Uncovering a novel interaction between SATB1- genome organizer and SET1A- histone H3K4 methyltransferase

Konstantinos Klaourakis

Senior Undergraduate Thesis

Department of Biology, University of Crete,

Foundation for Research & Technology- Hellas, Institute of Molecular Biology & Biotechnology

July 2016

Principal Investigator: Dr. Charalampos G. Spilianakis

Ι. Περίληψη

Στο εργαστηριό μας ερευνάμε την επίδραση της οργάνωσης της χρωματίνης και των επιγενετικών τροποποιήσεων στην ανάπτυξη κυτταρικών γενεαλογιών του ανοσοποιητικού συστήματος του ποντικού. Μια από τις σημαντικότερες δομές για τη χρωμοσωματική αρχιτεκτονική είναι το πυρηνικό ικρίωμα, πάνω στο οποίο προσδένονται περιοχές του DNA γνωστές ως matrix attachment regions (MARs) σχηματίζοντας βρόχους με τη βοήθεια πρωτεϊνών. Η πρωτεΐνη special AT-rich sequence-binding protein 1 (SATB1), η οποία εκφράζεται κυρίως στα Τ λεμφοκύτταρα, έχει σημαντική δράση στην οργάνωση της χρωματίνης, καθώς και στη ρύθμιση της μεταγραφής γονιδίων. Η SATB1 προσδένεται σε MARs όπου η αλληλουχία της μιας αλυσίδα έχει μόνο A, T, C νουκλεοτίδια και είναι επιρρεπής στο ξετύλιγμα της διπλής έλικας κάτω από υπερελίκωση. Επιπλέον, μεταλλαγές στη SATB1 σχετίζονται με αυτοανοσίες και καρκίνους. Από την άλλη μεριά, οι τροποποιήσεις των ιστονών αποτελούν ένα σημαντικό μηχανισμό ρύθμισης της μεταγραφής. Οι μόνο-, δι- και τρι-μεθυλιώσεις στη λυσίνης 4 της ιστόνης 3 (H3K4me1/2/3) έχουν συσχετιστεί με ενεργά εκφραζόμενα γονίδια. Τόσο στα ποντίκια, όσο και στον άνθρωπο τις τροποποιήσεις των ιστονών H3K4me1/2/3 πραγματοποιούν οι πρωτεΐνες SET1A/B και MLL1/2/3/4 μέσω του συμπλόκου COMPASS και COMPASS-like αντίστοιχα. Στο εργαστήριό μας σε προηγούμενα πειράματα έγινε ανοσοκατακρήμνιση και ανάλυση με φασματομετρία μάζας (IP-MS) σε θυμοκύτταρα ποντικών για να εντοπιστεί το πρωτεΐνωμα της SATB1. Τα αποτελέσματα των πειραμάτων έδειξαν μεταξύ άλλων την αλληλεπίδραση της SATB1 με το σύμπλοκο COMPASS και COMPASS-like. Έτσι διατυπώθηκε η υπόθεση ότι η SATB1 αλληλεπιδρά και οδηγεί το σύμπλοκο COMPASS σε συγκεκριμένες περιοχές του γονιδιώματος, ώστε να μεθυλιώσει την ιστόνη H3K4 και να επάγει τη μεταγραφή συγκεκριμένων γονιδίων. Εδώ ελέγχουμε την αλληλεπίδραση της SATB1 με την SET1A σε θυμοκύτταρα ποντικών, αρχικά μέσω διπλής χρώσης ανοσοφθορισμού, και στη συνέχεια με πειράματα συν-ανοσοκατακρήμνισης μαζί με στύπωμα κατά Western (co-IP-WB). Επιπλέον, για να ελέγξουμε τις αλλαγές στο πρότυπο μεθυλίωσης της H3K4, παρουσία και απουσία της SATB1, γίνανε πειράματα ανοσοκατακρήμνισης χρωματίνης μαζί με αλληλούχιση (ChIP-Seq) για την τροποποίηση της ιστόνης H3K4me2 σε C57BL/6 και CD4.Cre-*Satb1*$^{fl/fl}$ θυμοκύτταρα. Τέλος δημιούργησα μια διαδικτυακή

αυτοματοποιημένη ροή εργασιών που μπορεί να αναλύει ακατέργαστα ChIP-Seq δεδομένα.

I. Abstract

In our lab we study the effects of chromatin dynamics and epigenetic modifications in the development of immune-related cells. Nuclear matrix is highly involved in the chromatin architecture as DNA loci, known as matrix attachment regions (MARs), bind to it with the help of various proteins and form loops. The nuclear protein special AT-rich sequence-binding protein 1 (SATB1) is highly expressed in T lymphocytes and has an important function in the formation of chromatin loops and genome organization, as well as in the regulation of gene expression. SATB1 has been shown to bind MARs with a specific sequence in which the one strand consists of mixed A, T, C nucleotides and have a strong potential for unpairing when subjected to superhelical strain. Moreover, mutations of SATB1 lead to autoimmunity and various cancers. On the other hand histone modifications are one of the most important mechanisms of transcription regulation. Among the histone modifications mono-, di- and tri-methylation of lysine 4 of histone 3 (H3K4me1/2/3) have been connected with positive regulation of gene expression. In both mice and humans the histone modifications H3K4me1/2/3 are mediated by the proteins SET1A/B and MLL1/2/3 which are the catalytic domains of the COMPASS and COMPASS-like complexes respectively. Previous work in our lab has utilized immunoprecipitation coupled to mass spectrometry analysis (IP-MS) in murine thymocytes to identify SATB1's protein interactome. These experiments indicated, among other interactions, an interaction between SATB1 and COMPASS, as well as COMPASS-like complexes. Therefore, it was hypothesized that SATB1 interacts and guides the COMPASS complex to specific loci in order to methylate histone H3K4 and induce the expression of specific genes. This work tested the interaction of SATB1 with the histone methyltransferase SET1A in murine thymocytes, first by double immunofluorescence-sequential staining and then by co-immunoprecipitation coupled to Western blot (co-IP-WB). Moreover in order to identify changes of the histone methylation pattern in the absence of SATB1 we performed chromatin immunoprecipitation coupled with next generation sequencing (ChIP-Seq) experiments for the histone modification H3K4me2 in C57BL/6 and CD4.Cre-

Satb1$^{fl/fl}$ thymocytes. Finally I created an online automated workflow in order to analyze raw ChIP-Seq data.

Table of Contents:

II. Acknowledgements

First, I would like to sincerely thank the Assistant Professor Charalampos Spilianakis for all his guidance these four years. He trusted me without any obvious reason, other than my high ambitions, from the first year of my undergraduate studies and welcomed me that summer in his lab where I gained my first wet lab experience. Since then, in my four-year academic journey he has been encouraging and motivating me in both failures and successes. He taught me that failure is part of the learning process as "*each difficulty is for good, to make us better*" and after all with his assistance I was awarded the Amgen Scholarship. Looking back now I realize that without his continued support I would not be where I am today.

Secondly, I would like to thank the members of the lab Ms. Manuela Kapsetaki and Anta Salataj for their assistance in my experiments, as well as the rest of the lab members: Ms. Despina Tsoukatou, Chryssa Stathopoulou, Petros Tzerpos, Theodor Savvidis, Tomas Zelenka, Iro Skopa and Dimitris Metaxas-Mariatos for their academic advices and most importantly I have to thank them for making every day in the lab enjoyable, even when the experiments were not going as planned.

Finally, I would like to thank my family for supporting me and my dreams. They have always been there when I needed them and I am forever grateful for everything they have done for me.

III. Introduction

During interphase, DNA is non-randomly organized in a cell-type specific manner[1] in a way that enables each chromosome to occupy a distinct nuclear territory[2–5]. Within these territories chromatin is dynamic and as a result many interchromosomal contacts and intrachromosomal loops are formed[6]. Some of the major proteins that have been identified as genome architectural proteins are the CCCTC-binding factor (CTCF)[6,7], the Polycomb repressive complex 1 (PRC)[8–10], nuclear lamina proteins[11,12], and the special AT-rich sequence binding protein 1 (SATB1)[13–16]. Although most of these chromosomal contacts and loops may not have any biological significance and may be the result of sharing a common nuclear compartment, some of them have been linked to functional roles[17–19]. Numerous data suggest that the nuclear organization and the genome dynamics have a strong effect in normal cell development and differentiation[20,21], while in various diseases, such as in cancer, the chromatin destabilizes and the genome undergoes major conformational changes[22–25,16]. An example involves the differentiation of mature T lymphocytes ($CD4^+$ or $CD8^+$) to various subpopulations, such as from $CD4^+$ to Th1, Th2 or Th17[26]. The $CD4^+$ lineage commitment has been shown to be controlled by the cytokines that are being expressed at a given time, which are regulated by specific transcription factors, such as STAT6 and GATA3 that regulate the expression of cytokine genes *Il4*, *Il5* and *Il13* in Th2 cell lineage commitment[26–28]. Importantly the transcription regulation of these cytokines is controlled by inter- and intra-chromosomal interactions, as well as by epigenetic modifications[27–31].

The epigenetic modifications of chromatin have a major contribution to the regulation of gene expression, as well as to other important biological functions[32,33]. Furthermore, genome organization and epigenetic modifications are considered to be interdependent, although it has not been yet clearly shown whether the epigenetic modifications cause genome reorganization or/and vice versa[9,34,35]. One of the most studied epigenetic modifications is the methylation of histone H3 lysine 4 (H3K4me). The lysine 4 of histone 3 can be mono-, di- and tri-methylated (H3K4me1, H3K4me2 and H3K4me3 respectively) and all three H3K4me have been correlated with active transcription of genes. Finally, although they can be located globally in the genome, they are primarily identified in transcription start sites (TSS) of actively transcribed

genes, inside the coding region of active genes and in active or poised promoters or enhancers[32,33].

A. SET1A

SET1A is the catalytic histone methyltransferase protein of the COMPASS complex, a complex that specifically mono-, di- and tri-methylate's histone H3K4. COMPASS complex is composed of eight proteins, namely SET1A or SET1B, WDR5, RBBP5, ASH2L, CXXC1, HCFC1, WDR82 and DPY30. From them, the three are unique for COMPASS complex (SET1A or SET1B, CXXC1 and WDR82), while the other five can also be found in COMPASS-like complexes that have as a catalytic histone methyltransferase the MLL1, MLL2, MLL3 or MLL4 protein[36,37]. While MLL/COMPASS-like complexes have been associated with H3K4 methylation in a more gene-specific fashion, such as positive regulation of HOX expression[38] or activation of nuclear hormone receptors, SET1/COMPASS complex is more likely to control the majority of the global genome H3K4 methylation[39]. Moreover, studies have highlighted the importance of the multisubunit COMPASS complex in the stability and activity of both SET1A and SET1B histone methyltransferases. More specifically, the SET1 family is functional and stable only in the presence of the four highly conserved components, namely, WDR5, RBBP5, ASH2L and DPY30 (WRAD)[37,40,41]. On the other hand proteins of the complex such as CXXC1, HCFC1 and WDR82 are mostly needed for specific functions, such as the cross-talk between histone H2B ubiquitination and H3K4 methylation[40,42,43]. Overall, the COMPASS complex is a sophisticated network of proteins that functions in a way that accurately mono-, di- and tri- methylate histone H3K4 at specific loci and during defined developmental stages.

Histone H3K4me3 has been linked with the promoters and the TSS of transcriptionally active genes[44–46]. However, it has also been shown that H3K4me3 is dependent on monoubiquitylated histone H2B, specifically H2BK123ub1 in yeast[43,44] and H2BK120ub1 in human[47]. This cross-talk between the two histone modifications is probably due to the interaction of COMPASS complex with the H2Bub machinery-RAD6 and BRE1 proteins[42,43]. Moreover, because the depletion of H3K4me3 causes minimal transcriptional reduction[45], also the recruitment of H2Bub machinery takes

place after the recruitment of RNA Pol II[40] and the indications that COMPASS complex is recruited by the initiation Pol II machinery (Pol II CTD S5p)[33,48], it has been speculated that H3K4me3 does not directly promote activation of transcription[45]. As a result the question of the functional role of H3K4me3 has not yet been clarified, with the general idea being that it regulates downstream processes of transcription[45,47] and/or represents an epigenetic memory transcriptional activity of coding regions[48].

On the other hand, histone H3K4me2 appears to exist in higher levels in gene coding regions of transcriptionally active genes, rather than in promoters[43,49]. However, H3K4 dimethylation too displays the cross-talk with H2Bub and requires the prior existence of the H2Bub machinery in order for this histone modification to take place[43].

Finally, histone modification H3K4me1 marks predominately active (H3K4me1+, H3K27ac+) and poised (H3K4me1+, H3K27ac−) enhancers, and not promoters or TSS of transcriptionally active genes[50]. An enhancer is characterized as active or poised based on its ability to activate its linked gene(s)[51,52]. As a result active enhancers are linked with actively transcribed genes, while poised enhancers are at an inactive state that can be rapidly activated upon stimuli, such as lineage commitment[51,52]. Another difference between histone H3K4 monomethylation and H3K4 di-, tri-methylation is that its levels are only partially decreased in cells defective for H2B monoubiquitination, indicating that COMPASS complex can also monomethylate histone H3K4 independently of the H2Bub machinery[43].

B. SATB1

SATB1 is a global genome organizer that is highly expressed in thymocytes, while in non-stimulated Th1 and Th2 cells its expression is significantly reduced. SATB1 expression levels are also detectable in mouse embryonic stem cells (mESCs) where it has a functional role in cell differentiation.[53,54]

L.A. Dickinson et al. suggested that SATB1 binds in the minor groove of sequences with significant unwinding potential, high AT content and uneven distribution of guanine residues in the two strands (ATC sequence), while having little contact to DNA bases[55]. Matrix associated regions (MARs) are DNA sequences with high

binding affinity for the nuclear matrix and although they do not have a consensus sequence, they are usually highly AT-rich[56]. Moreover, as some MARs may define the ends of an active chromatin domain, they are often localized at the edges of transcription units[56]. Furthermore, T.K. Shigematsu et al. have proven that when subjected to negative superhelical strain, MARs have a great potential for unwinding[57]. Finally they demonstrated that SATB1 is a thymocyte-specific MAR-binding protein that binds only to double-stranded ATC sequences known as base-unpairing regions (BURs)[55].

T.K. Shigematsu et al. has shown by fluorescence *in situ* hybridization (FISH) analysis in Jurkat lymphoblastic cells that SATB1 binding sequences are located in loci resistant to extensive DNase 1 digestion and salt extraction[58]. These genomic regions are known as DNA "halos" and indicate genomic DNA loops[58]. Furthermore, it was shown by immunofluorescence in thymocytes that SATB1 has a three-dimensional cage-like pattern, which tightly surrounds heterochromatin[13]. As a result, it has been suggested that SATB1 organizes chromatin by shaping its binding sequences into the base of chromatin-loops[13,58].

J.D. Alvarez et al. demonstrated that SATB1 is very important for the regulation of gene expression during T-cell development. As a consequence, major dysregulation occurs to the gene expression profile of thymocytes from *Satb1*-knockout mice, with genes either being ectopically expressed or repressed[59]. Moreover, T-cell development in *Satb1*-knockout mice is blocked in the $CD4^+CD8^+$ double positive stage, while the few escaped peripheral $CD4^+$ T cells undergo apoptosis upon mitogen stimulation[59]. More recently, SATB1 has been shown to have a significant role in both positive and negative selection during T cell maturation, as well as in the establishment of immune tolerance[60]. As a consequence, deletion of SATB1 from all hematopoietic cells is closely linked to autoimmune symptoms[60]. On the other hand, strong evidence have connected the overexpression of SATB1 to poor prognosis of breast cancer[25,16]. The mechanism by which SATB1 has been assumed to advance tumor progression is by binding to BURs near gene loci and promoting activatory or repressive epigenetic modifications at metastasis-associated and tumor-suppressor genes respectively[25,16]. More specifically, T.K. Shigematsu et al. have proven that SATB1 interacts with histone modification enzymes, such as the histone acetylase p300 and the histone deacetylase HDAC1[16,15]. As a result, it is possible that SATB1 -

by forming chromatin-loops, provides to transcription complexes specific loci to attach and to epigenetically regulate the expression of nearby genes[16,15].

In this thesis I tried to answer the question that emerged from preliminary data: Does SATB1 physically interact with the histone methyltransferase COMPASS complex, and guides it to specific loci in order to epigenetically modify the transcription of nearby genes? (Fig.1)

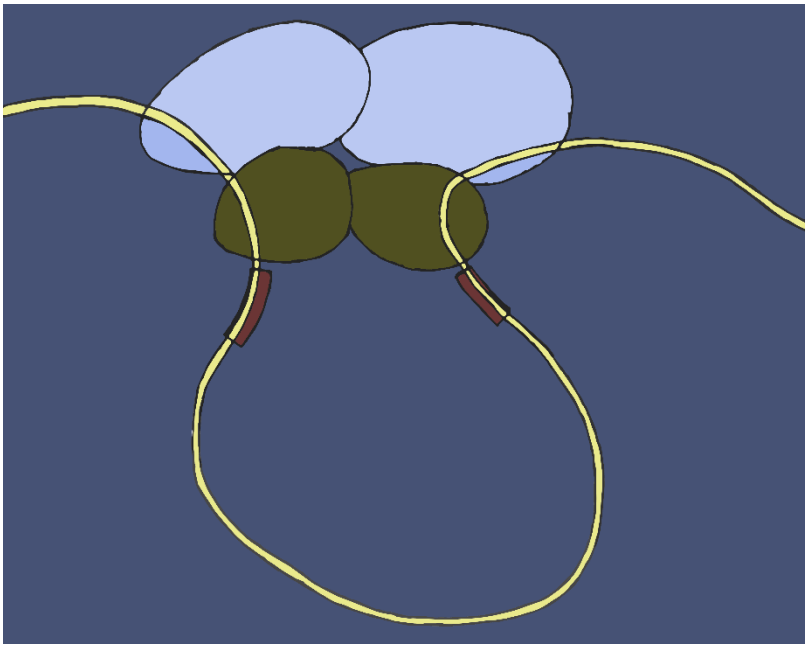


Fig.1 Hypothesis: SATB1 (green) physically interact with the histone methyltransferase COMPASS complex (blue), and guides it to specific loci in order to epigenetically modify the transcription of nearby genes (red).

IV. Materials and Methods

A. Thymocytes extraction

Every step of this protocol was done on ice. Mice 4 weeks old were sacrificed and their thymi were extracted. The thymi were placed in 10mL 1xPBS pH7.1 petri dish and homogenized with a cell strainer. The suspended thymocytes were poured into a nylon mesh filter and collected in a 15mL falcon. Then the thymocytes were centrifuged at 1400rpm for 5min at 4^{o}C and the cell pellet was retained. The cell pellet was resuspended in 10mL 1xPBS and the process was repeated one more time. The total number of thymocytes was counted with a Neubauer chamber.

10xPBS pH7.1: 80g NaCl, 2g KCl, 14.4g Na_2HPO_4, 2.4g KH_2PO_4

B. Double immunofluorescence-sequential staining

Every step of this protocol was done on ice, unless otherwise stated. First, coverslips with thickness No.0 were coated with poly-L-Lysine for exactly 5min. Immediately after the 5min incubation, poly-L-Lysine was decant and the wells were washed fast with 1xPBS. Then coverslips were left to dry completely at 65^{o}C. The thymocytes' extraction protocol was used for one C57BL/6 and one CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout mouse and the cells were divided in order to have 24 million cells in 12mL. 1 million cells were added on top of each coverslip and were let to set for 15min. After the cells were set at the bottom of the well, they were washed with 1xPBS and then fixed with 4%PFA/1xPBS for 10min. Immediately after the 10min, the cells were washed three times with 1xPBS for 5min and then permeabilized with 0.5%Triton/1xPBS for 8min. Immediately after the 8min incubation, the cells were washed three times with 1xPBS for 5min each and then 1xPBS was added. At this point the cells can be stored at 4^{o}C and the next steps do not require ice. Blocking buffer was added on the coverslips for 30min, while keeping them in an incubation chamber. Then the coverslips were washed three times with 1xPBS for 5 min and afterwards the two primary antibodies mixture were added on the coverslips for 1h, while keeping them in the incubation chamber. Again the coverslips were washed three times with 1xPBS for 5 min and then the two secondary antibodies mixture were added on the coverslips for 45min while keeping them in dark in the incubation

chamber. Once more the coverslips were washed two times with 1xPBS for 5 min while keeping them in dark. Finally, the coverslips were left to dry for a few minutes on hard paper and then mounted with a drop of DAPI in mounting media. The coverslips were left overnight at room temperature in dark, before they were scanned at a laser confocal microscope.

Blocking buffer: 7.5% BSA, 1xPBS.

Primary antibodies mixture: blocking buffer, 1:25 anti-SATB1 antibody (sc-5990), 1:200 anti-SET1A antibody (ab70378).

Secondary antibodies mixture: blocking buffer, 1:500 anti-rabbit 546 antibody, 1:500 anti-goat 488 antibody.

C. Co-Immunoprecipitation

Every step of this protocol was done on ice. The thymocytes' extraction protocol was used for two C57BL/6 mice and a total of 200 million thymocytes were extracted. The thymocytes were centrifuged at 1400rpm for 5min at 4^{o}C, the cell pellet was kept and was resuspended in 1mL lysis buffer for 35min at room temperature while gently shaking the tube. Then they were centrifuged at 14.000rpm for 15min at 4^{o}C and the supernatant was kept in a new tube. The total protein concentration was measured by the Bradford assay and was found 5,74μg/μL. 1mg was used for each immunoprecipitation reaction and the rest was kept in aliquots at -80^{o}C. Each of the 1mg protein extracts was diluted in a total volume of 1mL IP buffer and was stored at 4^{o}C until it was used. 50μL of Protein G Sepharose beads were used for 4μg antibody, as a result 100μL of 50% EtOH/50% Protein G Sepharose beads were used for each immunoprecipitation. Before using the Protein G Sepharose beads they were washed three times by centrifuging at 3000g for 2min at 4^{o}C, the supernatant was discarded and the beads were washed with IP buffer. After the third wash, 500μL of IP buffer were added to each of the 50μL Protein G Sepharose beads plus 4μg antibody (10μL SATB1 isoform from Spilianakis lab 400ng/μl or 10μL rabbit IgG from Santa Cruz 400ng/μL). The mixture of Protein G Sepharose beads with either SATB1 isoform or IgG was incubated for 4h at 4^{o}C while rotating. After the 4h incubation, the Protein G Sepharose/Antibody was centrifuged at 3000g for 2min at 4^{o}C, the supernatant was

discarded and the beads were washed three times with IP buffer. During the 4h incubation of the Sepharose beads with the antibodies, 100μL of Protein G Sepharose beads per immunoprecipitation were prepared again as described above in order to preclear the 1mg protein extracts that were previously stored at 4ºC. After the Sepharose beads were washed, the 1mg protein extracts were precleared for 1h at 4ºC on an end-to-end rotator. Then the Protein G Sepharose/protein extracts were centrifuged at 3000g for 2min at 4ºC and the precleared lysates were transferred to the Protein G Sepharose/Antibody that was previously washed. The 1mL mixtures (Sepharose/SATB1/protein extracts & Sepharose/IgG/protein extracts) were left overnight at 4ºC while rotating to immunoprecipitate. The following day the mixtures were centrifuged at 3000g for 2min at 4ºC, the flow through was stored at another tube and the Sepharose beads were washed with NETN buffer three times and centrifuged at 3000g for 2min at 4ºC. After the washes, the Sepharose beads were precipitated and the protocol continued with Western blot.

Lysis buffer: 50mM Tris pH8, 170mM NaCl, 0.5% NP-40, 50mM NaF, 1x Complete Protease Inhibitor, 1mM PMSF just prior to use.

IP buffer: 25mM Tris pH7.6, 150mM NaCl

NETN buffer: 10mM Tris pH8, 250mM NaCl, 5mM EDTA, 0.5% NP-40, 1mM PMSF just prior to use.

D. Western Blot

1. Western Blot

8% acrylamide gel was prepared and submerged in running buffer. Then the beads, 30μg flow-through and 100μg of input-protein extract were resuspended in 12.5μL 5x loading buffer. Then the four samples were boiled at 95ºC for 10min and immediately loaded on the gel. The gel was electrophoresed at 35Volt until the dye reached the separation gel and then at 150Volt until the dye reached the bottom of the gel. After the SDS-PAGE electrophoresis, the gel, PVDF membrane, sponges and papers were first submerged in cold transfer buffer for 10min and then they were sandwiched in the following order: sponge, 2x papers, gel, PVDF membrane, 2x papers and sponge. The sandwich was submerged in transfer buffer and the transfer was done overnight at

4ºC at 40Volts. The next day a quick wash was done to the membrane and subsequently it was incubated with blocking buffer for 1.5 hours at room-temperature with agitation. Following the blocking step the blocking buffer was discarded and the membrane was incubated with the primary antibody for 2 hours at room-temperature with agitation. Then, three washes were done to the membrane for 10min with TBS-T under agitation. Following the washes the membrane was incubated with the secondary antibody for 1hour at room-temperature under agitation. Finally, three washes were done to the membrane for 10min with TBS-T under agitation. Chemiluminescence followed.

10x Tris-Glycine pH8.3: 250mM Tris, 2M Glycine, no pH adjustment needed.

1x running buffer: 1x Tris-Glycine pH8.3, 0.1% SDS.

Transfer buffer: 0.1% SDS, 1x Tris-Glycine pH8.3, 20% Methanol.

5x loading buffer pH6.8: 10% SDS, 50% Glycerol, 100mM Tris, 80mM β-mercaptoethanol, 0.05%w/v Bromophenol Blue, 10mM EDTA.

8% SDS-PAGE gel: Separating gel: 4.5mL H_2O, 2.7mL acrylamide 30%/0.8% Bis, 2.6mL 1.5M Tris pH8.8, 100μL 10% SDS,100μL 10%APS, 10μL TEMED. **Stacking gel:** 3.0mL H_2O, 0.67mL acrylamide 30%/0.8% Bis, 1.25mL 0.5M Tris pH6.8, 50μL 10% SDS, 50μL 10% APS, 5μL TEMED.

10x TBS pH7.6: 500mM Tris, 9% NaCl, HCl until pH7.6.

TBS-T: 1x TBS, 0.1% Tween-20.

Blocking buffer: TBS-T, 5% milk.

Primary antibody: TBS-T, 1% milk, 1:1.000 rabbit anti-SET1A antibody (ab70378).

Secondary antibody: TBS-T, 1% milk, 1:20.000 anti-rabbit HRP antibody.

2. Harsh Stripping

First the stripping buffer was warmed at 50ºC and then the membrane was incubated in the stripping buffer for 45min at 50ºC under agitation. The stripping buffer was disposed in the fume hood and the membrane was rinsed under running distilled water tap for 1min. Then extensive washes were done to the membrane with TBS-T.

Blocking, incubation of the membrane with primary and secondary antibodies, as well as washes were done in the same way as in Western blot.

Stripping buffer: 20mL 10% SDS, 12.5mL 0.5M Tris-HCl pH6.8, 67.5mL distilled water, 0.8mL β-mercaptoethanol under the fume hood.

10x TBS pH7.6: 500mM Tris, 9% NaCl, HCl until pH7.6.

TBS-T: 1x TBS, 0.1% Tween-20.

Blocking buffer: TBS-T, 5% milk.

Primary antibody: TBS-T, 1% milk, 1:1.000 rabbit anti-SATB1 antibody (Lab isoform).

Secondary antibody: TBS-T, 1% milk, 1:20.000 anti-rabbit HRP antibody.

E. Native chromatin immunoprecipitation

1. Native chromatin immunoprecipitation

This protocol was based on S. Cuddapah, A. Barski, K. Cui et al. protocol[61]. Every step of this protocol was done on ice, unless otherwise stated. The thymocytes' extraction protocol was used for one C57BL/6 mouse 4 weeks old and a total of 110 million thymocytes were extracted. The thymocytes were divided in new falcons in order for each falcon to contain 30 million cells and were centrifuged at 1400rpm for 5min at 4^{o}C. Then the 30 million thymocytes were resuspended in 1mL MNase reaction buffer at room-temperature, 0.2units MNase mix from Sigma were added and were incubated for 16min at 37^{o}C. In order to stop the reaction, EGTA was added to a final concentration of 3mM and the tube was centrifuged at 1800rpm for 8min at 4^{o}C. The supernatant was discarded and the cell pellet was resuspended in 1mL sonication buffer. After transferring the cells in a 1.5mL tube, they were sonicated three times for 20sec with 20sec break on ice each time. Then they were centrifuged at 13.000rpm for 15min at 4^{o}C, the supernatant was kept and 20μL were taken from the 1mL of sonication buffer in order to check the shearing efficiency. The rest of the sample was stored in 5% glycerol at -80^{o}C and this sample is from now on going to be called soluble chromatin solution (SCS). After checking the shearing efficiency the SCS was defrosted and sonication buffer was added in order for the final volume to be 3mL,

and also Triton X-100 was added at a 1% final concentration. From the SCS 20μL were stored at -20ºC as an input. Then 10μg of anti-H3K4me2 antibody (stock 0.5μg/μL) were added in the SCS and was left rotating overnight at 4ºC. The following day the Dynabeads protein G were preblocked. First, the Dynabeads were homogenized and then 50μL of Dynabeads were added in 950μL 1xBSA/ 1%PBS. The beads were washed 3 times with 1xBSA/ 1%PBS for 3min at 4ºC while rotating. The beads were collected each time with a magnet on ice. After the last wash, the Dynabeads were incubated for 30min at 4ºC. Then the SCS was added to the Dynabeads and were incubated for 2hours at 4ºC while rotating. Following the 2hours incubation, the beads were washed for 3min at 4ºC while rotating the tube with the order of the following buffers, one for each wash: 1) low salt buffer, 2) high salt buffer, 3) lithium buffer and 4) TE buffer. The beads were collected each time with a magnet on ice. Finally, in order to reverse the crosslinks the beads were incubated overnight in 250μL elution buffer at 65ºC. The 20μL input was also incubated overnight in 250μL elution buffer at 65ºC. The next day 250μL elution buffer was added and then the tubes were incubated for 4hours at 65ºC. Afterwards, 100μg of RNase A (stock 100μg/μL) was added, as well as 250μL TE buffer and the tubes were incubated for 30min at 37ºC. Finally, 20μg of proteinase K (stock 10μg/μL) were added and the tubes were incubated for 1hour at 55ºC. At this point the tubes contained only DNA and in order to recover it Phenol-Chloroform extraction was conducted. First, 500μL PCI were added, then vortexed for about 1min, centrifuged at 14.000rpm for 5min at 4ºC and kept the aqueous phase. This step was done twice. Then 500μL chloroform were added, again vortexed for about 1min, centrifuged at 14.000rpm for 5min at 4ºC and kept the aqueous phase. Then 50μL 3M CH_3COONa, 40μg glycogen (20μg/μL) and 1mL 100% EtOH were added and the tubes were stored for 1hour at -80ºC. Then they were centrifuged at 14.000rpm for 15min at 4ºC, the supernatant was discarded and the pellet was resuspended in 1mL ice cold 80% EtOH. Finally, the tubes were centrifuged at 14.000rpm for 15min at 4ºC, the supernatant was discarded and the pellet was left to completely dry from the EtOH. Then the pellet was resuspended in 30μL dH_2O and the DNA concentration was measured with a Qubit Fluorometer. The DNA was stored at -80ºC until it was sent for sequencing. The ChIP-Seq libraries and analysis for our experiments were done in Dr. Dimitris Thanos lab at the Biomedical Research Foundation, Athens Academy.

MNase reaction buffer: 10mM Tris pH7.5, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$, 4% NP-40, 1mM PMSF just prior to use.

Sonication buffer: 50mM Hepes pH7.9, 150mM NaCl, 5mM EDTA, 0.5mM EGTA, 0.3% SDS, 0.1% Na-deoxycholate, 1mM PMSF just prior to use.

Low salt buffer: 20mM Tris pH8, 150mM NaCl, 2mM EDTA, 0.1%SDS, 1%Triton X-100, 1mM PMSF just prior to use.

High salt buffer: 20mM Tris pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, 1mM PMSF just prior to use.

Lithium buffer: 10mM Tris pH8, 250mM LiCl, 2mM EDTA, 1% NP-40, 1mM PMSF just prior to use.

TE buffer: 10mM Tris pH8, 1mM EDTA, 50mM NaCl.

Elution buffer: 50mM Tris pH8, 10mM EDTA, 1% SDS.

2. Shearing efficiency

20μL of chromatin were taken, lysis buffer was added until final a volume of 250μL and then incubated for 4hours at 65^{o}C. Afterwards, 100μg of RNase A (stock 100μg/μL) was added, as well as 250μL TE buffer and incubated the tube for 30min at 37^{o}C. Finally, 20μg of proteinase K (stock 10μg/μL) were added and were incubated for 1hour at 55^{o}C. At this point the tube contained only DNA and in order to recover it Phenol-Chloroform extraction was conducted the same way it was done for the ChIP. In the end, the pellet was resuspended in 20μL ddH_2O, where 1μL was used for quantitation, while the rest 19μL were loaded on a 2% agarose gel in order to visualize the shearing efficiency. We expected mono-di-tri-nucleosomes (170-340-520bp respectively) in order to have a good resolution in the sequencing.

Lysis buffer: 50mM Tris pH8, 10mM EDTA pH8, 0.2M NaCl, 1% SDS.

TE buffer: 10mM Tris pH8, 1mM EDTA.

PCI: Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)

F. Sequencing data analysis

The chromatin immunoprecipitation coupled with Sequencing (ChIP-Seq) workflow can be obtained from the following link:

https://usegalaxy.org/u/klaourakis/w/copy-of-chip-seq-analysis-1

The IP.fastq and input.fastq files were zipped with 7zip to .fastq.gz files and then uploaded to usegalaxy.org with the free FTP client, FileZilla. They were uploaded in usegalaxy.org by selecting in the Tools box: Get_Data -> Upload_File -> Choose FTP file. Finally, in Step_1 and Step_2 the IP and input files were imported respectively and then at the bottom of the page "run workflow" was clicked. The output files are: four FastQC files for the quality check of the IP and input sequencing before and after the trimming, two Flagstat files for the quality check of the IP and input mapping, MACS2 callpeaks in BED format for the summits and in tabular format for the peaks, bigWig for both the IP and input for visualization in genome browsers, a histogram for frequency statistics of MACS2 summits and finally a .bed file with the MACS2 peaks filtered for fold_enrichment>=5, plus a window of 500bp at both sides of the peak. The visualization of the bigWig files was done in the IGB and IGV genome browsers. Intersects were found from BedTools, with WindowBed.

GSE52071[62] , GSE66248[63] , GSE49847[64]

V. Results

A. A novel protein-protein interaction is identified

Previous immunoprecipitation experiments for SATB1 followed by mass-spectrometry analysis in our lab have identified about 290 proteins that potentially have a physical interaction with SATB1 (data not shown). Among the proteins that were identified, there were also six (SET1A, WDR5, RBBP5, ASH2L, CXXC1 and HCFC1) out of eight (WDR82 and DPY30) proteins from the SET1A COMPASS complex, a complex that mono-di-tri-methylate histone 3 at lysine 4 (H3K4me1, H3K4me2, H3K4me3).

In order to validate these data and prove that SATB1 has a physical interaction with SET1A we performed a series of experiments in C57BL/6 and CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout thymocytes. First we visualized this interaction *in vivo* by conducting double sequential immunofluorescence- these experiments were done in close collaboration with the PhD candidate Eralda Salataj. Although we did not conduct any quantification for the colocalization of the two signals, it was visible that the colocalization was significant and was a second indication that SATB1 and SET1A are found in close proximity inside the nucleus (Fig.2 and Sup.Fig.1). It was also clear that SET1A in thymocytes nuclei has a cage-like pattern circumscribing heterochromatin, similar to the published SATB1 distribution. Interestingly, in CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout thymocytes SET1A retained its cage-like pattern and did not collapse (Fig.3 and Sup.Fig.2). One possible explanation for which SET1A maintained this distribution could be that other proteins that interact with SET1A or one of the other seven proteins of the COMPASS complex has a cage-like pattern and are able to substitute SATB1 in this role. The immunofluorescence experiments are considered low resolution and cannot prove a direct physical interaction between the two proteins.

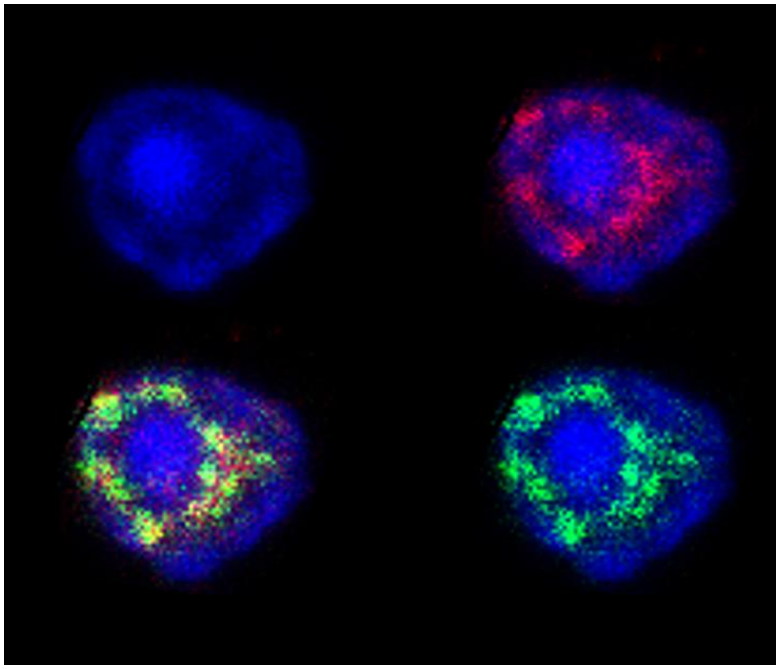


Fig.2 Double sequential immunofluorescence. C57BL/6 thymocytes stained for SATB1 (green), SET1A (red) and DAPI (blue). Both SATB1 and SET1A have a cage-like pattern that surrounds heterochromatin and colocalize. DAPI bright regions mark heterochromatin and faint areas euchromatin.

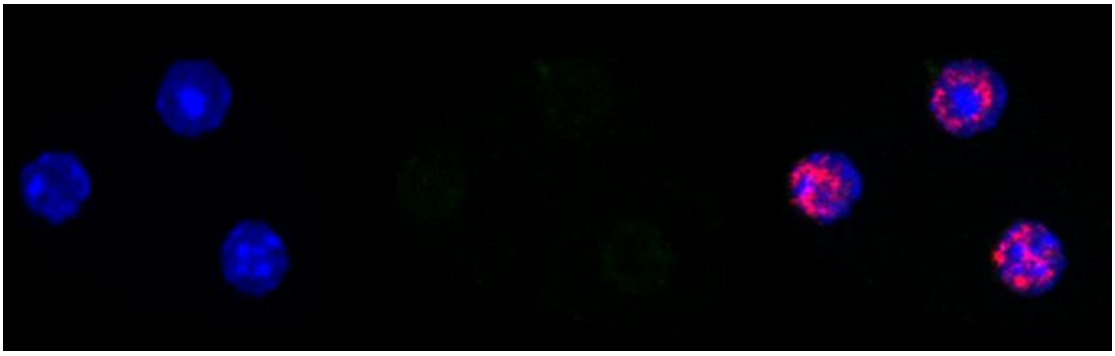


Fig.3 Double sequential immunofluorescence. CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout thymocytes stained for SATB1 (green), SET1A (red) and DAPI (blue). SET1A maintains the cage-like pattern even if SATB1 is not present. DAPI bright regions mark heterochromatin and faint areas euchromatin.

Next we wanted to prove with biochemical experiments that SATB1 and SET1A have a physical interaction. In order to do so we performed co-immunoprecipitation experiments for SATB1 followed by Western blot for SET1A. However, we were not able to precipitate SET1A together with SATB1, as SET1A was constantly observed in the flow-through (Fig.4). Moreover, when the membrane was stripped and we did Western blot for SATB1, SATB1 band was visible in the IP and not in the flow-through, indicating that the immunoprecipitation of SATB1 was successful (Sup.Fig.3). Although the result of this experiment cannot prove whether SATB1 and COMPASS complex interact or not, it indicates that SATB1 and SET1A probably do not have a direct physical interaction.

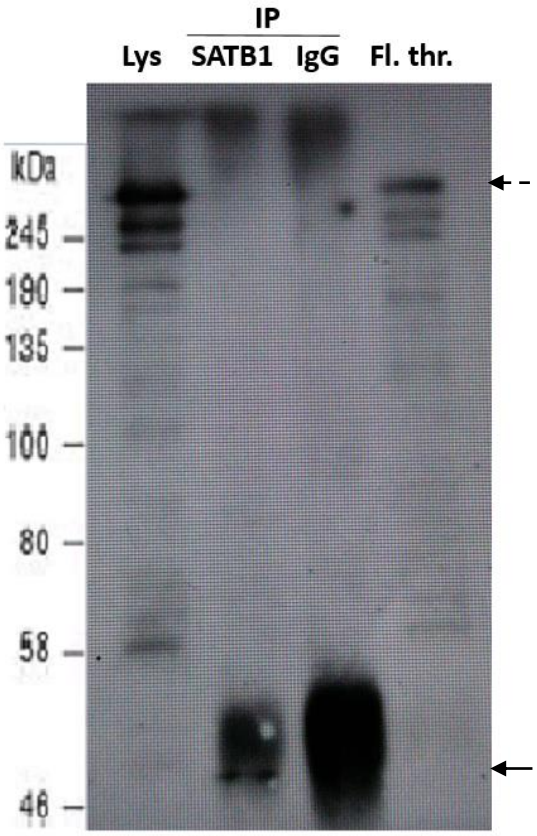


Fig.4 Co-immunoprecipitation for SATB1 followed by Western blot for SET1A. Lys: 100μg whole cell extract, SATB1: IP 1mg protein extract, IgG: IP 1mg protein extract, Flow through: 30μg protein extract after IP. Arrow with dashes: SET1A (260kDa). Solid arrow: antibody's heavy chain (50kDa). SET1A did not immunoprecipitate with SATB1 and can be clearly seen in the flow-through. This does not necessarily mean that SATB1 does not physically interact with SET1A.

B. Colocalization of SATB1, SET1A and H3K4me DNA binding sites

Even though we could not establish the physical interaction between SATB1 and SET1A with the co-IP followed by Western blot experiments, we were able to identify a colocalization in SATB1 and COMPASS complex DNA binding sites. More specifically we analyzed published raw ChIP-Seq data for SATB1, CXXC1, ASH2L and two histone modifications of the COMPASS complex, namely H3K4me1 and H3K4me3. CXXC1 is a protein found exclusively in the SET1A and SET1B COMPASS complexes[37], whereas ASH2L is a component found in all six COMPASS (SET1A, SET1B) and COMPASS-like (MLL1, MLL2, MLL3 and MLL4) complexes[65]. This analysis had low biological significance as the experiments for SATB1 and the two histone modifications were done in murine thymocytes[63,64], while for the CXXC1 and ASH2L in mouse embryonic stem cells (mESCs)[62]. Although SATB1 is expressed in mESCs and it is known to have a functional role in their differentiation[53,54], the comparison that we made can only give partial information regarding the colocalization of their DNA binding sites, as we compared different cell types. The data analysis was done using a publicly available workflow that we developed and is explained in the methods section. In table 1 we see the number of peaks that our analysis generated from the published ChIP-Seq data, as well as the number of peaks that colocalize between each protein, with a window of 1000 base pairs upstream and downstream of each peak's summit. There is a critical colocalization of the binding sites between SATB1 and the two histone modifications, and a relatively important colocalization between SATB1 and CXXC1 or ASH2L. However, we cannot claim that the two proteins bind simultaneously to the same DNA site, or that the one protein guides the other at these sites. Moreover, we see that the colocalization is not complete and this can be probably attributed to the difference of cell types that were used in the ChIP-Seq experiments. Finally, in table 2 are presented the number of peaks that colocalize between SATB1, CXXC1 and ASH2L (Fig.5), as well as SATB1, CXXC1, ASH2L and either H3K4me1 or H3K4me3. These results indicate that SATB1 and COMPASS complex might bind to the same DNA loci. However, this hypothesis has to be further explored with ChIP-Seq experiments in the same cell type, as well as with ChIP-MS approaches in order to confirm that the DNA binding occurs simultaneously at the same cell lineage.

proteins	SATB1	CXXC1	ASH2L	H3K4me1	H3K4me3
SATB1	21,074	4,536	2,522	15,620	16,181
CXXC1	-	6,617	3,215	5,772	10,657
ASH2L	-	-	6,799	3,707	5,404
H3K4me1	-	-	-	39,804	26,219
H3K4me3	-	-	-	-	26,513

Table.1 Peaks colocalization between SATB1, CXXC1, ASH2L, H3K4me1, H3K4me3. The numbers represent the number of intersects of the summit peaks plus 1000 base pairs upstream and downstream.

Table.2 Peaks colocalization of SATB1, CXXC1, ASH2L and either H3K4me1 or H3K4me3. The numbers represent the number of intersects of the summit peaks plus 1000 base pairs upstream and downstream.

SATB1 CXXC1 ASH2L	2,056
SATB1 CXXC1 ASH2L H3K4me1	1,374
SATB1 CXXC1 ASH2L H3K4me3	2,046

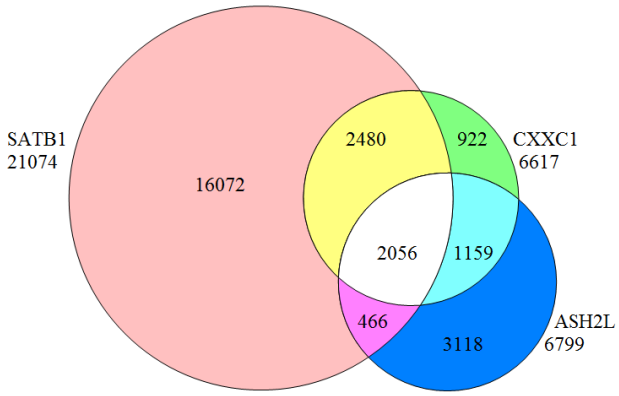


Fig. 5 Venn diagram showing the number of peaks that correlate between SATB1, CXXC1 and ASH2L. Although the CXXC1 and ASH2L ChIP-Seq experiments were done in different cell type from the SATB1 experiments, we can still see a major correlation of the peaks. This is an indicator that SATB1 and possibly the whole COMPASS complex, bind on the same DNA sites.

As a result of this analysis, the IP-MS and the double sequential IF experiments we wanted to test the hypothesis that SATB1, SET1A and the three histone H3K4 methylations have a statistical significant colocalization in their DNA binding sites inside the nucleus of murine thymocytes and whether the histone modification pattern

changes in SATB1-depleted thymocytes. Following this direction we performed ChIP-Seq for SATB1 and the histone modification H3K4me2 in C57BL/6 and CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout thymocytes, and at the time this thesis was being written additional ChIP-Seq experiments were being conducted in the lab. The ChIP-Seq analysis indicated that the reads were ambiguously mapped to repetitive and low complexity regions, as a result the libraries and the results that derived from them were considered problematic and unreliable (Sup.table1 & Sup.Fig.4). As a consequence more biological replicates have to be done in order to be able to produce statistical significant results.

VI. Discussion

In this thesis evidence have been provided for the interaction of the genome organizer SATB1 and the histone methyltransferase COMPASS complex, in murine thymocytes. Although the biochemical experiments failed to prove the physical interaction of SATB1 and the catalytic protein of the COMPASS complex- SET1A, they provided valuable information for future experiments.

First of all, as it was previously described, SATB1 in thymocytes has a three-dimensional cage-like pattern, which tightly surrounds heterochromatin[13]. Here we have shown that SET1A has the same cage-like pattern in thymocytes that colocalize with SATB1. What is more, SET1A cage-like pattern does not collapse in SATB1 conditional knock-out thymocytes. This suggests that SATB1 does not contribute to the nuclear distribution of SET1A, and if they indeed physically interact the function of this interaction may play a different role. Moreover, these results indicate that either SET1A is able- on its own or inside the COMPASS complex, to maintain the characteristic pattern, or it may interact with other proteins or transcription factors that also exhibit this nuclear distribution. There has been indication that other transcription factors also display the cage-like pattern that engulfs only euchromatin. One such example is the transcription factor BACH1[66] that was shown in our lab to have the aforementioned pattern and to possibly interact with SATB1 in murine thymocytes (Sup.Fig.5, data from Spilianakis lab).

Secondly, as it was mentioned before, the inability of the co-IP coupled with Western blot experiments to prove the existence of a physical interaction between SATB1 and SET1A does not prove that SATB1 and COMPASS complex do not indeed interact. The data that support this interaction are solid, as IP-MS for SATB1 precipitated six out of the eight proteins of the COMPASS complex, the co-IF indicated colocalization of SATB1 and SET1A and finally the ChIP-Seq analysis produced a good colocalization of peaks between SATB1, CXXC1 and ASH2L- two proteins of the COMPASS complex. As a result, we cannot yet conclude with confidence whether SATB1 interacts or not with the COMPASS complex, although a direct physical interaction with its catalytic protein- SET1A, is mostly improbable. It is also important to keep in mind that maybe SATB1 interacts directly with another protein of the COMPASS complex, and the precipitation of the SET1A and the other four

proteins is an indirect consequence of the complex. In the same context the ChIP-Seq and ChIP-MS experiments need to be conducted for SATB1 and SET1A in order to test if they bind simultaneously to the same DNA loci. These two types of experiments are going to provide novel and valuable information regarding this thesis hypothesis, as they will confirm or disprove the interaction and indicate if a common DNA localization pattern exists.

Moreover, during this thesis an easy-to-use without prior computational knowledge, publicly available workflow has been developed for ChIP-Seq analysis. Although this workflow has standard parameters that will work well with most ChIP-Seq data (Sup.Fig.6), it can be customized to match each experiment needs. However, if the sequencing quality is poor this workflow may not work at all, due to some strict parameters that poor sequencings do not fulfill.

In an expansion of the current hypothesis, it would be tempting to imagine that by organizing chromatin into loops, SATB1 is able to produce also clustering of non-coding loci, such as micro-RNA loci. Then, SATB1 could guide the methyltransferase COMPASS complex to the clustered micro-RNA loci in order to co-regulate their transcription. This could be a valid hypothesis, as it takes into account the functions of SATB1 as a chromatin organizer and transcription regulator[25,15] and the clustering of the micro-RNA loci that has been shown by our, as well as other, labs[67] (Sup.Fig.7, data from Spilianakis lab).

Finally, a question that is currently in the forefront of research and could also arise in the future of this research project is whether the genome organization guides the epigenetic modifications, or the epigenetic modifications control the genome organization, or maybe there is a combination of both. In our hypothesis context- and only if it is proven correct, we would have to ask whether SATB1 guides COMPASS complex to specific loci in order to methylate the histones, or the histone methylation produced by the COMPASS complex makes SATB1 create loops at specific loci in order to activate transcription of nearby genes. To address this interesting research endeavor we would have to conduct chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) for SATB1 in order to locate the chromatin interactions in which SATB1 is involved and then perform one of the chromosome conformation capture techniques in C57BL/6 and CD4.Cre-*Satb1*$^{fl/fl}$ thymocytes in order to uncover

the chromatin interactions that are formed specifically by SATB1. Last, by utilizing a catalytically inactive Cas9 (dCas9) fused with the catalytic site of a histone H3K4 demethylase- such as the LSD2, to demethylate histone H3K4 at specific sites and then conducting again one of the chromosome conformation capture techniques we could elucidate which modification is the cause and which the effect.

In summary, this work in combination with the preliminary data and the data that are currently being produced in our lab could uncover a novel regulatory mechanism of thymocyte maturation, which combines genome organization along with epigenetic modifications.

VII. References

1. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59–64 (2013).

2. Sehgal, N. *et al.* Large-scale probabilistic 3D organization of human chromosome territories. *Hum. Mol. Genet.* **25,** 419–36 (2016).

3. Bickmore, W. A. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14,** 67–84 (2013).

4. Dubé, J. C., Wang, X. Q. D. & Dostie, J. Spatial Organization of Epigenomes. *Curr. Mol. Biol. reports* **2,** 1–9 (2016).

5. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164,** 1110–21 (2016).

6. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137,** 1194–211 (2009).

7. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* **30,** 881–91 (2016).

8. Wani, A. H. *et al.* Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* **7,** 10291 (2016).

9. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47,** 1179–86 (2015).

10. Entrevan, M., Schuettengruber, B. & Cavalli, G. Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends Cell Biol.* (2016). doi:10.1016/j.tcb.2016.04.009

11. Amendola, M. & van Steensel, B. Nuclear lamins are not required for lamina-associated domain organization in mouse embryonic stem cells. *EMBO Rep.* **16,** 610–7 (2015).

12. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–13 (2010).

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

14. Cai, S., Lee, C. C. & Kohwi-Shigematsu, T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* **38,** 1278–88 (2006).

15. Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P. & Kohwi-Shigematsu, T. SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419,** 641–5 (2002).

16. Kohwi-Shigematsu, T. *et al.* Genome organizing function of SATB1 in tumor progression. *Semin. Cancer Biol.* **23,** 72–9 (2013).

17. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor

choice. *Cell* **126,** 403–13 (2006).

18. Apostolou, E. & Thanos, D. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **134,** 85–96 (2008).

19. Bacher, C. P. *et al.* Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat. Cell Biol.* **8,** 293–9 (2006).

20. Gómez-Díaz, E. & Corces, V. G. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.* **24,** 703–11 (2014).

21. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435,** 637–45 (2005).

22. Marella, N. V, Bhattacharya, S., Mukherjee, L., Xu, J. & Berezney, R. Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells. *J. Cell. Physiol.* **221,** 130–8 (2009).

23. Fritz, A. J. *et al.* Wide-scale alterations in interchromosomal organization in breast cancer cells: defining a network of interacting chromosomes. *Hum. Mol. Genet.* **23,** 5133–46 (2014).

24. Fritz, A. J. *et al.* Chromosomes at Work: Organization of Chromosome Territories in the Interphase Nucleus. *J. Cell. Biochem.* **117,** 9–19 (2016).

25. Han, H.-J., Russo, J., Kohwi, Y. & Kohwi-Shigematsu, T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* **452,** 187–93 (2008).

26. O'Shea, J. J. & Paul, W. E. Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327,** 1098–102 (2010).

27. Amsen, D., Spilianakis, C. G. & Flavell, R. A. How are TH1 and TH2 effector cells made? *Curr. Opin. Immunol.* **21,** 153–160 (2009).

28. Spilianakis, C. G. & Flavell, R. A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5,** 1017–1027 (2004).

29. Lee, G. R., Kim, S. T., Spilianakis, C. G., Fields, P. E. & Flavell, R. A. T Helper Cell Differentiation: Regulation by cis Elements and Epigenetics. *Immunity* **24,** 369–379 (2006).

30. Deligianni, C. & Spilianakis, C. G. Long-range genomic interactions epigenetically regulate the expression of a cytokine receptor. *EMBO Rep.* **13,** (2012).

31. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435,** 637–645 (2005).

32. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13,** 343–57 (2012).

33. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–95 (2011).

34. Huang, J., Marco, E., Pinello, L. & Yuan, G.-C. Predicting chromatin organization using histone marks. *Genome Biol.* **16,** 162 (2015).

35. Zhu, Y. *et al.* Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.* **7,** 10812 (2016).

36. Wu, M. *et al.* Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol. Cell. Biol.* **28,** 7337–44 (2008).

37. Lee, J.-H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* **280,** 41725–31 (2005).

38. Eissenberg, J. C. & Shilatifard, A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev. Biol.* **339,** 240–9 (2010).

39. Ernst, P. & Vakoc, C. R. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief. Funct. Genomics* **11,** 217–26 (2012).

40. Smith, E., Lin, C. & Shilatifard, A. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev.* **25,** 661–72 (2011).

41. Mersman, D. P., Du, H.-N., Fingerman, I. M., South, P. F. & Briggs, S. D. Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein stability, histone methylation, and gene expression. *J. Biol. Chem.* **287,** 2652–65 (2012).

42. Thornton, J. L. *et al.* Context dependency of Set1/COMPASS-mediated histone H3 Lys4 trimethylation. *Genes Dev.* **28,** 115–20 (2014).

43. Lee, J.-S. *et al.* Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131,** 1084–96 (2007).

44. Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15,** 163–76 (2005).

45. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20,** 259–66 (2013).

46. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–8 (2007).

47. Pavri, R. *et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125,** 703–17 (2006).

48. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11,** 709–19 (2003).

49. Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8695–700 (2002).

50. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21,** 1273–83 (2011).

51. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–83 (2011).

52. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–6 (2010).

53. Savarese, F. *et al.* Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev.* **23,** 2625–38 (2009).

54. Agrelo, R. *et al.* SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Dev. Cell* **16,** 507–16 (2009).

55. 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,** 631–45 (1992).

56. Wang, B., Dickinson, L. A., Koivunen, E., Ruoslahti, E. & Kohwi-Shigematsu, T. A novel matrix attachment region DNA binding motif identified using a random phage peptide library. *J. Biol. Chem.* **270,** 23239–42 (1995).

57. Kohwi-Shigematsu, T. & Kohwi, Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29,** 9551–60 (1990).

58. de Belle, I., Cai, S. & Kohwi-Shigematsu, T. The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J. Cell Biol.* **141,** 335–48 (1998).

59. 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,** 521–35 (2000).

60. Kondo, M. *et al.* SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J. Immunol.* **196,** 563–72 (2016).

61. Cuddapah, S. *et al.* Native Chromatin Preparation and Illumina/Solexa Library Construction. *Cold Spring Harb. Protoc.* **2009,** pdb.prot5237–pdb.prot5237 (2009).

62. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141,** 526–37 (2014).

63. 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,** 809–24 (2015).

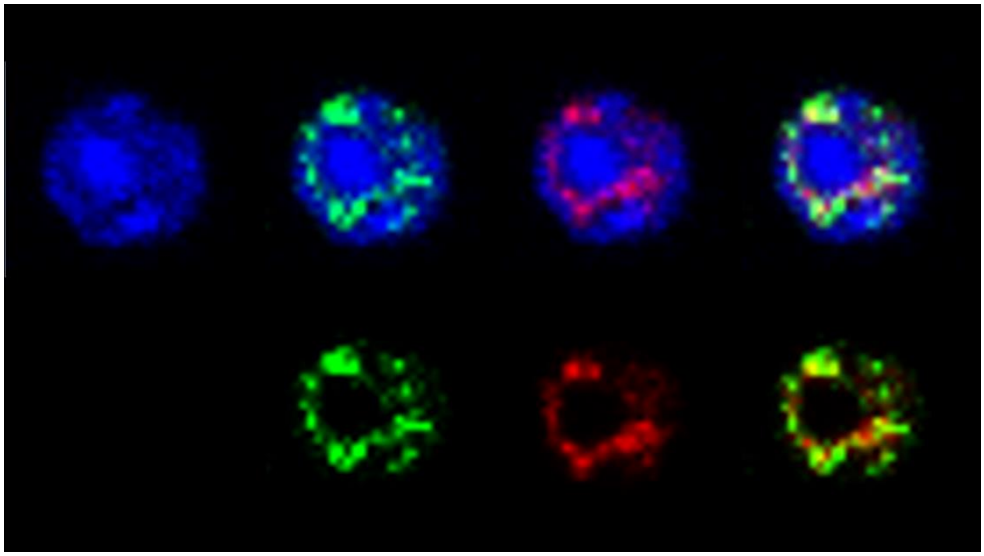
64. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515,** 355–64 (2014).

65. Ali, A., Veeranki, S. N. & Tyagi, S. A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. *Nucleic Acids Res.* **42,** 7611–24 (2014).

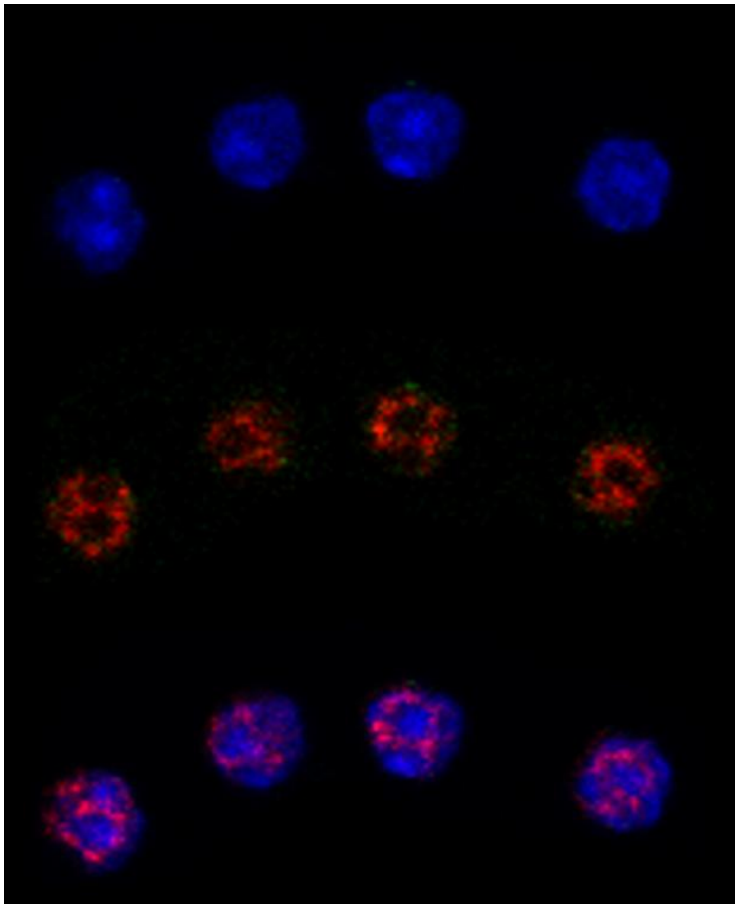
66. Zhou, Y., Wu, H., Zhao, M., Chang, C. & Lu, Q. The Bach Family of Transcription Factors: A Comprehensive Review. *Clin. Rev. Allergy Immunol.* **50,** 345–56 (2016).

67. Chen, D. *et al.* Dissecting the chromatin interactome of microRNA genes. *Nucleic Acids Res.* **42,** 3028–43 (2014).

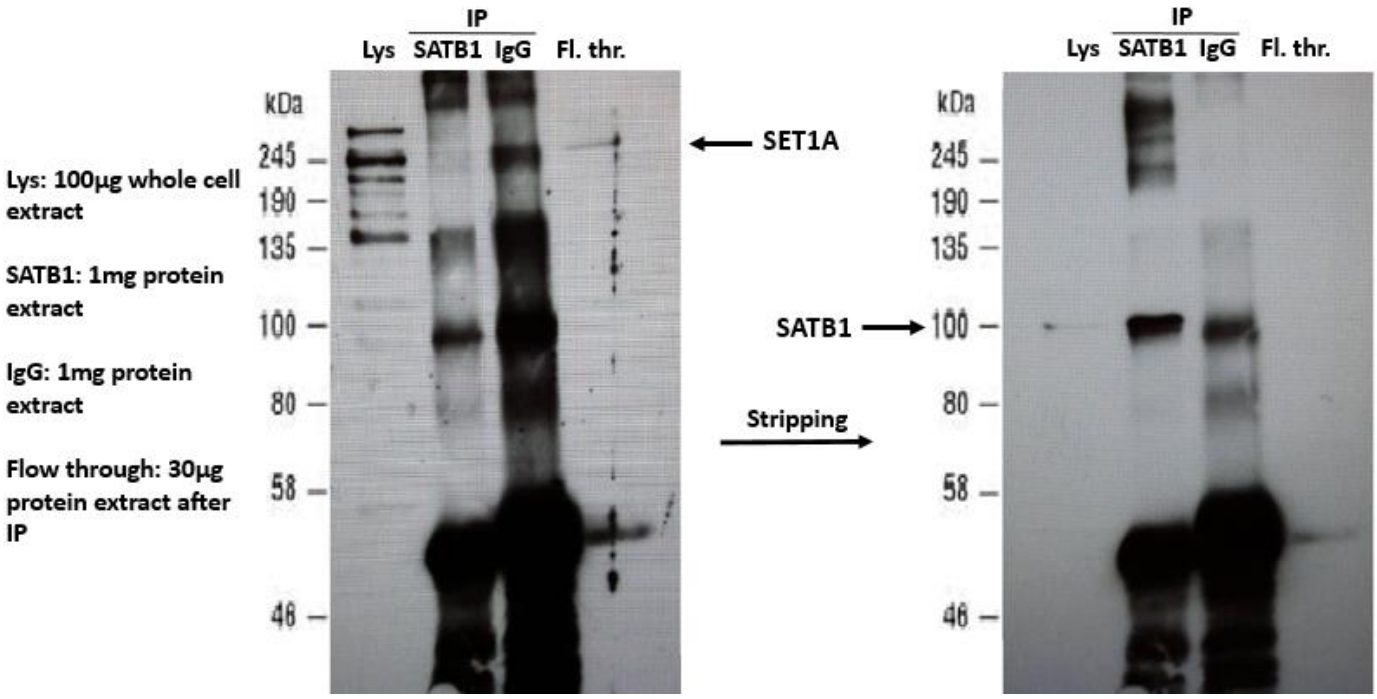
VIII. Supplementary



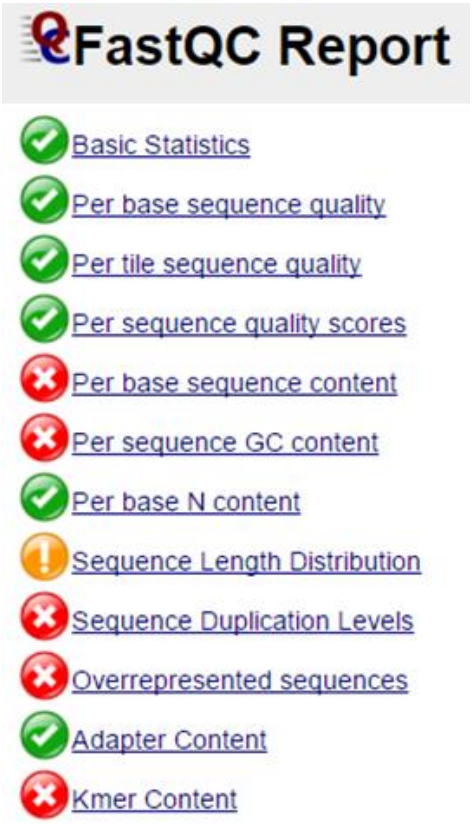
Sup.Fig.1 Double sequential immunofluorescence. C57BL/6 thymocytes stained for SATB1 (green), SET1A (red) and DAPI (blue). Both SATB1 and SET1A have a cage-like pattern that surrounds heterochromatin and colocalization. DAPI bright regions mark heterochromatin and faint areas euchromatin.



Sup.Fig.2 Double sequential immunofluorescence. CD4.Cre-*Satb1*$^{fl/fl}$ conditional knockout thymocytes stained for SATB1 (green), SET1A (red) and DAPI (blue). SET1A maintains the cage-like pattern even if SATB1 is not present. DAPI bright regions mark heterochromatin and faint areas euchromatin.



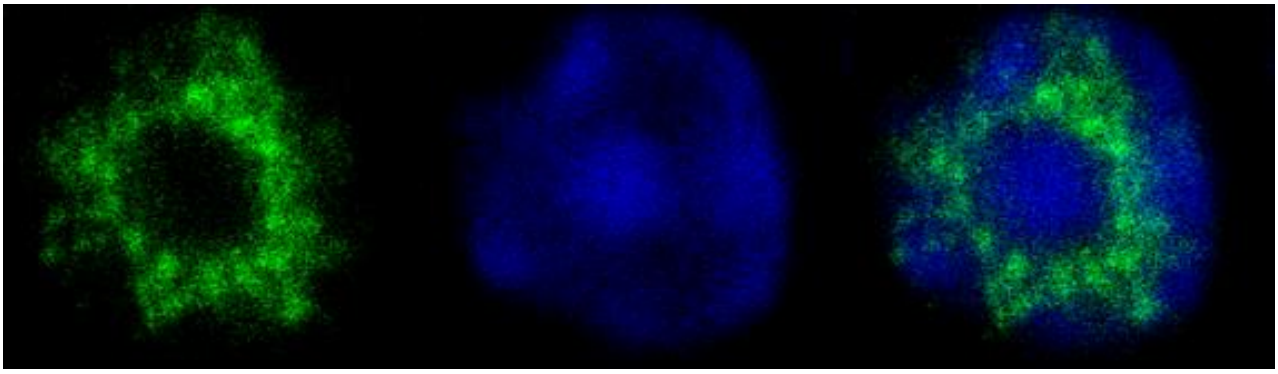
Sup.Fig.3 IP with SATB1 and Western blot with SET1A (left), stripping of the membrane and Western blot with SATB1 (right). SET1A appears in the flow-through. However, after membrane stripping and Western blot for SATB1, SATB1 band was visible in the IP and not in the flow-through, indicating that the immunoprecipitation of SATB1 was successful. Although there is a lot of background that gives more intense bands in the IgG than in the IP, we can see the opposite happening in the 100kD of SATB1 Western blot, proving that SATB1 has been precipitated.



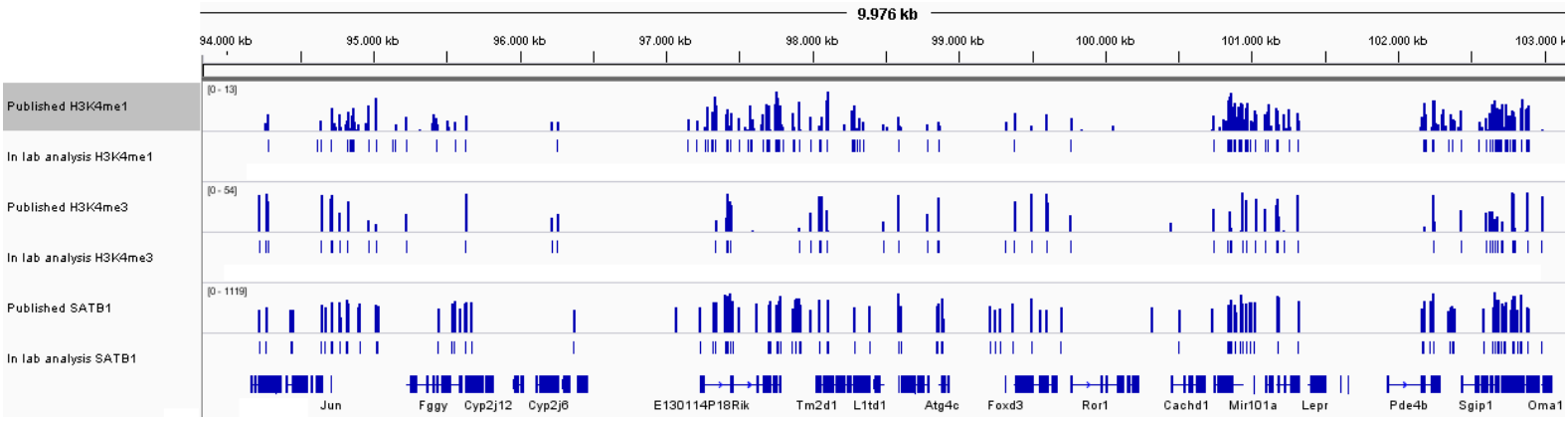
Sup.Fig.4 FastQC Report on H3K4me2 ChIP-Seq in murine thymocytes. This report shows that there are problems with the library. More specifically there are overrepresented sequences and the duplication levels are high, while the per base and GC content do not follow the predicted distribution. As a result the libraries were considered problematic and unreliable and further experiments need to be conducted in order to obtain statistical significant results.

Info	Number of Reads Sequenced	Number of Reads Aligned	Percentage of Aligned Reads	Number of Unique Reads	Percentage of Unique Reads	Number of Peaks
ChIP H3K4me2	20265741	541270	2,670862121	409127	75,58649103	587
Input	10270610	650507	6,333674436	586035	90,08896138	

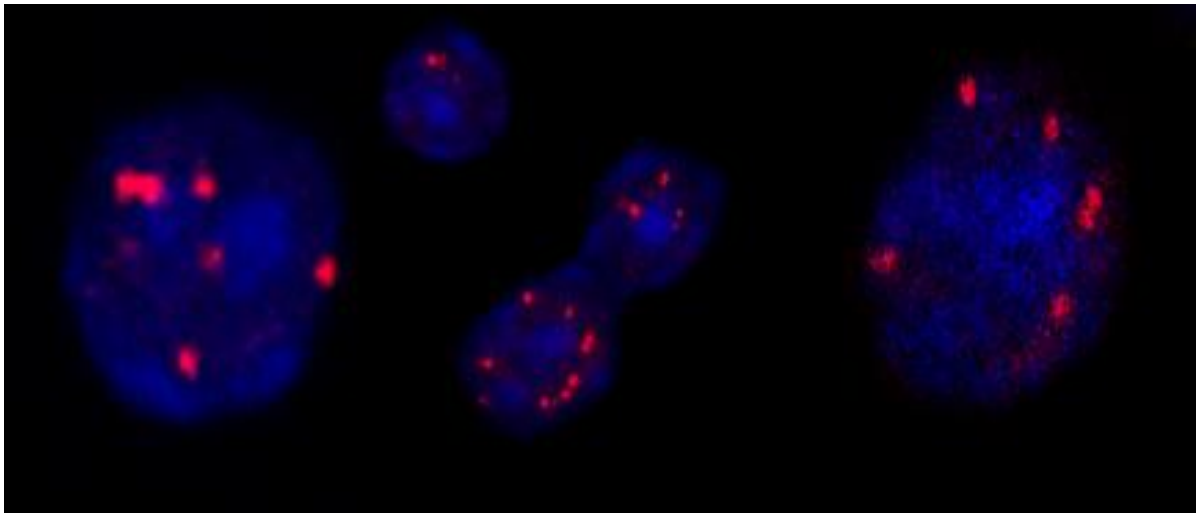
Sup.Table.1 H3K4me2 ChIP-Seq analysis. The low percentage of aligned reads and the also low final number of peaks provide an indication that the libraries are problematic and the analysis cannot be trusted. As a result more experiments should be done in order to provide statistical significant and unbiased results.



Sup.Fig.5 BACH1 cage-like pattern in C57BL/6 thymocytes. Unpublished data from Spilianakis lab. Transcription factor BACH1 displays the same cage-like pattern as SATB1 and SET1A in murine thymocytes. If SATB1 interacts with SET1A, could BACH1 also interact with SET1A and be the reason that SET1A holds the cage-like pattern even in the depletion of SATB1?



Sup.Fig.6 Example of comparison between published ChIP-Seq analysis[63,64] and in lab analysis of the same data using our own online workflow. Although the workflow will work well with most ChIP-Seq raw data, the user is able to customize most of the parameters. One example is to run first the FastQC alone in order to identify the encoding format and sequence length and then to optimize the settings of trimmomatic for trimming. Also, as this is a generalized easy-to-use online workflow it may produce errors if the library is problematic and the FastQC indicates poor quality of the sequences. As a result it must be used with caution from researches with no bioinformatics background. https://usegalaxy.org/u/klaourakis/w/copy-of-chip-seq-analysis-1



Sup.Fig.7 Clustering of the micro-RNA loci. Unpublished data from Spilianakis lab. Recent publications have correlated the spatial coordination of micro-RNAs to their expression and function. Moreover, our lab as well as others have noticed a special interaction of micro-RNA loci. Could this micro-RNA loci clustering be controlled by SATB1, and their co-expression related with epigenetic modifications, such as histone methylation?