Chromatin organizer SATB1 controls the cell identity of CD4$^+$ CD8$^+$ double-positive thymocytes by compacting super-enhancers

Delong Feng  
Southern Medical University

YanHong Chen  
Medical Genetic Institute of Henan Province, Henan Provincial Key Laboratory of Genetic Diseases and Functional Genomics, Henan Provincial People's Hospital, Zhengzhou, Henan

Ranran Dai  
Southern Medical University

Shasha Bian  
West China Hospital, Sichuan University

Wei Xue  
Southern Medical University

Yongchang Zhu  
Southern Medical University

Zhaoqiang Li  
Southern Medical University

Yiting Yang  
Southern Medical University

Yan Zhang  
Southern Medical University

Jiarui Zhang  
Southern Medical University

Jie Bai  
Henan Provincial People's Hospital

Litao Qin  
Henan Provincial People's Hospital

Yoshinori Kohwi  
University of California, San Francisco

Weili Shi  
Henan Provincial People's Hospital

Terumi Kohwi-Shigematsu  
University of California, San Francisco  https://orcid.org/0000-0002-5988-0669
Article

Keywords: SATB1, invariant nature kill T cells, double-positive thymocytes

Posted Date: November 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1069634/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

CD4+ and CD8+ double-positive (DP) thymocytes are at a crucial stage during the T cell development in the thymus. DP cells rearrange the T cell receptor gene Tcra to generate T cell receptors with TCRβ. Then DP cells differentiate into CD4 or CD8 single-positive (SP) thymocytes, Regulatory T cells, or invariant nature kill T cells (iNKT) according to the TCR signal. Chromatin organizer SATB1 is highly expressed in DP cells and plays an essential role in regulating Tcra rearrangement and differentiation of DP cells. Here we explored the mechanism of SATB1 orchestrating gene expression in DP cells. Single-cell RNA sequencing assay of SATB1-deficient thymocytes showed that the cell identity of DP thymocytes was changed, and the genes specifically highly expressed in DP cells were down-regulated. The super-enhancers regulate the expressions of the DP-specific genes, and the SATB1 deficiency reduced the super-enhancer activity. Hi-C data showed that interactions in super-enhancers and between super-enhancers and promoters decreased in SATB1 deficient thymocytes. We further explored the regulation mechanism of two SATB1-regulating genes, Ets2 and Bcl6, in DP cells and found that the knockout of the super-enhancers of these two genes impaired the development of DP cells. Our research reveals that SATB1 globally regulates super-enhancers of DP cells and promotes the establishment of DP cell identity, which helps understand the role of SATB1 in thymocyte development.

Introduction

T lymphocytes, a critical component of adaptive immunity, develop as thymocytes in the thymus. Most of T cells in human and mouse are αβ T cells expressing the TCR consisting of an α and β chain. The primary purpose of αβ T cell development in the thymus is to generate T lymphocytes with highly diverse T cell receptors (TCRs) to recognize a wide variety of foreign antigens and avoid responses to self-antigens at the same time. During differentiation into mature T cells, thymocytes go through several developmental stages, which can be characterized by the cell surface expression of CD4 and CD8 proteins. CD4+ CD8+ double positive (DP) thymocytes are at the center of αβ T cell development. DP cells generate the αβ TCR on the cell surface to recognize cortical epithelial cells expressing Class I or Class II MHC plus self-peptides for positive and negative selection. TCR signaling induces DP cells differentiation into CD4+ single positive (SP), CD8+ SP, regulatory T cells (T\textsubscript{reg}), or invariant natural killer T cells (iNKT).

The primary event in DP cells is Tcra rearrangement to generate TCRs with an affinity to self-peptide:MHC complexes, which requires that cells do Tcra rearrangement efficiently and sense the TCR signal correctly. Due to the low frequency of positive selection, the Tcra gene can do multiple rounds of Vα-Jα rearrangement on both alleles to increase the chance of passing positive selection. DP cells also up-regulate recombinase Rag1 and Rag2 expression for efficient rearrangement. Cell lifespan is another critical factor for generating diverse Tcra repertoire and the proper development of DP thymocytes. DP thymocytes can survive for an average of three to four days to allow Jα rearrangement sequentially from the 5’ end to the 3’ end of the Jα array. Depleting the factors that regulate the survival of DP thymocytes...
(e.g., ROR, an orphan nuclear factor, and YY1, Yin Yang 1) shortens the lifespan and reduces the usage of the 3' Jα segments\(^9\)–\(^{11}\). However, it is unknown that DP thymocytes orchestrate expression programs for efficient \(T_{cra}\) rearrangement and proper selection.

Special AT-rich binding protein 1 (SATB1) is a chromatin organizer that plays certain roles in many tissues, including skin\(^1\)\(^2\), tooth\(^1\)\(^3\)–\(^4\), and liver\(^1\)\(^5\), although it is not ubiquitously expressed. Studies also showed that SATB1 is involved in embryogenesis\(^1\)\(^6\)–\(^7\), neurogenesis\(^1\)\(^4\)–\(^8\), hematopoiesis\(^1\)\(^9\),\(^2\)\(^0\), and erythropoiesis\(^2\)\(^1\),\(^2\)\(^2\). Nevertheless, most of the studies on SATB1 focused on its role in T cell development, partially due to the highest SATB1 expression during thymocyte development, especially at the DP stage\(^6\),\(^2\)\(^3\)–\(^2\)\(^6\). Satb1 deleted mice displayed a smaller thymus, increased proportion of DP thymocytes, and fewer CD4/CD8 single positive cells, indicating a blockage at the DP stage during thymocyte development\(^2\)\(^3\). The development blockage may be due to insufficient \(T_{cra}\) rearrangement and impaired positive selection\(^6\),\(^2\)\(^7\). In post-selection thymocytes, SATB1 also plays roles in activating lineage-specifying genes, including ThPOK, Runx3, CD4, CD8, and Foxp3\(^2\)\(^5\),\(^2\)\(^8\). SATB1-deficient thymocytes display inappropriate T lineage determination after MHC I- and II-mediated selection, and deficient differentiation of regulatory T cells (\(T_{reg}\))\(^2\)\(^8\). These findings indicate the critical role of SATB1 in DP thymocytes, although the detail molecular mechanisms remain elusive.

SATB1 can activate or repress gene transcription by recruiting p300/CBP-associated factor (PCAF) or histone deacetylase, respectively\(^2\)\(^9\). SATB1 functions as a pioneer factor in establishing \(T_{reg}\) cell-specific super-enhancers, which is crucial for \(T_{reg}\) cell lineage specification in the thymus\(^2\)\(^8\). More studies focus on the roles of SATB1 in chromatin organization, which was suggested when it was identified as a nuclear matrix protein\(^6\),\(^2\)\(^1\),\(^2\)\(^4\),\(^2\)\(^5\),\(^3\)\(^0\),\(^3\)\(^1\). An early report showed that SATB1 induced a unique transcriptionally active chromatin structure at the T helper 2 (\(T_{H2}\)) cytokine locus during \(T_{H2}\) cell activation\(^2\)\(^4\). SATB1 directly controls the regulatory elements of several lineage-specifying genes, including \(Zbtb7b\) (encoding ThPOK), \(Runx3\), \(Cd4\), \(Cd8\), and \(Foxp3\)\(^2\)\(^5\). We also reported that SATB1 mediates the DP-specific interaction between the anti-silencer element and \(Rag1\) gene promoter for the high expression of \(Rag1\) and \(Rag2\) in DP thymocytes\(^6\). However, there is a lack of genome-wide study on SATB1 mediated chromatin architecture and its characteristics\(^2\)\(^9\).

In the present study, we analyzed the development of \(Satb1\) deficient thymocytes using single-cell RNA sequencing (scRNA-seq) technique and found that Satb1 deletion changed the cell identity of DP thymocytes. Further analysis showed that Satb1 plays an essential role in promoting the activity of super-enhancers by reorganizing chromatin interactions. We also analyzed the regulation and function of two genes encoding transcription factors BCL6 and ETS2, which are regulated by SATB1 and super-enhancers in DP thymocytes. These findings indicate that SATB1 controls the cell identity of double-positive thymocytes by reorganizing super-enhancers.

**Results**
SATB1 deletion changed the cell identity of CD4⁺ CD8⁺ double-positive thymocytes

To investigate the role of SATB1 in thymocyte development, we employed single-cell RNA sequencing on all thymocytes from 6-week-old female mice in which the Satb1 gene was deleted in hematopoietic stem cells using vav-cre transgene. A total of 13948 cells consisting of 6844 Satb1 deficient and 7104 control thymocytes passed the quality control criteria. Cells were separated into 16 different clusters according to gene expression and cell cycling progress (Fig. 1a and S1a). Our previous study showed that SATB1 regulates recombinase Rag1 expression in DP thymocytes. The scRNA-seq data showed that reduced Rag1 expression occurred on the level of a single cell and the number of cells (Fig. 1b and S1b-c). We observed a decreased DN2/3 thymocyte number, an increased DN4/ISP/DP cell number, and a reduced CD4/CD8 SP cell number in the Satb1 deficient thymus (Fig. 1a and S1d), which is consistent with flow cytometry analysis in the previous report. We noticed that most of the DP thymocytes from Satb1 deficient thymus were enriched in the cluster three, while the cells in the cluster three were rare in the wild type (Fig. 1a), suggesting that the cluster three was generated by Satb1 deletion. The result indicated that Satb1 deficiency changed the cell identity of DP thymocytes.

To explore how Satb1 deletion changed the cell identity, we analyzed the gene expression profile of Satb1 deficient DP thymocytes with independent bulk RNA-seq data (Fig. 1c and S2a). There were 576 downregulated genes and 928 upregulated genes in Satb1 deficient DP thymocytes (Fig. S2b-d and Table S1). Most of the upregulated genes were expressed in thymocyte development earlier stages like DN1 and DN2a, and the expression levels of these genes are higher in the earlier stages (Fig. 1d, e and S2e). To confirm SATB1 repressing the expression of the genes expressed explicitly in earlier stages, we did the Gene Set Enrichment Analysis (GSEA) with the DN1 and DN3 specific gene sets. The data showed that the DN1 and DN3 gene sets were significantly repressed by SATB1 (Fig. 1f). Most of the downregulated genes in Satb1 deficient DP thymocytes were expressed specifically in DP and SP thymocytes (Fig. 1d). The expressions of these genes were low in the earlier stages, especially from DN2b to ISP, and they were highly expressed in DP and SP (Fig. 1e). These data suggested that SATB1 controls the cell identity of DP thymocytes by activating the DP-specific genes and repressing the genes expressed in earlier stages.

**SATB1 binds and activates super-enhancers of DP thymocytes**

It has been showed that super-enhancers drive expression of genes that define cell identity. The previous study showed that SATB1 plays a role in activating super-enhancers in Foxp3⁺ regulatory T cells. To explore the role of SATB1 in organizing super-enhancers of DP thymocytes, we identified 246 super-enhancers of DP thymocytes using the algorithm ROSE with chromatin immunoprecipitation sequencing (ChIP-seq) data of histone H3 acetylated at Lys27 (H3K27ac) (Fig. 2a and Table S2). Many super-enhancers are associated with known DP signature genes, such as Tcra, Rag1, Cd8a, Cd4, etc. (Fig. 2a). Most of DP super-enhancer-associated genes are highly expressed in the DP stage during...
thymocyte development (Fig. 2b), suggesting that super-enhancers control the expression of cell identity genes in DP thymocytes.

We noticed that SATB1 binding sites overlapped with the active histone modification markers histone H3K4 monomethylation (H3K4me1), H3K4me3, H3K27ac, and the binding sites of the components of chromatin organization complex cohesion Rad21, Nipbl, and CTCF (Fig. 2c and S3a-d). It was reported that active promoters can produce false positive peaks in ChIP-seq experiments. The SATB1 ChIP-seq experiment in Satb1 deficient thymocytes from Satb1$f/f$. CD4-cre mice showed that the SATB1 binding in active regions were specific in SATB1-expressing WT cells (Fig. 2c). The SATB1 signals were enriched in gene promoter regions and super-enhancers (Fig. 2d and S3e-f). The Satb1 deletion impaired both traditional enhancers and super-enhancers (Fig. 2e and S3g). The average intensities of H3K27ac and H3K4me1 in super-enhancers were reduced in Satb1 deficient DP cells (Fig. 2e and S3h). Most of the genes associated with the lost and maintained super-enhancers were downregulated in Satb1 deficient DP thymocytes (Fig. 2f-h and S3i). The genes associated with super-enhancers were more sensitive to Satb1 deletion than typical enhancers (TEs) associating genes (Fig. 2o). These data suggested that SATB1 regulates the DP signature genes by activating super-enhancers.

SATB1 clusters are associated with cell identity genes.

The SATB1 binding sites were clustered in the genome and displayed a super-enhancer like distribution (Fig. 2c). We did the ROSE analysis with SATB1 ChIP-seq data and identified 743 SATB1 clusters (Fig. 3a). The SATB1 clusters associated genes consisted of many DP signature genes, such as Tcra, Rag2, Ets1, Cd8a, Cd4, etc. The GSEA analysis showed that SATB1 clusters associated genes had high level of active histone modification H3K27ac (Fig. 3b). Satb1 deletion downregulated most of the SATB1 clusters associated genes (Fig. 3c), suggesting that SATB1 clusters tended to activate the expression of their associated genes. We analyzed the expression of the SATB1 clusters associated genes during thymocyte development. Most of these genes are specifically expressed in the DP stage and their expression are significantly lower in other stages than in DP thymocytes (Fig. 3d). The Gene Ontology (GO) term enrichment analysis showed that these genes enriched in DP associating functions, such as T cell activation and V(D)J recombination (Fig. 3e). These results indicated that SATB1 formed clusters to regulate the DP signature genes.

SATB1 mediated the chromatin interactions of super-enhancers.

The previous report showed that SATB1 regulated V(D)J recombinase Rag1 and Rag2 expression in DP thymocytes by mediating an enhancer-promoter interaction. We performed Hi-C experiments with sorted DP thymocytes from Satb1$f/f$ and Satb1$f/f$. vav-cre mice. Visualization of Hi-C data revealed some alterations of chromatin organization in Satb1 deficient DP thymocytes (Fig. S4). Satb1 deletion induced a modest increase in diagonal interactions of most. The contact density decaying curves showed that the interactions less than 2Mb in length slightly increased in Satb1 deficient thymocytes (Fig. S4b). Satb1 deletion has few effects on compartments (Fig. S5a). Satb1 deletion didn't change the number of TADs
but slightly reduced sizes of TADs, especially the TADs containing SEs (Fig. S5b). The TAD strength almost kept the same level in Satb1 deficient cells (Fig. S5c). The data suggest that the role of SATB1 on the organization of chromatin is in some specific regions rather than the whole genome.

To characterize chromatin interactions affected by Satb1 deletion, we generated a chromatin interaction matrix with a 50 kb resolution and identified significantly differential interactions using the Bioconductor package multilHiCcompare \(^{34}\). We compared 766,111 chromatin interactions and found 500 significantly increased interactions and 411 decreased interactions. Most of the distances between two anchors of decreased interactions are around 10kb, less than increased interactions (Fig. 4a). In addition, there are high SATB1, H3K27ac, and CTCF ChIP-seq signals in the anchor regions of decreased interactions (Fig. 4b). These data showed that, unlike CTCF and Cohesin, which mediate long-range chromatin interactions, SATB1 mainly mediates chromatin relatively short-range chromatin interactions (from tens to 100kb).

We then explored the chromatin loop structure using the program Fit-Hi-C, a tool for assigning statistical confidence estimates to intra-chromosomal contact maps \(^{35}\). Satb1 deletion increased loop numbers but reduced the loop strength (Fig. 4c and S5d). We also notified the reduced strength of enhancer-promoter loops and intra-SE loops (Fig. 4c and S5d). The loop strength increased in the promoters of the upregulated genes and decreased in the downregulated genes (Fig. S5e). The loop sizes also reduced in Satb1 deficient thymocytes (Fig. 4d). Reduced loop strength in SE may decrease activity of SE. We then analyzed the loop numbers, H3K27ac levels, associated gene expressions of SEs. 80% of SEs had a reduced loop strength and 66% a reduced loop number (Fig. S5f). The loop numbers and loop strength were significantly correlated (Fig. S5f). The loop numbers and strength of SEs were correlated with H3K27ac and gene expression (Fig. 4e and S5g to h), suggesting that Satb1 deletion impaired the organization and activity of SEs. To further validate the relationship of SATB1 binding and chromatin looping, we analyzed the loop strength of SATB1 clusters. The aggregate pile-up analysis showed the reduced loop strength of SATB1 cluster regions in Satb1 deficient thymocytes (Fig. 4f). These data indicated that SATB1 promotes internal interactions of SEs and interactions between SEs and promoters to affect gene expression by forming compacted chromatin organization.

**SATB1 regulated transcription factors Bcl6 and Ets2 by mediating chromatin topology.**

SATB1 regulates the DP signature genes, including \(Tcra\), \(Rag1/2\), \(Cd4\), and \(Ets1\), most of which have been proved to play an essential role in DP thymocytes. We also observed that some genes, like \(Bcl6\) and \(Ets2\), were regulated by SATB1 and super-enhancers in DP thymocytes (Fig. 1c). We further explored the regulation of the \(Bcl6\) and \(Ets2\) genes in DP thymocytes. B-cell lymphoma 6 protein (Bcl6) is a zinc finger transcription repressor and is found to be frequently translocated in diffuse large B cell lymphoma \(^{36-39}\). It is a master transcription factor for the differentiation of Follicular Helper T cells (Tfh) \(^{40}\). ETS proto-oncogene 2 (Ets2) belongs to the ETS family of transcription factors and is involved in stem cell development, cell senescence and death, tumorigenesis, and thymocyte development \(^{41-45}\). These two genes were reported involved in thymocyte development, but their regulations remain elusive \(^{43,46}\).
We observed that Satb1 occupied the loci of the Bcl6 and Ets2 genes in DP thymocytes (Fig. 5a and 5b). There are super-enhancers located at the Bcl6 upstream and Ets2 downstream, respectively (Fig. 5a and 5b). These two super-enhancers span more than 100kb and were characterized of H3K27 acetylation, Satb1, cohesin, and CTCF binding (Fig. 5a and 5b). More important, the super-enhancers and Bcl6 or Ets2 are in a topology associating domain (TAD) or sub-TAD. The Hi-C data showed that the chromatin interactions in the two loci, including super-enhancers and promoters, decreased dramatically in Satb1 deficient DP thymocytes (Fig. 5a, 5b, S6a, S6b, and S6c). 3C-HTGTS data revealed that Bcl6 or Ets2 promoter had strong interactions with the super-enhancer in WT DP thymocytes and the interactions reduced in Satb1 deficient cells (Fig. 5c and 5d). The H3K27 acetylation of the super-enhancers and promoters also reduced dramatically (Fig. 5c and 5d), indicating the reduced chromatin interactions the activities of super-enhancers and promoters. RNA-seq and qPCR data showed the reduced expression of Bcl6 and Ets2 in Satb1 deficient thymocytes (Fig. 1c, 5e, and S1c). Analysis of Bcl6 and Ets2 expression during thymocyte development showed that these two genes expression peaks are in DP stages, a similar pattern as Satb1 (Fig. 5f). These results suggested that the specifically high expressions of Bcl6 and Ets2 in DP thymocytes were regulated by super-enhancers and SATB1.

**Super-enhancer regulates Ets2 expression in the thymus**

To explore the role of the super-enhancer in Ets2 (named as Ets2-SE) expression in thymocyte, we generated Ets2-SE knockout mice in which the 166kb region containing the Ets2-SE was deleted. The Ets2-SE deletion dramatically reduced the Ets2 expression in thymocytes (Fig. 6a). The 4C assay showed that the Ets2 promoter had much less interactions with the whole locus including upstream and downstream regions (Fig. 6b). The result indicated that the SE regulates Ets2 expression in thymocytes.

Then we analyzed the thymocyte development of the Ets2-SE deleted mice. The cell numbers of thymi reduced by 69% (n=7) in the Ets2-SE−/− mice (Fig. 6c). Flow cytometric analysis showed that Ets2-SE deletion slightly increased the percentage of CD4−CD8− double negative 4 (DN4) thymocytes and reduced DP thymocytes (Fig. 6d), indicating the defective transition from DN to DP. We analyzed the cell viability of thymocytes in culture and the result showed that Ets2-SE deleted thymocytes had a shorter lifespan (Fig. 6e), which might explain the defective development. It was reported that short lifespan of DP thymocytes caused impaired Tcra rearrangement\(^9, 10\). We analyzed Tcra rearrangement using a single primer pair targeting C region of the Tcra gene during 5’rapid amplification of cDNA ends (5’ RACE)\(^47\). MiXCR immune repertoire analysis program was used for Ja and Va usage\(^48\). The Ets2-SE deletion didn’t affect the Ja and Va usage (Fig. S7a-b). These results indicated that the super-enhancer controls Ets2 expression in thymocytes, which plays a role in the DN-to-DP transition and DP lifespan during thymocyte development.

**The Bcl6-SE regulated Bcl6 expression and Tcra rearrangement in thymocytes**

To confirm the role of the Bcl6-SE in Bcl6 regulation in thymocytes, we deleted a 119.2kb region (chr16:24146914-24266171) containing the Bcl-SE in mice. The deletion reduced the Bcl6 expression
around fifty folds in thymocytes and dramatically changed the chromatin conformation of the locus (Fig. 7a). The cell numbers of thymocytes decreased in Bcl6-SE homozygous mice (Fig. 7b). The proportion of DN, DP, and SP populations was not affected by the Bcl6-SE deletion (Fig. 7c). Within the DN population, the percentage of DN3 was increased significantly (Fig. 7c), which is consistent with the observation in the conditional Bcl6 knockout mice with a lck-cre transgene. We also analyzed T cells in spleen, mesenteric lymph nodes, inguinal lymph nodes, and auxiliary lymph nodes (Fig. S8a-b). We only observed the increased ratio of CD4<sup>+</sup>/CD8<sup>+</sup> T lymphocytes in inguinal lymph nodes of Bcl6-SE deleted mice (Fig. S8a-b). The Bcl6-SE deletion didn’t affect cell lifespan (Fig. 7d).

We also detected Tcra rearrangement using 5′ RACE sequencing. We noticed that the usages of the proximal Vα genes increased slightly and the usages of the distal Vα genes reduced in thymocytes of Bcl6-SE mice (Fig. S8c). To show the difference more clearly, we combined Vα genes into five groups: the proximal, the proximal repeats, the central repeats, the distal repeats, and the distal. The usages of the proximal and the proximal repeat increased and the central repeat, the distal repeat, and the distal decreased (Fig. 7e). Consistent with the abnormal usage of the Vα genes, the Ja usage data showed an abnormal pattern with increased 5′ Ja and reduced 3′ Ja usage in the Bcl6-SE deleted thymocytes (Fig. 7f). Taken together, the results indicated that the super-enhancer regulates Bcl6 expression and plays a role in normal T cell development and Tcra rearrangement.

**Discussion**

DP cells are at a critical stage of T cell development, which mainly undergo two biological processes: 1) generate highly diverse T cell receptors through Tcra rearrangement and performing positive and negative selection simultaneously; 2) determine the direction of differentiation according to the TCR signal, producing CD4<sup>+</sup> SP, CD8<sup>+</sup> SP, T<sub>reg</sub>, and iNKT, respectively. Some transcription factors such as TCF-1<sup>49,50</sup>, E proteins<sup>51</sup>, c-Myb<sup>52,53</sup>, and ROR<sub>γ</sub><sup>9</sup>, are involved in regulating DP cells. However, most of the factors only participate in one of biological processes, rather than acting as a master regulator. SATB1 regulates Tcra rearrangement<sup>6</sup>, positive and negative selection<sup>27</sup>, and lineage decision in DP cells<sup>25</sup>, which makes it as a good candidate of the master regulator of DP thymocytes. Here we provided evidence that SATB1 controls the DP cell identity in a single cell transcription profile, although SATB1 deficient DP cells still highly express CD4 and CD8. Furthermore, the regulatory effect of SATB1 on DP cell identity is specific because SATB1 deletion does not change the transcription programs of thymocytes at other stages such as DN2/3 and CD4<sup>+</sup>/CD8<sup>+</sup> SP.

Since the concept of super-enhancer was proposed, many studies have supported the role of super-enhancers in regulating cell identity genes<sup>32,54</sup>. Our data also showed that super-enhancers control genes involved in Tcra rearrangement and positive/negative selection. It was reported that SATB1 acts as a pioneer molecular in establishing T<sub>reg</sub> cell-specific super-enhancers<sup>28</sup>. However, SATB1 has a high occupancy at super-enhancer regions and mediates interactions within super-enhancers and between
super-enhancers and promoters in DP cells, suggesting that SATB1 regulates DP cell identity genes by reorganizing super-enhancers.

SATB1 regulates many genes related to DP thymocyte function, including *Rag1, Tcra, Cd4, Cd8a* etc. Here we showed that transcription factors Bcl6 and Ets2 are controlled by SATB1 and super-enhancers in DP thymocytes. Super-enhancer knockout mice confirmed that the high expressions of Bcl6 and Ets2 in DP cells play an essential role in DP cells. The research on the dominant-negative truncated Ets2 transgenic mice and a phosphomutant Ets2 (T72A) transgenic mice showed that Ets2 plays an essential role of Ets2 in thymocyte development. In this study, Ets2-SE mice have a similar phenotype, indicating that Ets2 high expression regulated by the super-enhancer is critical for the development and survival of DP cells. Bcl6 is a transcriptional repressor that plays an essential role in the germinal center response and is also involved in leukemogenesis. Recent studies have shown that conditional deletion of Bcl6 lead to defective differentiation of DN to DP and abnormal activation of Notch signaling in DP cells. Consistent with the previous report, Bcl6-SE knockout mice also displayed reduced cell number of thymocytes and defective transition from DN to DP. We also noticed that *Tcra* rearrangement was impaired, and the mechanism remains elusive. These results support the notion that SATB1 orchestrates DP thymocyte function-related genes through reorganizing super-enhancers.

In summary, this study explored the mechanism by which SATB1 controls the cell identity of DP thymocytes and provides evidence that SATB1 promotes the intra-interactions of the super-enhancers, augments the super-enhancer activity, and then enables the high expression of cell identity genes. Thus, SATB1 maintains the cell identity of DP cells and ensures the normal development of thymocytes.

### Methods

#### Mouse

Satb1^{fl/fl} vav-cre^{+} mice were generated as previously described, and used in this study as Satb1cKO mice. The Bcl6 SE^{-/-} and ETS2 SE^{-/-} mice were generated using CRSIPR-Cas9 system by Beijing Vitalstar Biotechnology. The deleted regions were chr16: 24146828-24266085 and chr16: 95745432-95912361, respectively. All experiments involving mice were performed using protocols approved by Southern Medical University Animal Studies Committee. Animals were housed and bred in a specific pathogen-free animal facility.

#### scRNA-seq Library Construction and data processing

The single-cell library was constructed using the ChromiumTM Controller and ChromiumTM Single Cell 3' Reagent Version 2 Kit (10×Genomics, Pleasanton, CA) according to the manufacturer’s instructions. The final libraries were sequenced using the Illumina Hiseq 4000 (BGI-Shenzhen, China).

For each sample, the cleaned data was generated by Cell Ranger (v2.2.0) and filtered for the low-quality reads and unrelated sequence. The data was aligned to mouse mm10 reference genome. Data merging,
thresholding, normalization, principal component analysis, clustering analysis, visualization, differential gene expression analysis, and cell cycle phases analysis were carried out in Seurat (v2.3.4) according to their recommended steps (https://satijalab.org/seurat). In details, cells were sorted based on the barcodes and the unique molecular identifiers (UMIs) were counted per gene for each cell. In total, 8872-9283 (averagely 9077) cells were captured for individual libraries, and 1161-1621 (averagely 1391) genes were detected with UMIs per cell. Cells having total mitochondria-expressed genes beyond 10% were eliminated, along with cells expressing less than 500 or greater than 3000 total genes. After this, we performed global normalization using the SCTransform function in Seurat Seurat (v2.3.4). These pre-processed data were then analyzed to identify variable genes and principal component analysis.

For further analysis, UMAP were used for dimensionality reduction. Cells were represented in a two-dimensional UMAP plane, and clusters were identified and annotated according to the previously published canonical immune markers. The cell cycle phase score was calculated for each cell using the Seurat function CellCycleScoring. Significance of differential expression was calculated using the Wilcoxon rank-sum test.

**RNA isolation and bulk RNA-Seq.**

The sorted DP cells were isolated using a TRIzol. The total RNA was quantified and qualified by Agilent 2100 Bioanalyzer and NanoDrop2000. 1µg of total RNA was used for following library preparation. The poly(A) mRNA isolation was performed using Poly(A) mRNA Magnetic Isolation Module. First strand cDNA was synthesized using ProtoScript II Reverse Transcriptase and the second-strand cDNA was synthesized using Second Strand Synthesis Enzyme Mix (New England Biolabs). The purified double-stranded cDNA was subjected to end repair, 3'-dA tailing, and adapter ligation. Size selection of adapter ligated DNA were performed prior to PCR amplification. The ligated DNA was then amplified by 10-15 cycles with Illumina P5/P7 primers. The PCR products were cleaned up using AMPure XP beads (Beckman Coulter), validated using an Qsep100 and quantified by Qubit3.0 Fluorometer (Invitrogen). Libraries were a 2×150bp paired end (PE) sequenced on a NextSeq 550 on an Illumina HiSeq instrument. Three biological replicates were performed in Satb1WT and Satb1cKO DP T cells.

Fastq files were processed by Cutadapt(v1.18)(http://code.google.com/p/cutadapt) to be high quality clean data, then clean data were aligned to mouse genome (mm10) by Hisat2 (http://ccb.jhu.edu/software/htsat2/index.shtml). The abundance or the coverage of each transcript was determined by read counts and normalized using the number of reads per kilobase exon per million mapped reads (RPKM). Genome browser tracks in bigWig format were generated using bamCoverage function of deetools (v3.5.0) (https://deeptools.readthedocs.io/en/develop).

Differential expression genes analysis was performed using DESeq2 (v1.30.0) with default setting. The increased genes were defined as log2FC > 1 and adjusted p < 0.05, and the decreased genes as log2FC < −1 and adjusted p < 0.05. To verify the rationality of the DEGs, the volcano plot and heatmaps were draw using ggplot2 package in R. To cluster the samples and calculate the correlation coefficients between the
samples, a Spearman correlation test was applied, and the results was visualized using R package pheatmap.

Using WebGestalt (http://www.webgestalt.org) database on DEG set to obtain all Gene Ontology (GO) terms and KEGG pathways, accompanied by number of genes in that GO-term and pathway, enriched p-value, and FDR. Only the GO-term and pathway with a FDR value < 0.05 were considered as significantly enriched.

RNA-seq data from mouse T cell precursors in different developmental stages including DN1, DN2a, DN2b, DN3, and DP, CD4^SP, CD8^SP (Gene Expression Omnibus accession: GSE109125) were used to create DN1 and DN3 gene sets. Of the RNA-seq dataset, genes that were differentially up-regulated (P < 0.05 and Log2 fold change (FC) > 1) between DN1 versus DP, DN3 versus DP were used as gene sets for GSEA. GSEA was run on all expressed WT_DP versus Satb1cKO_DP RNA-seq genes, which were ranked by log2 FC value.

**ChIP-seq**

ChIP-seq data were either generated in this study or downloaded from public resource. The raw data were processed and analyzed according to the following procedure. First, SRA files were converted to fastq format, then aligned to mouse genome (mm10) using Bowtie2 (v2.3.5.1) (http://bowtie-bio.sourceforge.net/bowtie2). PCR duplicated fragments were filtered by Picard (http://broadinstitute.github.io/picard). Then, filtered reads were mapped. Peaks were identified by Homer (v4.10.4) (http://homer.ucsd.edu/homer). FRiP (Fragments Ratio in Peaks) value was calculated using bedtools (v2.29.2) (https://bedtools.readthedocs.io/en/latest). We used deepTools (v3.5.0) to generate bigWig file with RPKM normalization. Enriched peaks region was used as input to DESeq2(v1.30.0) to find differential peaks from ChIP-Seq data as well as normalized the data. Peaks were annotated by using R package ChIPSeeker(v1.26.0) (http://bioconductor.org/packages/devel/bioc/vignettes/ChIPseeker/inst/doc/ChIPseeker.html). For mapping peaks to gene features, we identified distribution of the peaks of each ChIP-Seq data across the genome. Promoter was defined as a region within ±3 kb from the TSS, and peaks without being mapped to promoter, upstream, intron, or exon were considered as intergenic target loci. Profiles were obtained on a region of ±3 kb from the center of peaks, and average scores were plotted to generate averaged read density around peaks using ngsplot (v2.41) (https://code.google.com/p/ngsplot/).

**Identification of super-enhancers**

Super-enhancers (SEs) were identified using the rank ordering of super-enhancers (ROSE) algorithm (http://younglab.wi.mit.edu/super_enhancer_code.html). H3K27ac peaks were used to define enhancers, followed by further filtering based on the criteria: briefly, peaks located within ±3kb region of TSSs were excluded. The remaining H3K27ac peaks were defined as putative enhancers. Enhancers located within ±12.5 kb regions of each other were stitched together, scored, and ranked based on H3K27ac ChIP-Seq signals. Enhancers were plotted with enhancer rank versus enhancer density, and all enhancer regions above the inflection point of the curve were defined as SEs. An analogous procedure was used to define
SE regions by enrichment of Satb1 at enhancers. Super-enhancers and typical enhancers were assigned to the genes using the default parameters of the ROSE algorithm. ChIP-seq signal at SE regions were plotted as averaged profiles in ngsplot (v2.41). GSEA was used to identify how SE-gene sets distribute in gene lists ranked by either gene expression fold change values or H3K27ac ChIP-seq enrichment on enhancers. GO analysis for SE was performed as previous described.

Hi-C

We performed *in situ* Hi-C with 5-10 million cells. Cells were crosslinked at a final concentration of 1% formaldehyde and in ice bath for 10 min and quenched by 200 mM glycine for 5 min at room temperature. After two washes with cold 1×PBS, cells were pelleted and kept at -80°C until use. Subsequently, crosslinked cells were lysed, and chromatin was digested with 150 U Mbol overnight at 37°C. Mbol was inactivated at 62°C for 20 min, DNA fragment ends were biotinylated with biotin-dCTP using Klenow large fragment for 60-90 min at 37°C, and the sample was diluted, and proximity ligated for four hours at room temperature. Crosslinked DNA was reversed by addition of SDS, proteinase K, and NaCl, and allowed to incubate overnight at 68°C. DNA was purified by phenol/chloroform, followed by ethanol precipitation, and resuspended in 100ul nuclease-free dH₂O. DNA was treated with T4 DNA polymerase to remove unligated biotinylated ends and sheared to 300-500bp by sonication subsequently, the ligated junctions were pulled down with Dynabeads MyOne Streptavidin C1 magnetic beads. End-repair, dA-tailing, adapter-ligation, and PCR amplification were performed on biotinylated DNA fragments bound to beads. After purification, libraries were sequenced on an Illumina HiSeq 2000 platform to obtain 150 bp paired-end reads.

For each sample, reads were obtained following quality filtering and adaptor trimming using fastp (v0.20.0) (https://github.com/OpenGene/fastp). Hi-C mapping, filtering, correction, and binning were performed with HiC-Pro (v2.11.1) (https://github.com/nservant/HiC-Pro). The paired-end reads were mapped to the mouse mm10 genome. Self-circle ligation, dangling ends, re-ligation, and the other dumped types were filtered out with HiC-Pro after mapping. We generated raw contact matrices at 10-kb, 20-kb, 50-kb, 100-kb, 500-kb, 1-Mb resolutions. For raw contact matrix correction, we used the iterative correction method (ICE). The hicpro2juicebox.sh utility was used to convert the allValidPairs output of the pipeline into Juicebox hic format at fragment resolution. Visualization of Hi-C contact matrices was done with Juicerbox. Differential analysis and visualization of local interactions from Hi-C data were obtained using HiTC (v1.34.0) R package (http://bioconductor.org/packages/devel/bioc/html/HiTC.html).

A/B compartment analysis was performed at 250-kb resolution using a publicly available script (matrix2compartment.pl). The script can be accessed through GitHub (https://github.com/dekkerlab/cworld-dekker). And then, we used the first principal component (PC1) to predict regions of active (A compartments) and inactive chromatin (B compartments). We generated 250-kb tracks and correlated A/B compartments with H3K27ac and H3K27me3 ChIP-Seq data for each cell type using the H3K27ac mark as an indicator of A compartments and the H3K27me3 mark as an indicator of B compartments. We identified the regions with changes (A-A, B-B, and A-B/B-A) in sign of the PC1 value between Satb1WT and Satb1cKO DP cells as A/B compartment switched regions.
The TAD structure (insulation/boundaries) was defined by the insulation score. The matrices which were used to calculate the insulation score were normalized by ICE method for discarding the bias of raw matrices. And then, insulation score was computed at 10-kb resolution. A publicly available script (matrix2insulation.pl) was used to detect TAD boundaries, with the following options: ‘--is 600000 --ids 2500000 --im mean --bmoe 3’. TADs were called using normalized Hi-C matrices at 10-kb resolution with insulation2tads.pl. The script can be accessed through GitHub (https://github.com/dekkerlab/cworld-dekker).

We generated a chromatin interaction matrix with a 50 kb resolution using the Hi-C data, and the differential chromatin interactions between Satb1WT and Satb1CKO DP cells were calculated using the Bioconductor package multHiCcompare (v1.8.0), which provides functions for the joint normalization and detection of differential chromatin interactions in our two Satb1WT and two Satb1CKO replicates experiments. To identify the significantly differential chromatin interactions, we used an adjusted p-value of 0.05 and a log fold change of 0. The chromosomal sequencing was divided into 50 kb bins for the normalized signal file of the Satb1WT ChIP-Seq. Once the bin attained by ChIP-Seq were located within the differential chromatin interactions, and the value in the corresponding bin of the Hi-C matrix indicated the normalized count.

Chromatin loops were called using Fit-Hi-C (v2.0.7). Input files of Fit-Hi-C were created by using a publicly available script (hicpro2fithic.py) from HiC-Pro. Next, loops were called using fixed-size bin resolutions from 10 to 20 kb in both cell types. Briefly, significant interaction loops (p <= 0.05 and contact frequencies >= 5) were identified through jointly modeling the contact probability using raw contact frequencies and ICE normalization vectors with the Fit-Hi-C algorithm. Enhancer-promoter loops were annotated using pgltools (v2.2.0). The anchors of loops were intersected with promoters and enhancers. Promoters were defined as ±5 kb windows of the TSS of all expressed genes, enhancers are defined using enhancer dataset of DP cells (http://www.enhanceratlas.org/downloadv2.php). Super-enhancer loops were defined by H3K27ac-defined SE sets obtained through ROSE software.

The format of Hi-C data, .hic, was converted into .cool files using hic2cool (v0.8.3) software (https://github.com/4dn-dcic/hic2cool). Hi-C matrices in cool format were used to generate genome-wide aggregate plots at TADs and loops of Satb1WT and Satb1cKO DP cells detected by Hi-C. We used coolpup.py (v0.9.5) to pile-up normalized Hi-C signals at a 10-kb resolution at loops previously identified and plotted 500-kb upstream and downstream of the loop anchor coordinates (https://github.com/open2c/coolpuppy). The local rescaled pileups of TADs annotated using insulation score valleys used above in DP cells were analyzed at a 10-kb resolution. We plotted them using plotpup.py (v0.9.5).

4C

Chromatin was crosslinked for 10 minutes at room temperature in 1X PBS/10% FBS containing 4% formaldehyde. Crosslinking was blocked by glycine addition. Following cell lysis, nuclei pellets were
resuspended in 1.2X Buffer2 (NEB), followed by SDS addition. The samples were incubated while shaking for 60 minutes at 37°C. TritonX-100 was added to quench SDS. Crosslinked chromatin was primarily digested using MboI. Ligation was performed in the presence of T4 DNA Ligase in diluted conditions. Chromatin reverse crosslinking was carried out using proteinase K, phenol-chloroform and DNA were purified by ethanol precipitation. Secondary restriction was performed using NlaIII. Secondary ligation was carried out with T4 DNA Ligase in diluted conditions. Ligated DNA products were then extracted using phenol-chloroform with ethanol precipitation and purified. 4C-seq library preparation was achieved by inverse PCR using 100 ng template DNA (for 10 reactions in total). The viewpoint-directed inverse PCR primers carrying Illumina P5 or P7 sequencing adapters as below: Ets2 forward primer: 5'-GTGACTGGAGTTCCAGACTGTTGCTCTTCCGATCTgcccgggacagcgtggtc-3'; Ets2 verse primer: 5'-TCTTTCCCTACACGCAGCCTCTTCCGATCTcgagcacttaggcgctcatc-3'.

### 3C-HTGTS

3C-HTGTS libraries were prepared with FACS-sorted DP cells. In brief, 10 million cells were cross-linked with 1% formaldehyde at room temperature for 10 min, quenched with glycine (final concentration 0.125 M) on ice for 5 min. Cell were lysed and followed by the addition 200U MboI to digest the chromatin overnight at 37°C with gentle shaking. MboI was inactivated by adding 10% SDS to a final concentration of 1.5% and incubating at 37°C for 30 min. To reduce the SDS concentration, the solution was diluted with T4 ligase buffer containing 1% Triton X-100, followed by incubation at 37°C for one hour. T4 ligase (New England Biolabs) was added and incubated for overnight at 16°C. Crosslinking were reversed and samples were treated with proteinase K and RNase A prior to DNA extraction with 1:1 phenol-chloroform and precipitation with ethanol. The 3C libraries were sonicated to 300-500bp on a Qsonica Bioruptor Sonicator. Sonicated DNA was linearly amplified with a biotinylated primer (Bcl6 5'-TTACCATTGCTCCGACGCAG-3', Ets2: GGACCTGCAGACAGCCTAAC) that anneals the promoters of Bcl6 or Ets2. The biotin-labeled single-stranded DNA products were enriched with streptavidin C1 beads, and followed by 3' ends ligation with the bridge adaptor. The adaptor-ligated products were amplified through nested PCR using a nested primer and an adaptor-complementary primer (Nest primers: Bcl6: 5'-CTGGAGTTCAGACGTGCTCTTCCGATCTTCCGCCGGCGCAGAATGCCT-3', Ets2: 5'-CTGGAGTTCAGACGTGCTCTTCCGATCCTGACTATAGGGCACGCGTGG-3'). And a final PCR for another 10-15 cycles of amplification with P5 and P7 was performed. After purification, the final libraries were sequenced on an Illumina HiSeq 2000 platform to obtain 150 bp paired-end reads.

3C-HTGTS fastq data were filtered by removing adaptor and low-quality reads with using fastp (v0.20.0). Filtered reads were extracted from the sequence file after quality control through Cutadapt (v1.18), Pear (v0.9.6), etc.. Pair-end reads containing NestPrimer or AdapterPrimer were obtained, and then the reads were filtered by searching restriction sites sequences. The remaining reads were mapping to mouse genome mm10 with bowtie2 (v2.3.5.1). The mapping reads were filtered the duplicated reads, self-ligation reads, relegation reads and dumped reads. For visualization, we convert the final bam files into bedGraph file. The signal peak bedGraph file was obtained by post-comparison filtering, signal statistics, and
standardization. We normalized bedGraph file using CPM (Counts Per Million in cis) normalization method, and visualized on IGV. Finally, we organized the results report and visualized it with R.

**FACS analysis**

Thymus, spleen, mesenteric lymph nodes, inguinal lymph nodes, and auxiliary lymph nodes from 6-8 weeks mice were ground in MACS buffer (1×PBS, 0.5% BSA, 2mM EDTA) and filtered with 40 um nylon mesh. Red blood cells were lysed in RBC buffer (Biolegend, USA) for 10 minutes at room temperature. FACS analysis was performed on the BD FACS Aria3 machine. DP cells were defined as CD4+CD8+. Cells were gated on CD4-CD8- and DN cells were defined as followed: DN1 CD44+CD25-, DN2 CD44+CD25+, DN3 CD44-CD25+, and DN4 CD44-CD25-. The B cells were defined by the expression of CD19 and the T lymphocytes were defined by the CD3 expression. Cells from the spleen, mesenteric lymph nodes, inguinal lymph nodes, and auxiliary lymph nodes were gated with CD3 and the mature T cells were defined as CD4+ or CD8+. For Bcl6 staining, cells were then fixed and permeabilized using True-Nuclear Transcription Factor Buffer Set (Biolegend, USA) after cell surface staining.

**5' RACE sequencing**

The 1ug RNA was used for the reverse transcription (RT). The cDNA libraries were performed with oligo(dT) primer by 5'Rapid amplification of complementary DNA ends (5'RACE). During the reverse transcription reaction, A SMART oligonucleotide was ligated to the 5'end of each cDNA. The TCR library is then prepared by touchdown PCR amplification using a set of primers: a forward primer to the adaptor and a reverse primer to the constant region of the TCR. Touchdown PCR was performed under the conditions of 95°C for 3 min, followed by 15 cycles of 98°C for 15 s, 65°C (decrease 0.7°C/cycles) for 15 s and 72°C for 1 min, 20 cycles of 98°C for 15 s, 60°C for 15 s and 72°C for 1 min and one cycle 10 min at 72°C. PCR products, ranged from 300-700bp were gel purified by DNA Gel Purification Kit. Illumina paired end adapters were ligated to TCR libraries, followed by 2rd round PCR with Illumina P5/P7 adaptor primers under 30s at 95°C, 15s at 65°C, 60s at 72°C for 10-15 cycles, Plus a final extenaion for 10 min at 72°C. Ampure bead purified PCR product. Finally, the TCR libraries were sequenced on an Illumina HiseqXtenPE150 (2 × 150 nt).

Single-cell pellets of thymus were subjected to RNA extraction using phenol and chloroform. 500ng mRNA were then reverse transcribed according to previous reports. Next, Tcra and Tcrb sequencing amplicons were amplified using the primers described below: Forward primer: TGAACCTTAAGCAGTGGTATCAACGCAGAG, Tcra: 5'-CAGGGTCAGGGTTCTGGATA, Tcrb: 5'-TCTGATGGCTCAAACACAGC. The library for Illumina sequencing was prepared using DNA Library Prep Kit for Illumina (NovoNGS, Shanghai, China).

The 5'RACE raw data were filtered by fastp and the adapter sequences, contamination, and low-quality reads were removed. T cell receptor beta chain V, D, and J gene identification, CDR3 sequence extraction in clean reads were performed using MiXCR. The corresponding germline sequences were mapped to the reference sequences derived from international ImMunoGeneTics (IMGT) information.
References

1. Proudhon, C., Hao, B., Raviram, R., Chaumeil, J. & Skok, J.A. Long-Range Regulation of V(D)J Recombination. *Adv Immunol* **128**, 123–182 (2015).

2. Dutta, A., Zhao, B. & Love, P.E. New insights into TCR beta-selection. *Trends Immunol* (2021).

3. Shih, H.Y., Hao, B. & Krangel, M.S. Orchestrating T-cell receptor alpha gene assembly through changes in chromatin structure and organization. *Immunol Res* **49**, 192–201 (2011).

4. Marx, A. *et al.* Thymus and autoimmunity. *Semin Immunopathol* **43**, 45–64 (2021).

5. Carico, Z. & Krangel, M.S. Chromatin Dynamics and the Development of the TCRalpha and TCRdelta Repertoires. *Adv Immunol* **128**, 307–361 (2015).

6. Hao, B. *et al.* An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J Exp Med* **212**, 809–824 (2015).

7. Naik, A.K., Byrd, A.T., Lucander, A.C.K. & Krangel, M.S. Hierarchical assembly and disassembly of a transcriptionally active RAG locus in CD4(+)CD8(+) thymocytes. *J Exp Med* **216**, 231–243 (2019).

8. Wang, R. *et al.* Transcription factor network regulating CD(+)CD8(+) thymocyte survival. *Crit Rev Immunol* **31**, 447–458 (2011).

9. Guo, J. *et al.* Regulation of the TCRalpha repertoire by the survival window of CD4(+)CD8(+) thymocytes. *Nat Immunol* **3**, 469–476 (2002).

10. Chen, L., Foreman, D.P., Sant'Angelo, D.B. & Krangel, M.S. Yin Yang 1 Promotes Thymocyte Survival by Downregulating p53. *J Immunol* **196**, 2572–2582 (2016).

11. Wang, R. *et al.* T cell factor 1 regulates thymocyte survival via a RORgammat-dependent pathway. *J Immunol* **187**, 5964–5973 (2011).

12. Fessing, M.Y. *et al.* p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* **194**, 825–839 (2011).

13. Zhang, Y. *et al.* Fluorosed mouse ameloblasts have increased SATB1 retention and Galphaq activity. *PLoS One* **9**, e103994 (2014).

14. den Hoed, J. *et al.* Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders associated with SATB1 dysfunction. *Am J Hum Genet* **108**, 346–356 (2021).

15. Gong, J. *et al.* Hepatic SATB1 induces paracrine activation of hepatic stellate cells and is upregulated by HBx. *Sci Rep* **6**, 37717 (2016).

16. Goolam, M. & Zernicka-Goetz, M. The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. *Development* **144**, 1450–1461 (2017).

17. Li, Z. *et al.* N(6)-methyladenine in DNA antagonizes SATB1 in early development. *Nature* **583**, 625–630 (2020).

18. Wang, F. *et al.* Positive feedback between RNA-binding protein HuD and transcription factor SATB1 promotes neurogenesis. *Proc Natl Acad Sci U S A* **112**, E4995-5004 (2015).
19. Will, B. et al. Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nat Immunol* **14**, 437–445 (2013).

20. Satoh, Y. et al. The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. *Immunity* **38**, 1105–1115 (2013).

21. Wen, J. et al. SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. *Blood* **105**, 3330–3339 (2005).

22. Gong, H. et al. SATB1 regulates beta-like globin genes through matrix related nuclear relocation of the cluster. *Biochem Biophys Res Commun* **383**, 11–15 (2009).

23. Alvarez, J.D. et al. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev* **14**, 521–535 (2000).

24. Cai, S., Lee, C.C. & Kohwi-Shigematsu, T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* **38**, 1278–1288 (2006).

25. Kakugawa, K. et al. Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. *Cell Rep* **19**, 1176–1188 (2017).

26. Feng, D., Li, Z., Qin, L. & Hao, B. The role of chromatin organizer Satb1 in shaping TCR repertoire in adult thymus. *Genome*, 1–12 (2021).

27. Kondo, M. et al. SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J Immunol* **196**, 563–572 (2016).

28. Kitagawa, Y. et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat Immunol* **18**, 173–183 (2017).

29. Zelenka, T. & Spilianakis, C. SATB1-mediated chromatin landscape in T cells. *Nucleus* **11**, 117–131 (2020).

30. Dickinson, L.A., Joh, T., Kohwi, Y. & Kohwi-Shigematsu, T. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**, 631–645 (1992).

31. Cai, S., Han, H.J. & Kohwi-Shigematsu, T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* **34**, 42–51 (2003).

32. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

33. Jain, D., Baldi, S., Zabel, A., Straub, T. & Becker, P.B. Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Res* **43**, 6959–6968 (2015).

34. Stansfield, J.C., Cresswell, K.G. & Dozsmorov, M.G. multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics* **35**, 2916–2923 (2019).

35. Ay, F., Bailey, T.L. & Noble, W.S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res* **24**, 999–1011 (2014).

36. Gioulbasani, M. et al. The transcription factor BCL-6 controls early development of innate-like T cells. *Nat Immunol* **21**, 1058–+ (2020).
37. Ye, B.H. et al. Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma. *Science* **262**, 747–750 (1993).

38. Kerckaert, J.P. et al. LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. *Nat Genet* **5**, 66–70 (1993).

39. Yang, H. & Green, M.R. Epigenetic Programming of B-Cell Lymphoma by BCL6 and Its Genetic Deregulation. *Front Cell Dev Biol* **7**, 272 (2019).

40. Nurieva, R.I. et al. Bcl6 mediates the development of T follicular helper cells. *Science* **325**, 1001–1005 (2009).

41. Gallant, S. & Gilkeson, G. ETS transcription factors and regulation of immunity. *Arch Immunol Ther Exp (Warsz)* **54**, 149–163 (2006).

42. Anderson, M.K., Hernandez-Hoyos, G., Diamond, R.A. & Rothenberg, E.V. Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage. *Development* **126**, 3131–3148 (1999).

43. Zaldumbide, A., Carlotti, F., Pognonec, P. & Boulukos, K.E. The role of the Ets2 transcription factor in the proliferation, maturation, and survival of mouse thymocytes. *J Immunol* **169**, 4873–4881 (2002).

44. Fisher, I.B., Ostrowski, M. & Muthusamy, N. Role for Ets-2(Thr-72) transcription factor in stage-specific thymocyte development and survival. *J Biol Chem* **287**, 5199–5210 (2012).

45. Liu, M. et al. Role of Ets Proteins in Development, Differentiation, and Function of T-Cell Subsets. *Med Res Rev* **36**, 193–220 (2016).

46. Solanki, A. et al. The transcriptional repressor Bcl6 promotes pre-TCR-induced thymocyte differentiation and attenuates Notch1 activation. *Development* **147** (2020).

47. Zhao, H. et al. A role of the CTCF binding site at enhancer Ealpha in the dynamic chromatin organization of the Tcra-Tcrd locus. *Nucleic Acids Res* **48**, 9621–9636 (2020).

48. Bolotin, D.A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380–381 (2015).

49. Ioannidis, V., Beermann, F., Clevers, H. & Held, W. The beta-catenin–TCF-1 pathway ensures CD4(+)CD8(+) thymocyte survival. *Nat Immunol* **2**, 691–697 (2001).

50. Wang, F. et al. Exploring the stage-specific roles of Tcf-1 in T cell development and malignancy at single-cell resolution. *Cell Mol Immunol* **18**, 644–659 (2021).

51. Jones, M.E. & Zhuang, Y. Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. *Immunity* **27**, 860–870 (2007).

52. Yuan, J., Crittenden, R.B. & Bender, T.P. c-Myb promotes the survival of CD4+CD8+ double-positive thymocytes through upregulation of Bcl-xL. *J Immunol* **184**, 2793–2804 (2010).

53. Hu, T., Simmons, A., Yuan, J., Bender, T.P. & Alberola-Ila, J. The transcription factor c-Myb primes CD4+CD8+ immature thymocytes for selection into the iNKT lineage. *Nat Immunol* **11**, 435–441 (2010).
54. Heinz, S., Romanoski, C.E., Benner, C. & Glass, C.K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144–154 (2015).

**Declarations**

**Data availability**

The raw sequence data of single-cell RNA-Seq, bulk RNA-seq, ChIP-seq, ATAC-seq, Hi-C, 4C-seq, 3C-HTGTS, and 5’RACE reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database under the accession number: GSE182995.

GEO accession codes (or SRA accession number) of the published data used in this study are as follows: H3K4me3 ChIP-seq of CD4+ CD8+ DP thymocytes, GSE21207; Rad21 and Nipbl ChIP-seq, GSE48763; CTCF ChIP-seq, GSE141223; Satb1 ChIP-seq, GSE90635; RNA-seq of T cell development, GSE109125; H3K27ac ChIP-seq of Satb1WT and Satb1cKO DP thymocytes, DRP003376; H3K4me1 ChIP-seq of Satb1WT and Satb1cKO DP thymocytes, DRP003376; The remaining data are available within the article, supplementary information or available from the authors upon request. Source data are provided with this paper.

**Acknowledgements**

We thank Michael Krangel from Department of Immunology Duke University for Satb1 cKO mice, Hongfen Shen and Qianbing Zhang for Flow cytometry, Zhongxi Huang and Jiahong Wang for research computing. This work was supported by the National Natural Science Foundation of China (31970836 and 32170885 to Bingtao Hao, 81801549 to Litao Qin) and the National Institutes of Health grant (NIH R01 ES023854 to Terumi Kohwi-Shigematsu).

**Contributions**

D.F. designed, performed, and analyzed the following experiments: scRNA-seq, bulk RNA-seq, Hi-C, 4C-seq, 3C-HTGTS. Y.C. performed FACS and 5’RACE of Bcl6-SE mice. R.D. performed bioinformatics analysis of ChIP-seq, Hi-C, 3C-HTGTS, and 5’RACE. S.B. designed the study, performed FACS and 5’RACE of Ets2-SE mice. W.X. performed FACS, ChIP-seq, 3C-HTGTS library preparation. Y.Zhu performed 3C-HTGTS library preparation. Z.L. performed 4C-seq, ATAC-seq and ChIP-seq analysis. J.B. performed 5’RACE library preparation. Y.Zhang provided technical support of Hi-C library preparation. Y.Y. provided technical support, reagents, and conceptual advice. J.Z. assisted with bioinformatical analysis. L.Q. provided technical support, reagents, and conceptual advice. Y.K. did control ChIPseq experiment with SATB1 in KO thymocytes. S.W. provided technical support and conceptual advice. T.K-S. provided Satb1 conditional knockout mouse and did critical discussion toward manuscript preparation. S.L. supervised the project. B.H. wrote the manuscript with helps from D.L., Y.C., R.D., S.B., X.W. and T.K. All authors read and approved the final version of the manuscript.
Satb1 controls cell identity of DP thymocytes by repressing early expressing genes and activating DP specific genes. a) Two-dimensional UMAP plot of thymocytes single-cell transcriptomes in vav-cre x Satb1fl/fl (Satb1cKO) and Satb1fl/fl (Satb1WT) mice. Each dot represents one cell and different areas

Figure 1
identified by unsupervised clustering. b) Single cell RNA-seq of Rag1 gene expression in difference cell subsets from Satb1cKO and Satb1WT mice. c) Volcano plot showing bulk RNA-seq differentially expressed genes (Satb1cKO vs Satb1WT) in DP thymocytes. The x- and y-axis represent log2(fold change) and -log10(p-value), respectively. Some genes are annotated. d) Heat map and e) boxplot showing relative gene expression of the differentially expressed genes during the differentiation from DN1 to SP thymocytes. f) GSEA of the differentially expressed genes (Satb1cKO vs Satb1WT) enriched for DN1- or DN3-specific genes.
Figure 2

SATB1 binds and activates super-enhancers of DP thymocytes a) Hockey stick plots based on H3K27ac signals in Satb1WT DP thymocytes using the ROSE algorithm. Some super-enhancer-associated genes are highlighted. b) Heat map (left) and boxplot (right) showing relative gene expression of super-enhancer-associated genes during the differentiation from DN1 to SP thymocytes. c) Normalized H3K27ac, H3K4me3, CTCF, Rad21, NIPBL and SATB1 ChIP-seq profiles at the Cd4 locus in WT DP cells and the SATB1 ChIP-seq in SATB1 deficient thymocytes (Satb1f/f x CD4-cre). d) Line plot displaying H3K27ac, H3K4me1, and Satb1 ChIP-seq signal density at super-enhancer regions. e) Line plot displaying H3K27ac signal density in Satb1WT vs Satb1cKO at DP super-enhancer regions. f) GSEA of the differentially expressed genes enriched for the super-enhancer associated geneset or lost super-enhancer associated geneset. g) The relative expression of super-enhancer-associated genes in Satb1WT and Satb1cKO DP. Three biological replicates of each sample were analyzed. h) Relative expression of genes associated with super-enhancers gained, maintained, or lost in Satb1 KO DP. * p value < 0.05, *** p value < 0.001 by Kruskal–Wallis test, followed by Dunn’s multiple comparisons test. o) Relative expression of the super-enhancer (SE) or traditional enhancer (TE) associated genes in Satb1cKO DP cells. **** p value < 0.0001 by two side Student's t test.
Figure 3

SATB1 clusters are associated with cell identity genes. a) Clusters ranked by Satb1 ChIP-seq signals in DP thymocytes using the ROSE algorithm. b) GSEA of the ranked gene list from H3K27ac signal or c) the differentially expressed genes enriched for the top 500 genes associated with Satb1 clusters. d) Heat map (up) and boxplot (bottom) showing relative gene expressions of Satb1-cluster-associated genes during the differentiation from DN1 to SP thymocytes. **** p value < 0.0001 by Kruskal–Wallis test,
followed by Dunn’s multiple comparisons test. e) Gene ontology analysis on the genes associated with super-enhancers by enrichment for H3K27ac or Satb1. The top seven biological processes ranked by significance (−log(q-value)) are depicted.

Figure 4

SATB1 mediated the chromatin interactions of super-enhancers a) Boxplot showing the distance between two anchors of significantly changed interactions (50kb bin) in Satb1 deficient DP thymocytes. The gain
contains significant increased interactions, the loss significant decreased interactions, the maintain no significant changed interactions. The interactions with 0 bp distance were deleted. The Wilcoxon signed-rank test was used. b) Boxplot of ChIP-seq signals in anchors of differential chromatin interactions (50kb bin). The Wilcoxon signed-rank test was used. c) Aggregate analysis and d) Boxplot of loop sizes of all loops, enhancer-promoter (E-P) loops, and super-enhancer (SE) loops identified from Hi-C data of Satb1WT and Satb1cKO thymocytes. Loops are 10 kb resolution. **** p value < 0.0001 by two side Student's t test. e) Pearson's correlation analysis of super-enhancer associated H3K27ac signals (left) or relative gene expression (right) with loop numbers in super-enhancers. f) Aggregate analysis of loops in SATB1 ChIP-seq clusters.
SATB1 regulates transcription factors Bcl6 and Ets2 by mediating promoter-enhancer interactions. a) and b) In situ Hi-C chromatin interaction heatmaps (up) of Satb1WT and Satb1cKO DP thymocytes for the Bcl6 (a) and Ets2 locus (b). H3K27ac, H3K4me1, H3K4me3, Satb1, Rad21, Nipbl, and CTCF ChIP-seq tracks (bottom) of Satb1WT cells for each locus. c) and d) ChIP-seq and 3C-HTGTS at the Bcl6 and Ets2 loci. The 3C-HTGTS bait is at the Bcl6 or Ets2 promoters. e) Relative expression of Bcl6 and Est2 in...
Satb1WT and Satb1cKO DP thymocytes detected by reverse-transcripted quantitative PCR. The expressions are normalized to the Actb gene and then to WT. The data represent mean ± SD of three experiments. * p value < 0.05, ** p value < 0.01 by two side Student’s t test. f) The expression profiles of Satb1, Bcl6, and Ets2 during the differentiation of DN1 into DP thymocytes.
Super-enhancer regulates Ets2 expression in the thymus

a) Relative Ets2 expression in thymocytes from Ets2 super-enhancer knockout (Ets2 SE-/-) and wild-type (WT) mice was analyzed using reverse-transcription qPCR. The expressions are normalized to the Actb gene and then to WT. The data represent the mean ± SD of three experiments. *** p < 0.001 by two side Student’s t test.

b) 4C assay from the viewpoint of the Ets2 promoter in thymocytes from WT or Ets2 SE-/- mice. The data are presented as reads per million mapped reads (RPM).

c) Comparison of wild-type and Ets2-SE knockout total thymocytes. Results are the mean ± SD of seven WT and KO mice. ** p < 0.01 by two side Student’s t test.

d) Flow cytometry of thymocytes from WT and Ets2 SE-/- mice. Right, average frequency of thymocyte subsets. Results are the mean ± SD of three WT and KO mice. * p < 0.05, *** p < 0.001 by two side Student’s t test.

e) Thymocytes were cultured from 0 to 40 hours in RPMI 1640 medium with 10% FBS at 5% CO2 and 37 °C. MTS assay was performed to detect the cell viabilities. Each experiment was repeated three times and the data are the mean ± SD of three WT and KO mice. *** p < 0.001 by two-way ANOVA.
Figure 7

Super-enhancer regulates Bcl6 expression in the thymus a) Relative Bcl6 expression and 3C-HTGTS analysis of thymocytes from Bcl6 super-enhancer knockout (Bcl6 SE-/-) and wild-type (WT) mice. The expressions are normalized to the Actb gene and then to WT. The data represent the mean ± SD of three experiments. *** p < 0.001 by two side Student’s t test. b) Comparison of WT, Bcl6-SE KO heterozygous, and homozygous total thymocytes. Results are the mean ± SD of five mice for each genotype. ** p < 0.01
by two side Student's t test. c) Flow cytometry of thymocytes from WT and Bcl6 SE/- mice. Right, average frequency of thymocyte subsets. Results are the mean ± SD of three WT and KO mice. *** p < 0.001 by two side Student's t test. d) Thymocytes were cultured from 0 to 48 hours in RPMI 1640 medium with 10% FBS at 5% CO2 and 37 °C. MTS assay was performed to detect the cell viabilities. Each experiment was repeated three times and the data are the mean ± SD of three WT and KO mice. e) Relative Va and f) Jα usages were determined by deep-sequencing of Tcra transcripts amplified by 5’RACE of WT and Bcl6 SE/- thymocytes, respectively. The relative Va or Jα usages were calculated by dividing the number of the clonotypes containing the Va or Jα genes by the total clonotype number. The Va usage is the sum of the usage frequency of all Va genes in the region.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1DEGinSatb1cKODPxlsx
- TableS2H3K27acsuperenhancer.xlsx
- TableS3Satb1cluster.xlsx
- 2021FengDLSupFig.docx