SATB1 undergoes isoform-specific phase transitions in T cells

Tomas Zelenka[1,2], Petros Tzerpos[1,3], George Panagopoulos[4], Konstantinos C. Tsolis[5], Dionysios-Alexandros Papamatheakis[1,2], Vassilis M. Papadakis[2], David Stanek[6], Charalampos Spilianakis[1,2,#]

[1] Department of Biology, University of Crete, GR70013 Heraklion, Crete, Greece

[2] Institute of Molecular Biology and Biotechnology—Foundation for Research and Technology Hellas, GR70013 Heraklion, Crete, Greece

[3] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, HU-4032 Hungary (current address)

[4] Biomedical Research Foundation, Academy of Athens, 4 Soranou Efessiou Street, Athens 11527, Greece

[5] Department of Microbiology and Immunology, KU Leuven, Rega Institute, Leuven, Belgium

[6] Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic

[#] Corresponding author, email: spiliana@imbb.forth.gr

Abstract

Intracellular space is demarcated into functional membraneless organelles and nuclear bodies via the process of phase separation. Phase transitions are involved in many functions linked to such bodies as well as in gene expression regulation and other cellular processes. In this work we describe how the genome organizer SATB1 utilizes its prion-like domains to undergo phase transitions. We have identified two SATB1 isoforms with distinct biophysical behavior and showed how phosphorylation and interaction with nuclear RNA, impact their phase transitions. Moreover, we show that SATB1 is associated with transcription and splicing, both of which evinced deregulation in *Satb1* knockout mice. Thus, the tight regulation of different SATB1 isoforms levels and their post-translational modifications are imperative for SATB1's physiological roles in T cell development while their deregulation may be linked to disorders such as cancer.

Keywords: SATB1, phase separation, prion, poly-Q, T cells, thymocytes, nuclear matrix, chromatin organization, immunity, cancer.

Introduction

Phase transitions and more specifically liquid-liquid phase separation (LLPS) are processes involved in the organization of membraneless organelles and they were proposed to regulate many cellular processes[1]. The ability of a protein to undergo phase transitions depends on its structure together with environmental cues. LLPS is mostly driven by weak, multivalent interactions between macromolecules; especially protein-protein interactions assembling into high order oligomers, or weak multivalent interactions between intrinsically disordered regions (IDRs) and often accompanied by interactions with RNA[2,3]. Low-complexity prion-like domains (PrLDs) and RNA binding domains are known features that affect a protein's biophysical properties[4,5]. A special characteristic of prion-like proteins is the presence of a poly-Q domain[6]. Poly-Q harboring proteins and generally proteins with PrLDs are known to undergo phase separation[4,5,7–10]. However, these proteins are prone to abnormalities and their phase transitions frequently result in fibrous amyloids or other rigid aggregates, which are associated with a number of pathologies[8,11]. Despite often having negative connotation, PrLDs most certainly have some important physiological roles; from regulating the solubility of proteins, to processes linked to LLPS, such as gene expression regulation and others[12].

In general, the regulatory mechanisms of phase transitions are poorly understood. There are several factors contributing to regulation of phase transitions, including post-translational modifications and alternative splicing[13]. The latter is gaining attention as a robust tissue-specific regulatory mechanism of gene expression, allowing production of functionally diverse proteins from a single gene[14]. More specifically, alternative splicing allows subtle changes to LLPS propensity as well as more dramatic inclusion or exclusion of whole IDRs or other regions, defining a protein's LLPS properties. A regulatory role of alternative splicing in LLPS has

already been described for BRD4[15], TAU[16,17], hnRNPDL[18] and other proteins, collectively constituting examples of distinct mechanisms on how alternative splicing affects phase transitions[13].

Recently, we have identified the PrLD-containing protein SATB1 and investigated its function as a mediator of the 3D enhancer network in developing T cells[19]. In this work, we focused on the physiological and pathological implications of SATB1's phase transitions. Apart from its physiological roles in T cells, where it is predominantly expressed[20], SATB1 has been largely studied in the context of various cancers[21–23]. Apart of its function as a tissue-specific genome organizer[19,24], it also modulates nuclear architecture by recruiting chromatin modifying protein complexes[25,26]. Here we identified two full-length SATB1 isoforms that display different biophysical properties, further emphasizing the importance of alternative splicing, as an additional regulatory layer for T cell development. Both SATB1 isoforms bear two PrLDs and a short poly-Q domain, collectively providing a phase separation potential to SATB1. We showed that the interaction with RNA controls SATB1's solubility and phase transition behavior. Moreover, SATB1's phosphorylation in a serine residue S635 regulates its phase transitions and DNA binding ability. As a phase separating protein, SATB1 participated in transcription and splicing and its mislocalization to cytoplasm resulted in protein aggregation. The deregulated production of the two isoforms and the altered phase transitions of SATB1 can contribute to the severity of pathologies such as cancer.

Results

Identification of a novel SATB1 protein isoform

SATB1 protein consists of an N-terminal part containing the ULD (ubiquitin-like oligomerization domain; similar in sequence to the PDZ domain) and CUTL (CUT repeat-like)

domains which are responsible for its self-oligomerization[27–29] and the interaction with other proteins[25,26,30]. The CUT1, CUT2 domains and a homeodomain are responsible for SATB1's DNA binding activity and are all localized to its C-terminus (Fig. 1a). Between CUT2 and the homeodomain, there is a glutamine-rich region following the predicted extra peptide (EP) of the long SATB1 isoform [UniProt: long SATB1 isoform E9PVB7_Mouse (795 aa), short SATB1 isoform Q60611 (764 aa)]. To validate the presence of this extra peptide of 31 amino acids (and thus the existence of a long SATB1 isoform), we have generated a custom-made antibody targeting specifically the peptide originating from the extra *Satb1* exon. Moreover, we have also developed an antibody targeting the N-terminal part of SATB1 (the first 330 amino acid residues; Fig. 1a). None of the commercially available antibodies can distinguish between the two SATB1 isoforms *in vivo*, as they target parts of SATB1 shared by both isoforms. We have also designed primers for RT-qPCR experiments in a way that the forward primer either hybridized at the 3'-end of exon10 and simultaneously at the 5'-end of exon11 (Fwd1) or entirely within the extra exon (Fwd2) of the long isoform, to specifically amplify the short or long *Satb1* transcripts, respectively (Fig. 1b). We found that in murine thymocytes the short *Satb1* transcripts were about 3-5-fold more abundant compared to the long *Satb1* transcripts (Fig. 1c). Additionally, we have performed stranded-total-RNA-seq experiments in thymocytes and detected a similar difference of 5.6-fold enrichment of the short over the long *Satb1* isoform (Fig. 1d). Note also the presence of seven additional annotated transcripts, three of which may encode for truncated protein isoforms and four untranslated transcripts which all may have an unknown regulatory function. It should be noted that when investigating the SATB1 protein levels, we have to bear in mind that the antibodies targeting the N-terminus of SATB1 protein cannot discriminate between the short and long isoforms, thus we can only compare the amount of the long SATB1 isoform to

the total SATB1 protein levels (Fig. 1e). Additional validation of the specificity for the custom-made SATB1 antibody is available in Extended Data Fig. 1a. Based on this data, our conclusion is that in murine thymocytes a long SATB1 isoform can be detected at both mRNA and protein levels.

SATB1 is a nuclear protein and, utilizing confocal microscopy, it was previously shown to display a cage-like pattern in thymocytes[26]. We revisited this question and employed super-resolution microscopy approaches to study SATB1's subnuclear localization. Our results indicated that the previously described cage-like pattern of SATB1[26], at high resolution, rather resembled, often almost perfectly shaped, individual spheres which may or may not be interconnected (Fig. 1f). An apparent emerging question, upon the detection of these SATB1 speckles, was whether they are involved in the regulation of physiological processes such as transcription.

SATB1 co-localizes with sites of active transcription

Each super-resolution microscopy technique comes with certain limitations, hence for our applications we compared the results from STED (Stimulated Emission Depletion; provides high resolution, lower risk of bleed-through) with 3D-SIM (3D Structured Illumination Microscopy; allows higher throughput, broader variety of fluorophores and DAPI staining, less photobleaching). In search for functional differences between the two SATB1 isoforms, we probed their subnuclear localization *in vivo*, utilizing 3D-SIM super-resolution microscopy (Fig. 2a). The quantification of SATB1 speckles in four nuclear zones, derived based on the relative intensity of DAPI staining, highlighted the localization of SATB1 mainly to the regions with medium to low DAPI staining, apparently correlated with euchromatin and interchromatin regions (zones 3 & 4, Fig. 2a, b). A similar distribution of the SATB1 signal could also be seen

from the fluorocytogram of the pixel-based colocalization analysis between the SATB1 and DAPI signals (Extended Data Fig. 1b). The major difference between the localization of the different SATB1 isoforms, uncovered by the use of the two different antibodies, relied in the heterochromatin areas (zone 1), where the long isoform was less frequently present. Given the localization of SATB1 to the nuclear zones with estimated transcriptional activity (Fig. 2b, zone 3), we investigated the potential association between SATB1 and transcription. We unraveled the localization of SATB1 isoforms and the sites of active transcription labeled with 5-fluorouridine, using both STED (Fig. 2c) and 3D-SIM (Extended Data Fig. 1c). Based on qualitative evaluation we found that the SATB1 signal was found in close proximity to nascent transcripts.

We further performed co-immunoprecipitation experiments (utilizing an antibody targeting all SATB1 protein isoforms) coupled to mass spectrometry analysis (IP-MS) to determine SATB1's interactome in murine thymocytes (Extended Data Fig. 2a-b; Supplementary File 1). Unbiased k-means clustering of the significantly enriched SATB1 interactors revealed two major clusters consisting mostly of proteins involved in transcription (blue cluster 1; Fig. 2d and Extended Data Fig. 2c) and splicing (yellow cluster 2; Fig. 2d and Extended Data Fig. 2c). Both transcription and splicing were also significantly enriched terms in the gene ontology enrichment analysis of SATB1 interactors (Fig. 2e; Supplementary File 2), indicating that SATB1 plays an important role in thymocyte RNA production and processing.

To further investigate the functional link between SATB1 and gene transcription and splicing, we performed stranded-total-RNA-seq experiments in wild type (WT) and *Satb1*$^{fl/fl}$*Cd4*-Cre^{+} (*Satb1* cKO) murine thymocytes, with SATB1 being specifically depleted at the CD4^{+}CD8^{+} T cells during thymocytes development[19]. Consistent with SATB1's localization at transcription sites, we identified a vast deregulation with 1,641 differentially expressed genes

(Fig. 2f) and 2,014 genes with altered splicing efficiency (Extended Data Fig. 2d-e; Supplementary File 3-4). We should note that the extent of deregulation of spliced genes was directly correlated with SATB1 binding (Extended Data Fig. 2d). Similarly, the transcription of multiple genes was directly affected by SATB1 binding as we previously demonstrated[19].

In summary, we found that in the primary thymocyte nucleus, SATB1 localized in the euchromatin and interchromatin regions in close proximity to actively transcribed genes, interacted with protein complexes involved in the regulation of gene transcription and mRNA splicing and upon SATB1 depletion, both physiological processes of transcription and splicing were deregulated.

SATB1 protein is phase separated *in vivo*

SATB1 has a predicted phase separation potential[5]. Here we demonstrated its connection to transcription and found that it forms almost perfectly spherical speckles (Fig. 1f), markedly resembling phase separated transcriptional condensates[31,32]. Thus, we next investigated whether the observed nuclear SATB1 speckles could be a result of phase transitions. To probe differences in phase separation in mouse primary cells, without any intervention to SATB1 structure and expression, we first utilized 1,6-hexanediol treatment, which was previously shown to dissolve the LLPS droplets[33]. If SATB1 was indeed undergoing phase transitions into the transcriptional condensates, we would expect to detect a sensitivity of the SATB1 speckles to the hexanediol treatment, followed by decreased colocalization with fluorouridine-marked sites of active transcription. The long SATB1 isoform speckles evinced such sensitivity as demonstrated by a titration series with increasing concentrations of 1,6-hexanediol treatment followed by 3D-SIM super-resolution microscopy analysis (Fig. 3a, b). Moreover, analysis with STED microscopy revealed that the long SATB1 isoform was more sensitive to hexanediol than collectively all

SATB1 protein isoforms (Fig. 3c and Extended Data Fig. 3a). To quantify the difference in colocalization between SATB1 and fluorouridine-labeled sites of active transcription, as well as the changes upon hexanediol treatment, we used a pixel-based colocalization analysis (Fig. 3d and Extended Data Fig. 3b-d). Based on these results, the long SATB1 isoform displayed a higher association with sites of active transcription and this association was more sensitive to the hexanediol treatment than collectively all SATB1 isoforms.

To expand our study regarding the effect of 1,6-hexanediol in dissociating SATB1 speckles, we employed Raman spectroscopy, a non-invasive label-free approach, which is able to detect changes in chemical bonding. Raman spectroscopy was already used to predict global transcriptomic profiles from living cells[34], thus we reasoned that it may also be used to study phase separation. We measured Raman spectra in thymocytes of both WT and *Satb1* cKO animals and compared them with spectra from cells upon 1,6-hexanediol treatment. Principal component analysis of the resulting Raman spectra clustered the treated and non-treated *Satb1* cKO cells together, while the WT cells clustered separately (Fig. 3e). The WT 1,6-hexanediol treated cells clustered between the aforementioned groups. These results suggested that the principal component 1 (PC1) could discriminate WT and *Satb1* cKO cells based on chemical bonding provided by SATB1 protein. These bonds were probably enriched for weak interactions responsible for phase transitions that are susceptible to hexanediol treatment. This shifted the cluster of WT treated cells towards the *Satb1* cKO cells. However, the remaining covalent bonds differentiated the WT samples from *Satb1* cKO cells where all bonds provided by SATB1 were absent. This experiment demonstrated that Raman spectroscopy can be utilized as a non-invasive research approach to unravel the global chemical state of a cell and its changes upon various

conditions. Moreover, together with the light microscopy experiments, it allowed us to investigate SATB1's biophysical behavior *in vivo*.

The N-terminus of SATB1 undergoes phase transitions

To further support our findings that SATB1 undergoes phase transitions *in vivo*, we utilized the light-inducible CRY2-mCherry optoDroplet system[35] (Fig. 4a). This tool allows studying the phase separating properties of individual parts of a protein in a controlled manner. The CRY2-mCherry chimeric constructs, encoding part of a protein of interest are transiently transfected into cells. Following illumination by blue light, the photo-responsive protein CRY2 is induced to homo-oligomerize while bringing the protein molecules of interest into spatial proximity. Only proteins with phase separating properties will then form droplets, which are visualized by accumulation of the mCherry signal. First, we selected the N-terminal part of SATB1 (1-298 aa), which is known for its oligomerization potential and its involvement in protein-protein interactions[25–30]. Note, that the multivalent nature of a protein indicates its predisposition to undergo LLPS[1–3]. Moreover, the selected region of SATB1 includes part of a prion-like domain (Fig. 4a), which makes the N-terminus of SATB1 a prospective candidate for phase transitions. Transfected NIH-3T3 cells were activated with a 488 nm laser every two seconds to initiate CRY2 association and mCherry signal was detected. For the analysis, cells were grouped into two categories based on the original mCherry relative fluorescence, reflecting the transfection efficiency, as this directly affects the speed of droplet assembly[35]. The construct encompassing the N-terminus of SATB1 formed optoDroplets at a similar or even faster rate than the positive control FUS (Fig. 4b), a popular model for LLPS studies[35]. The low SATB1 concentration group of cells evinced the first droplets ~10 seconds upon induction, compared to FUS, where the first droplets appeared after ~14-20 seconds. Cells with higher SATB1

expression displayed even faster droplet assembly, where both FUS and the N-terminus of SATB1 started forming droplets at ~6 seconds upon activation. The droplets resembled phase separated condensates based on their round shape and ability to fuse (Fig. 4c). The droplet assembly rate for both the N-terminal part of SATB1 and FUS proteins (Fig. 4d), highlights the direct correlation between protein concentration and the speed of droplet assembly. We detected a higher number of FUS optoDroplets that were formed, however the N-terminus SATB1 optoDroplets grew in bigger sizes (Fig. 4b and Extended Data Fig. 4a). FUS optoDroplets disassembled entirely after 100 and 250 seconds of inactivation time (488 nm laser off), for the low and high protein concentration groups, respectively. In contrary, the optoDroplets formed by the N-terminus of the SATB1 construct displayed a much slower disassembly rate and often even after ten minutes of inactivation time there were some rigid structures still present (Fig. 4e). The group of cells with lower concentration of the SATB1 N-terminus displayed a gradual decrease in droplet size during inactivation time, unlike the high concentration group (Extended Data Fig. 4b). Considering the original faster assembly of SATB1 droplets, we hypothesized that they also turned more quickly into long-lasting aggregates, which would also explain the differences between low and high concentration groups. FRAP experiments confirmed the rigidity of the SATB1 droplets having around 43% of an immobile fraction compared to FUS with only 17% of an immobile fraction (Fig. 4f). However, the mobile fraction of SATB1 droplets manifested a halftime of recovery 3.3 seconds, much lower than the halftime of 11.4 seconds of FUS droplets, supporting the hypothesis that the N-terminus constructs were generally over-activated.

The N-terminus of SATB1 is shared by both the short and long SATB1 isoforms, so with the use of the optoDroplet system we investigated whether the two full length SATB1 isoforms could undergo phase transitions similar to the shorter N-terminal part. Upon activation, both

full-length isoforms displayed visible ultrastructural changes which were reversible upon inactivation (Fig. 5a). These did not resemble the optoDroplets detected for FUS and the N-terminus of SATB1. In this respect, although full length SATB1 bears a nuclear localization signal (aa 20-40), we identified a fraction of cells where the protein was localized in the cytoplasm, in the form of rigid aggregates (Fig. 5b). Computational analysis, using the algorithm *cat*GRANULE[36], of the protein sequence for both SATB1 isoforms indicated a higher propensity of the long SATB1 isoform to undergo phase transitions with a propensity score of 0.390, compared to 0.379 for the short isoform (Fig. 5c). This difference was dependent on the extra peptide of the long isoform. Out of the 31 amino acids comprising the extra peptide, there are six prolines, five serines and three glycines – all of which contribute to the low complexity of the polypeptide region[3]. Since cytoplasmic protein aggregation has been previously described for proteins containing poly-Q domains and generally PrLDs, we next generated truncated SATB1 constructs encoding two of its IDR regions, the poly-Q domain (Fig. 1a) and in the case of the long SATB1 isoform also the extra peptide neighboring the poly-Q domain (Fig. 1a and 4a). The constructs containing the extra peptide of the long SATB1 isoform frequently resulted in protein aggregates which did not respond to CRY2 activation (Fig. 5d). The constructs derived from the short SATB1 isoform did not form almost any aggregates or only in a very small fraction of cells (Fig. 5d). These results indicated a difference in the biophysical behavior of the two SATB1 isoforms.

Modes of regulation of SATB1 phase transitions

To determine why SATB1 forms aggregates in the cytoplasm, we next investigated the potential means of regulation for the SATB1 phase transitions. Phase separating properties of prion-like proteins are regulated by the presence of nuclear RNA, which keeps them in a soluble form when

in the nucleus[37]. However, currently it is not known whether SATB1 interacts with RNA, except for a study predicting its interaction with the Xist lncRNA[38]. Our data, involve SATB1 to both transcription and splicing (Fig. 2 and 3). Moreover, SATB1 was originally discovered as a protein of the nuclear matrix[39], so its association with RNA is plausible. To test this hypothesis, we performed nuclear matrix extraction experiments, including RNase treatment, to investigate the potential association of SATB1 with RNA. A fraction of SATB1 remained in the nucleus even after the extraction of cells with high concentrations of salt or treatment with DNase, though the treatment with RNase alone depleted SATB1 from the nucleus (Fig. 6a). To further validate these findings, we performed RNA immunoprecipitation experiments utilizing antibodies against the SATB1 protein and we detected the association of SATB1 with lncRNAs such as Xist, Malat and Neat, which constitute great examples of structural nuclear RNAs (Fig. 6b). To further support our hypothesis that nuclear RNA keeps the IDR constructs of the long SATB1 isoform soluble, we forced their localization to the nucleus by fusing them to the SV40 nuclear localization signal. These IDR constructs remained soluble (Fig. 6c) and they did not respond to the CRY2 activation even after 20 minutes of activation, in support of our hypothesis.

Another factor affecting the phase transitions of proteins includes their post-translational modifications. More specifically, the phosphorylation of the PrLD of FUS was shown to prevent its aggregation[40,41]. Similarly, the efficient disassembly of stress granules and other membraneless organelles was dependent on phosphorylation[42–44]. SATB1 harbors many verified and predicted phosphorylation sites[20]. The selected IDR region of SATB1 (Fig. 4a) contained multiple serine residues which are targets of phosphorylation. Based on DNA affinity chromatography coupled to mass spectrometry, we identified a novel SATB1 phosphorylation site, S635 (S666 in the long isoform, Fig. 6d). To investigate the impact of phosphorylation in

position S635, which directly follows the PrLD, we generated a S635A SATB1 mutant, abolishing phosphorylation at this site (Fig. 4a). We performed DNA affinity purification assays, utilizing protein extracts from transiently transfected cells with either the wild type or mutated SATB1 and found that this mutant had abrogated DNA binding ability (Extended Data Fig. 4c). Even more importantly, a CRY2-mCherry construct with the mutated part (S635A) of the short SATB1 isoform developed aggregates in the cytoplasm (Fig. 6e), similar to those of the cytoplasmic long isoform (Fig. 5d). This observation suggested that phosphorylation of SATB1 plays an important role in its phase transitions, affecting DNA binding and therefore the regulatory effects of SATB1 on genome organization and transcriptional gene regulation.

Furthermore, a regulatory mechanism coordinating the production of the short and long *Satb1* isoforms could also control SATB1's phase transitions – given their differential propensity to undergo phase transitions. Such a mechanism could rely on the differential promoter usage as recently reported[45,46]. Our stranded-total-RNA-seq experiments revealed a series of long isoform transcripts originating from various *Satb1* promoters, though indistinguishable from promoters priming the short isoforms (Fig. 6f). Therefore, we reasoned that a more plausible hypothesis would be based on the regulation of alternative splicing. We should note that we have identified a binding site specific to the long SATB1 isoform[19] right at the extra exon of the long isoform (Fig. 6f), suggesting a potential autoregulatory role for the expression of the long SATB1 isoform. The study of alternative splicing based on our RNA-seq analysis revealed a deregulation in the usage of the extra exon of the long *Satb1* isoform (the only *Satb1* exon affected) in *Satb1* cKO cells (deltaPsi = 0.12, probability = 0.974; Supplementary File 4), further suggesting that SATB1 itself is able to control the levels of the short and long *Satb1* isoforms. A possible mechanism controlling alternative splicing is based on its kinetic coupling with transcription. Several studies

indicated how chromatin acetylation and generally increased chromatin accessibility may lead to exon skipping due to enhanced RNA polymerase II elongation[47,48]. We hypothesized that the regulation of *Satb1* gene may include similar mechanisms, possibly even being regulated by a negative feedback loop via the long SATB1 isoform binding, resulting in increased chromatin accessibility at the extra exon and thus exon skipping. This hypothesis is supported by the increased usage of the extra exon in the absence of SATB1 in *Satb1* cKO (Supplementary File 4). To further address this, we utilized the TCGA breast cancer dataset (BRCA) as a cell type expressing SATB1[49]. ATAC-seq experiments for a series of human patients with aggressive breast cancer[50] revealed differences in chromatin accessibility at the extra exon of the *SATB1* gene (data not shown), suggesting its potential involvement in alternative splicing regulation according to the "kinetic coupling" model. Indeed, the increased chromatin accessibility at the extra exon (allowing faster read-through by RNA polymerase) was positively correlated with the expression of the short *SATB1* isoform and slightly negatively correlated with the long *SATB1* isoform (Fig. 6g). Moreover, we investigated whether the differential expression of *SATB1* isoforms was associated with poor disease prognosis. Worse pathological stages of breast cancer and expression of *SATB1* isoforms displayed a positive correlation for the long isoform but not for the short isoform (Fig. 6h). This was further supported by worse survival of patients with increased levels of long *SATB1* isoform and low levels of estrogen receptor (Extended Data Fig. 4d). These observations not only supported the existence of the long *SATB1* isoform in humans, but they also shed light at the potential link between the regulation of *SATB1* isoforms production and their involvement in pathological conditions.

Discussion

In this work we described the existence of two SATB1 protein isoforms in both human and mouse and investigated their biophysical properties. We demonstrated that SATB1 in primary murine thymocytes was mainly localized in the transcriptionally active nuclear zone and co-localized with sites of active transcription. Unfortunately, we could not directly compare differences between the long and short SATB1 isoforms *in vivo*. However, the long isoform-specific protein droplets evinced higher colocalization with fluorouridine-labeled sites of active transcription as well as higher sensitivity to hexanediol, compared to the droplets detected by antibodies targeting collectively all SATB1 isoforms. Following these observations, we can hypothesize that the long SATB1 isoform was more frequently associated with phase transitions and thus also active transcription and splicing.

Furthermore, we used genetically engineered optogenetic constructs to investigate the biophysical properties of different SATB1 polypeptides. The N-terminus of SATB1, shared by both SATB1 isoforms, evinced rapid phase transitions in the CRY2-mCherry optoDroplet heterologous system. We analyzed transiently transfected cells, separately, based on the relative recombinant protein concentration. Notably the low-expressing cells transfected with the N-terminus of SATB1 displayed even faster droplet assembly compared to the positive control FUS. Correspondingly, in the high-expressing cells, SATB1 tended to form larger condensates, generally displaying signs of over-saturation and thus conversion into more rigid gel-like structures[1], known from PrLD proteins. Even though the coalescing droplets and other aspects resembled liquid-liquid phase separation, with the mounting controversy[1,51–53] and other plausible biophysical models available[51,54,55], we would refrain from the strict association between SATB1 and LLPS itself until a specific molecular toolbox for LLPS research is defined, especially for the *in vivo* systems. Our current approach could not clearly discriminate whether

overexpression, turning SATB1 into long-lasting aggregates, was solely due to the internal properties of SATB1 or due to the artificial multivalency provided by the CRY2 optogenetic system. SATB1 contains two PrLD regions, one of which is included at the 3' end of the SATB1 N-terminal construct and likely facilitates the protein's phase transitions. The second PrLD directly extends the region harboring the extra peptide of the long SATB1 isoform, which by itself contributed to slightly higher phase separation propensity of the long isoform and its aggregation in the cytoplasm. However, the forced nuclear localization of the second PrLD with the extra peptide of the long isoform, even at very high concentrations, did not indicate any droplet formation. This would suggest that the N-terminal part of SATB1 with its ability to form multivalent interactions was necessary for SATB1's phase transitions. Therefore, there is only a subtle difference in the biophysical properties of the short and long SATB1 isoforms permitting the tight regulation of SATB1's phase transitions.

There may exist several means of regulation of SATB1's phase transitions. We showed that the nuclear RNA environment contributes to the solubility of nuclear SATB1, resulting in protein aggregation when mislocalized into the cytoplasm. Apart from nuclear RNA, buffering a protein's phase transitions[37], two studies described the role of nuclear-import receptors in chaperoning and disaggregating aberrant phases of PrLD containing proteins[56,57]. The latter does not seem to apply to SATB1, as SATB1 possesses a unique NLS lacking the clusters of basic residues, thus being dissimilar to known nuclear localization signals. This suggests that there might be other factors involved in its nuclear transport machinery[58]. Moreover, we described a novel phosphorylation site (S635) in murine SATB1 protein. The mutation S635A resulted in SATB1's aggregation and abruption of its DNA binding ability. Thus, it is reasonable to hypothesize that phosphorylation is also crucial for SATB1's interactions with RNA, having a

synergistic effect on SATB1's phase transitions. SATB1 is the substrate for a great number of post-translational modifications[20], hence they likely serve important roles in the regulation of physiological processes, including the modulation of phase transitions. A functional association between SATB1 and PML bodies was already described in Jurkat cells[59]. We should note that PML bodies represent another example of phase separated nuclear bodies[60] associated with SATB1. Targeting of SATB1 into PML bodies depends on its phosphorylation status; when phosphorylated it remains diffused, whereas unphosphorylated SATB1 is localized to PML bodies[61]. This is in line with the phase separation model as well as with our results from S635A mutated SATB1, having a phosphorylation blockade that induces its phase transitions.

There is an increasing body of evidence that PrLDs could serve important biological functions such as regulation of a protein's phase transitions behavior[12]. The predicted saturation concentration (i.e. the concentration threshold that has to be crossed to trigger phase separation) of human SATB1 is 114 μM, which falls at the lower end of identified proteins with phase separation potential[5]. However, the low saturation concentration would allow a precise regulation of SATB1's phase transitions just by modulating its expression levels. Moreover, given the slightly different biophysical properties of the two SATB1 isoforms, the controlled production of either the short or the long SATB1 isoform would establish yet another layer of regulation. We have recently described the SATB1-mediated promoter-enhancer network which controls the expression of important master regulator genes in developing T cells[19]. Here we hypothesize, that the increasing protein levels of SATB1 and/or the long-to-short isoform ratio could promote the incorporation of the regulated genes into the transcriptional condensates, as it was also described for other factors[62–64]. Therefore, SATB1 could be a great example where phase transitions are utilized for its gene expression regulation properties. Such a mechanism of action would only be

possible in cell types with high SATB1 concentration, such as in murine $CD4^+CD8^+$ double positive T cells or possibly in embryonic stem cells. Moreover, the chromatin epigenetic landscape could represent another regulatory layer affecting the relative isoform abundance.

Overall, our observations set a new ground for future research as the biophysical properties of SATB1 may determine its pathological behavior in a number of disorders. Deregulation of any of the described mechanisms, such as the altered RNA binding capacity of SATB1, deregulation of its multivalency by mutations or by post-translational modifications, deregulation of its expression levels, or changes to the epigenetic landscape, ultimately leading to a different proportion of the long and short SATB1 isoforms, could all result in physiological aberrations. Future studies should draw their attention to isoform-specific expression levels of SATB1, ideally coupled with microscopy, to reveal its localization pattern in cells. Moreover, there is a large knowledge gap in the identification and validation of post-translational modifications of SATB1 and their underlying function. Here we provided experimental proof for RNA binding by SATB1, however it is not yet clear which domain is capable for RNA binding and under which circumstances. Lastly, more focus should be brought to potential mutations in the susceptible poly-Q region of SATB1. Mutations of the poly-Q region may be difficult to identify, however they could lead to the extension of the poly-Q stretch, ultimately facilitating aggregate formation and thus a pathological situation as known for other PrLD-bearing proteins[8,65].

Author contributions

T.Z. and C.S. designed the study. T.Z. cloned the CRY2 constructs and transfected the cells, performed the microscopy experiments, RNA-seq experiments and the computational analyses. P.T. created the *Satb1* cKO mouse, cloned the original *Satb1* transcripts and generated custom-made antibodies. G.P. performed the SATB1 co-immunoprecipitation and mass spectrometry experiments. K.C.T. analyzed the mass spectrometry data. D.A.P. performed the western blot experiments. V.M.P. performed and analyzed the Raman spectroscopy experiments. D.S. provided resources for the super-resolution microscopy and consulted the imaging experiments and the analysis for alternative splicing. T.Z. wrote the original manuscript. C.S. supervised the work, obtained funding and corrected the manuscript. All authors read, discussed and approved the manuscript.

Declaration of interests

The authors declare no competing interests.

Data availability

RNA-seq experiments are deposited in Gene Expression Omnibus database under accession number GSE173470. The other datasets generated and/or analyzed during the current study are available upon request.

Material and methods

Satb1 isoforms cloning

Murine thymocytes total RNA was used for reverse transcription. The cDNA was amplified utilizing primers specific for mouse *Satb1*: SATB1-fwd (5′-GCC AGA TCT ATG GAT CAT TTG AAC GAGGC-3′, with integrated restriction enzyme site for BglII and SATB1-rev (5′-GCC CTG CAG TCA GTC TTT CAA GTC GGCAT-3′, with an integrated restriction enzyme site for PstI. The PCR product was cloned in a TOPO TA vector (Thermo Fisher Scientific, pCR™ II-TOPO®) and 20 clones with *Satb1* inserts were analyzed upon restriction enzyme digestion and Sanger sequencing indicating the presence of *Satb1* cDNAs with different length. BLAST search indicated that the cloned cDNAs encoded a short SATB1 isoform of 764 amino acids (UniProtKB – Q60611) and a long SATB1 isoform of 795 amino acids (UniprotKB – E9PVB7).

Custom SATB1 antibodies production

The first 330 amino acids of SATB1 short isoform (UniProtKB - Q60611, 1-764aa) fused to a 6x histidine tag at the N-terminus were expressed in *E. coli* and the antigen was utilized for the immunization of New Zealand rabbits and the production of SATB1 anti-sera (Davids Biotechnologie). Rabbit polyclonal immunoglobulin specific to SATB1 N-terminal 330 amino acids were purified with column chromatography with the immobilized antigen used for immunization. The long SATB1-specific antisera were produced by Davids Biotechnologie upon immunization of New Zealand rabbits with the GKGESRGVFLPSLLTPAPWPHAA peptide (corresponding to the extra peptide identified in the long SATB1 isoform) and subsequent purification of rabbit polyclonal peptide-specific immunoglobulins using affinity purification.

Animals and isolation of thymocytes

All experiments were conducted in accordance with the Laboratory Animal Care and Ethics Committee of IMBB-FORTH. Animal work was approved by the IMBB Institutional Animal Care and Ethics Committee. All the experiments were performed on mice with C57BL/6 background. The generation of *Satb1*$^{fl/fl}$*Cd4*-Cre^{+} mouse was previously described[19]. The animals used for the experiments were 4-8 weeks old. Primary thymocytes were resuspended by rubbing and passing the thymus through a 40 μm cell strainer (Falcon, 352340) in 1X PBS buffer. Cells were washed twice with 1X PBS, centrifuged at 500 g, at 4°C for 5 minutes, resuspended in 10 ml of 1X PBS and both steps were repeated.

RT-qPCR

Isolated thymocytes were resuspended in 1 ml of TRIzol Reagent (Sigma Aldrich, T9424) and RNA was isolated according to the manufacturer's protocol. Isolated RNA (5 μg) was decontaminated by incubation with 4 units of DNase I (NEB, M0303L) at 37°C for 30 minutes. The reaction was stopped by adding 10 μl of DNase Inactivation Reagent (Ambion, 8174G) and incubating at 70°C for 10 minutes. The supernatant was transferred to a new tube and RNA was precipitated by adding 10 μg of Linear Acrylamide (Ambion, AM9520), 2.5X volumes of 100% Ethanol and incubated at –80°C for 30 minutes. The supernatant was removed and the pellet was washed twice with 75% Ethanol. The air-dried pellets were resuspended in RNase-free water and incubated at 55°C for 15 minutes to dissolve. For reverse transcription, 1 μg of RNA was combined with 1 μl of oligo dTs (100 pmol/μl), 1 μl dNTPs (10 mM) and topped up with ddH_2O to 13 μl. Samples were denatured at 65°C for 5 minutes and transferred on ice. Samples were combined with 4 μl 5X Minotech RT assay buffer, 1 μl 0.1 M DTT, 1 μl RNase inhibitor (40 U/μl) and 1 μl Minotech RT enzyme (Minotech, 801-1). The mixture was incubated at 45°C for 60

minutes and then heat-inactivated at 70°C for 15 minutes. The original transcripts were quantified with qPCR, using SYBR Select Master Mix (Applied Biosystems, 4472908) according to the manufacturer's instructions with the following primers: long *Satb1* isoform (fwd: TTCTTACCAAGCCTGCTGACC, rev: TGGTCTTCTGCCGGTTTTCC), all short *Satb1* isoforms (fwd: CAGAACAGATTCAGCAGCAG, rev: TGGTCTTCTGCCGGTTTTCC) and *Dnmt1* (fwd: GCAGCCTATGAAGATTGAAGAG, rev: CCTGGCTAGATAGAGTGGTTTG).

Stranded-total-RNA sequencing

Freshly isolated thymocytes from female animals were resuspended in 1 ml of TRIzol Reagent (Invitrogen, 15596026) and RNA was isolated according to manufacturer's protocol. The aqueous phase with RNA was transferred into a tube and combined with 10 µg of Linear Acrylamide (Ambion, AM9520), 1/10 of sample volume of 3M CH3COONa (pH 5.2), 2.5 volumes of 100% Ethanol and tubes were mixed by flipping. Samples were incubated at –80°C for 40 minutes. Samples were brought to 0°C and centrifuged at 16,000 g, at 4°C for 30 minutes. The supernatant was removed and the pellet was washed twice with 75% ethanol. The air-dried pellets were resuspended in 40 µl RNase-free water and incubated at 55°C for 15 minutes to dissolve. To remove any residual DNA contamination, RNase-free DNase Buffer was added to samples until 1X final concentration together with 20 units of DNase I (NEB, M0303L) and incubated at 37°C for 20 minutes. Samples were then additionally purified using the RNeasy Mini Kit (Qiagen, 74104) according to the manufacturer's protocol. The quality of RNA was evaluated using the Agilent 2100 Bioanalyzer with Agilent RNA 6000 Nano Kit (Agilent Technologies, 5067-1511). Libraries were prepared using the Illumina TruSeq Stranded Total RNA kit with ribosomal depletion by Ribo-Zero Gold solution from Illumina according to the manufacturer's protocol and sequenced on an Illumina HiSeq 4000 (2x 75 bp).

Raw reads were mapped to the mm10 mouse genome using HISAT2[66]. Only mapped, paired reads with a map quality >20 were retained. Transcripts were assembled with StringTie[67] using an evidence-based Ensembl-Havana annotation file. Transcripts and genes were summarized using featureCounts[68] and statistically evaluated for differential expression using DESeq2[69]. When application required an intra-sample transcript comparison, DESeq2 values were further normalized to the gene length.

The list of exons was extracted from the mouse GENCODE gene set (version M20) using a customized script. Overlapping exons were merged and final exon-exon and intron-exon lists were generated. The lists were converted to 2 bp junctions (1 bp covering the first exon and 1 bp covering the second exon or the intron sequence). Both lists were intersected and purified to retain only paired exon-exon and intron-exon junctions from the same genes, i.e. embedded or overlapping genes and other unconventional cases were filtered out. Coverage of the junctions was assessed using the bedtools multicov -q 10 -s -split command and utilizing the HISAT2 mapped bam files from the RNA-seq analysis. The resulting exon-exon and exon-intron files from both WT and *Satb1* cKO were combined, junctions were paired and the unpaired junctions were filtered out. Junctions were split into the 5-prime and 3-prime end of the exon. Splicing efficiency was estimated for entire genes after summing up the coverage of all junctions from the same gene. Genes with overall exon-exon WT coverage (all exon-exon junctions together for all biological replicates) lower than 50 were filtered out. The splicing efficiency was determined as the intron/exon coverage ratio and its changes as a difference between *Satb1* cKO – WT intron/exon ratios. Moreover, the SATB1 binding sites' dataset (GSE173446[19]) was intersected with the junctions to search for a potential dependency of splicing on the SATB1 presence (Extended Data Fig. 2d). The SATB1 binding propensity was calculated as the sum of fold

change enrichments of all SATB1 peaks (for each peak summit against random Poisson distribution with local lambda, based on MACS2)[70] overlapping each gene and normalized to 1 kbp.

Alternative splicing was analyzed using the Whippet program[71] with --biascorrect flag and using both options: with and without the --bam flag (using the HISAT2 mapped bam files from the RNA-seq analysis, ensuring coverage of unannotated splice-sites and exons). In both cases, the mouse GENCODE gene set (version M11) was used to build the index file. The results from the whippet-delta script were filtered for results with probability >=0.9 and absolute delta psi value >=0.1. Out of 150,382 and 130,380 events it yielded 3,281 (2,015 genes) and 1,717 (1,077 genes) events that passed the filtering step for the bam-file-supplemented and without the bam-file analyses, respectively. Moreover, the analysis of back-splicing (circular RNA) revealed 639 events (519 genes) being underrepresented and 443 events (374 genes) overrepresented in the *Satb1* cKO.

Western blot

Whole cell protein extracts from female mice thymi were electrophoresed on a 10% SDS polyacrylamide gel. Proteins were transferred to a nitrocellulose membrane at 320 mA for 1.5 hours at 4°C. The membrane was blocked with 5% BSA at RT for 1 hour and then washed three times with TBST, 5 minutes each. The membrane was incubated with antibodies against the long isoform of SATB1 (1:200; Davids Biotechnologie, custom-made), all isoforms of SATB1 (1:50; Santa Cruz Biotechnology, sc-5990) and beta Actin (1:5,000; ORIGENE, TA811000) for 1 hour at room temperature in 1% BSA/TBS-T. The detection was performed using goat anti-rabbit HRP (1:5,000; Jackson, 111-035-003), donkey anti-goat HRP (1:5,000; Jackson, 705-035-003)

and goat anti-mouse HRP (1:5,000; Jackson, 115-035-146) for the long isoform of SATB1, all isoforms of SATB1 and beta Actin, respectively, all in 1% BSA/TBS-T.

SATB1 co-immunoprecipitation coupled to mass spectrometry

Thymocyte protein extracts were prepared from eight male 4-6 weeks old C57BL/6 mice. Thymocytes were incubated at room temperature for 30 minutes with rotation, in 1 ml lysis buffer (150 mM NaCl, 50 mM Tris pH=7.5, 5% Glycerol, 1% Nonidet P-40, 1 mM $MgCl_2$, 200 U benzonase, 1X Proteinase Inhibitors, 1 mM PMSF). 24 mg of protein extract were precleared using 100 µl of Dynabeads® Protein G (Life Technologies, 10004D) at 4°C for 1.5 hours on an end-to-end rotator. 50 µl of beads pre-blocked with 15 µg rabbit anti-SATB1 or rabbit IgG sera (Santa Cruz Biotechnology, C2712) at 4°C for 4 hours rotating and crosslinked at room temperature for 30 minutes with tilting/rotation using cross-linker Bis(sulfosuccinimidyl) suberate BS3 (Thermo Fisher Scientific, 21580). Precleared lysates were subjected to immunoprecipitation by incubation with Dynabeads crosslinked with anti-SATB1 or IgG at 4°C for 30 minutes. Beads were washed three times with buffer I (150 mM NaCl, 50 mM Tris pH=7.5, 5% Glycerol, 0.05% Nonidet P-40, 1mM PMSF), twice with Nonidet P-40 free Buffer II (150 mM NaCl, 50 mM Tris pH=7.5, 5% Glycerol, 1 mM PMSF) and boiled at 95°C for 5 minutes.

Proteins co-Immunoprecipitated with SATB1 were resolved via SDS-PAGE in a 10% polyacrylamide gel. The gel was stained with silver nitrate compatible with mass spectrometry fixation: 50% MeOH, 12% acetic acid, 0.05% Formalin for 2 hours and washed with 35% ethanol for 20 minutes for 3 times. Sensitization followed with 0.02% $Na_2S_2O_3$ for 2 minutes and washes performed with H_2O for 5 minutes 3 times. Staining was performed with 0.2% $AgNO_3$, 0.076% formalin for 20 minutes and washes were with H_2O for 1 minute 2 times. Development

took place with 6% Na_2CO_3, 0.05% formalin, 0.0004% $Na_2S_2O_3$. Staining was stopped with 50% MeOH, 12% acetic acid for 5 minutes.

Ten pairs of protein bands (anti-SATB1, IgG) were excised from silver-stained polyacrylamide gels and cut into small pieces. Gel pieces were covered with destain solution (30 mM potassium ferricyanide, 100 mM sodium thiosulfate) and vortexed for 10 minutes in order to remove silver nitrate. For the reduction and alkylation of the cysteine residues, the gel pieces were covered with 100 µl 10 mM DTT and shaked for 45 minutes at 56°C followed by 45 minutes shaking at room temperature with 100 µl 55 mM iodoacetamide. Proteins were digested in 30 µl of diluted Trypsin solution and incubated overnight at 37°C. The next day, the supernatant was collected. For the extraction from the gel matrix of generated peptides, gel pieces were shaken for 20 minutes first in 50 µl 50% acetonitril (ACN) and finally in 50 µl 0.1% TFA/50% ACN. Peptide solutions were centrifuged by Speed Vac until dry powder remained and then analyzed by means of liquid chromatography combined with mass spectrometry analysis (nanoLC-MS/MS), as previously described[72].

Peptide intensities of each experiment were normalized based on the mean difference of their distribution (log2 intensity) with the reference sample (SATB1 IP-MS 2nd biological replicate). Missing values were imputed using the mean values of each peptide within a given study group (SATB1 IP-MS or the control – IgG IP-MS), plus random noise equal to the mean observed variability in the dataset. Dataset variability was calculated as the standard deviation of the difference between all the biological replicates of a given group and the group mean. Protein abundance was calculated using the sum of peptide intensities. Differentially abundant proteins were selected after comparing the t-test p-value (using log2 intensities) and the fold change for each protein. P-values were adjusted for multiple hypothesis testing error using the

Benjamini-Hochberg method. Threshold for selection of differentially abundant proteins was an adjusted p-value < 0.05 and a fold change > 2x. Proteins that were uniquely identified in the SATB1 IP-MS dataset and proteins commonly identified with a significantly higher abundance in the SATB1 IP-MS replicates were considered for the downstream analysis as enriched proteins.

Protein-protein interaction data were downloaded from the STRING database[73]. STRING database mappings from Ensemble protein IDs to Uniprot IDs were used to map the interaction data with the experimental results. Protein-protein interaction network was build using the NetworkX Python library[74]. To cluster network nodes, vector representations were initially extracted for each node, using the node2vec algorithm[75] and followed by k-means clustering. GO and pathway enrichment analysis were performed using g:Profiler tool[76] using the default settings.

Identification of SATB1 phosphosite S635

DNA affinity chromatography coupled with mass spectrometry was performed as previously described[77,78] to identify associated proteins. A region of 151 bp, part of the RHS6 DNase I hypersensitive site of the TH2-LCR[79,80] was cloned into TOPO-TA vector using following primers: RP1-fwd CTTCCATTTGCTGGCCTCT; RP2-rev CCTGCATCTGATACTTAGACGTA. Two types of biotinylated oligonucleotides were utilized. Targeted fragment was biotinylated by PCR amplification with the following primers: biotinT7-fwd Biotin-d(CTCGAGTAATACGACTCACTATAGG); M13-rev AACAGCTATGACCATG. The same primers were used to prepare oligonucleotide for the negative selection (preclearing), by amplification from the TOPO-TA vector without the RHS6 region. 1 mg of M-280 Streptavidin Dynabeads (Invitrogen, 11205D) were resuspended and washed three times according to the manufacturer's instructions. 100 µl of 2X Binding & Wash buffer (10 mM Tris pH=7.5, 1 mM

EDTA, 2 M NaCl) containing 10 μg of the biotinylated oligonucleotide in an equal volume of water was added. The beads with the oligonucleotide were incubated for 15 minutes with rotation at room temperature. Beads were collected on a magnet and further washed with 1X Binding & Wash buffer and once with DNA binding buffer (20 mM HEPES pH=7.9, 0.1 mM KCl, 0.2 mM EDTA, 0.01% NP-40, 0.5 mM DTT, 0.2 mM PMSF, 10% Glycerol) supplemented with 0.1 mg/ml oligo dI:dC and poly dI:dC. Thymocyte nuclear extracts (5 mg) were first negatively selected by the "preclearing" biotinylated oligonucleotide and then positively selected utilizing the biotinylated RHS6 sequence. The biotinylated oligonucleotides used for the positive selection of bound thymocyte proteins were boiled and electrophoresed on a denaturing polyacrylamide gel. Upon Coomassie blue silver staining, individual slices of the positively selected samples, the negatively selected samples and washes were utilized in parallel for nanoLC-MS/MS analysis as previously described[72]. Subsequent protein identification was done using Mascot and Sequest software.

Immunofluorescence experiments and super-resolution microscopy

Freshly isolated thymocytes were incubated for 60 minutes (producing the best results compared to 30 and 240 minutes) at 37°C in complete DMEM (GIBCO, 11995-073) medium with 10% serum with 2 mM 5-Fluorouridine (FU; Sigma Aldrich, F5130) or directly attached on poly-D-lysine-coated coverslips. For super-resolution microscopy, high precision Zeiss squared coverslips of 1.5H thickness (0.170+-0.005 mm; Marienfeld Superior, 0107032) were treated with 1 M HCl overnight, rinsed with ddH2O and then stored in absolute ethanol. Right before the experiment, coverslips were coated by dipping in 0.1 mg/ml poly-D-lysine solution (Sigma Aldrich, P6407). Attached cells were washed once with 1X PBS. Coverslips to be treated with 1,6-hexanediol (Sigma Aldrich, 804308) were incubated for 5 minutes at room temperature with

the 1,6-hexanediol solution, carefully washed twice with 1X PBS and then fixed for 10 minutes on ice with 4% formaldehyde (Pierce, 28908) in 1X PBS. Fixed cells were permeabilized with 0.5% Triton-X in 1X PBS for 5 minutes on ice. Cells were washed three times with 1X PBS for 5 minutes each and blocked for 30 minutes at room temperature with Blocking Buffer [0.4% acetylated BSA (Ambion, AM2614) in 4X SSC] in a humidified chamber. Cells were incubated for 1.5 hours at room temperature with primary antibodies (see the table below) in Detection Buffer (0.1% acetylated BSA, 4X SSC, 0.1% Tween 20) in a humidified chamber. The excess of primary antibodies was washed away by three washes, for 5 minutes each, with Wash Buffer (4X SSC, 0.1% Tween 20). Cells were incubated for 60 minutes at room temperature with the secondary antibodies in Detection Buffer (0.1% acetylated BSA, 4X SSC, 0.1% Tween 20) in a humidified chamber. The excess was washed away by three washes, for 5 minutes each, with Wash Buffer (4X SSC, 0.1% Tween 20). The coverslips for 3D-SIM experiments were stained for 5 minutes at room temperature with DAPI in 1X PBS (0.85 μg/ml) and then washed three times in 1X PBS. The coverslips for STED experiments were directly mounted with 90% glycerol + 5% (w/v) n-propyl gallate non-hardening medium and sealed with a transparent nail polish.

Summary of antibodies used in super-resolution microscopy experiments:

	Primary antibody			Secondary antibody **STED**			Secondary antibody **3D-SIM**		
Factor	**Vendor**	**Cat. No.**	**Dil.**	**Name**	**Species specificity**	**Dil.**	**Name**	**Species specificity**	**Dil.**
long SATB1 isoform	Davids Biotechnologie	n/a	1:500	Abberior Star Red or Alexa Fluor 555	goat anti-rabbit	1:100 or 1:500	Alexa Fluor 488	goat anti-rabbit or donkey anti-rabbit	1:500
all SATB1 isoforms	Davids Biotechnologie	n/a	1:500	Abberior Star Red	goat anti-rabbit	1:100	Alexa Fluor 488	goat anti-rabbit	1:500
all SATB1 isoforms	Santa Cruz Biotechnology	sc-5990	1:200	–	–	–	Alexa Fluor 488	donkey anti-goat	1:500
5-fluorouridine	Sigma Aldrich	B2531	1:250	Abberior Star 580	goat anti-mouse	1:100	Alexa Fluor 568	goat anti-mouse	1:500

The STED experiment images were acquired using an inverted microscope DMi8 with laser scanning confocal head Leica TCS SP8, equipped with STED 3X module, 100x/1.4 oil immersion objective and 660 nm continual and 775 pulse depletion lasers. Depletion lasers were set to 60% and typically only a single representative z-stack from the center of each cell was scanned to reduce the effect of photobleaching and to maximize the lateral resolution, with line accumulation set to 8 and speed of 400 Hz. Raw images were deconvoluted using the Huygens Professional imaging software. The final images were analyzed using custom-made macros in Fiji software[81]. Cells from the representative middle z-stack were manually selected for the analysis and the integrated density was calculated by the measure tool of Fiji. The pixel-based colocalization analysis of selected cells was performed using the Coloc2 plugin. The presented images have artificially increased contrast and brightness for better visualization.

The 3D-SIM experiment images were acquired using the DeltaVision OMX V4 imaging platform equipped with PlanApo N 60x/1.42 oil immersion objective, pco.edge 5.5 sCMOS cameras and 405, 445, 488, 514, 568 and 642 nm lasers. Spherical aberration was minimized using immersion oil with RI 1.516 for sample acquisition. Images were acquired over the majority of the cell volume in z-dimension with 15 raw images per plane (five phases, three angles), providing ~115-135 nm lateral and ~340-350 nm axial resolution for 488/568 nm lasers, respectively. Image reconstruction and deconvolution was performed by the SoftWoRx software using channel-specific OTFs with Wiener filter set to 0.01 for DAPI and 0.001 for the other channels. For data analysis, mostly a single representative z-stack from the center of the cell was manually selected to minimize any effect of artefacts. The final images were analyzed using custom-made macros in Fiji software[81]. Cells were manually selected for the analysis and then the exact nuclear borders were drawn based on the DAPI signal. This was used to calculate the

integrated density by the measure tool of Fiji as well as to perform the pixel-based colocalization analysis using the Coloc2 plugin. The DAPI zonation was achieved using the IsoData Classifier in combination with the machine learning Trainable Weka Segmentation plugin. The selection of the resulting zones was overlaid with SATB1 speckles, identified as points using Find Maxima function in Fiji with prominence set to 2000 or 5000 depending on the biological replicate. For each cell, the number of points in each zone was converted to its relative proportion in respect to the total number of speckles in the cell. The presented images have artificially increased contrast and brightness for better visualization.

Nuclear matrix preparation

Freshly isolated thymocytes were attached to poly-D-lysine coated coverslips. Soluble proteins were removed by extraction in cytoskeletal buffer (CSK: 10 mM Pipes pH 6.8, 300 mM sucrose, 100 mM NaCl, 3 mM $MgCl_2$, 1 mM EGTA, 2 mM VRC) supplemented with 0.5% Triton X-100 for 3.5 minutes at 4°C. The first group of cells was cross-linked with 4% PFA in 1X PBS for 10 minutes at 4°C, followed by three washes in 70% Ethanol. The other groups were subjected to DNase treatment in 300 µl DNase Buffer (10 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$) combined with DNase I (10 U = 5 µl; NEB, M0303L) at 37°C for 40 minutes. The last group was subjected to RNase A treatment (10 µl = 100 µg; Qiagen, 1007885) in PBS at 37°C for 40 minutes. After the enzymatic treatment, nucleic acid fragments were extracted by a 5 minute treatment on ice with CSKpremix (without Triton X-100) supplemented with 250 mM Ammonium Sulfate, followed by one 1X PBS wash and then they were fixed as previously described. Alternatively, they were further extracted with CSKpremix with 2M NaCl, followed by one 1X PBS wash and then fixed. After fixation, cells were washed twice with 1X PBS. Samples were blocked for 30 minutes with Blocking Buffer as previously described and stained

with an anti-SATB1 antibody targeting all isoforms (Santa Cruz Biotechnology, sc-5990) and a donkey anti-goat Alexa Fluor 546 secondary antibody. Coverslips were mounted with hardening medium ProLong Gold with DAPI (Invitrogen, P36935). Images were taken using the inverted microscope DMI6000 CS with laser scanning confocal head Leica TCS SP8, equipped with 63x/1.40 oil immersion objective. Images were scanned at 600 Hz speed, unidirectionally with frame average set to 3. In the experiment with double DNase+RNase treatment, no cells were detectable using fluorescence microscopy, hence the remaining cellular shells were identified in bright field. The final images were analyzed using the Fiji software[81]. The presented images have artificially increased contrast and brightness for better visualization.

Raman spectroscopy

Isolated thymocytes were washed once with 1X PBS and then incubated in 10% 1,6-hexanediol (Sigma Aldrich, 240117) in 1X PBS solution (Sigma Aldrich, P4417) for 10 minutes at room temperature, side by side with a non-treated control. Samples were fixed with 1% formaldehyde (final conc.; Pierce, 28908) for 10 minutes and then quenched with glycine to a 0.125 M final concentration and incubated at room temperature for 5 minutes, while rocking. Cells were washed five times with 1X PBS. For measurements, a drop of cells was loaded on a stainless-steel microscopic slide (Renishaw, UK) and imaged through an Olympus 60x water immersion objective lens with a numerical aperture (NA) of 1.2 and a working distance of 280 μm (UPLSAPO60XW/1.2, Olympus, Tokyo, Japan). Raman measurements were made by a modified Raman Spectrometer (LabRAM HR, HORIBA Scientific, Lille, France). The excitation laser line used had a central wavelength at 532 nm and the resulting laser power on the sample was ~10 mW. The laser spot spatial diameter was approximately 0.6 μm, with an axial length of about 0.5 μm. A grating of 600 groves was used that resulted in a Raman spectral resolution of

around 2 cm^{-1}. A temperature-controlled stage (PE120-XY, Linkam) was coupled with the microscope stage that kept cells temperature constant. Spectral calibration was performed with a SiO_2 reference sample, presenting a single peak at 520.7 cm^{-1}. Raman signal was acquired in the spectral range between 300 and 3,150 cm^{-1}. Acquisition time was set to 10 seconds with a spectral accumulation of 3 spectra/point. All Raman spectra acquired followed this processing methodology: a) cosmic rays were removed by an internal function (Despike, LabSpec LS6, Horriba); b) background signal was subtracted from the raw Raman spectral data using a polynomial function; c) all spectra were normalized in intensity by an internal function (Unit Vector, LabSpec LS6, Horriba). PCA was performed using the Raman PCA Tool[82], publicly available through (https://sourceforge.net/projects/ramantoolset/).

Construction of CRY2-mCherry SATB1 recombinant proteins

For all CRY2 constructs, the pCMV-CRY2-mCherry vector was used (a gift from Won_Do Heo; Addgene plasmid # 58368). As a positive control, the N-terminal part of FUS protein was used from the plasmid pHR-FUS_N-mChr-CRY2WT (a gift from Clifford Brangwynne; Addgene plasmid # 101223). The N-terminal part of SATB1 was cloned into the pCMV-CRY2-mCherry vector using a single BspEI restriction enzyme site from the full-length construct. The IDR part of all *Satb1* isoforms (incl. the S635A mutated version) and also the N-terminal part of *FUS* was PCR amplified from plasmids, generating new restriction enzyme sites for XhoI and BamHI, using the following primers: for *Satb1* fwd: GTCTACTCGAGCACGAAAGGAAGAGGACCCC, rev: GATTGGATCCTTACACGGAAATTTTGGTTCGTG, for *FUS* fwd: GTCTACTCGAGCAATGGCCTCAAACGATTATACC, rev: GATTGGATCCTTATCCACGGTCCTGCTGTCCATAG. The sequence encoding the SV40 NLS

signal peptide PKKKRKV was isolated from a lab's plasmid using BspEI and SalI enzymes and inserted into CRY2-mCherry-IDR-SATB1 constructs between BspEI and XhoI sites, embedding it between mCherry and the IDR part of SATB1. Bacteria were grown in Luria-Bertani Broth at 37°C and plasmid DNA was isolated using the Nucleobond Xtra midi kit (MACHEREY-NAGEL, 740410), according to the manufacturer's instructions. PCR reactions were performed utilizing the Phusion High-Fidelity PCR Master Mix (NEB, M0531L) to prevent any mutations. For all restriction reactions the respective NEB reagents were used according to the manufacturer's instructions. Ligation reactions were performed using T4 DNA Ligase (Minotech, 202-2) according to the manufacturer's instructions. Ligation products were transformed into DH5a competent cells via 42°C heat shock for 45 seconds. All constructs were sequence-verified using Sanger sequencing in Macrogen Europe.

Tissue cultures and transient transfection

Cells were cultured in the complete growth medium DMEM (Gibco, 11995-073) supplemented with 10% Fetal Bovine Serum (Sigma Aldrich, F4135) and with 2 mM L-glutamine (Gibco, 25030) in a 5% CO_2 incubator at 37°C, until 70% confluency. Before the experiment, cells were washed in Dulbecco's PBS (DPBS; Gibco, 14200067) without Mg^{2+} and Ca^{2+}. Cells were treated for 3-5 minutes at 37°C in 1X trypsin solution until detached. Trypsin was quenched by complete growth medium and cells collected. For optoDroplet experiments, 800 ng of plasmid DNA was combined with 50 µl of Opti-MEM I Reduced Serum Medium and pipetted onto a glass bottom dish (MatTek, P35GC-0-14-C). Next, 2 µg of Lipofectamine 2000 (Invitrogen, 11668-027) was carefully mixed with 50 µl Opti-MEM I Reduced Serum Medium and it was incubated for 5 minutes at room temperature. Lipofectamine/OptiMEM mix was slowly and drop-wise added to the DNA/OptiMEM mixture on the glass bottom dish and incubated for 30 minutes at room

temperature. 50,000 of NIH/3T3 ATCC CRL-1658 cells were added to the DNA/Lipofectamine/OptiMEM mixture and they were incubated at 37°C in a humidified 5% CO_2 incubator for 6 hours. After six hours, 2 ml of the growth medium with antibiotics (penicillin/streptomycin to the final concentration 100 I.U./ml / 100 μg/ml) was added and continued transfection until ready to assay (24-48 h post transfection).

Live cell microscopy

Live cell imaging was performed in a live imaging box at 37°C with 5% CO_2, using the inverted microscope DMI6000 CS with laser scanning confocal head Leica TCS SP8, equipped with a 63x/1.40 oil immersion objective. To ensure the minimal pre-activation effect, cells were located with dim bright field illumination. In all cases presented, only a single plane in the center of the cell was scanned. CRY2 activation was achieved by 488 nm Argon laser illumination every two seconds, complemented with DPSS 561 nm laser illumination for mCherry detection. During the inactivation period, only the 561 nm laser was used to detect the mCherry signal. The scanning parameters were: PinholeAiry 1.5 AU, frame average 2 and speed to 600 Hz, scanned unidirectionally. Images were analyzed using custom-made macros in Fiji software[81] and data were processed in Bash and R. Briefly, images were corrected for movement using the StackReg plugin[83]. After basic pre-processing, droplets were identified in each cell for each time point using the Analyze Particles function of Fiji. The area of droplets was used to determine the droplet size. The number of droplets was normalized to the size of a cell. The images presented have artificially increased contrast and brightness for better visualization. For the FRAP experiments, the scanning speed was set to 1,000 Hz, bidirectionally (0.54 s / scan) and every time a selected point was photobleached for 300 ms. Bleach point mean values were background subtracted and corrected for fluorescence loss using the intensity values from the entire cell. The

data were then normalized to mean pre-bleach intensity and fitted with exponential recovery curve in Fiji or in frapplot package in R.

Structure of SATB1 and the use of predictors

SATB1 domains used in Fig. 1a were previously described[20]. The prediction of IDR regions in Fig. 4a is based on PONDR VL3 score (http://www.pondr.com/) and it is shown only for the long SATB1 isoform. Similarly, the sequence of the long SATB1 isoform was analyzed using the hidden-Markov model algorithm PLAAC to predict prion-like domains[84], using background frequencies from *S. cerevisae*. For better lucidity, the background probabilities are not shown.

Breast cancer data processing

Breast invasive carcinoma cohort data (BRCA) from the TCGA were obtained from the FireBrowse repository of the Broad Institute of MIT & Harvard (http://firebrowse.org). For the correlation experiment between chromatin accessibility and RNA expression, the ATAC-seq bigwig files were downloaded[50] and signal from a region encompassing the extra exon of the long *SATB1* isoform (chr3:18351000-18351700) was extracted using multiBigwigSummary from the deeptools package[85]. The signal was correlated with summarized values of the two short isoforms (uc003cbh.2 and uc003cbj.2) and separately with the long isoform (uc003cbi.2) RNA-seq values (level 3, isoform-specific normalized RSEM dataset, version 2016_01_28). Patient TCGA-A2-A0ET-01A was removed as an outlier. The same RNA-seq values were also used for correlation with the metadata on the pathological stage of cancer patients. Next, we analyzed the recurrent copy number alterations using the R package GAIA[86], using the SNP6 level 3 segmented sCNA minus germline CNV dataset. The output file with tumor samples was further processed by the GenomicRanges package[87] and annotated by the findOverlaps function.

Survival plots were constructed using the survival package in R[88], utilizing the normalized RNA-seq data and clinical metadata from TCGA. RNA-seq data were transformed into z-scores as follows: z = [(value gene X in tumor Y) – (mean gene X in all tumors with CNV segment mean value <|0.2|)] / (standard deviation X in all tumors with CNV segment mean value <|0.2|). Patients were divided into groups with low and high estrogen receptor levels. Moreover, patients were divided based on the relative expression levels of each tested isoform into UPregulated, DOWNregulated and unchanged groups, using the z-score threshold 1.65 (~ P = 0.0495). Survival plots for individual *SATB1* isoforms were plotted. Short isoform uc003cbh.2 and long isoform uc003cbi.2 were used for visualization. Different z-score thresholds were also tested. We reasoned that normalization to the normal tissue samples would bias the results due to possible infiltration of T cells (predominantly expressing *SATB1*).

Acknowledgements

We would like to thank Despina Tsoukatou, Manouela Kapsetaki and Joseph Papamatheakis for fruitful discussions. This work was supported by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF) Research Funding Program ARISTEIA [MIRACLE 42], by FONDATION SANTE (X-COAT) and by Chromatin3D-H2020-MSCA-ITN (GA642934). Author V.M.P acknowledges the support of Stavros Niarchos Foundation within the framework of the project ARCHERS ("Advancing Young Researchers' Human Capital in Cutting Edge Technologies in the field of Systems Biology Approaches and Personal Genomics for Health and Disease Treatment"). We acknowledge the Light Microscopy Core Facility, IMG CAS, Prague, Czech Republic, supported by MEYS (LM2018129, CZ.02.1.01/0.0/0.0/18_046/0016045) and RVO: 68378050-KAV-NPUI,

for their support with the super-resolution imaging and image analysis presented herein. We thank the personnel of the Proteomics Facility at the Institute of Molecular Biology and Biotechnology, M. Aivaliotis and N. Kountourakis, for the Mass spectrometry analysis of tryptic peptides by nanoflow liquid chromatography with tandem mass spectrometry. Moreover, the results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
2. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
3. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
4. Gotor, N. L. *et al.* RNA-binding and prion domains: the Yin and Yang of phase separation. *Nucleic Acids Res.* **48**, 9491–9504 (2020).
5. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688-699.e16 (2018).
6. Pearce, M. M. P. & Kopito, R. R. Prion-like characteristics of polyglutamine-containing proteins. *Cold Spring Harb. Perspect. Med.* **8**, (2018).
7. Langdon, E. M. *et al.* mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).
8. Peskett, T. R. *et al.* A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation. *Mol. Cell* **70**, 588-601.e6 (2018).
9. Zhang, H. *et al.* RNA controls polyQ protein phase transitions. *Mol. Cell* **60**, 220–230 (2015).

10. Zhang, S., Hinde, E., Parkyn Schneider, M., Jans, D. A. & Bogoyevitch, M. A. Nuclear bodies formed by polyQ-ataxin-1 protein are liquid RNA/protein droplets with tunable dynamics. *Sci. Rep.* **10**, 1–18 (2020).
11. Williams, A. J. & Paulson, H. L. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* **31**, 521–528 (2008).
12. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* **294**, 7128–7136 (2019).
13. Bratek-Skicki, A., Pancsa, R., Meszaros, B., Lindt, J. V. & Tompa, P. A guide to regulation of the formation of biomolecular condensates. *FEBS J.* **287**, 1924–1935 (2020).
14. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
15. Han, X. *et al.* Roles of the BRD4 short isoform in phase separation and active gene transcription. *Nat. Struct. Mol. Biol.* **27**, 333–341 (2020).
16. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* **8**, 275 (2017).
17. Boyko, S., Surewicz, K. & Surewicz, W. K. Regulatory mechanisms of tau protein fibrillation under the conditions of liquid–liquid phase separation. *Proc. Natl. Acad. Sci.* **117**, 31882–31890 (2020).
18. Batlle, C. *et al.* hnRNPDL phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation. *Cell Rep.* **30**, 1117-1128.e5 (2020).
19. Zelenka, T. *et al.* The 3D enhancer network of the developing T cell genome is controlled by SATB1. *bioRxiv* 2021.07.09.451769 (2021) doi:10.1101/2021.07.09.451769.
20. Zelenka, T. & Spilianakis, C. SATB1-mediated chromatin landscape in T cells. *Nucleus* **11**, 117–131 (2020).
21. Froemberg, A., Engeland, K. & Aigner, A. The Special AT-rich Sequence Binding Protein 1 (SATB1) and its role in solid tumors. *Cancer Lett.* **417**, 96–111 (2018).
22. Naik, R. & Galande, S. SATB family chromatin organizers as master regulators of tumor progression. *Oncogene* **38**, 1989–2004 (2019).

23. Sunkara, K. P., Gupta, G., Hansbro, P. M., Dua, K. & Bebawy, M. Functional relevance of SATB1 in immune regulation and tumorigenesis. *Biomed. Pharmacother.* **104**, 87–93 (2018).

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. Yasui, D., Miyano, M., Cai, S. T., Varga-WEisz, P. & Kohwi-Shigematsu, T. SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419**, 641–645 (2002).

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

27. Galande, S., Dickinson, L. A., Mian, I. S., Sikorska, M. & Kohwi-Shigematsu, T. SATB1 cleavage by caspase 6 disrupts PDZ domain-mediated dimerization, causing detachment from chromatin early in T-cell apoptosis. *Mol. Cell. Biol.* **21**, 5591–5604 (2001).

28. Wang, Z. *et al.* The structural basis for the oligomerization of the N-terminal domain of SATB1. *Nucleic Acids Res.* **40**, 4193–4202 (2012).

29. Wang, Z. *et al.* Crystal structure of the ubiquitin-like domain-CUT repeat-like tandem of special AT-rich sequence binding protein 1 (SATB1) reveals a coordinating DNA-binding mechanism. *J. Biol. Chem.* **289**, 27376–27385 (2014).

30. Notani, D. *et al.* N-terminal PDZ-like domain of chromatin organizer SATB1 contributes towards its function as transcription regulator. *J. Biosci.* **36**, 461–469 (2011).

31. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

32. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).

33. Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* **3**, e201702000010 (2017).

34. Kobayashi-Kirschvink, K. J. *et al.* Linear regression links transcriptomic data and cellular Raman spectra. *Cell Syst.* **7**, 104-117.e4 (2018).

35. Shin, Y. *et al.* Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell* **168**, 159-171.e14 (2017).

36. Bolognesi, B. *et al.* A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* **16**, 222–231 (2016).

37. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).

38. Agostini, F., Cirillo, D., Bolognesi, B. & Tartaglia, G. G. X-inactivation: quantitative predictions of protein interactions in the Xist network. *Nucleic Acids Res.* **41**, e31 (2013).

39. 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).

40. Ding, X. *et al.* Amyloid-forming segment induces aggregation of FUS-LC domain from phase separation modulated by site-specific phosphorylation. *J. Mol. Biol.* **432**, 467–483 (2020).

41. Monahan, Z. *et al.* Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* **36**, 2951–2967 (2017).

42. Rai, A. K., Chen, J.-X., Selbach, M. & Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018).

43. Shattuck, J. E., Paul, K. R., Cascarina, S. M. & Ross, E. D. The prion-like protein kinase Sky1 is required for efficient stress granule disassembly. *Nat. Commun.* **10**, 3614 (2019).

44. Wippich, F. *et al.* Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **152**, 791–805 (2013).

45. Khare, S. P. *et al.* NF-κB signaling and IL-4 signaling regulate SATB1 expression via alternative promoter usage during Th2 differentiation. *Front. Immunol.* **10**, (2019).

46. Patta, I. *et al.* Dynamic regulation of chromatin organizer SATB1 via TCR-induced alternative promoter switch during T-cell development. *Nucleic Acids Res.* **48**, 5873–5890 (2020).

47. Schor, I. E., Rascovan, N., Pelisch, F., Alló, M. & Kornblihtt, A. R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci.* **106**, 4325–4330 (2009).

48. Zhou, H.-L. *et al.* Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci.* **108**, E627–E635 (2011).

49. Han, H.-J., Russo, J., Kohwi, Y. & Kohwi-Shigematsu, T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* **452**, 187–193 (2008).

50. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).

51. A, P. & Weber, S. C. Evidence for and against liquid-liquid phase separation in the nucleus. *Non-Coding RNA* **5**, 50 (2019).

52. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).

53. Mir, M., Bickmore, W., Furlong, E. E. M. & Narlikar, G. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**, (2019).

54. Erdel, F. & Rippe, K. Formation of chromatin subcompartments by phase separation. *Biophys. J.* **114**, 2262–2270 (2018).

55. McSwiggen, D. T. *et al.* Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* **8**, e47098 (2019).

56. Guo, L. *et al.* Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell* **173**, 677-692.e20 (2018).

57. Yoshizawa, T. *et al.* Nuclear import receptor inhibits phase separation of FUS through binding to multiple sites. *Cell* **173**, 693-705.e22 (2018).

58. Nakayama, Y., Mian, I. S., Kohwi-Shigematsu, T. & Ogawa, T. A nuclear targeting determinant for SATB1, a genome organizer in the T cell lineage. *Cell Cycle* **4**, 1099–1106 (2005).

59. Kumar, P. P. *et al.* Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* **9**, 45–56 (2007).

60. Banani, S. F. *et al.* Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).

61. Tan, J.-A. T., Song, J., Chen, Y. & Durrin, L. K. Phosphorylation-dependent interaction of SATB1 and PIAS1 directs SUMO-regulated caspase cleavage of SATB1. *Mol. Cell. Biol.* **30**, 2823–2836 (2010).

62. Cai, D. *et al.* Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. *Nat. Cell Biol.* **21**, 1578–1589 (2019).

63. Lu, Y. *et al.* Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression. *Nat. Cell Biol.* **22**, 453–464 (2020).

64. Zamudio, A. V. *et al.* Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* **76**, 753-766.e6 (2019).

65. Gruber, A. *et al.* Molecular and structural architecture of polyQ aggregates in yeast. *Proc. Natl. Acad. Sci.* **115**, E3446–E3453 (2018).

66. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

67. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

68. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

70. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

71. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H. & Blencowe, B. J. Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol. Cell* **72**, 187-200.e6 (2018).

72. Stratigi, K. *et al.* Spatial proximity of homologous alleles and long noncoding RNAs regulate a switch in allelic gene expression. *Proc. Natl. Acad. Sci.* **112**, E1577–E1586 (2015).

73. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

74. Hagberg, A., Swart, P. & S Chult, D. *Exploring network structure, dynamics, and function using networkx*. https://www.osti.gov/biblio/960616-exploring-network-structure-dynamics-function-using-networkx (2008).

75. Grover, A. & Leskovec, J. node2vec: scalable feature learning for networks. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 855–864 (ACM, 2016). doi:10.1145/2939672.2939754.

76. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

77. Yaneva, M. & Tempst, P. Affinity capture of specific DNA-binding proteins for mass spectrometric identification. *Anal. Chem.* **75**, 6437–6448 (2003).

78. Yaneva, M. & Tempst, P. Isolation and mass spectrometry of specific DNA binding proteins. *Methods Mol. Biol. Clifton NJ* **338**, 291–303 (2006).

79. 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).

80. Williams, A. *et al.* Hypersensitive site 6 of the Th2 locus control region is essential for Th2 cytokine expression. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6955–6960 (2013).

81. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

82. Candeloro, P. *et al.* Raman database of amino acids solutions: a critical study of Extended Multiplicative Signal Correction. *Analyst* **138**, 7331–7340 (2013).

83. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **7**, 27–41 (1998).

84. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).

85. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).

86. Morganella, S., Pagnotta, S. M. & Ceccarelli, M. Finding recurrent copy number alterations preserving within-sample homogeneity. *Bioinforma. Oxf. Engl.* **27**, 2949–2956 (2011).

87. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

88. Therneau, T. M. & Grambsch, P. M. *Modeling survival data: extending the Cox model.* (Springer, 2000).

89. Costes, S. V. *et al.* Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.* **86**, 3993–4003 (2004).

Figure legends

Fig. 1 | Identification of two SATB1 protein isoforms in murine thymocytes.

a, SATB1 protein consists of the following domains and structural features: ULD – ubiquitin-like domain, CUTL – CUT-like domain, CUT1 and CUT2 domains, EP – the peptide encoded by the predicted extra exon of the long *Satb1* isoform, Q – compositional bias represented by a poly-Q domain and a stretch of prolines, HD – homeodomain. In this study, we used two custom-made antibodies (Davids Biotechnology): an antibody targeting the first 330 amino acids of all isoforms (cannot discriminate between the two protein isoforms) and the long isoform antibody specifically targeting the extra peptide of the long isoform. **b**, Design of the RT-qPCR experiment to quantitate relative expression of the short and long *Satb1* isoforms in murine thymocytes. **c**, RT-qPCR results for the relative mRNA levels of the two *Satb1* isoforms. Values from different replicates were normalized to *Dnmt1* mRNA levels. Data are the mean ± s.d. **d**, Comprehensive list of quantified *Satb1* isoforms based on stranded-total-RNA-seq experiment in thymocytes. Data are the mean ± s.d. **e**, Western blot analysis for the detection of the long SATB1 isoform (custom antibodies, Davids Biotechnology) and all SATB1 isoforms (Santa Cruz, sc-5990) in murine thymocytes. The specificity of SATB1 antibodies was validated by the comparison with *Satb1*$^{fl/fl}$*Cd4*-Cre^{+} (SKO) murine thymocytes. The green arrow indicates the expected SATB1 band (~100 kD). The 80 kD band could represent SATB1 lacking post-translational modifications. **f**, Confocal and STED microscopy images indicating the subnuclear SATB1 cage-like pattern. The super-resolution microscopy unveils that the cage-like pattern is actually composed of individual, round 40-60 nm large droplets.

Fig. 2 | SATB1 is localized in the active nuclear zone and sites of active transcription.

a, 3D-SIM immunofluorescence experiment utilizing two different SATB1 antibodies to visualize the nuclear localization pattern of the long isoform compared to all SATB1 isoforms. Scale bar 2 µm. **b**, Quantification of the 3D-SIM immunofluorescence images. Nuclei of primary murine thymocytes were categorized into four zones based on the intensity of DAPI staining and SATB1 speckles in each zone were counted. Images used represented a middle z-stack from the 3D-SIM experiments. The graph depicts the differences between long and all SATB1 isoforms' zonal localization in the nucleus. The horizontal lines inside violin plots represent the 25th, 50th and 75th percentiles. Red circles represent the mean ± s.d. *P* values by Wilcoxon rank sum test. **c**, STED microscopy indicating that SATB1 droplets co-localize with sites of active transcription (60 minutes pulse 5-fluorouridine treatment). Scale bar 1 µm. **d**, STRING network of all significantly enriched SATB1 interacting proteins clustered using the k-means method (left). The full network including protein names is available in Extended Data Fig. 2c. The blue cluster 1 and yellow cluster 2 were mostly enriched for transcription and splicing-related factors, respectively. A more detailed network of selected transcription and splicing-related factors is provided (right). **e**, Gene ontology enrichment analysis of SATB1-interacting proteins revealed factors involved in transcription and splicing. **f**, Volcano plot from stranded-total-RNA-seq experiment displaying the differentially expressed genes (FDR < 0.05) in *Satb1* cKO thymocytes.

Fig. 3 | SATB1 forms phase separated droplets *in vivo*.

a, 3D-SIM immunofluorescence microscopy on 1,6-hexanediol-treated cells. Increasing concentration of 1,6-hexanediol gradually solubilized SATB1 droplets. Scale bar 5 µm. **b**, Quantification of results in **a** showing a significant decrease of SATB1 signal, indicating its sensitivity to hexanediol. **c**, Comparison of immunofluorescence signal based on the antibody

targeting all the SATB1 isoforms (Santa Cruz Biotechnology, sc-5990) and only the long isoform (custom, Davids Biotechnology) upon hexanediol treatment, detected by STED microscopy. **d**, The left graph displays Pearson correlation coefficients (PCC) derived from pixel-based colocalization analysis between long and/or all SATB1 isoforms and fluorouridine-stained sites of active transcription, with and without 10% 1,6-hexanediol treatment. SATB1 scans rotated by 90° served as a negative control for the colocalization with FU. The right graph depicts the Costes P-values[89] derived from randomly shuffled chunks of analyzed images (100 randomizations). The grey line indicates a 0.95 level of significance. Images used for the analysis were generated by STED microscopy. Complete results of the colocalization analysis for both STED and 3D-SIM experiments, including Manders‘ coefficients are depicted in Extended Data Fig. 3b-d. **e**, Principal component analysis of Raman spectra from WT and *Satb1* cKO thymocytes with and without 10% 1,6-hexanediol treatment. In **b**, **c** and **d**, the horizontal lines inside violin plots represent the 25th, 50th and 75th percentile. Red circles represent the mean ± s.d. *P* values by Wilcoxon rank sum test.

Fig. 4 | The N-terminus of SATB1 forms phase separated droplets.

a, Graphical representation of all CRY2-mCherry constructs generated. The graph displays the IDR prediction by VL3 PONDR score in magenta and the prediction of prion-like domains by PLAAC algorithm in green. **b**, Formation of optoDroplets upon 488 nm pulse illumination in cells transiently transfected with CRY2-mCherry-N-terminus_SATB1 constructs as well as with negative (vector) and positive control constructs (part of FUS protein). Two groups of cells were identified based on the original relative expression levels of the recombinant proteins. Scale bar 5 µm. **c**, OptoDroplets evinced a round shape and the ability to coalesce. **d**, Number of optoDroplets normalized to the cell size at each time point upon activation, to visualize the speed

of droplet assembly and its dependence on the original concentration of the protein. **e**, Number of optoDroplets normalized to the cell size at each time point of inactivation (488 nm laser was off) to visualize the speed of droplet disassembly and its dependence on the original concentration of the protein. **f**, FRAP experiment utilizing either the N-terminus of SATB1 construct or the FUS positive control construct. The fitted curves were generated by the exponential recovery curve fitting function of Fiji. In **d**, **e** and **f**, the error bars represent the s.e.m.

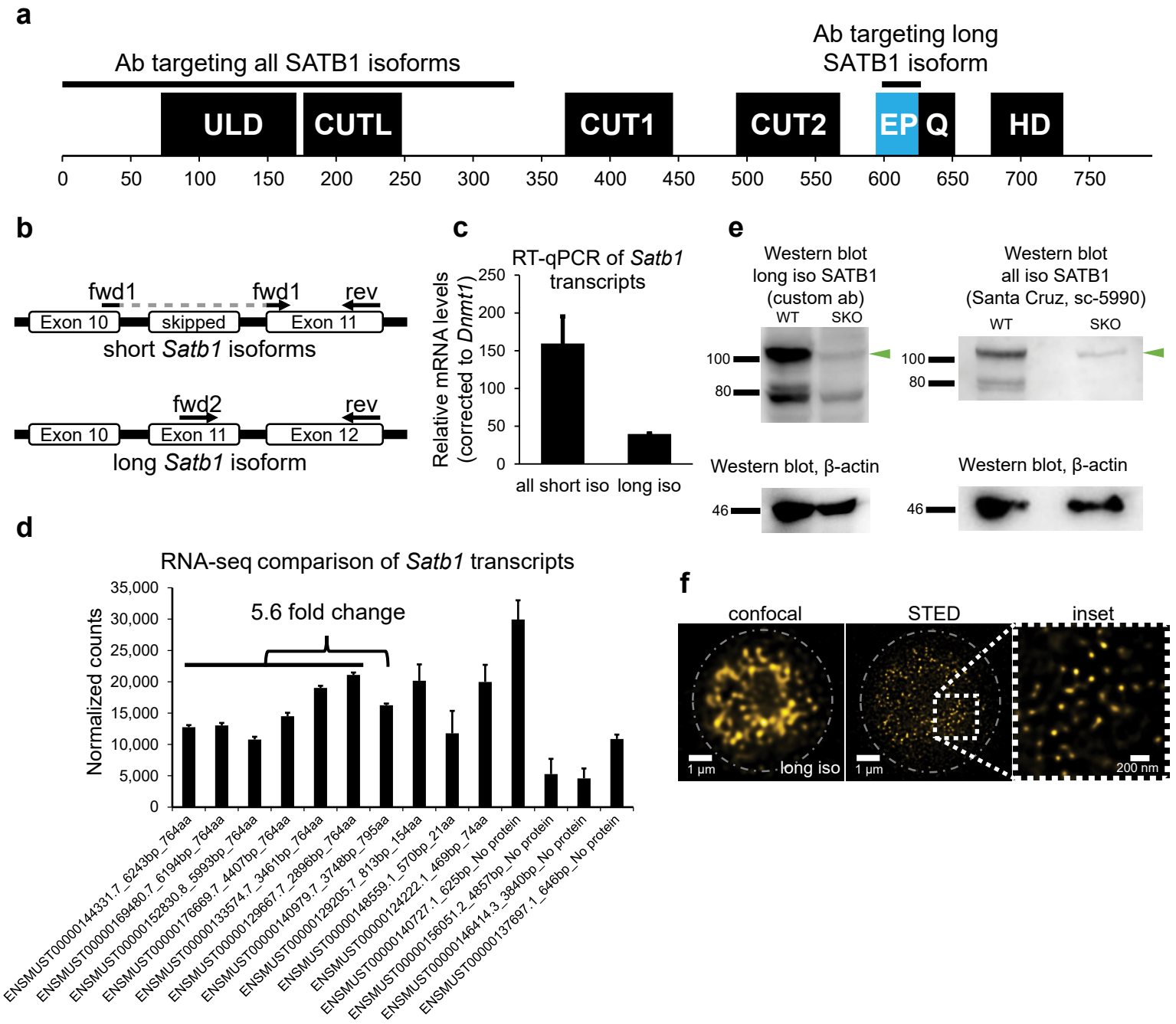
Fig. 5 | Discrete biophysical properties of SATB1 protein isoforms.

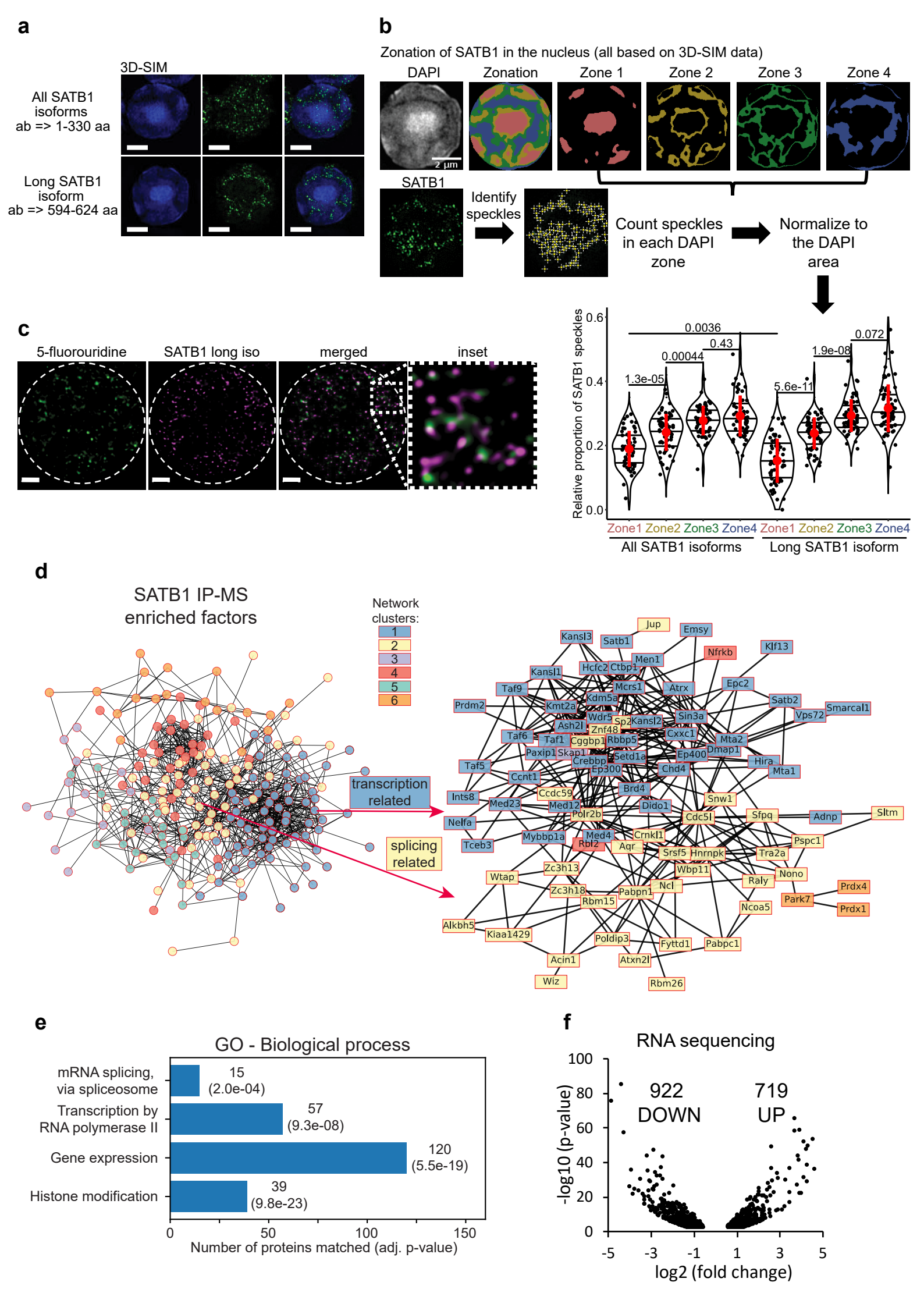
a, CRY2-mCherry constructs expressing both the short and the long full length SATB1 isoforms underwent reversible ultrastructural changes, resembling phase transitions. Scale bar 5 µm. **b**, In a fraction of cells, the full-length long SATB1 isoform was mislocalized outside the nucleus which triggered the formation of rigid aggregates. Scale bar 5 µm. **c**, SATB1 long isoform has an increased propensity to undergo phase transitions due to the presence of the extra peptide which displays compositional bias (enrichment in S, G, Q, P, E, K and R amino acids). The propensity score to form granules was predicted by the *cat*Granule algorithm. **d**, CRY2-mCherry constructs encompassing only the two SATB1 IDR regions and the poly-Q domain, but not the NLS signal, displayed much higher rates of cytoplasmic aggregation for the long SATB1 isoform compared to the short isoform. Green arrows indicate cells with long SATB1 isoform aggregates.

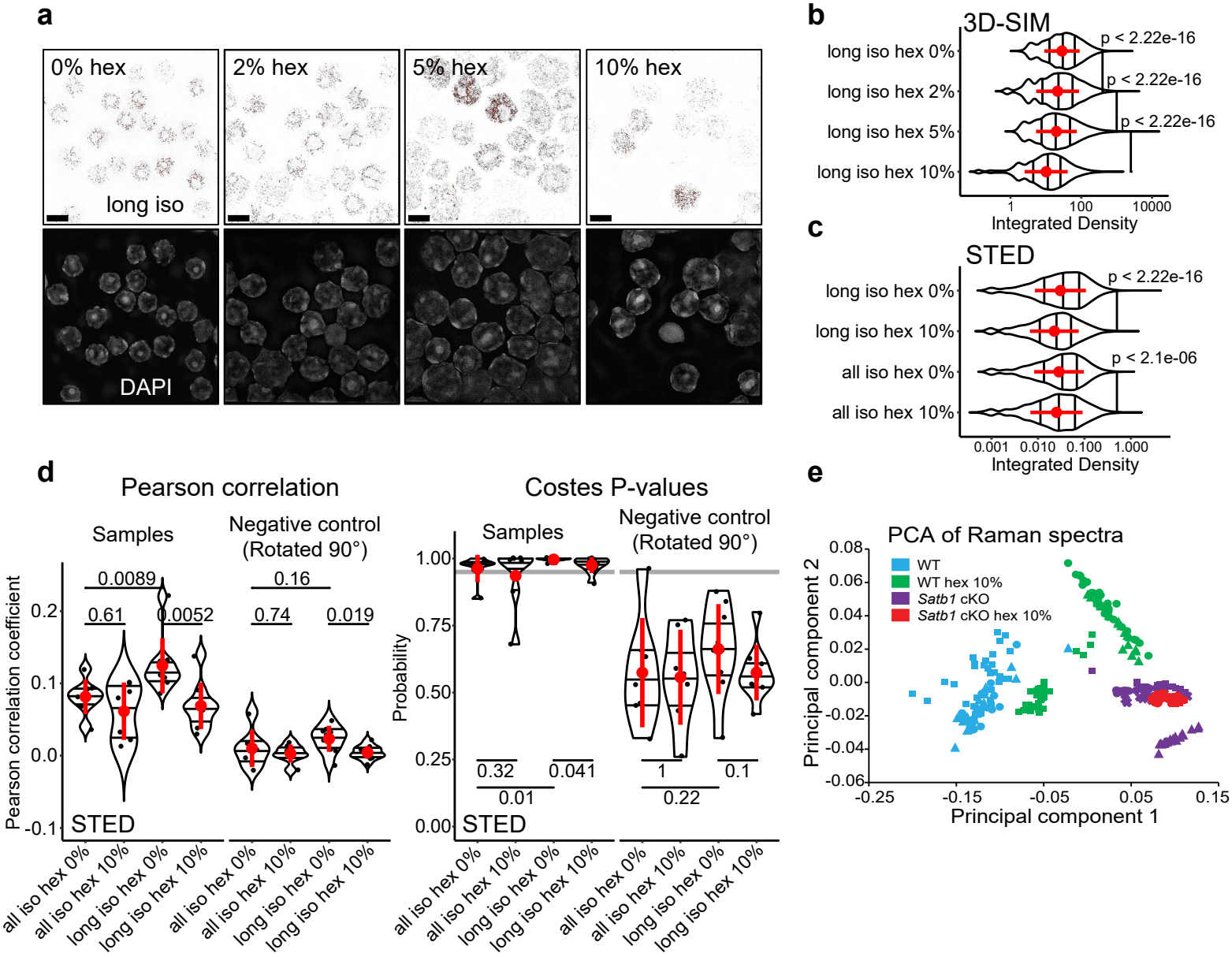
Fig. 6 | Modes of regulation of SATB1 phase transitions.

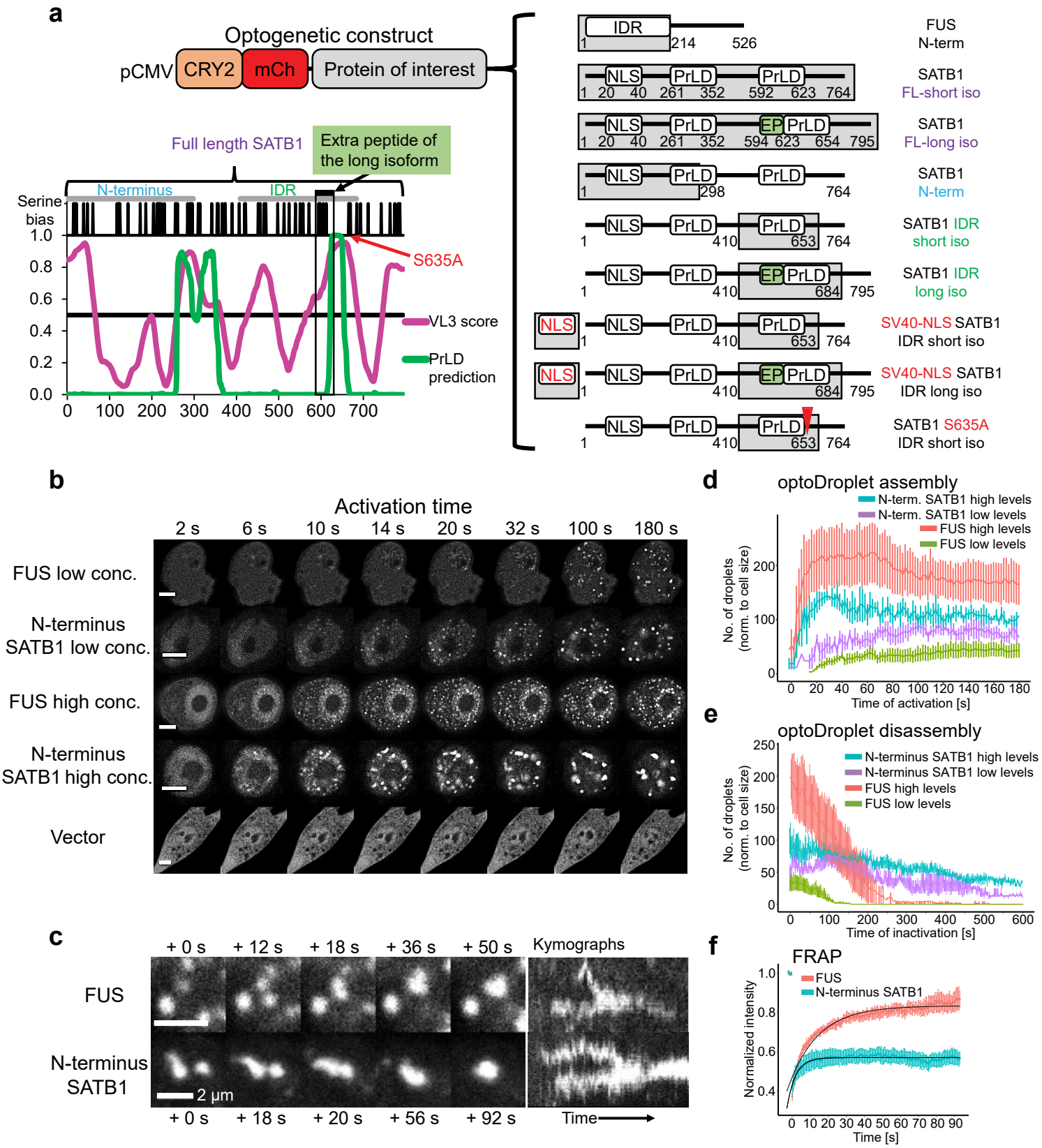
a, Nuclear matrix extraction coupled to immunofluorescence analysis revealed that a fraction of SATB1 protein remained in the cell nucleus even after DNase treatment and high-salt extraction. However, RNase A treatment almost completely depleted SATB1 from the nucleus, indicating a high level of association between SATB1 and nuclear RNA. **b**, SATB1 co-immunoprecipitation

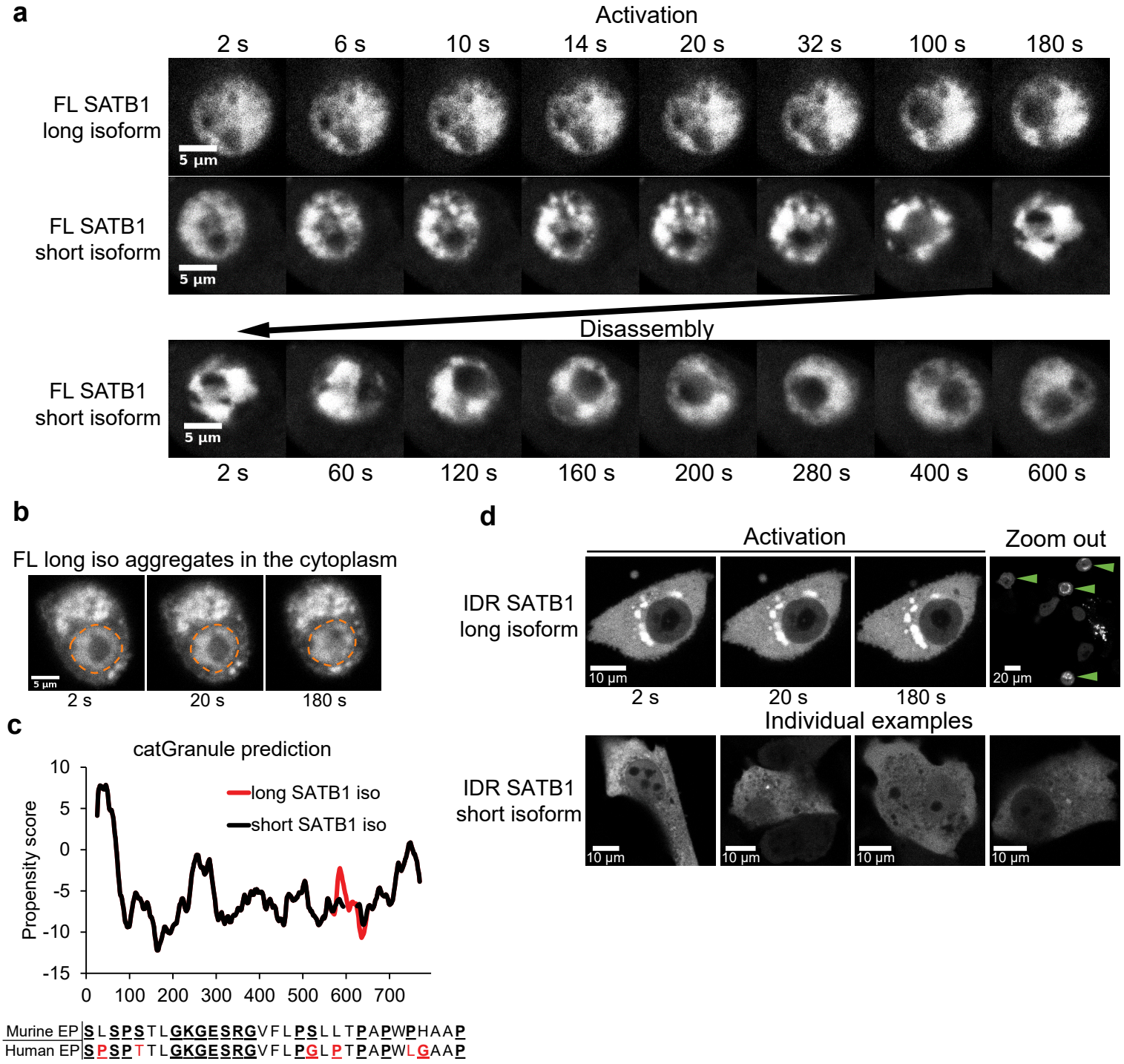
confirmed the association of SATB1 with the known nuclear lncRNAs Xist, Malat and Neat. Data are mean ± s.d. **c**, Forced nuclear localization of CRY2-mCherry constructs harboring the long SATB1 isoform IDR and the SV40 NLS displayed no protein aggregation in response to 488 nm laser even after 20 minutes of activation. Scale bar 5 µm. **d**, DNA affinity chromatography coupled to mass spectrometry using murine thymocytes protein extracts and an RHS6 biotinylated probe as a bait uncovered SATB1 S635 phosphosite. **e**, The S635A mutation, preventing phosphorylation of the short SATB1 isoform highly increased the aggregation propensity of the corresponding CRY2-mCherry-IDR construct. Green arrows indicate cells with the SATB1 aggregates. **f**, Murine *Satb1* transcripts reconstructed based on stranded-total-RNA-seq data. Both long and short *Satb1* isoforms were produced from multiple promoters. The long SATB1 isoform binding sites were retrieved from GSE173446[19]. **g**, Increased ATAC-seq signal in human breast cancer patients is positively correlated with the expression of the short *SATB1* isoform and negatively correlated with the long isoform expression. **h**, The expression of the long *SATB1* isoform is associated with higher pathological stages of human breast cancer, unlike the short isoform. The horizontal lines inside violin plots represent the 25th, 50th and 75th percentiles. Red circles represent the mean ± s.d. *P* values by Wilcoxon rank sum test.

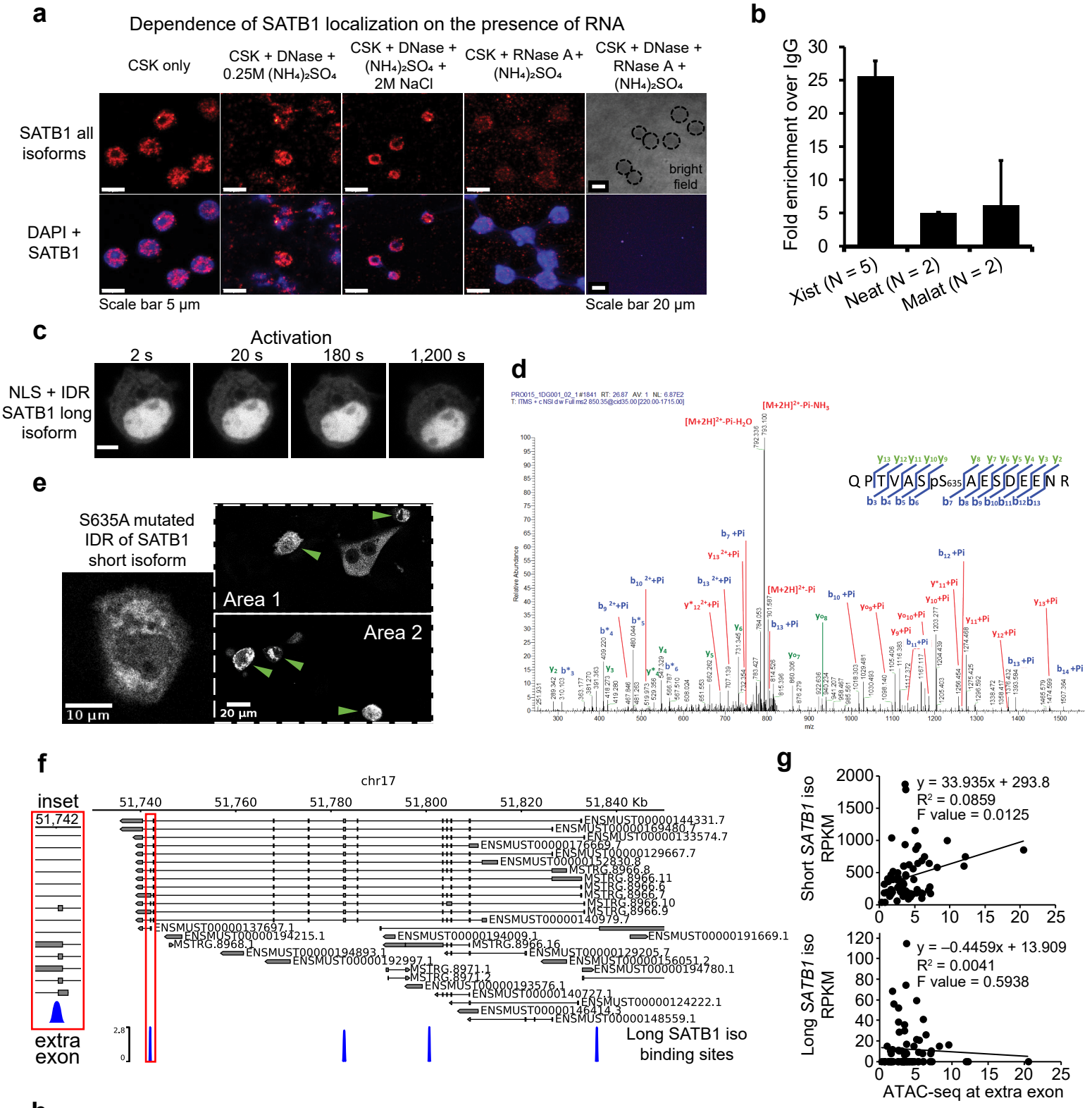

a
Ab targeting all SATB1 isoforms
Ab targeting long SATB1 isoform
ULD
CUTL
CUT1
CUT2
EP
Q
HD
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750
b
fwd1
fwd1
rev
Exon 10
skipped
Exon 11
short Satb1 isoforms
fwd2
rev
Exon 10
Exon 11
Exon 12
long Satb1 isoform
c
RT-qPCR of Satb1 transcripts
Relative mRNA levels (corrected to Dnmt1)
250 200 150 100 50 0
all short iso
long iso
e
Western blot long iso SATB1 (custom ab)
WT
SKO
100
80
Western blot all iso SATB1 (Santa Cruz, sc-5990)
WT
SKO
100
80
Western blot, β-actin
46
Western blot, β-actin
46
d
RNA-seq comparison of Satb1 transcripts
Normalized counts
35,000 30,000 25,000 20,000 15,000 10,000 5,000 0
5.6 fold change
ENSMUST00000144331.7_6243bp_764aa
ENSMUST00000169480.7_6194bp_764aa
ENSMUST00000152830.8_5993bp_764aa
ENSMUST00000176669.7_4407bp_764aa
ENSMUST00000133574.7_3461bp_764aa
ENSMUST00000129667.7_2896bp_764aa
ENSMUST00000140979.7_3748bp_795aa
ENSMUST00000129205.7_813bp_154aa
ENSMUST00000148559.1_570bp_21aa
ENSMUST00000124222.1_469bp_74aa
ENSMUST00000140727.1_625bp_No protein
ENSMUST00000156051.2_4857bp_No protein
ENSMUST00000146414.3_3840bp_No protein
ENSMUST00000137697.1_646bp_No protein
f
confocal
STED
inset
1 µm
long iso
1 µm
200 nm


a
3D-SIM
All SATB1 isoforms ab => 1-330 aa
Long SATB1 isoform ab => 594-624 aa
b
Zonation of SATB1 in the nucleus (all based on 3D-SIM data)
DAPI
Zonation
Zone 1
Zone 2
Zone 3
Zone 4
2 µm
SATB1
Identify speckles
Count speckles in each DAPI zone
Normalize to the DAPI area
Relative proportion of SATB1 speckles
0.0036
0.43
0.00044
1.3e-05
0.072
1.9e-08
5.6e-11
Zone1 Zone2 Zone3 Zone4 Zone1 Zone2 Zone3 Zone4
All SATB1 isoforms
Long SATB1 isoform
c
5-fluorouridine
SATB1 long iso
merged
inset
d
SATB1 IP-MS enriched factors
Network clusters:
1
2
3
4
5
6
transcription related
splicing related
e
GO - Biological process
mRNA splicing, via spliceosome
15 (2.0e-04)
Transcription by RNA polymerase II
57 (9.3e-08)
Gene expression
120 (5.5e-19)
Histone modification
39 (9.8e-23)
Number of proteins matched (adj. p-value)
f
RNA sequencing
922 DOWN
719 UP
-log10 (p-value)
log2 (fold change)


a
0% hex
2% hex
5% hex
10% hex
long iso
DAPI
b
3D-SIM
long iso hex 0%
long iso hex 2%
long iso hex 5%
long iso hex 10%
p < 2.22e-16
p < 2.22e-16
p < 2.22e-16
1
100
10000
Integrated Density
c
STED
long iso hex 0%
long iso hex 10%
all iso hex 0%
all iso hex 10%
p < 2.22e-16
p < 2.1e-06
0.001
0.010
0.100
1.000
Integrated Density
d
Pearson correlation
Samples
Negative control (Rotated 90°)
Pearson correlation coefficient
0.0089
0.61
0.0052
0.16
0.74
0.019
STED
Costes P-values
Probability
0.32
0.041
0.01
1
0.1
0.22
STED
all iso hex 0%
all iso hex 10%
long iso hex 0%
long iso hex 10%
e
PCA of Raman spectra
WT
WT hex 10%
Satb1 cKO
Satb1 cKO hex 10%
Principal component 2
Principal component 1


a
Optogenetic construct
pCMV
CRY2
mCh
Protein of interest
Full length SATB1
Extra peptide of the long isoform
N-terminus
IDR
Serine bias
S635A
VL3 score
PrLD prediction
IDR
NLS
PrLD
EP
FUS N-term
SATB1 FL-short iso
SATB1 FL-long iso
SATB1 N-term
SATB1 IDR short iso
SATB1 IDR long iso
SV40-NLS SATB1 IDR short iso
SV40-NLS SATB1 IDR long iso
SATB1 S635A IDR short iso
b
Activation time
2 s
6 s
10 s
14 s
20 s
32 s
100 s
180 s
FUS low conc.
N-terminus SATB1 low conc.
FUS high conc.
N-terminus SATB1 high conc.
Vector
c
+ 0 s
+ 12 s
+ 18 s
+ 36 s
+ 50 s
Kymographs
FUS
N-terminus SATB1
2 µm
+ 0 s
+ 18 s
+ 20 s
+ 56 s
+ 92 s
Time
d
optoDroplet assembly
N-term. SATB1 high levels
N-term. SATB1 low levels
FUS high levels
FUS low levels
No. of droplets (norm. to cell size)
Time of activation [s]
e
optoDroplet disassembly
N-terminus SATB1 high levels
N-terminus SATB1 low levels
FUS high levels
FUS low levels
No. of droplets (norm. to cell size)
Time of inactivation [s]
f
FRAP
FUS
N-terminus SATB1
Normalized intensity
Time [s]


a
Activation
2 s
6 s
10 s
14 s
20 s
32 s
100 s
180 s
FL SATB1 long isoform
FL SATB1 short isoform
5 µm
Disassembly
FL SATB1 short isoform
2 s
60 s
120 s
160 s
200 s
280 s
400 s
600 s
b
FL long iso aggregates in the cytoplasm
2 s
20 s
180 s
c
catGranule prediction
long SATB1 iso
short SATB1 iso
Propensity score
Murine EP SLSPSTLGKGESRGVFLPSLLTPAPWPHAAP
Human EP SPSPTTLGKGESRGVFLPGLPTPAPWLGAAP
d
Activation
Zoom out
IDR SATB1 long isoform
10 µm
20 µm
2 s
20 s
180 s
Individual examples
IDR SATB1 short isoform


a
Dependence of SATB1 localization on the presence of RNA
CSK only
CSK + DNase + 0.25M (NH₄)₂SO₄
CSK + DNase + (NH₄)₂SO₄ + 2M NaCl
CSK + RNase A + (NH₄)₂SO₄
CSK + DNase + RNase A + (NH₄)₂SO₄
SATB1 all isoforms
DAPI + SATB1
bright field
Scale bar 5 µm
Scale bar 20 µm
b
Fold enrichment over IgG
Xist (N = 5)
Neat (N = 2)
Malat (N = 2)
c
Activation
2 s
20 s
180 s
1,200 s
NLS + IDR SATB1 long isoform
d
Q P T V A S pS635 A E S D E E N R
e
S635A mutated IDR of SATB1 short isoform
Area 1
Area 2
10 µm
20 µm
f
chr17
inset
51,742
51,740
51,760
51,780
51,800
51,820
51,840 Kb
extra exon
Long SATB1 iso binding sites
g
Short SATB1 iso RPKM
y = 33.935x + 293.8
R² = 0.0859
F value = 0.0125
Long SATB1 iso RPKM
y = –0.4459x + 13.909
R² = 0.0041
F value = 0.5938
ATAC-seq at extra exon
h
Short SATB1 iso expression
Long SATB1 iso expression
0.38
0.16
0.61
0.97
0.18
0.15
0.41
0.73
t1
t2
t3
t4
NA