-864 e814 (2019).
80. T. R. Luperchio, Sauria, M. EG., Wong, X., Gaillard, M-C., Tsang, P., Pekrun, K., Ach, R.A., Yamada, N.A., Taylor, J., Reddy, K.L., Chromosome Conformation Paints Reveal the Role of Lamina Association in Genome Organization and Regulation. *bioRxiv*, **doi:** https://doi.org/10.1101/481598
81. S. Chen, T. R. Luperchio, X. Wong, E. B. Doan, A. T. Byrd, K. Roy Choudhury, K. L. Reddy, M. S. Krangel, A Lamina-Associated Domain Border Governs Nuclear Lamina Interactions, Transcription, and Recombination of the Tcrb Locus. *Cell Rep* **25**, 1729-1740 e1726 (2018).

82. D. Peric-Hupkes, W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels, B. van Steensel, Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603-613 (2010).
83. H. Zhao, E. G. Sifakis, N. Sumida, L. Millan-Arino, B. A. Scholz, J. P. Svensson, X. Chen, A. L. Ronnegren, C. D. Mallet de Lima, F. S. Varnoosfaderani, C. Shi, O. Loseva, S. Yammine, M. Israelsson, L. S. Rathje, B. Nemeti, E. Fredlund, T. Helleday, M. P. Imreh, A. Gondor, PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Mol Cell* **59**, 984-997 (2015).
84. T. Lucas, T. L. Hafer, H. G. Zhang, N. Molotkova, M. Kohwi, Discrete cis-acting element regulates developmentally timed gene-lamina relocation and neural progenitor competence in vivo. *Dev Cell* **56**, 2649-2663 e2646 (2021).
85. I. Solovei, M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, J. Guck, B. Joffe, Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356-368 (2009).
86. J. Kind, L. Pagie, H. Ortabozkoyun, S. Boyle, S. S. de Vries, H. Janssen, M. Amendola, L. D. Nolen, W. A. Bickmore, B. van Steensel, Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-192 (2013).
87. Q. Bian, N. Khanna, J. Alvikas, A. S. Belmont, beta-Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *J Cell Biol* **203**, 767-783 (2013).
88. X. Wong, T. R. Luperchio, K. L. Reddy, NET gains and losses: the role of changing nuclear envelope proteomes in genome regulation. *Curr Opin Cell Biol* **28**, 105-120 (2014).
89. J. C. Harr, T. R. Luperchio, X. Wong, E. Cohen, S. J. Wheelan, K. L. Reddy, Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J Cell Biol* **208**, 33-52 (2015).
90. X. Wong, V. E. Hoskins, A. J. Melendez-Perez, J. C. Harr, M. Gordon, K. L. Reddy, Lamin C is required to establish genome organization after mitosis. *Genome Biol* **22**, 305 (2021).
91. L. A. Watson, L. H. Tsai, In the loop: how chromatin topology links genome structure to function in mechanisms underlying learning and memory. *Curr Opin Neurobiol* **43**, 48-55 (2017).
92. J. Close, H. Xu, N. De Marco Garcia, R. Batista-Brito, E. Rossignol, B. Rudy, G. Fishell, 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).
93. S. Galande, L. A. Dickinson, I. S. Mian, M. Sikorska, T. Kohwi-Shigematsu, 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).
94. L. A. Dickinson, C. D. Dickinson, T. Kohwi-Shigematsu, 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).
95. K. Nakagomi, Y. Kohwi, L. A. Dickinson, T. Kohwi-Shigematsu, A novel DNA-binding motif in the nuclear matrix attachment DNA-binding protein SATB1. *Mol Cell Biol* **14**, 1852-1860. (1994).
96. Z. Wang, X. Yang, S. Guo, Y. Yang, X. C. Su, Y. Shen, J. Long, 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).
97. P. Pavan Kumar, P. K. Purbey, C. K. Sinha, D. Notani, A. Limaye, R. S. Jayani, S. Galande, Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. *Mol Cell* **22**, 231-243 (2006).
98. D. Notani, K. P. Gottimukkala, R. S. Jayani, A. S. Limaye, M. V. Damle, S. Mehta, P. K. Purbey, J. Joseph, S. Galande, Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. *PLoS Biol* **8**, e1000296 (2010).

99. D. Notani, P. L. Ramanujam, P. P. Kumar, K. P. Gottimukkala, C. Kumar-Sinha, S. Galande, N-terminal PDZ-like domain of chromatin organizer SATB1 contributes towards its function as transcription regulator. *J Biosci* **36**, 461-469 (2011).
100. T. L. Stephen, K. K. Payne, R. A. Chaurio, M. J. Allegrezza, H. Zhu, J. Perez-Sanz, A. Perales-Puchalt, J. M. Nguyen, A. E. Vara-Ailor, E. B. Eruslanov, M. E. Borowsky, R. Zhang, T. M. Laufer, J. R. Conejo-Garcia, SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. *Immunity* **46**, 51-64 (2017).
101. D. Yasui, M. Miyano, S. Cai, P. Varga-Weisz, T. Kohwi-Shigematsu, SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419**, 641-645. (2002).
102. J. Thakur, S. Henikoff, Architectural RNA in chromatin organization. *Biochem Soc Trans* **48**, 1967-1978 (2020).
103. D. Michieletto, N. Gilbert, Role of nuclear RNA in regulating chromatin structure and transcription. *Curr Opin Cell Biol* **58**, 120-125 (2019).
104. S. Galande, T. Kohwi-Shigematsu, Poly(ADP-ribose) polymerase and Ku autoantigen form a complex and synergistically bind to matrix attachment sequences. *J Biol Chem* **274**, 20521-20528. (1999).
105. W. M. Liu, F. K. Guerra-Vladusic, S. Kurakata, R. Lupu, T. Kohwi-Shigematsu, 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).
106. M. Levo, J. Raimundo, X. Y. Bing, Z. Sisco, P. J. Batut, S. Ryabichko, T. Gregor, M. S. Levine, Transcriptional coupling of distant regulatory genes in living embryos. *Nature* **605**, 754-760 (2022).
107. P. J. Batut, X. Y. Bing, Z. Sisco, J. Raimundo, M. Levo, M. S. Levine, Genome organization controls transcriptional dynamics during development. *Science* **375**, 566-570 (2022).
108. G. Mohana, J. Dorier, X. Li, M. Mouginot, R. C. Smith, H. Malek, M. Leleu, D. Rodriguez, J. Khadka, P. Rosa, P. Cousin, C. Iseli, S. Restrepo, N. Guex, B. D. McCabe, A. Jankowski, M. S. Levine, M. C. Gambetta, Chromosome-level organization of the regulatory genome in the Drosophila nervous system. *Cell* **186**, 3826-3844 e3826 (2023).
109. L. Lizana, Y. B. Schwartz, The scales, mechanisms, and dynamics of the genome architecture. *Sci Adv* **10**, eadm8167 (2024).
110. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557-572 (2004).
111. C. Lazaris, S. Kelly, P. Ntziachristos, I. Aifantis, A. Tsirigos, HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC Genomics* **18**, 22 (2017).
112. A. Kloetgen, P. Thandapani, P. Ntziachristos, Y. Ghebrechristos, S. Nomikou, C. Lazaris, X. Chen, H. Hu, S. Bakogianni, J. Wang, Y. Fu, F. Boccalatte, H. Zhong, E. Paietta, T. Trimarchi, Y. Zhu, P. Van Vlierberghe, G. G. Inghirami, T. Lionnet, I. Aifantis, A. Tsirigos, Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet* **52**, 388-400 (2020).
113. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
114. A. Tsirigos, N. Haiminen, E. Bilal, F. Utro, GenomicTools: a computational platform for developing high-throughput analytics in genomics. *Bioinformatics* **28**, 282-283 (2012).
115. M. Imakaev, G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, L. A. Mirny, Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003 (2012).
116. Y. Gong, C. Lazaris, T. Sakellaropoulos, A. Lozano, P. Kambadur, P. Ntziachristos, I. Aifantis, A. Tsirigos, Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun* **9**, 542 (2018).
117. N. C. Durand, M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).

1485 118. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H.
1486 Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-
1487 regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589
1488 (2010).
1489 119. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R.
1490 M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome*
1491 *Biol* **9**, R137 (2008).
1492 120. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: enabling
1493 browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).
1494 121. J. M. Zullo, I. A. Demarco, R. Pique-Regi, D. J. Gaffney, C. B. Epstein, C. J. Spooner, T. R.
1495 Luperchio, B. E. Bernstein, J. K. Pritchard, K. L. Reddy, H. Singh, DNA sequence-dependent
1496 compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, 1474-1487
1497 (2012).
1498 122. A. Favorov, L. Mularoni, L. M. Cope, Y. Medvedeva, A. A. Mironov, V. J. Makeev, S. J.
1499 Wheelan, Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS*
1500 *Comput Biol* **8**, e1002529 (2012).
1501 123. A. Akalin, V. Franke, K. Vlahovicek, C. E. Mason, D. Schubeler, Genomation: a toolkit to
1502 summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127-1129 (2015).
1503 124. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features.
1504 *Bioinformatics* **26**, 841-842 (2010).
1505

FIGURE 1

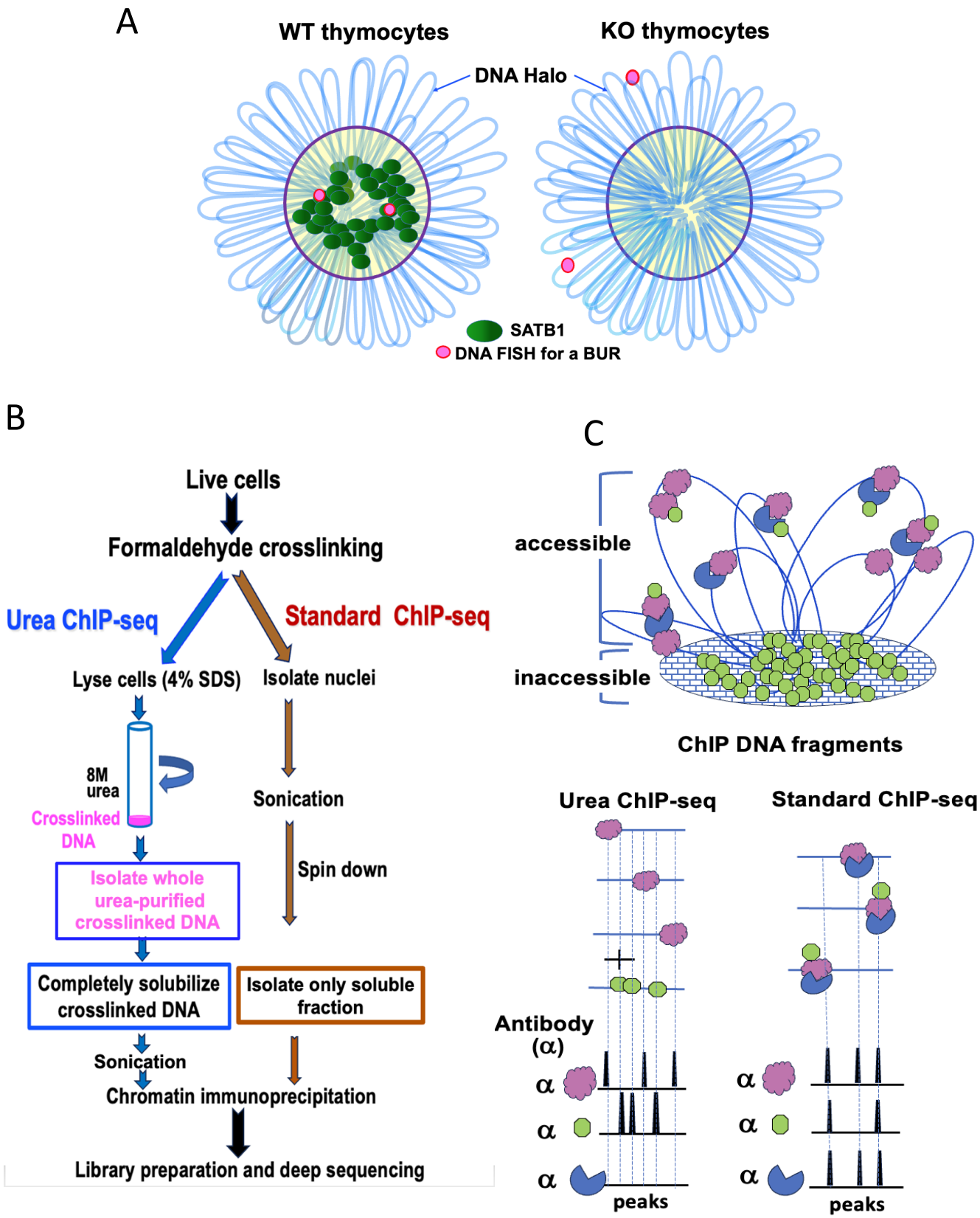

A
WT thymocytes
KO thymocytes
DNA Halo
SATB1
DNA FISH for a BUR
B
Live cells
Formaldehyde crosslinking
Urea ChIP-seq
Standard ChIP-seq
Lyse cells (4% SDS)
Isolate nuclei
8M urea
Crosslinked DNA
Sonication
Spin down
Isolate whole urea-purified crosslinked DNA
Completely solubilize crosslinked DNA
Isolate only soluble fraction
Sonication
Chromatin immunoprecipitation
Library preparation and deep sequencing
C
accessible
inaccessible
ChIP DNA fragments
Urea ChIP-seq
Standard ChIP-seq
Antibody (α)
α
α
α
peaks
α
α
α
peaks

FIGURE 2

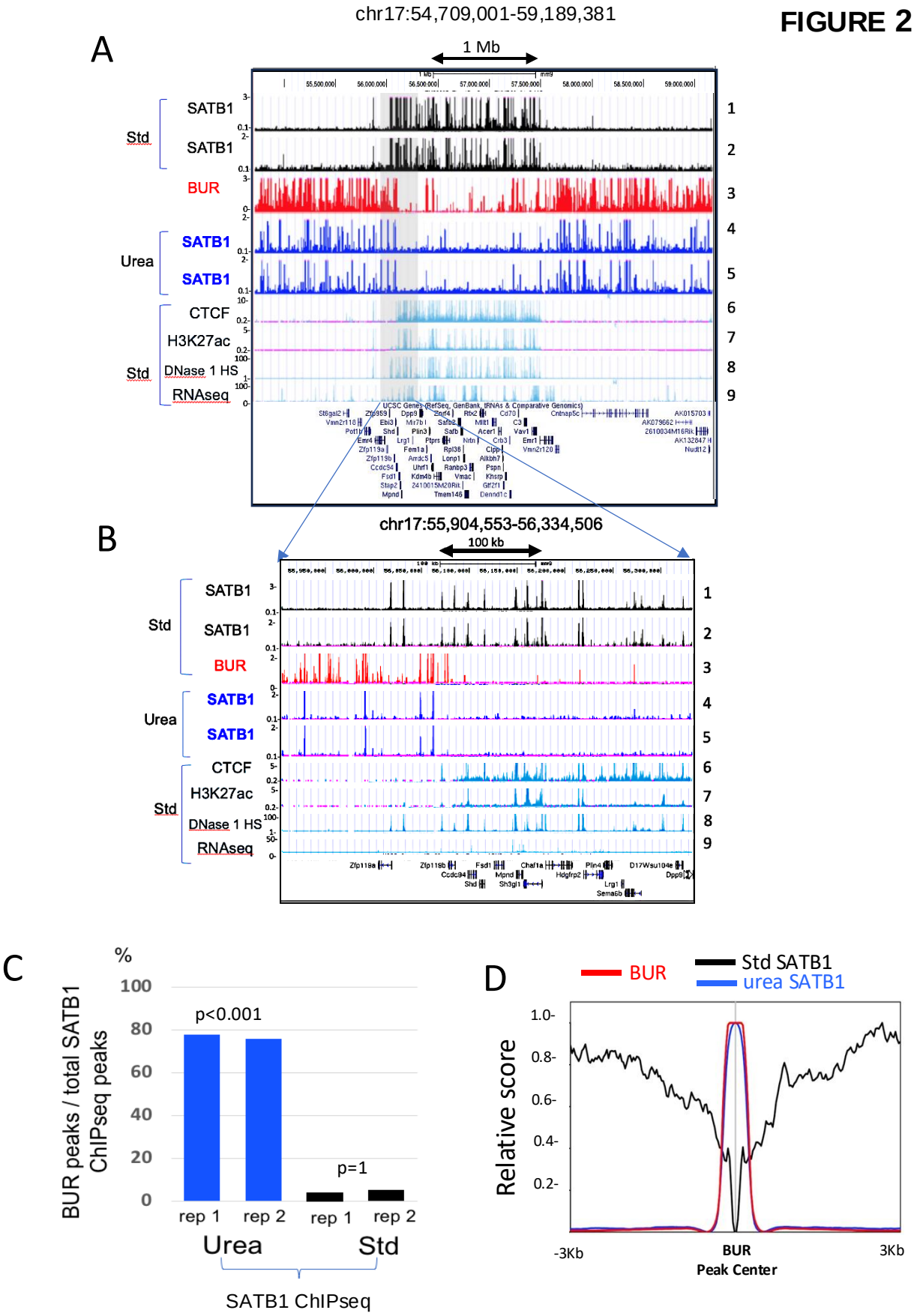

chr17:54,709,001-59,189,381
A
1 Mb
Std
SATB1
SATB1
BUR
Urea
SATB1
SATB1
Std
CTCF
H3K27ac
DNase 1 HS
RNAseq
1
2
3
4
5
6
7
8
9
B
chr17:55,904,553-56,334,506
100 kb
Std
SATB1
SATB1
BUR
Urea
SATB1
SATB1
Std
CTCF
H3K27ac
DNase 1 HS
RNAseq
C
%
BUR peaks / total SATB1 ChIPseq peaks
100
80
60
40
20
0
p<0.001
p=1
rep 1
rep 2
rep 1
rep 2
Urea
Std
SATB1 ChIPseq
D
BUR
Std SATB1
urea SATB1
Relative score
1.0
0.8
0.6
0.4
0.2
-3Kb
BUR
Peak Center
3Kb

FIGURE 3

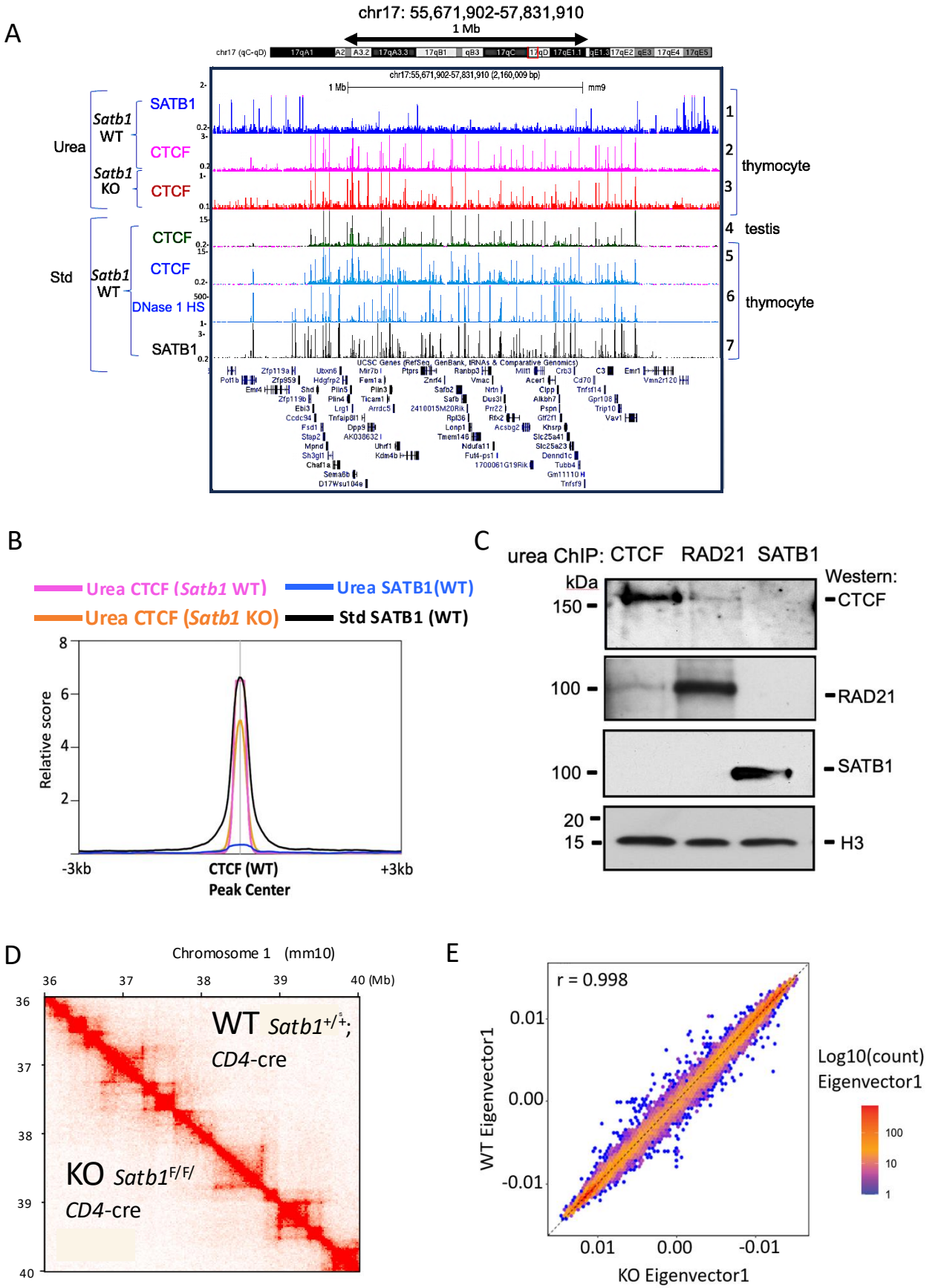
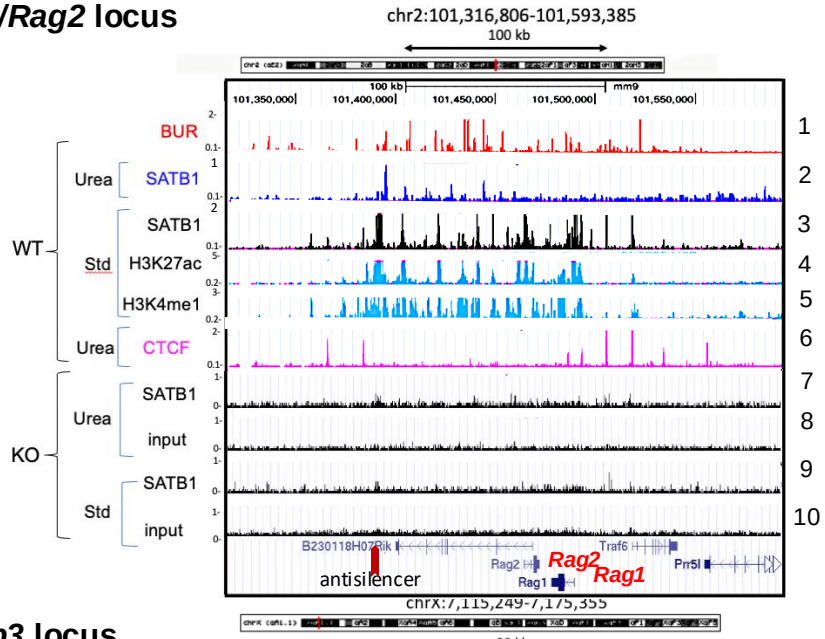
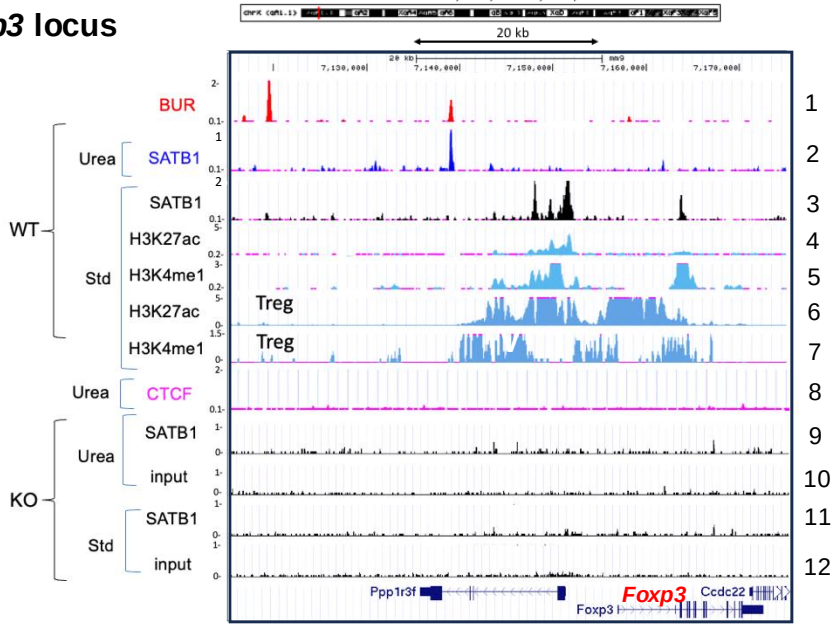

A
chr17: 55,671,902-57,831,910
1 Mb
Urea
Satb1 WT
SATB1
CTCF
Satb1 KO
CTCF
Std
Satb1 WT
CTCF
CTCF
DNase 1 HS
SATB1
thymocyte
testis
thymocyte
B
Urea CTCF (Satb1 WT)
Urea SATB1(WT)
Urea CTCF (Satb1 KO)
Std SATB1 (WT)
Relative score
-3kb
CTCF (WT)
Peak Center
+3kb
C
urea ChIP: CTCF RAD21 SATB1
kDa
Western:
150
CTCF
100
RAD21
100
SATB1
20
15
H3
D
Chromosome 1 (mm10)
36
37
38
39
40 (Mb)
WT Satb1+/+; CD4-cre
KO Satb1F/F/ CD4-cre
E
r = 0.998
WT Eigenvector1
KO Eigenvector1
Log10(count)
Eigenvector1
100
10
1

FIGURE 4

***Rag1/Rag2* locus**



***Foxp3* locus**



***Runx3* locus**

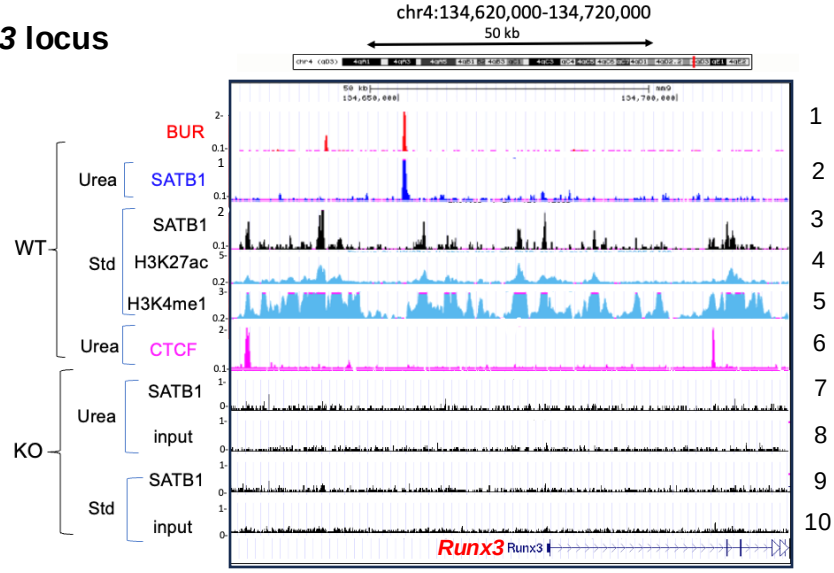


FIGURE 5

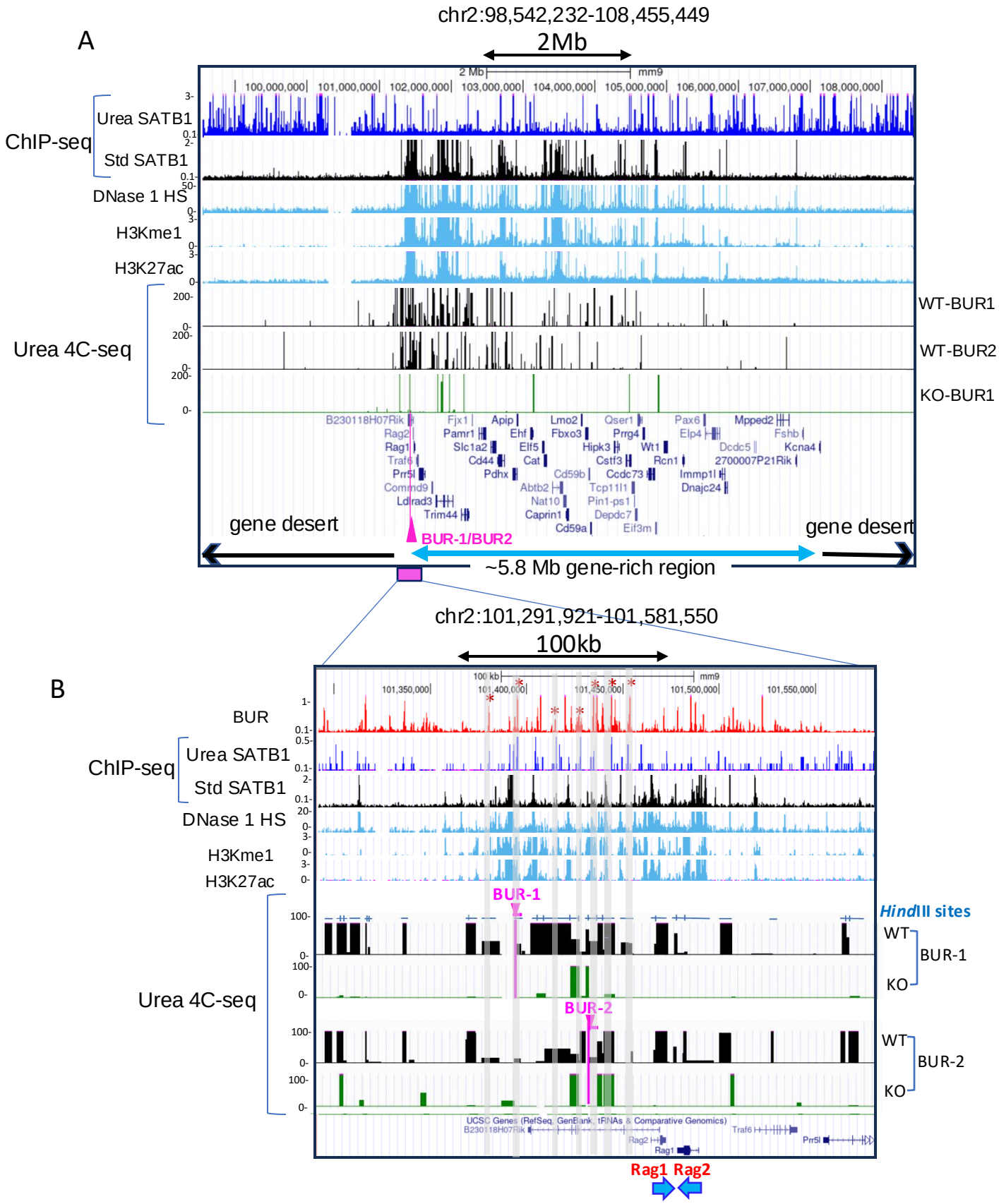

A
chr2:98,542,232-108,455,449
2Mb
ChIP-seq
Urea SATB1
Std SATB1
DNase 1 HS
H3Kme1
H3K27ac
Urea 4C-seq
WT-BUR1
WT-BUR2
KO-BUR1
gene desert
BUR-1/BUR2
gene desert
~5.8 Mb gene-rich region
B
chr2:101,291,921-101,581,550
100kb
BUR
ChIP-seq
Urea SATB1
Std SATB1
DNase 1 HS
H3Kme1
H3K27ac
Urea 4C-seq
BUR-1
BUR-2
HindIII sites
WT
KO
BUR-1
WT
KO
BUR-2
Rag1 Rag2

FIGURE 6

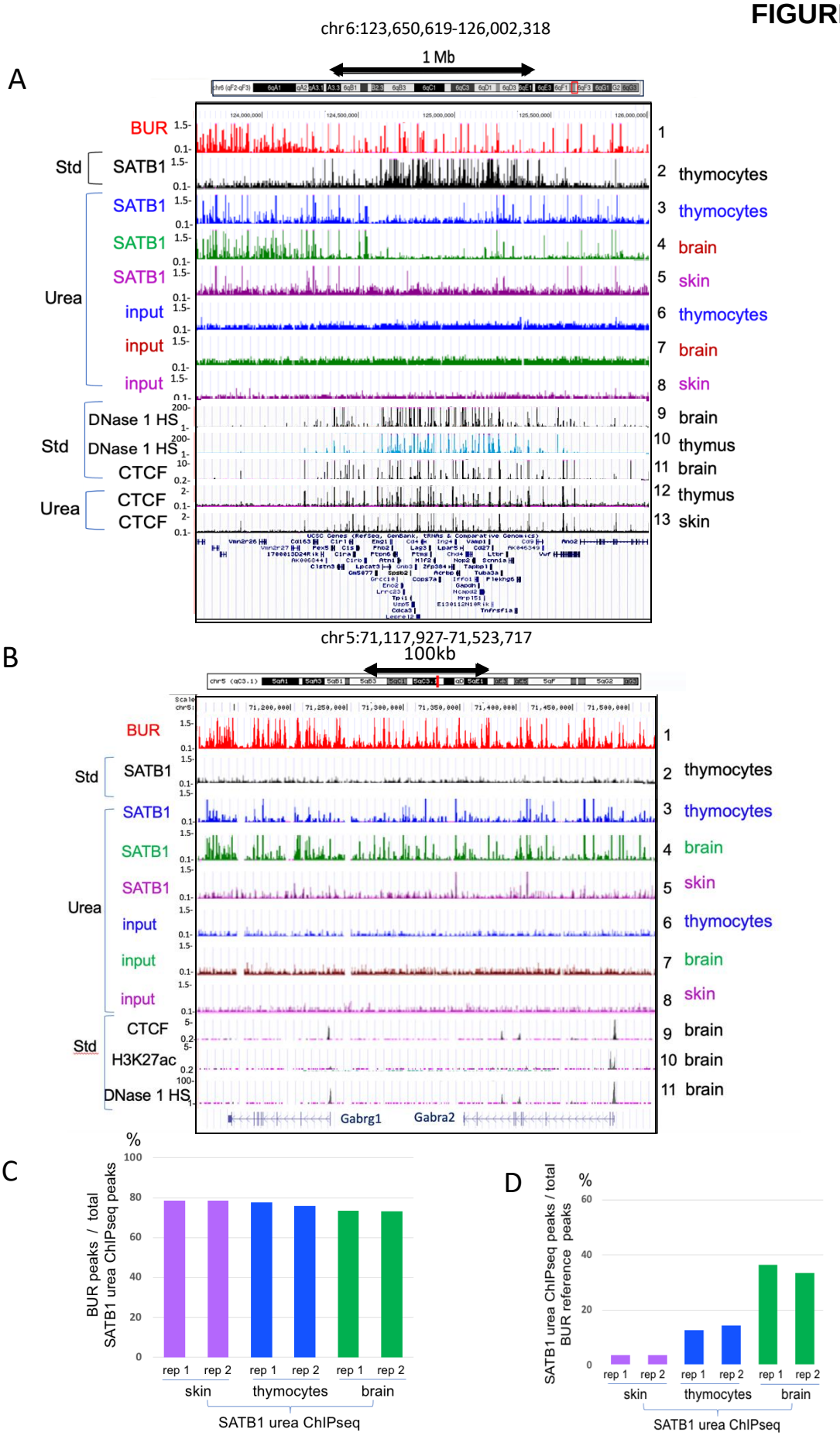

A
chr6:123,650,619-126,002,318
1 Mb
BUR
Std
SATB1
Urea
SATB1
SATB1
SATB1
input
input
input
Std
DNase 1 HS
DNase 1 HS
CTCF
Urea
CTCF
CTCF
1
2 thymocytes
3 thymocytes
4 brain
5 skin
6 thymocytes
7 brain
8 skin
9 brain
10 thymus
11 brain
12 thymus
13 skin
B
chr5:71,117,927-71,523,717
100kb
BUR
Std
SATB1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

FIGURE 7

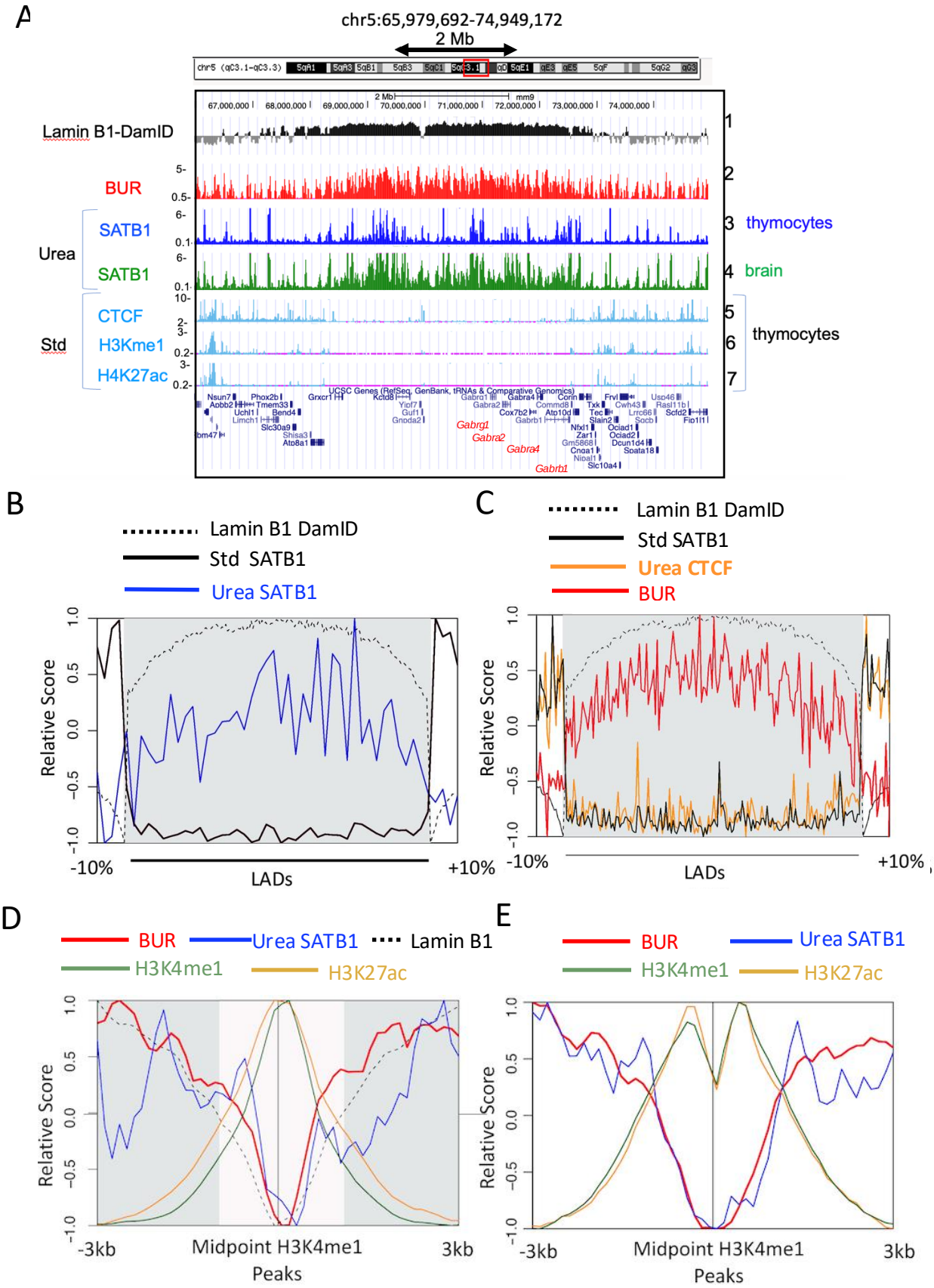
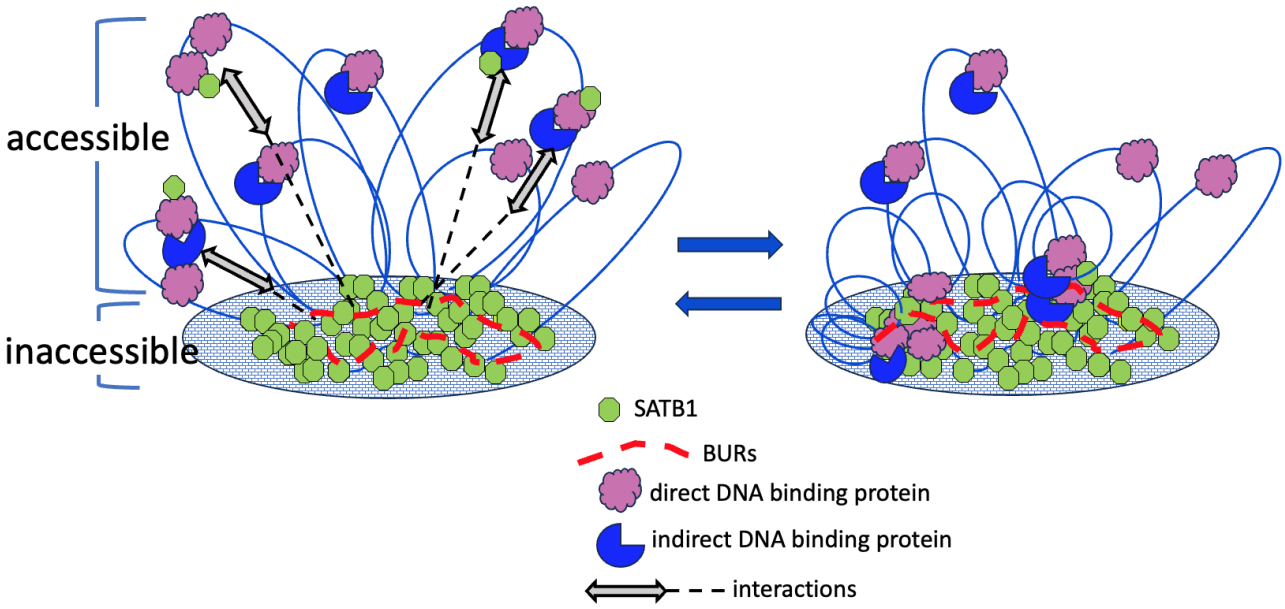

A
chr5:65,979,692-74,949,172
2 Mb
Lamin B1-DamID
BUR
Urea
SATB1
SATB1
Std
CTCF
H3Kme1
H4K27ac
thymocytes
brain
thymocytes
B
Lamin B1 DamID
Std SATB1
Urea SATB1
Relative Score
-10%
LADs
+10%
C
Lamin B1 DamID
Std SATB1
Urea CTCF
BUR
Relative Score
-10%
LADs
+10%
D
BUR
Urea SATB1
Lamin B1
H3K4me1
H3K27ac
Relative Score
-3kb
Midpoint H3K4me1
Peaks
3kb
E
BUR
Urea SATB1
H3K4me1
H3K27ac
Relative Score
-3kb
Midpoint H3K4me1
Peaks
3kb

FIGURE 8


accessible
inaccessible
SATB1
BURs
direct DNA binding protein
indirect DNA binding protein
interactions

Supplement figures and tables

Identification	Title	page
Figure 3--supplement figure 1	Similar DNA binding profiles for PRC2 core subunits obtained by stardard ChIP-seq and Urea ChIP-seq	1
Figure 3--supplement figure 2	SATB1 does not directly co-bind chromatin with CTCF and RAD21 or has a role in TAD formation	2
Figure 4--supplement figure 1	Standard ChIP-seq for SATB1 produces non-specific peaks in *Satb1*$^{-/-}$ thymocytes	3
Fig. 5--supplement figure 1	Urea 4C-seq protocol	4
Figure 5--supplement figure 2	Long-range chromatin interactions from BUR-1 traverse TADs	5
Figure 5—supplement figure 3	Long-distance interactions covering distal gene-rich regions by BUR-1	6
Figure 5--supplement figure 4	Genes within the gene-rich region (chr2:98,542,232-108,455,449) that exhibit SATB1-dependent expression	7
Figure 7--supplement figure 1	BURs and BUR-bound SATB1 largely co-map with LaminB1-DamID	8
Figure 7--supplement figure 2	Some SATB1-bound BURs located in SATB1-dependent gene loci are outside of LADs	9
Figure 7--supplement figure 3	Interaction sites from BUR-1 or BUR-2 are mostly confined to the inter-LAD region	10
Supplement Table 1 and Table 2	Supplement Table 1. Intersection analysis of ChIP-seq samples Supplement Table 2. ChIP-seq sample ID and peak number	11
Supplement Table 3 and Table 4	Supplement Table 3. Intersection analysis of WT and KO ChIP-seq samples Supplement Table 4. WT and KO ChIP-seq sample ID and total peak number	12

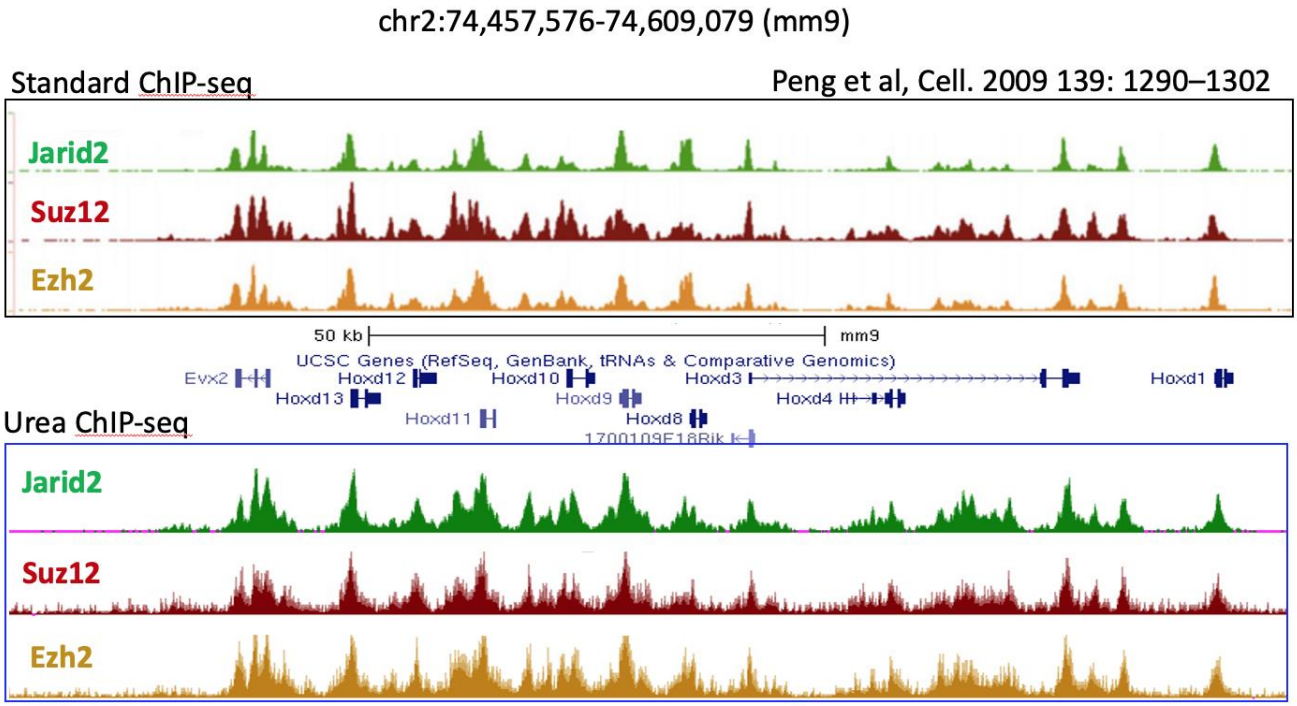


Figure 3--supplement figure 1. **Similar DNA binding profiles for PRC2 core subunits obtained by stardard ChIP-seq and Urea ChIP-seq**

The DNA binding profiles of PRC2 core subunits, Jarid2, Suz12, and Ezh2 obtained by standard ChIP-seq (64) [Peng et al., Cell, 2009, GEO repository (GSE18776)] were compared against the DNA binding profiles of these proteins by urea ChIP-seq. Their binding profiles in the *Hoxd* cluster region on chromosome 2 (mm9) are shown. Their DNA binding profiles by urea ChIP-seq and by standard ChIP-seq are very similar.

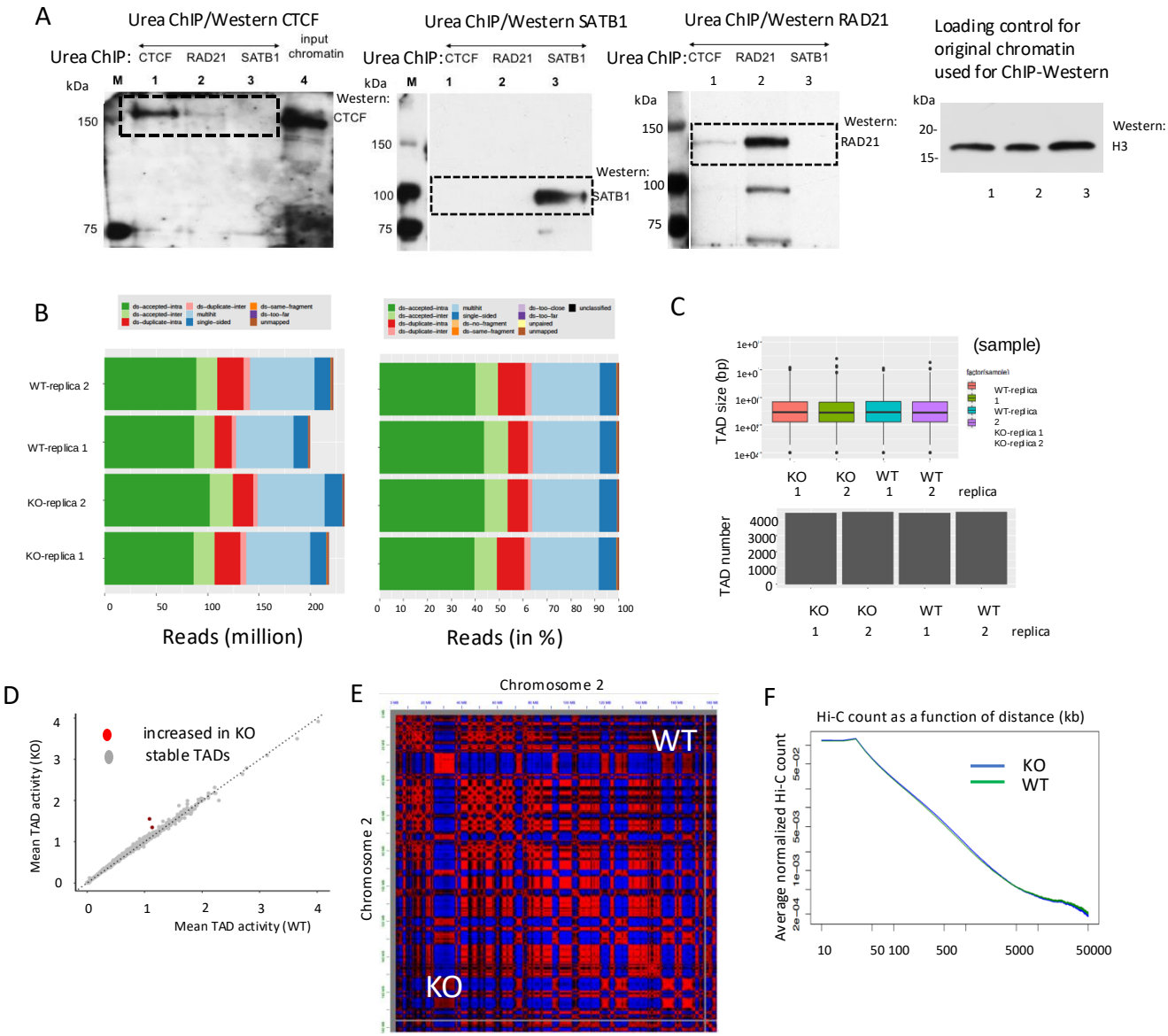


Figure 3--supplement figure 2. **SATB1 does not directly co-bind chromatin with CTCF and RAD21 or has a role in TAD formation.**

(A) Urea ChIP-Western studies for CTCF, RAD21 and SATB1. Original data for Figure 2C is shown. The excised areas are indicated by dotted boxes (first three panels). Western blot with the anti-H3 antibody shows that the original urea-purified chromatin samples before ChIP were similar in amount (right panel). **(B)** Read alignment statistics for Hi-C datasets, as absolute reads (left) and relative reads (in %, right). "ds.accepted.intra" are all intra-chromosomal reads used for all downstream analyses. **(C)** Distribution of TAD sizes and number of TAD calls per sample using hic-ratio TAD caller at 10kb bin resolution and the insulation score of 100kb windows. **(D)** Scatter plot of pair-wise comparisons of TAD activities (the sum of normalized Hi-C chromatin contacts inside the TAD in WT) between WT and KO thymocyte samples, showing well-conserved intra-TAD activity in KO thymocytes. **(E)** The Pearson correlation estimates computed between the counts for every pair of 10kb bins on chromosome 2 shows that WT (control: upper right triangle) and KO (observed:lower left triangle) thymocytes have no difference in compartmentalization. The compartment A (red) and compartment B (blue). **(F)** The scaling plots displaying how the Hi-C contacts decrease as a function of genomic distance.

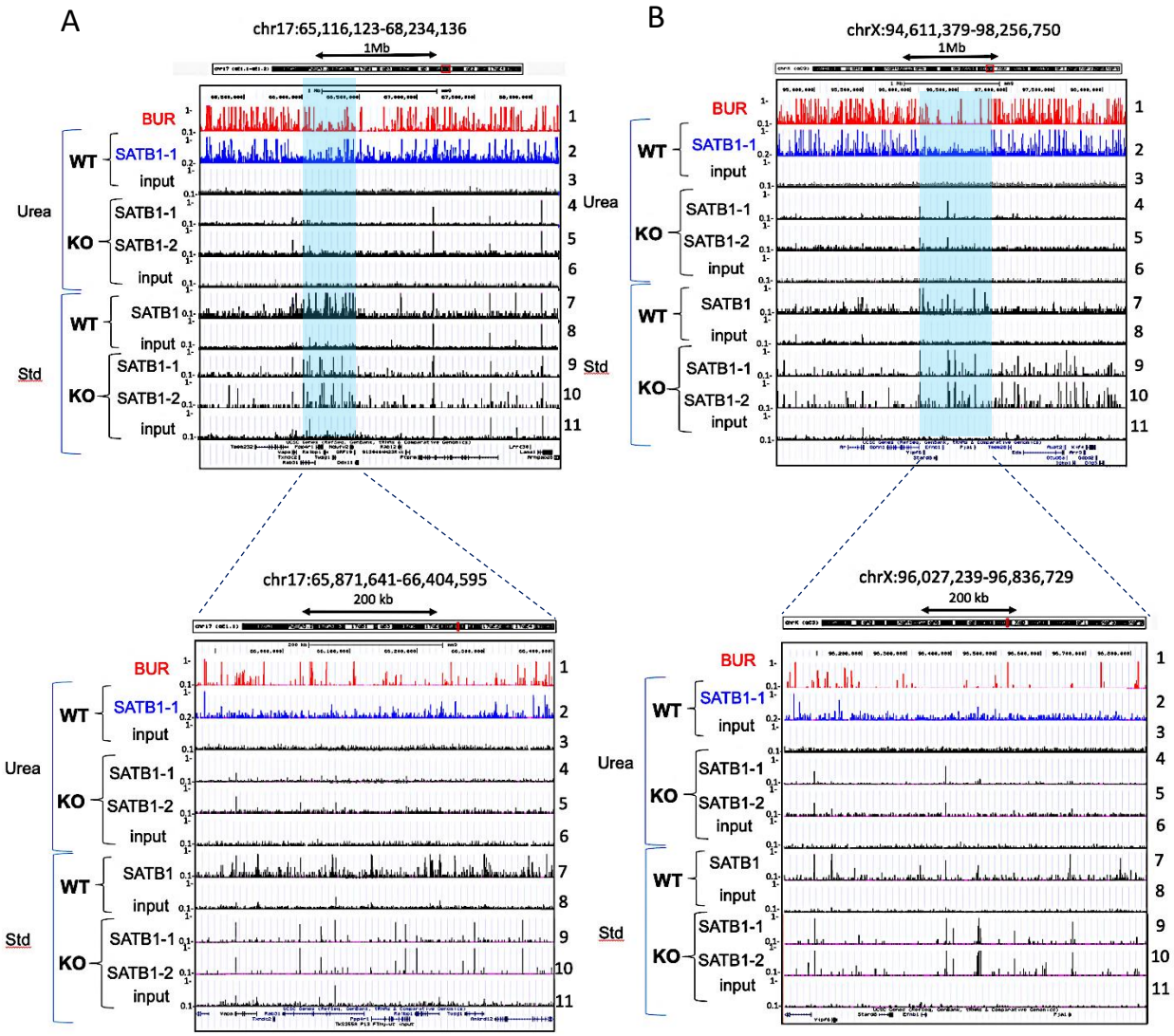


***Figure 4--supplement figure 1*. Standard ChIP-seq for SATB1 produces non-specific peaks in *Satb1*$^{-/-}$ thymocytes. (A)** Urea ChIP-seq (Urea) and std ChIP-seq (Std) profiles for SATB1 and input chromatin in a randomly chosen, 3.12 Mb region in chromosome 17 shows appearance of some peaks in input chromatin from std ChIP-seq that match with peaks for SATB1 in std ChIP-seq (top). For *Satb1*$^{-/-}$ thymocytes (KO), using either anti-SATB1 antibody 1583 (rabbit polyclonal, SATB1-1) or rabbit monoclonal ab109122 (SATB1-2), urea and std ChIP-seq experiments for SATB1 were performed. With std ChIP-seq, multiple peaks were generated in KO thymocytes in the similar regions observed in WT thymocytes (panel A track 7 versus 9 and 10), while urea ChIP-seq produced only few peaks (panel A top, track 2 versus 4 and 5). A zoom-in view of blue-highlighted SATB1-clustered regions (bottom) shows that for standard ChIP-seq, multiple peaks in KO thymocytes co-mapped with peaks in WT thymocytes, but others did not (panel A bottom, track 7 versus 9 and 10). **(B)** Another randomly chosen 3.45 Mb region in X chromosome (panel B top) and its zoom-in view (panel B bottom) shows high incidence of std ChIP-seq peaks in KO thymocytes co-mapping with std ChIP-seq peaks in WT thymocytes (tracks 7 versus 9 and 10).

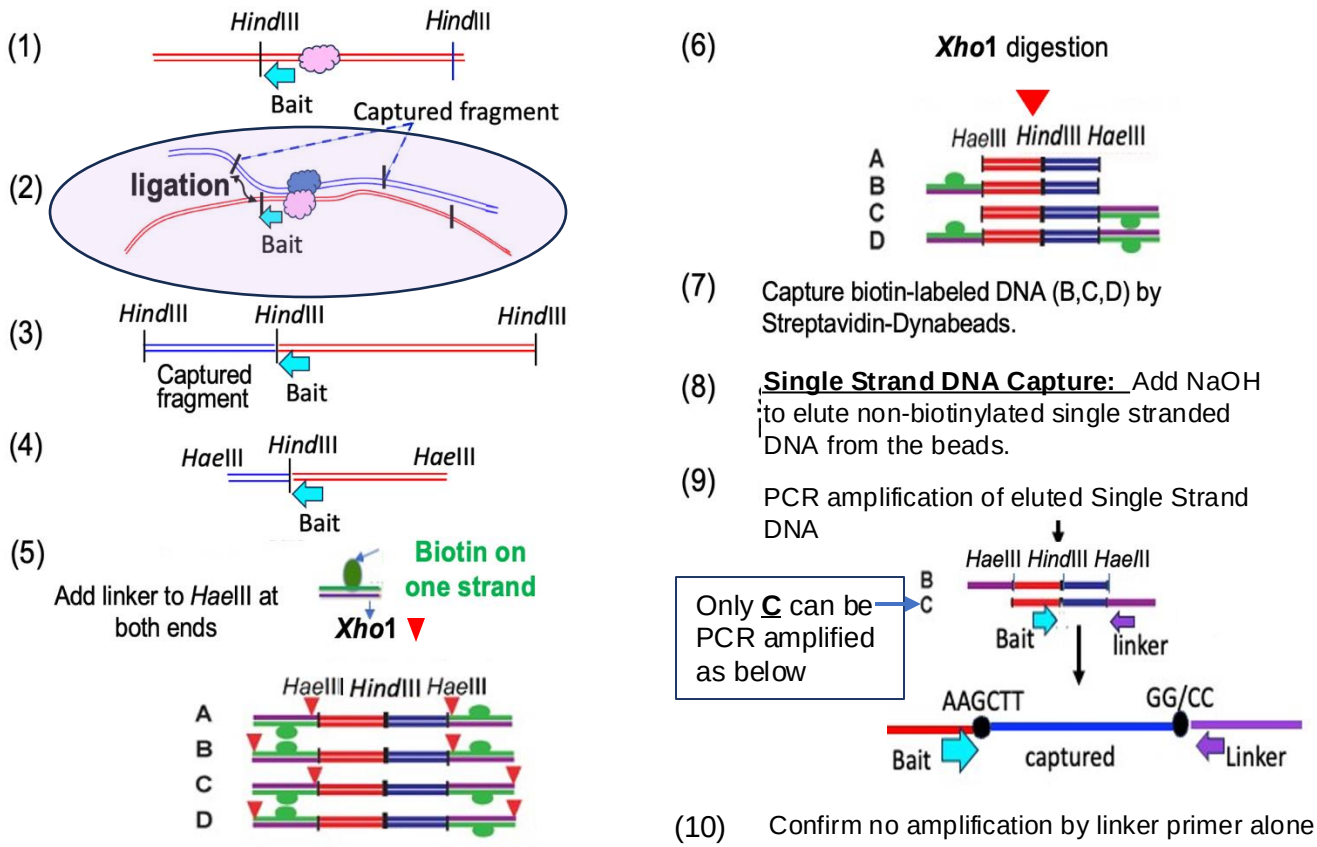


Fig. 5--supplement figure 1. **Urea 4C-seq protocol. Starting with urea-ultracentrifugation purified formaldehyde crosslinked chromatin**: **(1)** A bait sequence is designed adjacent to a *Hind*III site of a *Hind*III–*Hind*III fragment containing SATB1 bound site (e.g. BUR-1). **(2)** Ligation is performed to ligate the original SATB1-bound fragment (bait) with other fragments whose close proximity was fixed by formaldehyde at the time of crosslinking and presumably retained after urea purification. The fixed chromatin structure with proximity is illustrated by pink shade. **(3)** Chromatin is decrosslinked and purified. **(4)** Ligated genomic DNA fragments are digested with *Hae*III (secondary restriction enzyme). **(5)** A Linker fragment containing an *Xho*1 site and Biotin labeled on one strand is blunt-end ligated to *Hae*III sites of the ligated genomic DNA fragments shown in step 4 to generate A, B, C, D double stranded fragments. The *Xho* 1 sites are indicated by red arrowhead. **(6)** The ligated products are digested with *Xho*1.
(7) Biotin-labeled double stranded DNA (B, C, and D) are captured by Streptavidin-Dynabeads. **(8)** The beads are treated with NaOH to convert DNA to single strand DNA. This will result in dissociation and elution of non-biotinylated single strand DNA (derived from B and C) from the Dyanbeads. **(9)** Using a specific primer including the *Hind*III of the bait sequence at the end and another primer from the Linker, captured DNA fragments are PCR amplified. **(10)** Control experiment is performed to confirm that Linker primer alone does not generate any amplified products. Once this is confirmed, the experiment is expected to contain minimum background signals from non-specific sequences.

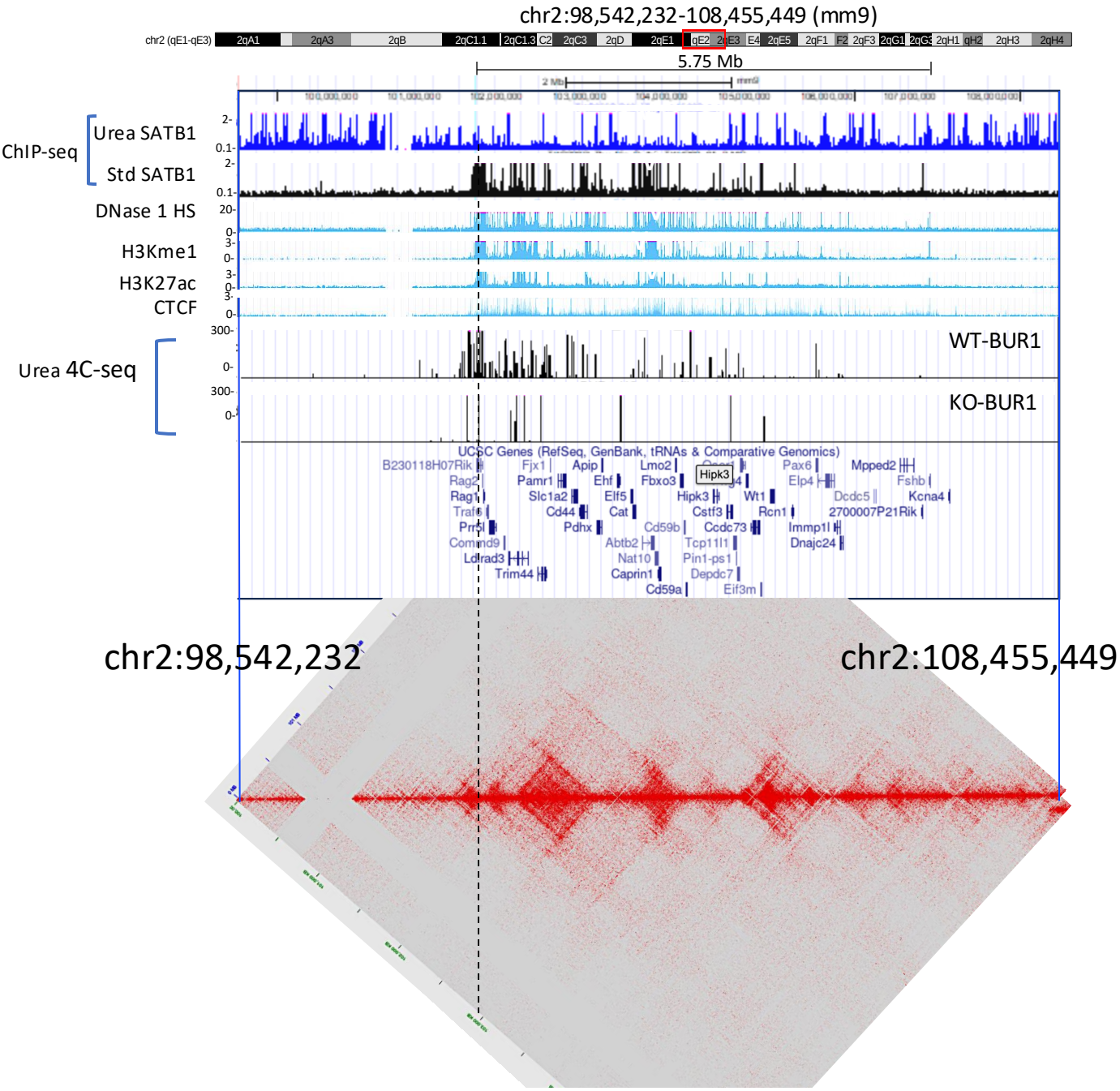


Figure 5--supplement figure 2. **Long-range chromatin interactions from BUR-1 traverse TADs**. BUR-1 interactions determined by urea4C-seq for wild type (WT) and SATB1-KO thymocytes covers the entire 5.75 Mb gene-rich region and these interactions bypass multiple small TADs revealed by the Hi-C heatmap (mm10) generated in the same region. Dotted line indicates the position of BUR-1

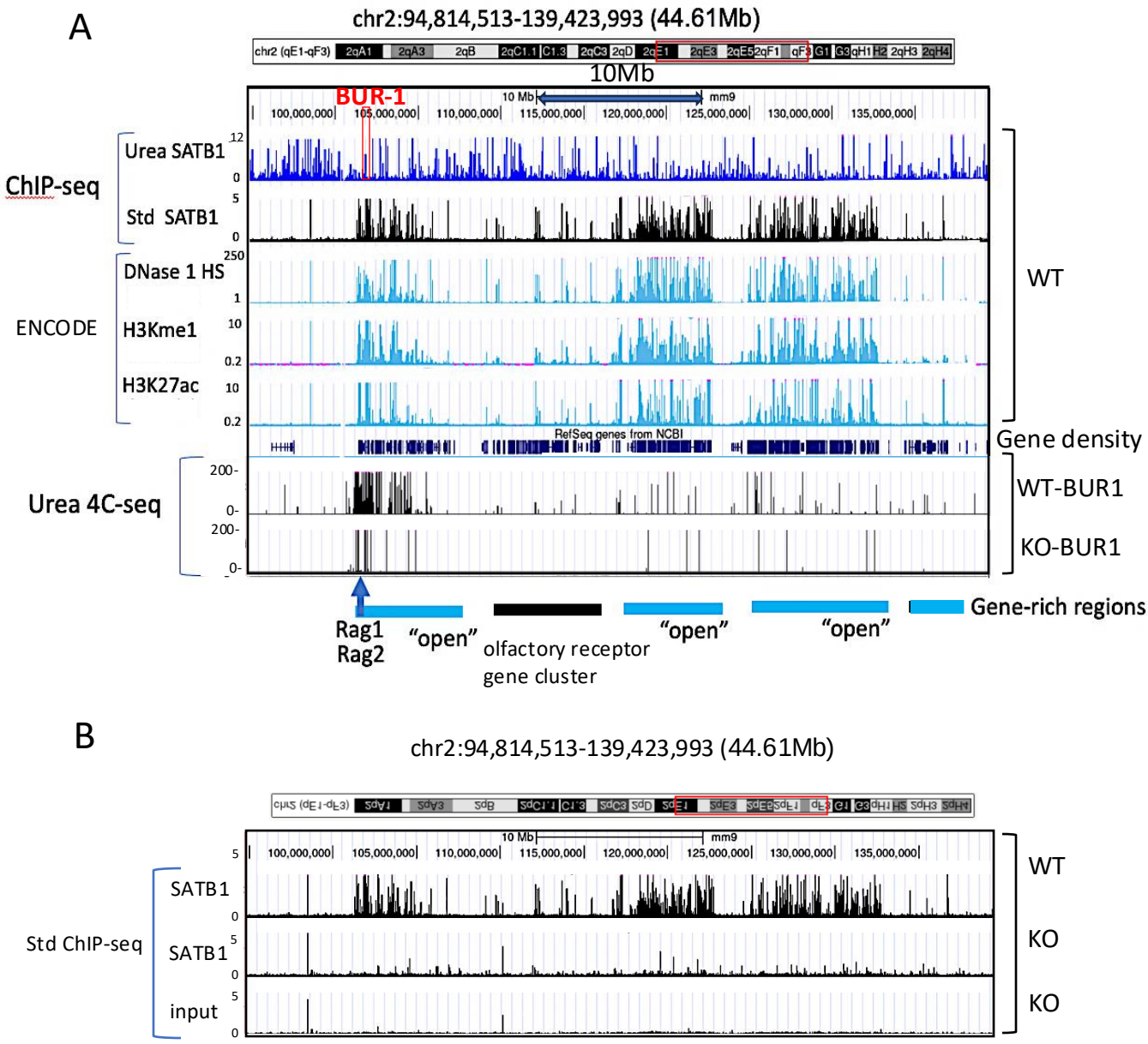


Figure 5—supplement figure 3. **Long-distance interactions covering distal gene-rich regions by BUR-1**. **(A)** A wide view (44.61 Mb) covering distal gene-rich regions shows BUR-1 (red bracket) interacts with distal gene-rich DNase 1 HS regions marked by clusters of H3Kme1 and H3K27a (indicated by blue horizontal bars). These interactions were greatly reduced in SATB1 deficient thymocytes (KO). A gene-rich region (indicated by a black horizontal bar) that does not correspond to DNase 1 HS is enriched in olfactory receptor genes. DNase 1 hypersensitive (HS) regions, which indicate highly accessive chromatin (or "**open**" chromatin). H3Kme1 and H3K27ac data are from ENCODE (LICR) and DNase 1 HS data are from ENCODE (University of Washington). **(B)** SATB1 peaks identified by standard (Std) ChIP-seq in this region are mostly real and only detected in wildtype (WT) and not in SATB1-null (KO) thymocytes.

chr2:98,542,232-108,455,449

	Gene Name	log2FoldChange	KO vs WT	p value
SATB1 activated	Rag2	-1.395	0.380	2.29E-14
	Fjx1	-1.138	0.454	0.0074
	Rag1	-0.973	0.509	8.57E-09
	Pamr1	-0.657	0.634	0.0865
	Mpped2	-0.541	0.688	0.1919
	Cd44	-0.521	0.697	0.0299
	Abtb2	-0.502	0.706	0.0055
	Cd59a	-0.369	0.774	0.3787
SATB1 repressed	Rcn1	0.472	1.387	0.0339
	Depdc7	0.518	1.432	0.2204
	Apip	0.543	1.457	0.0224
	Prr5l	0.676	1.598	0.0787
	Prrg4	1.032	2.045	0.0001

Figure 5--supplement figure 4. **Genes within the gene-rich region (chr2:98,542,232-108,455,449) that exhibit SATB1-dependent expression.** Genes showing differential expression for SATB1 with $p<0.05$ are shown in red (SATB1 activated) and in blue (SATB1 repressed). Data are from Zelenka et al., *Nat Commun*.,13, 6954, 2022 (54).

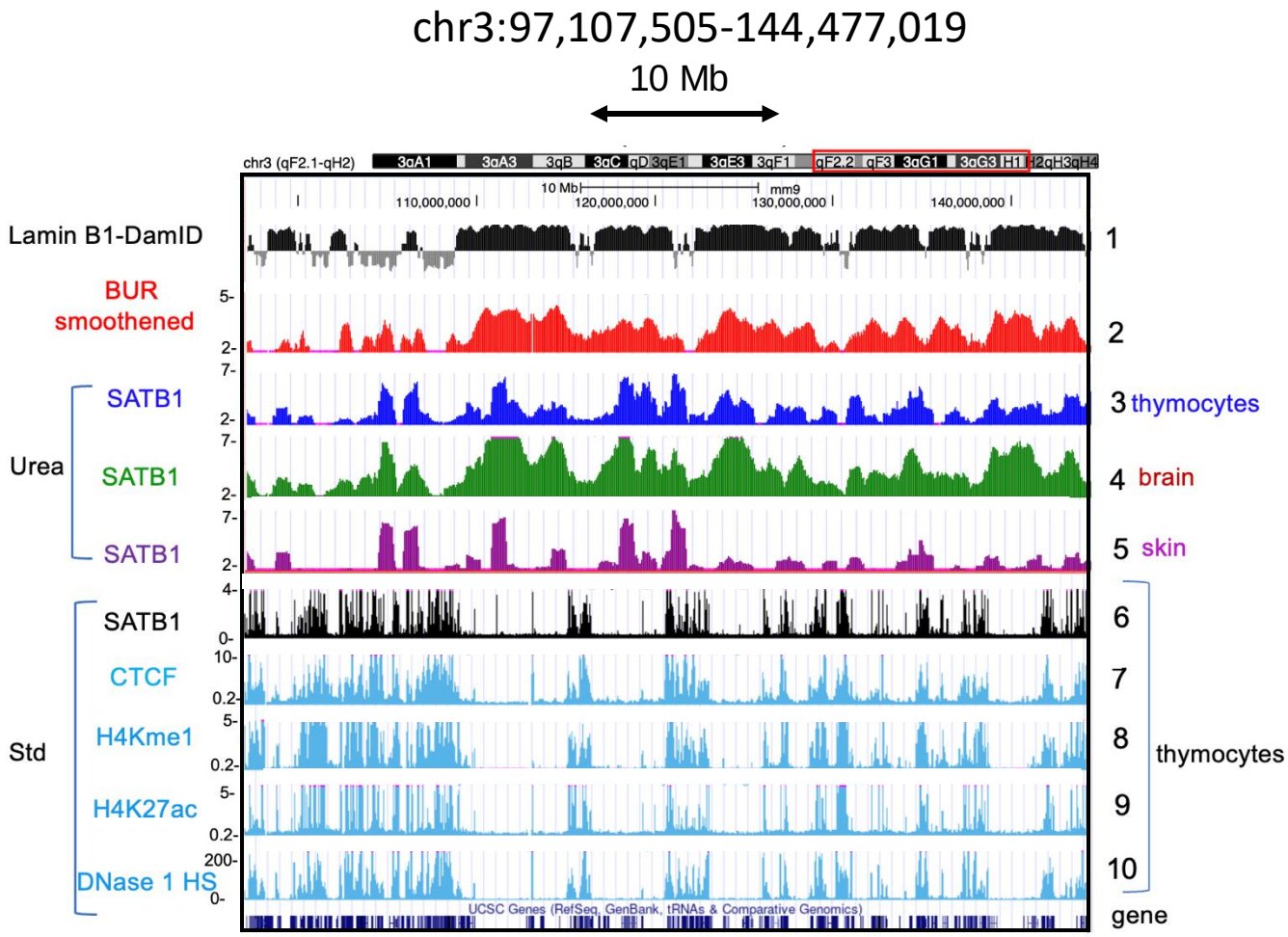


Figure 7--supplement figure 1 . **BURs and BUR-bound SATB1 largely co-map with LaminB1-DamID.** The BUR distribution assessed by the BUR reference map (tracks 2) was compared against Lamin B1-DamID (track 1) over a 47.4Mbp region in chromosome 3. Lamin B1-DamID data are from Hao et al., *J. Exp. Med*, 212, 809, 2015 (45). The BUR distribution shows very similar overall distribution of LADs mapped by Lamin B1-DamID. BUR-binding frequency of SATB1 in thymocytes (track 3) is lower than that in brain (track 4). Distribution of BURs (track 2) and SATB1-binding regions determined by urea ChIP-seq (Urea) (track 3, 4, and 5) largely overlap with LADs (track 1). These regions mostly avoid DNase hypersensitive regions (HS) (track 10), which overlap with SATB1-binding sites determined by standard ChIP-seq (Std) (track 6), CTCF-binding sites (track 7), enhancer marks (H3Kme1 and H3K27ac (tracks 8 and 9, respectively). Data for CTCF, H3Kme1 and H3K27ac are from ENCODE (LICR) and for DNase 1 hypersensitive sites (HS) is from ENCODE (University of Washington).

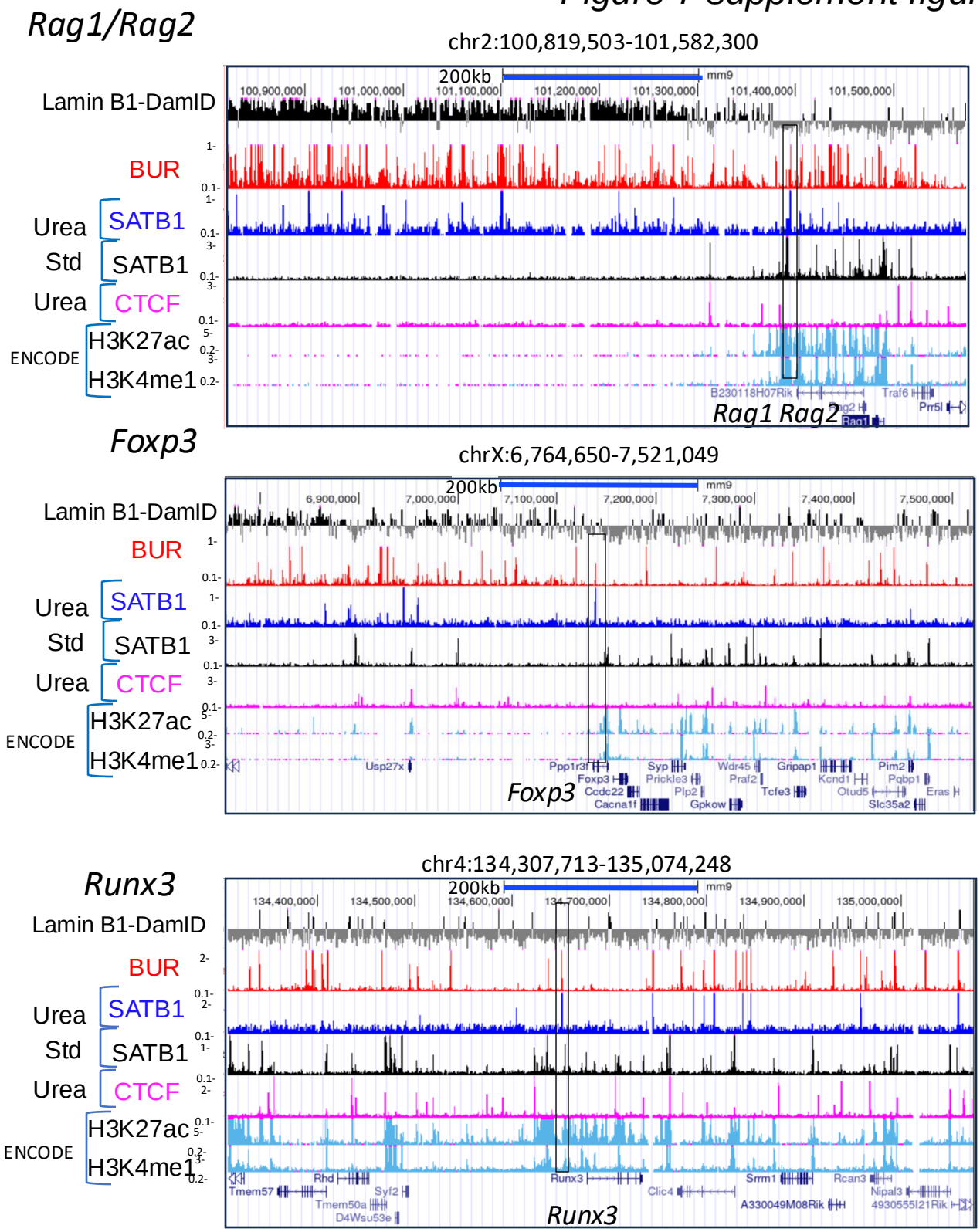


***Figure 7--supplement figure 2*. Some SATB1-bound BURs located in SATB1-dependent gene loci are outside of LADs.** SATB1-bound BURs shown in Figure 4 (indicated by black box) in the *Rag1/Rag2, Foxp3, and Runx3* gene loci are found outside LADs mapped by LaminB1-DamID. These are close to H3Kme1 and H3K27ac sites. H3Kme1 and H3K27ac data are from ENCODE (LICR).

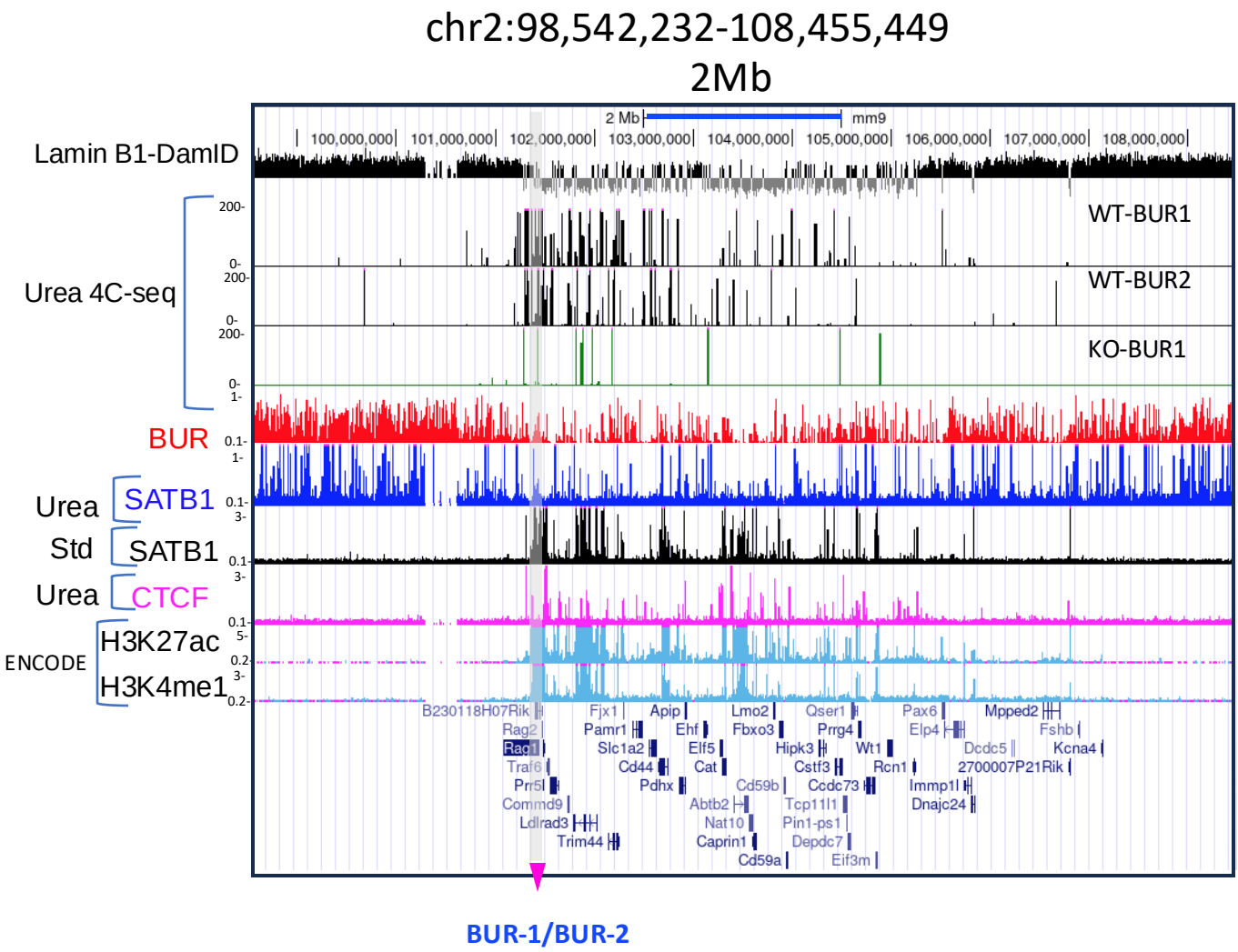


***Figure 7--supplement figure 3*. Interaction sites from BUR-1 or BUR-2 are mostly confined to the inter-LAD region.** The region containing BUR-1 and BUR-2 as baits is light gray shaded with a pink arrowhead. This region is near the border of a LAD and an inter-LAD region, shown by LADs identified by Lamin B-DamID. Interaction profiles from either BUR-1 and BUR-2 obtained in wild-type thymocytes (WT) are found in a region surrounded by LADs. These interactions are greatly reduced in Satb1-KO thymocytes (KO). Many interaction sites are close to H3Kme1 and H3K27ac sites. H3Kme1 and H3K27ac data are from ENCODE (LICR).

Supplement Table 1. Intersection analysis of ChIP-seq samples

	BUR-L	UREA SATB1 thymus	UREA SATB1 thymus	Standard SATB1 thymus	Standard SATB1 thymus	UREA SATB1 brain	UREA SATB1 brain	UREA SATB skin	UREA SATB1 skin	UREA CTCF thymus	UREA CTCF skin	Standard CTCF thymus
%	A	B	C	D	E	F	G	H	I	J	K	L
A-120A		12.68	14.34	0.77	0.60	36.41	33.54	3.69	3.62	0.25	0.58	0.60
B- 355B	77.71		85.02	1.12	0.86	92.07	90.28	24.13	23.56	0.77	1.00	0.86
C-355C	75.87	73.38		0.96	0.75	92.39	91.74	20.82	20.17	0.68	0.93	0.77
D-328B	3.83	0.91	0.90		47.92	1.28	1.11	1.93	1.91	22.35	23.31	36.39
E-408B	5.23	1.23	1.25	85.04		1.72	1.57	3.96	3.59	18.08	21.99	37.71
F-345A	73.58	30.35	35.28	0.52	0.40		78.12	7.80	7.59	0.31	0.58	0.56
G-345B	73.25	32.16	37.86	0.49	0.39	84.42		8.33	8.09	0.30	0.56	0.56
H-404B	78.54	83.69	83.65	8.30	9.59	82.06	81.13		81.06	2.35	12.85	1.78
I-404E	79.49	84.45	83.79	8.47	8.98	82.60	81.45	83.78		2.38	12.17	1.70
J-346D	2.96	1.47	1.52	52.97	24.15	1.78	1.61	1.29	1.27		83.53	60.22
K-403C	4.36	1.23	1.34	35.64	18.94	2.18	1.95	4.57	4.19	53.87		42.86
L-Encode LICR	2.68	0.63	0.65	33.00	19.27	1.24	1.14	0.37	0.35	23.04	25.42	

Supplement Table 1. Intersection of peaks were determined by GeneTool between two samples of choice. Percent of peaks in each sample on the left column intersected with peaks in each of the samples on the top row are indicated.

Supplement Table 2. ChIP-seq sample identification and peak number

		Sample ID	Total #Peaks
A. BUR-L		A-120A	238380
B. UREA SATB1 thymus	replica 1	B-355B	38882
C. UREA SATB1 thymus	replica 2	C-355C	45049
D. Std SATB1 thymus	replica 1	D-328B	48175
E. Std SATB1 thymus	replica 2	E-408B	27147
F. UREA SATB1 brain	replica 1	F-345A	117963
G. UREA SATB1 brain	replica 2	G-345B	109154
H. UREA SATB1 skin	replica 1	H-404B	11210
I. UREA SATB1 skin	replica 2	I-404E	10846
J. UREA CTCF thymus	replica 1	J-346D	20323
K. UREA CTCF skin	replica 1	K-403C	31509
L. Std CTCF thymus	Encode	L-Encode LICR	53124

Supplement Table 2. Identification of each sample and their total number of peaks are shown.

Supplement Table 3. Intersection analysis of WT and KO ChIP-seq samples

	urea ChIP-seq					standard ChIP-seq				
	TKS355B	TKS400B	TKS400D	TKS340A	TKS356A	TKS408B	TKS407D	TKS407B	TKS390A	TKS407C
	SATB1 ChIP	Phantom (KO SATB1 ChIP)	Phantom (KO SATB1 ChIP)	Urea Input	Urea Input	SATB1 ChIP	Phantom (KO SATB1 ChIP)	Phantom (KO SATB1 ChIP)	Standard Input	Standard Input
	UREA	UREA antibody 1	UREA antibody 2	WT	KO	STANDARD	Standard Antibody 1	Standard Antibody 2	WT	KO
TKS355B		0.0	0.0	0.4	0.4	0.8	0.1	0.1	0.5	0.3
TKS400B	1.0		78.5	33.6	6.1	85.7	89.7	93.8	53.0	15.3
TKS400D	0.8	66.1		25.4	4.2	88.9	93.2	93.2	41.5	11.8
TKS340A	10.7	1.4	1.3		51.2	55.6	7.8	10.6	79.2	39.3
TKS356A	17.7	0.4	0.3	90.8		76.1	8.3	12.3	93.4	59.4
TKS408B	1.2	0.3	0.3	4.1	3.1		3.6	4.5	12.7	3.0
TKS407D	2.3	3.1	3.8	6.7	4.3	36.4		90.9	9.6	4.1
TKS407B	1.3	1.3	1.5	3.5	2.3	18.2	37.2		4.9	2.2
TKS390A	3.6	0.7	0.6	25.2	16.9	52.0	3.6	4.6		14.3
TKS407C	16.7	1.2	1.1	75.7	63.9	79.8	8.3	12.0	86.5	

Supplement Table 3: Intersection of peaks were determined by GeneTool between two samples of choice. Percent of peaks in each sample on the left column that intersected with peaks in each sample on the top row are indicated. Important sets of data for comparison are highlighted in blue for urea ChIP-seq and pink for standard ChIP-seq.

Supplement Table 4. WT and KO ChIP-seq sample identification and total peak number

Sample ID	Description	Total # Peaks
TKS355B	SATB1 ureaChIP-seq	38882
TKS400B	KO SATB1 ureaChIP-seq 1	98
TKS400D	KO SATB1 ureaChIP-seq 2	118
TKS340A	WT Input ureaChIP-seq 1	2381
TKS356A	KO Input ureaChIP-seq 1	1372
TKS408B	SATB1 standard ChIP-seq	27147
TKS407D	KO SATB1 standard ChIP-seq 1	2884
TKS407B	KO SATB1 standard ChIP-seq 2	7003
TKS390A	WT Input Standard ChIP-seq 1	7377
TKS407C	KO Input Standard ChIP-seq 1	1224

Supplement Table 4: Identification of each sample and their total number of peaks are shown.