

DSS1 is required for proper Integrator–PP2A function

Received: 12 August 2024

Accepted: 16 June 2025

Published online: 05 July 2025

Check for updates

Congling Xu^{1,5}✉, Qian-Xing Zhou^{2,5}, Hai Zheng^{2,5}, Aixia Song^{2,5},
Wen-Ying Zhao¹, Ting-Ting Xu¹, Yan Xiong², Yi-Jie Zhang¹, Zixuan Huang^{2,3},
Yanhui Xu²✉, Jingdong Cheng³✉ & Fei Xavier Chen^{2,4}✉

Integrator–PP2A (INTAC) is a highly modular complex orchestrating the transition of paused RNA polymerase II into productive elongation or promoter-proximal premature termination, with its loss resulting in transcription dysregulation and genome instability. Here, we identify human DSS1—a flexible 70-residue protein found in multiple functionally diverse complexes including the 26S proteasome—as an integral subunit of the INTAC backbone. Structural analysis of DSS1–INTAC, both alone and in association with paused polymerase, demonstrates intimate interactions between DSS1 and the INTAC backbone. We identify tryptophan 39 of DSS1 as being critical for interacting with INTAC and find that its mutation disrupts DSS1’s interaction with INTAC, while maintaining DSS1’s interaction with the proteasome. This substitution not only impairs INTAC-dependent transcriptional regulation, but also reveals that INTAC is DSS1’s major chromatin-bound form. Together, our findings reveal a role for DSS1 in supporting the structure and regulatory functions of INTAC.

Transcription by RNA Polymerase II (Pol II) in metazoans is a complex and finely tuned process that includes a series of phases, starting with initiation, moving through promoter-proximal pausing and release, progressing to productive elongation, and culminating in termination^{1–4}. The process begins with the assembly of the pre-initiation complex at promoters, where general transcription factors recruit Pol II, leading to the melting of double-stranded DNA and synthesis of a nascent transcript^{5–7}. Pol II then enters a promoter-proximal paused state within 200 bp downstream of the transcription start sites. After pausing, Pol II can either progress to productive elongation for RNA synthesis or be terminated prematurely, making way for the subsequent transcription cycle^{8–11}.

Promoter-proximal pausing of Pol II is stabilized by the negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF), with NELF anchoring Pol II near its pausing sites and DSIF preventing premature termination and proteolysis of Pol II subunits^{12–20}. The transition into elongation is predominantly facilitated by P-TEFb composed of kinase CDK9 and a regulatory subunit cyclin T1, which phosphorylates Pol II, NELF, and DSIF among its targets^{1,2,21,22}. Conversely, the decision for premature termination of promoter-proximal Pol II is significantly influenced by the recently identified Integrator–PP2A (INTAC) complex^{23–25}.

The multisubunit INTAC complex, with its dual catalytic activities, is composed of six distinct modules. Among these modules, the backbone and shoulder modules provide a structural framework,

¹Shanghai Key Laboratory of Anesthesiology and Brain Functional Modulation, Clinical Research Center for Anesthesiology and Perioperative Medicine, Translational Research Institute of Brain and Brain-Like Intelligence, Shanghai Fourth People’s Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China. ²Fudan University Shanghai Cancer Center, Shanghai Key Laboratory of Medical Epigenetics, Shanghai Key Laboratory of Radiation Oncology, Institutes of Biomedical Sciences, Fudan University, Shanghai, China. ³Minhang Hospital & Institutes of Biomedical Sciences, Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism, Fudan University, Shanghai, China. ⁴Shanghai Academy of Natural Sciences (SANS), Fudan University, Shanghai, China. ⁵These authors contributed equally: Congling Xu, Qian-Xing Zhou, Hai Zheng, Aixia Song. ✉e-mail: conglingxu@tongji.edu.cn; xuyh@fudan.edu.cn; cheng@fudan.edu.cn; feixchen@fudan.edu.cn

enabling the positioning of the RNA endonuclease and phosphatase modules on opposite sides of this scaffold²⁵. This structural arrangement allows the decoupling of the two catalytic modules, which exert differential impacts on the fate of paused Pol II and gene expression²⁶. The SOSS module, featuring the single-stranded DNA (ssDNA)-binding protein SSB1/2, targets ssDNA within R-loops, thereby facilitating the recruitment of INTAC to R-loop-enriched genomic regions²⁷. The auxiliary/arm module of INTAC mediates the recognition of transcription factors, and it has some INTAC-independent roles in the regulation of transcription^{28–32}.

Integrator was identified through the purification of deleted in split hand/split foot 1 protein (DSS1) by Shiekhhattar and colleagues two decades ago³³. DSS1, a homolog of suppressor of exocyst mutations 1 (SEM1) from yeast, is characterized by its small size and flexibility, being intrinsically disordered on its own^{34,35}. In addition to being a potential component of Integrator and INTAC, DSS1/SEM1 also associates with several other complexes, including the 26S proteasome and TREX-2 in yeast and mammals, and additionally in mammals with BRCA2 complexes^{35–40}. Within the proteolysis pathway, DSS1/SEM1 is not only integrated into the 19S regulatory complex, contributing to the efficient assembly of the 26S proteasome, but it might also act as a ubiquitin receptor^{36,37,41,42}. The incorporation of DSS1/SEM1 into TREX-2 or BRCA2 occurs independently of its association with the 26S proteasome. In the TREX-2 complex, DSS1 directly binds and stabilizes PCID2, which facilitates the formation of a platform for mRNA binding and export from the nucleus^{38,39,43}. Moreover, as an integral and conserved partner of the BRCA2 complex, DSS1 is crucial for the conformational dynamics and stability of BRCA2, ensuring its function in homologous recombination repair of double-strand DNA breaks (DSBs)^{40,44,45}. Despite the defined functions of DSS1 in some of its complexes, the interaction with Integrator/INTAC, as well as its functions in the steps of transcriptional regulation executed by this complex, are unexplored.

In this study, we highlight the role of DSS1 as a component of the INTAC backbone module, that is important for maintaining the stability of INTAC to support its diverse functions in transcription regulation.

Results

DSS1 associates with the INTAC complex

To determine whether DSS1 genuinely constitutes a component of the Integrator or INTAC complex, we engineered a cell line with inducible expression of Flag-tagged DSS1 using the HEK293 Flp-In T-REX system, followed by Flag-affinity purification and mass spectrometry (Fig. 1a). Our analysis of the proteomics data revealed the enrichment of all INTAC subunits, including phosphatase module subunits PP2A-A and PP2A-C, the presence of which distinguishes INTAC from Integrator alone (Fig. 1b and Supplementary Data 1)⁴⁶. Further analysis also retrieved several known DSS1-containing complexes, such as the 26S proteasome, TREX-2, and BRCA2 complexes (Fig. 1c). To verify the endogenous interaction between DSS1 and INTAC, we integrated a Flag tag, along with a FKBP12^{F36V} degradation tag (dTAG) for rapid depletion, at the endogenous DSS1 locus in the human colorectal cancer cell line DLD-1 (Fig. 1d). Co-immunoprecipitation with a Flag antibody in this modified DSS1-Flag-dTAG cell line successfully co-purified all examined INTAC subunits across different modules (Fig. 1e). Reciprocally, endogenous immunoprecipitation experiments targeting INTS3 and INTS5 further substantiated their association with DSS1 (Fig. 1f). These data support a close physical association of DSS1 with INTAC.

Structure of a DSS1-containing INTAC complex

To elucidate the molecular mechanisms of the interaction between DSS1 and INTAC, we determined the structure of the reconstituted DSS1-INTAC complex using cryo-electron microscopy (cryo-EM)

single-particle analysis. The cryo-EM map was refined to an overall resolution of 4.1 Å, with the maps of each module improved to a resolution ranging from 3.7 to 4.2 Å by multibody refinement (Fig. 1g, Supplementary Fig. 1, and Supplementary Data 4). A structural model was built by fitting previously determined structures of INTAC⁴⁶ into the cryo-EM maps followed by manual adjustment (Fig. 1h, Supplementary Fig. 1, and Supplementary Data 4).

In our DSS1-INTAC structure, the backbone and shoulder modules of INTAC establish a cruciform scaffold, positioning the RNA endonuclease and phosphatase modules on opposite sides (Fig. 1g, h), which is consistent with previously described INTAC configurations^{46,47}. The SOSS and auxiliary modules of INTAC, noted for their dynamic nature^{27,48}, were not captured in our visualization. Notably, we identified an additional density within the backbone module closely interacting with INTS7. This allowed us to model the central and C-terminal segments of DSS1, spanning alanine 36 to methionine 67, into our structure (Fig. 1g, h). The association of DSS1 and INTS7 is characterized by extensive hydrophobic interactions, situating DSS1 within a groove formed by α -helices of INTS7 (Fig. 1i). Furthermore, DSS1's placement is secured by the arch-shaped structure of INTS2, which likely restricts DSS1's exposure to external proteins (Fig. 1g–i). Notably, a density corresponding to DSS1 can also be found in previously reported structural maps of INTAC and INTAC-PEC^{25,46,49}, supporting the notion that endogenous DSS1 stably associates with INTAC. These structural insights suggest that DSS1 is an integral component of the INTAC backbone module.

In addition to the DSS1-INTAC structure presented in this study, structures of DSS1 within the 26S proteasome, TREX-2, and BRCA2 complexes have been previously characterized^{35–40}. To understand the molecular mechanism underlying the multifaceted functions of DSS1, we compared its configurations across these different complexes. DSS1 presents strikingly unique structural adaptations in each context (Supplementary Fig. 2a–d). While DSS1 tends to adopt α -helical structures at different positions of its C terminus in the various contexts, its N-terminal and central regions display remarkable flexibility, allowing for diverse structural accommodations to fit the interfaces of its interacting partners (Supplementary Fig. 2a–d). This structural compliance of DSS1 could underly its ability to perform analogous scaffolding roles in otherwise unrelated complexes that are implicated in various biological processes.

The genomic distribution of DSS1 is similar to that of INTAC

We next sought to delineate the cellular distribution of DSS1. We performed cellular fractionation followed by DSS1 immunoprecipitation and mass spectrometry analysis. The results suggest that DSS1 is mainly associated with the 26S proteasome in the cytoplasm, while being more enriched in INTAC-PEC on chromatin (Supplementary Fig. 2e and Supplementary Data 2). Time-course treatment with FKBP12^{F36V} degrader, dTAG^V-I⁵⁰, resulted in a marked depletion of endogenous DSS1 in DSS1-Flag-dTAG cells (Fig. 2a). Immunofluorescence analysis revealed that endogenous DSS1 can be localized to the nucleus, and the immunofluorescence signal was reduced upon dTAG treatment, confirming a nuclear pool of DSS1 (Fig. 2b, c). We therefore asked whether DSS1 associates with chromatin-bound INTAC. Initial attempts of Flag ChIP-seq in DSS1-Flag-dTAG cells did not yield significant enrichment (data not shown), potentially due to crosslinking constraints that obscure antibody recognition by sequestering DSS1 within the hydrophobic pocket of the backbone module. Consequently, we utilized Flag CUT&Tag in DSS1-Flag-dTAG cells treated with DMSO or dTAG for 24 h. This approach identified 54,757 DSS1-bound genomic regions in DMSO-treated cells, which were significantly decreased following dTAG treatment (Fig. 2d, e). The control IgG samples showed markedly lower signal intensity, further validating the specificity of the Flag CUT&Tag

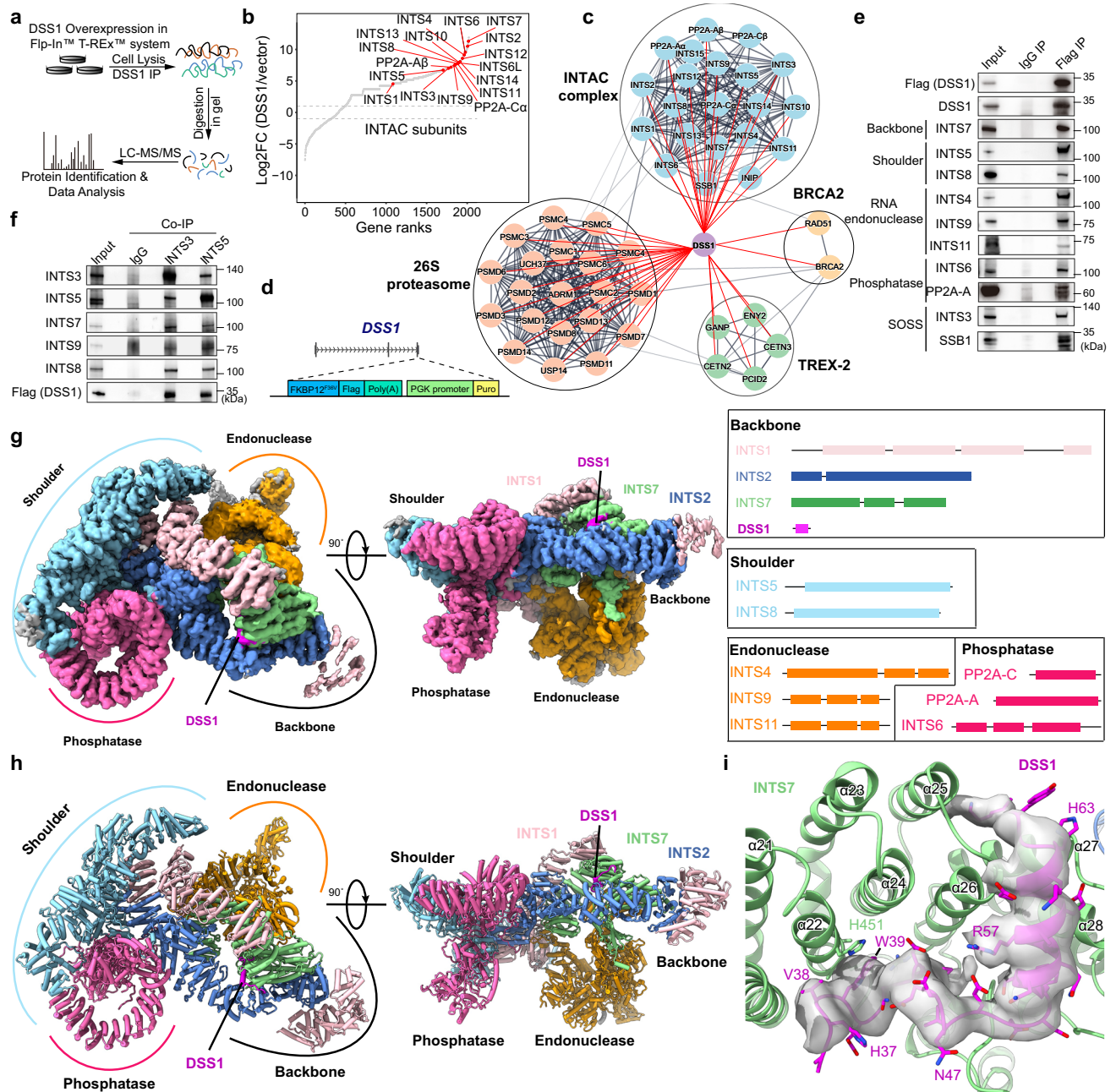


Fig. 1 | DSS1 is an integral component of the INTAC backbone module. **a** The workflow of DSS1 purification and mass spectrometry (MS) using Flp-In T-REx system in HEK293 cells. **b** Proteins identified by MS were ranked by the enrichment in purification of DSS1 versus vector. INTAC subunits are labeled. **c** STRING protein-protein interaction network of DSS1 interactome combined with the proteomics data of DSS1. Red lines indicate the interactions identified in our DSS1 MS data. **d** Schematic of the generation of human DSS1-Flag-dTAG DLD-1 cells. **e** Co-IP analysis of Flag (DSS1) followed by western blotting in DSS1-Flag-dTAG cells. Data represent three biological independent experiments with similar results. Source

data are provided as a Source Data File. **f** Co-IP analysis of endogenous INTS3 and INTS5 in DSS1-Flag-dTAG cells. Data represent three biological independent experiments with similar results. Source data are provided as a Source Data File. **g, h** Overall cryo-EM map (**g**) and the molecular model (**h**) of the DSS1-INTAC complex in two different views, with subunit surfaces/models colored as schematic module (right of **g**). Notably, the cryo-EM map is a composite map derived from multibody refinement. **i** A close-up view of DSS1 (magenta) in DSS1-INTAC was shown. DSS1 is represented as sticks, surrounded by a transparent density map, indicating its nice fitting.

signals (Fig. 2d, e). These DSS1-associated genomic regions included distal promoters, as well as intergenic and intragenic regions (Fig. 2f), aligning with the distribution patterns of INTAC subunits^{24,26,51}. To evaluate the differential binding of DSS1 across various genomic regions, we calculated the average occupancy of DSS1 at each genomic element. Notably, although the number of DSS1-bound promoters and distal intergenic regions were comparable (14,586 vs. 15,694), we found that DSS1 occupancy is higher on average at promoters compared with other genomic regions (Fig. 2g). Pearson

correlation analysis demonstrated a strong correlation between the genomic localization of DSS1 and other INTAC subunits, as well as active chromatin markers of promoters and enhancers (Supplementary Fig. 2f). Moreover, DSS1 occupancy is comparable to that of other INTAC subunits, particularly at promoter regions that have the highest enrichment of INTAC (Fig. 2g-i and Supplementary Fig. 2g). Together, these results suggest a significant nuclear localization of DSS1 and a genomic distribution that is congruent with other INTAC subunits.

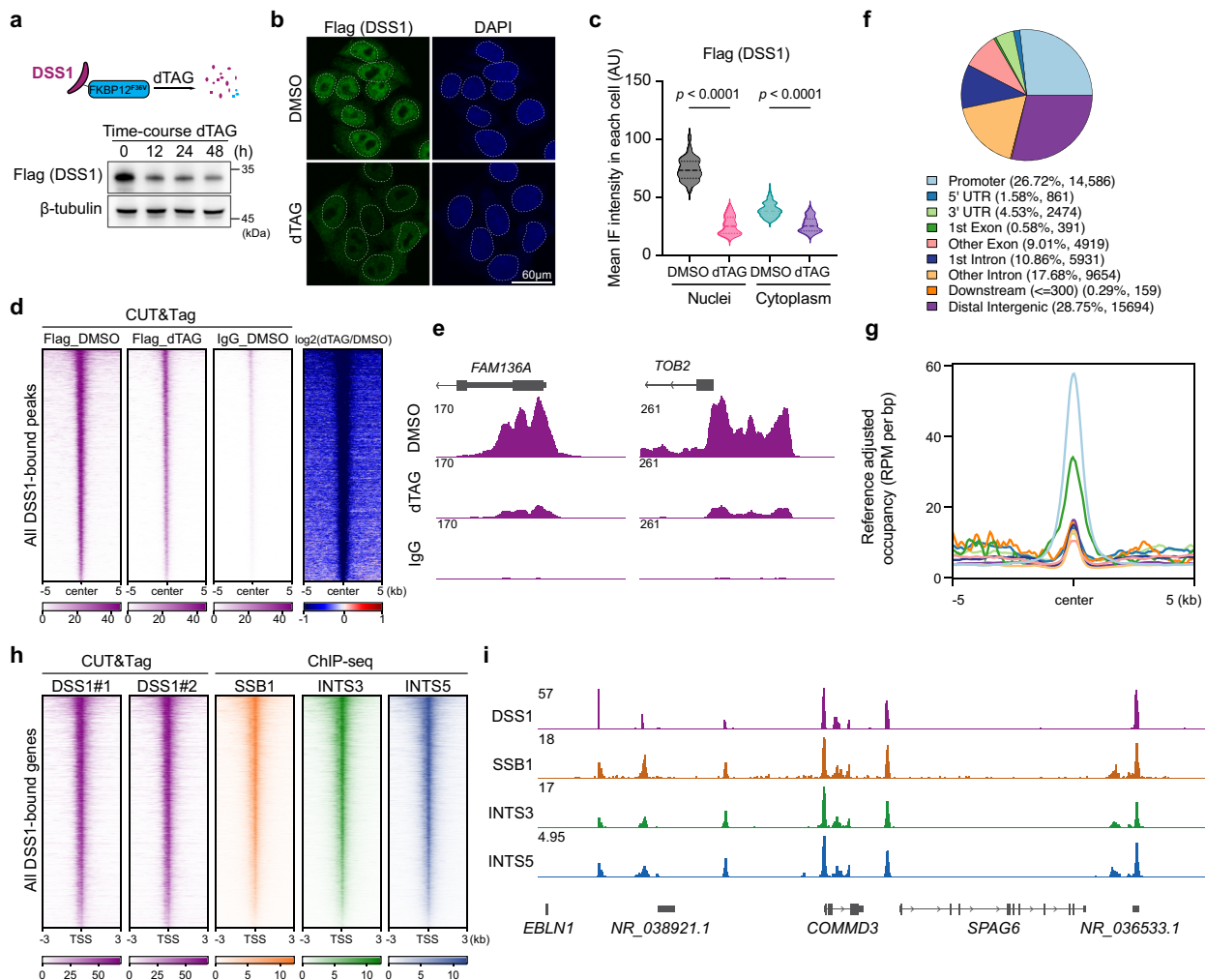


Fig. 2 | Similar genomic distribution of DSS1 and INTAC. **a** DSS1 degradation induced by time-course dTAG treatment in DSS1-Flag-dTAG cells. β -tubulin serves as loading control. Data represent three biological independent experiments with similar results. Source data are provided as a Source Data File. **b** Representative images showing the localization of endogenous DSS1 (green) along with the DAPI signal (blue) after 24-h DMSO or dTAG treatment in DSS1-Flag-dTAG cells. Data represent three biological independent experiments with similar results. **c** Quantification of mean immunofluorescence (IF) intensity of DSS1 in nuclei and cytoplasm of cells. (DMSO/nuclei: $n = 110$; dTAG/nuclei: $n = 83$; DMSO/cytoplasm, $n = 111$; dTAG/cytoplasm, $n = 63$). AU means arbitrary units. Outliers were

excluded by ROUT method, $Q = 10$ %. P values were calculated using two-tailed unpaired t -tests. (Both the $p < 1e-15$). Source data are provided as a Source Data file. **d** Flag (DSS1) CUT&Tag signals over 10 kb regions centered on Flag (DSS1) peaks in DSS1-Flag-dTAG cells treated with DMSO or dTAG for 24 h. **e** Representative track examples showing DSS1 occupancy in DSS1-Flag-dTAG cells treated with DMSO or dTAG for 24 h. **f** Pie chart showing DSS1-associated genomic regions. **g** Metaplot of DSS1's average occupancy at each genomic element. **h** Occupancy of DSS1, SSB1, INTS3, and INTS5 over 6 kb regions centered on the TSS of DSS1 target gene promoters. **i** Representative track examples showing the occupancy of DSS1, SSB1, INTS3, and INTS5.

Selective disruption of the DSS1-INTAC interaction without affecting proteasomal DSS1

Our detailed structural analysis revealed significant hydrophobic interactions between DSS1 and the backbone subunit INTS7, which highlighted the importance of a tryptophan residue at position 39 (W39) within DSS1 for its association with INTAC (Fig. 3a and Supplementary Fig. 3a, b). Immunoprecipitation assays demonstrated that substituting W39 in DSS1 with the positively charged arginine (W39R) significantly disrupted its interaction with INTAC without affecting its association with proteasome subunits such as PSMD4 and PSMD14 (Fig. 3b). Furthermore, mutations of residues within INTS7's hydrophobic pockets similarly compromised INTAC's interaction with DSS1 (Supplementary Fig. 3c, d).

To assess the functional implications of the DSS1-INTAC interaction, we overexpressed HA-GFP-tagged wild-type (WT) DSS1 and the W39R mutant in DSS1-Flag-dTAG cells (Fig. 3c).

Notably, the overexpression of either DSS1-WT or DSS1-W39R led to a more rapid degradation of the endogenous DSS1 compared to cells transfected with an empty vector (Fig. 3d, e). This effect implies the involvement of DSS1 in supporting proteasomal activities independently of its association with INTAC. Supporting this notion, upon inhibition of protein synthesis with cycloheximide, the depletion of DSS1 extended the half-life of proteins typically subject to rapid turnover, such as CDC6 and C-MYC (Supplementary Fig. 3e)⁵². Remarkably, reintroducing either DSS1-WT or the INTAC-interaction-deficient DSS1-W39R mutant restored the degradation rates of these proteins to their original levels (Supplementary Fig. 3e). These findings indicate that DSS1 is integrated into INTAC and the proteasome through distinct interfaces, with the W39 residue being crucial for its association with INTAC but dispensable for its role in proteasome-mediated degradation.

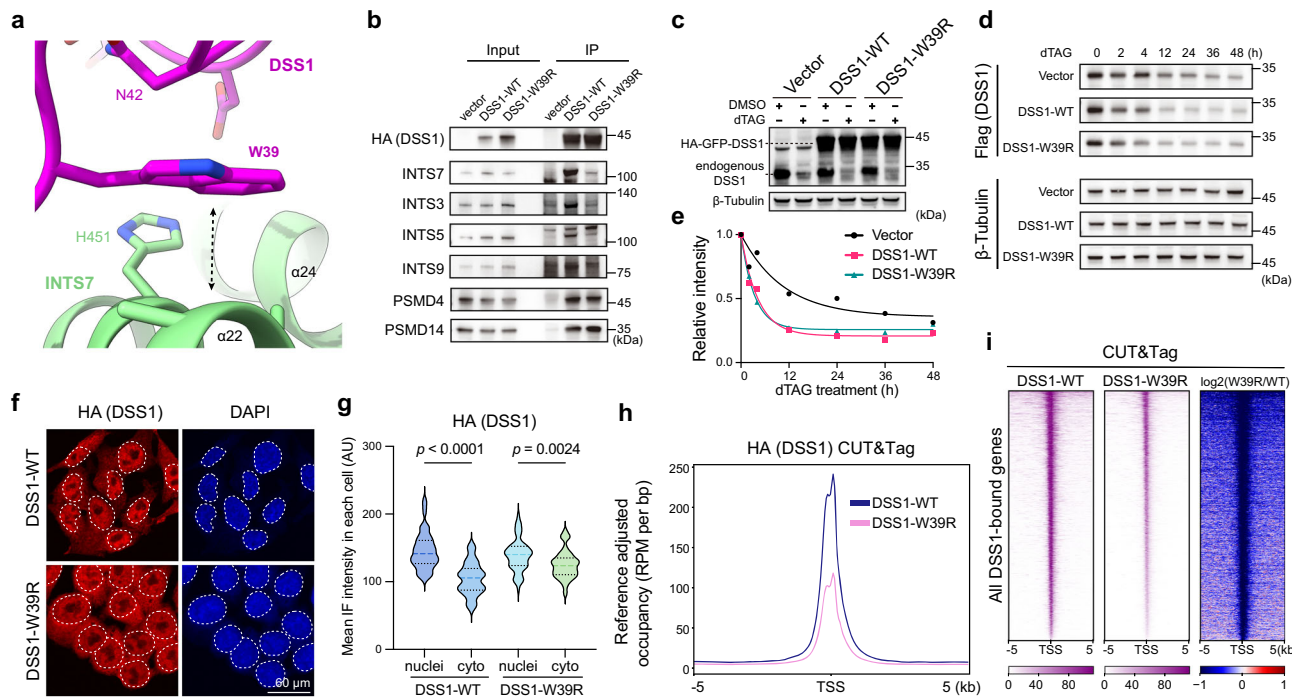


Fig. 3 | The DSS1-INTAC interaction is essential for the optimal chromatin recruitment of DSS1.

a Detailed structural view of the interaction sites between INTAC and DSS1. Residue W39 stacks with the backbone of the helix 22 of INTS7, indicated by the bidirectional arrow. **b** Co-IP analysis of HA (DSS1) followed by western blotting. PSMD4 and PSMD14 are 26S proteasome subunits. The experiment was repeated independently 3 times with similar results. Source data are provided as a Source Data File. **c** Overexpression of DSS1-WT, DSS1-W39R, or vector in DSS1-Flag-dTAG cells treated with DMSO or dTAG, followed by western blotting of DSS1. β -tubulin serves as loading control. Data represent three biological independent experiments with similar results. Source data are provided as a Source Data File. **d** Western blotting of Flag (DSS1) in DSS1-WT, DSS1-W39R, or vector overexpressing DSS1-Flag-dTAG cells with time-course dTAG treatment. β -tubulin serves as loading control. Data represent three biological independent experiments with similar results. Source data are provided as a Source Data File. **e** The quantified and normalized band intensities in **(d)**. Exponential one phase decay model was

utilized to fit nonlinear regression curve for Flag (DSS1) band intensities. Samples derive from the same experiment and that gels/blots in **(d)** were processed in parallel. Source data are provided as a Source Data file. **f** Representative images showing the localization of overexpressed DSS1-WT or DSS1-W39R (red) along with the DAPI signal (blue) after endogenous DSS1 degradation in DSS1-Flag-dTAG cells. Data represent two biological independent experiments. **g** Quantification of mean IF intensity of HA (DSS1) in nuclei and cytoplasm of DSS1-Flag-dTAG cells with DSS1-WT or DSS1-W39R. (WT/nuclei: $n = 91$; WT/cytoplasm: $n = 90$; W39R/nuclei, $n = 82$; W39R/cytoplasm, $n = 82$). AU means arbitrary units. Outliers were excluded by ROUT method, $Q = 10\%$. Multiple comparison after One-way-ANOVA was used to calculate p value. (In DSS1-WT, $p = 4.53e-13$, in DSS1-W39R, $p = 2.42e-3$). Source data are provided as a Source Data file. **h**, **i** Metaplot (**h**) and heatmaps (**i**) showing HA (DSS1) CUT&Tag signals over 10 kb regions centered on the TSS of target genes in dTAG-treated DSS1-Flag-dTAG cells expressing DSS1-WT or DSS1-W39R.

The DSS1-INTAC interaction is required for the optimal chromatin recruitment of DSS1

Since perturbations of the proteasome could potentially impact the cellular levels of any protein and therefore have unpredictable and uninterpretable consequences, we compared the effects of expressing DSS1-WT and DSS1-W39R to differentiate the role of DSS1 within INTAC from its proteasomal activity. To this end, we expressed DSS1-WT or DSS1-W39R in DSS1-Flag-dTAG cells that were treated with dTAG to deplete endogenous DSS1. Immunofluorescence analysis revealed that DSS1-WT is predominantly localized to the nucleus, whereas DSS1-W39R was equally distributed across both nuclear and cytoplasmic compartments (Fig. 3f, g). To evaluate their genomic distribution, we performed CUT&Tag for both DSS1-WT and DSS1-W39R in cells depleted for endogenous DSS1. Despite similar expression levels (Fig. 3c), DSS1-WT showed significantly higher occupancy on chromatin than DSS1-W39R on average (Fig. 3h). Heatmap analysis shows a widespread reduction in genomic occupancy for DSS1-W39R compared with DSS1-WT (Fig. 3i), which can be seen on example genes (Supplementary Fig. 3f). These findings indicate that the DSS1-INTAC interaction is required for the proper nuclear localization and chromatin association of DSS1.

Structural analysis of the DSS1-containing INTAC-paused elongation complex (PEC)

The intimate interaction between DSS1 and INTAC seen in our structural studies, along with their similar genomic distributions, implies that DSS1 is likely present in most or all forms of INTAC. We and colleagues previously reported a structure of INTAC in association with the paused elongation complex (PEC)^{46,47}, which typically is comprised of the paused form of Pol II, NELF, and DSIF. Consistent with our hypothesis, purifications of DSS1 were enriched for subunits of Pol II, NELF, and DSIF (Supplementary Fig. 4a, b). Endogenous co-IP further corroborated the association of DSS1 with these PEC components (Supplementary Fig. 4c).

To better ascertain the incorporation of DSS1 into INTAC in the context of PEC, we reconstituted a complex containing DSS1-INTAC and PEC, followed by cryo-EM single-particle analysis. The resulting map reached an overall resolution of 4.6 Å, with individual sub-complexes improved to resolutions from 4.2 to 4.3 Å through multibody refinement. The structural model was fitted by integrating previously determined structures of the INTAC-PEC⁴⁶ into the cryo-EM maps, followed by manual adjustments (Supplementary Fig. 1 and Supplementary Data 4).

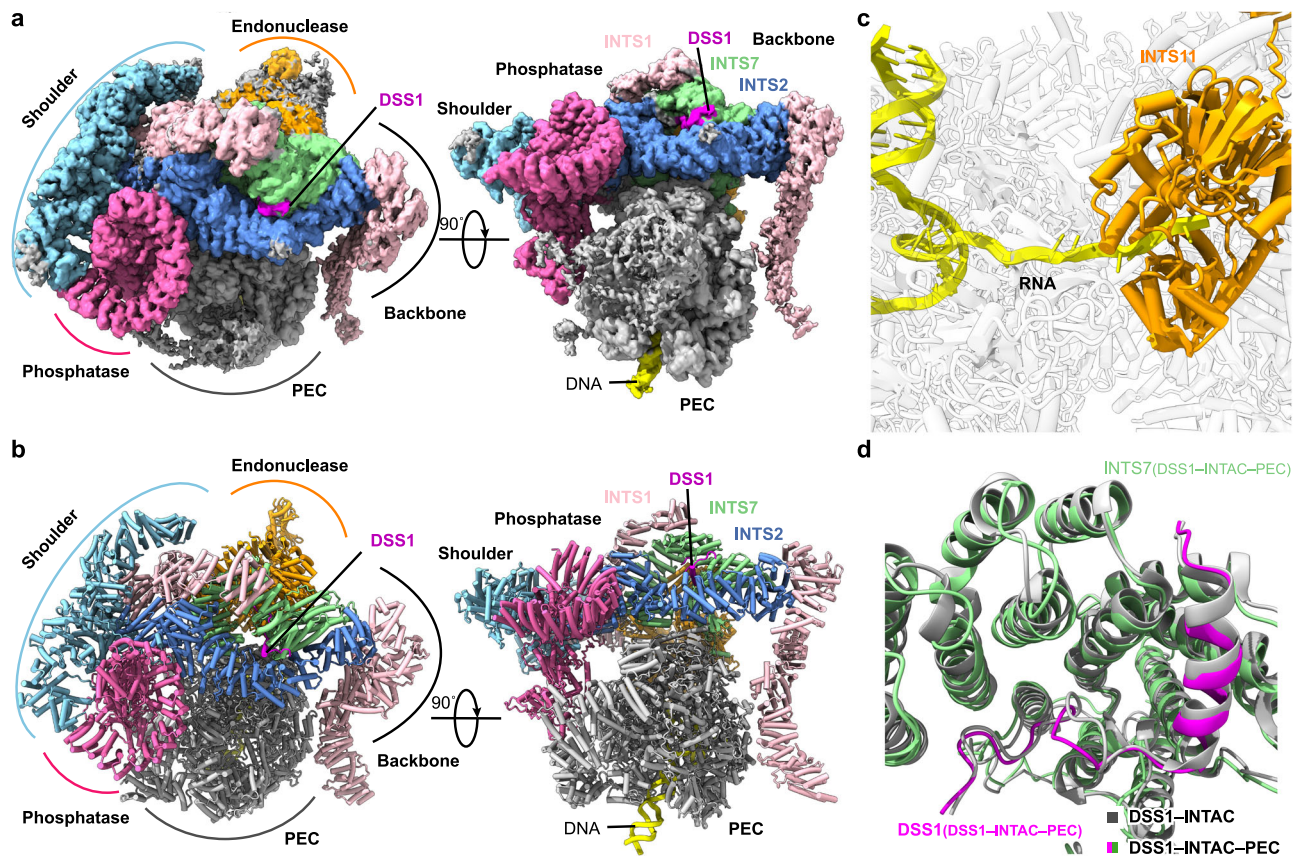


Fig. 4 | Structure of DSS1-INTAC-PEC. a, b Cryo-EM map (a) and molecular model (b) of DSS1-containing INTAC-PEC shown in two different views. Notably, the cryo-EM map is a composite map derived from multibody refinement. c Structural model

showed a close-up view of the mRNA (yellow) inserted into the catalytic center of INTS11 (orange). d Structural model showing the conformation of DSS1 in DSS1-INTAC (Gray) and DSS1-INTAC-PEC (INTS7: green, DSS1: magenta).

In our DSS1-INTAC-PEC structure, we observed an extra density corresponding to the N-terminus of INTS1, which is not visible in the standalone INTAC structure, in addition to the integration of the PEC components, including Pol II, NELF, and DSIF (Fig. 4a, b). This “tail” of INTS1 extends from the bottom of the backbone module and loops back to interact with PEC, consistent with its conformation observed in the previously published INTAC-PEC structures^{46,49}. Overall, INTAC forms close contacts with the PEC, and features extensive interacting interfaces between them (Fig. 4a, b).

Examination of the DSS1-INTAC-PEC structure shows single-stranded RNA extending from the DNA-RNA hybrid within Pol II, and its 5' end inserting into the catalytic pocket of INTS11, the enzymatic subunit of the INTAC endonuclease module (Fig. 4c). Overlaying INTS11 from the DSS1-containing INTAC and INTAC-PEC reveals a transition of the RNA endonuclease center from a closed, inactive state to an expanded, active conformation upon association with PEC (Supplementary Fig. 4d). Moreover, the phosphatase module exhibited structural reconfigurations within the INTAC-PEC framework (Supplementary Fig. 4e, f), likely driven by the association of the phosphatase module, specifically INTS6, with NELF^{46,47}. Notably, the conformation of DSS1, along with the backbone module, remains largely unaltered within INTAC and the INTAC-PEC assemblage (Fig. 4d). These findings support the notion that DSS1 is an integral subunit of INTAC, likely functioning as part of the INTAC backbone module.

DSS1 loss impairs the stability of INTAC and its function in transcription

We next sought to determine whether DSS1 regulates the integrity of the INTAC complex by performing rescue experiments with DSS1-

WT and DSS1-W39R, the latter of which affects DSS1's interaction with INTAC but not the proteasome (Fig. 3b and Supplementary Fig. 3e). Specifically, we overexpressed either DSS1-WT or DSS1-W39R, followed by degradation of the endogenous DSS1. Western blotting showed a notable reduction in the levels of several subunits of INTAC—such as INTS7 (backbone), INTS5 (shoulder), INTS4 (endonuclease), and INTS3 (SOSS)—upon introduction of DSS1-W39R compared to DSS1-WT in cells depleted of endogenous DSS1 (Fig. 5a, b and Supplementary Fig. 5a). Using INTS3 as an example, we employed calibrated ChIP-seq with reference exogenous genome (ChIP-Rx) to estimate the chromatin occupancy of INTAC. Heatmap analysis showed a pervasive decline in INTS3 occupancy for all DSS1 bound peaks, including promoters (Fig. 5c and Supplementary Fig. 5b), an effect that can also be seen on example genes (Fig. 5d). These results suggest that DSS1 loss likely impairs the stability of INTAC.

To determine the role of DSS1 in supporting INTAC function in transcription, we analyzed genome-scale Perturb-seq results and found that depleting different modules of INTAC elicits distinct changes of gene expression (Fig. 5e)⁵³. Notably, the transcriptional changes following DSS1 depletion closely align with those seen in cells lacking subunits of the backbone and shoulder modules. To further clarify the role of DSS1 within INTAC, we compared gene expression changes triggered by DSS1 depletion to those caused by depletion of key subunits from all known DSS1-containing complexes. We found that the transcriptional profile of DSS1-depleted cells most closely resembled that of cells lacking INTAC backbone (INTS1/2/7) or shoulder (INTS5/8) subunits (Supplementary Fig. 5c), and that the differentially expressed genes in DSS1-depleted cells overlapped

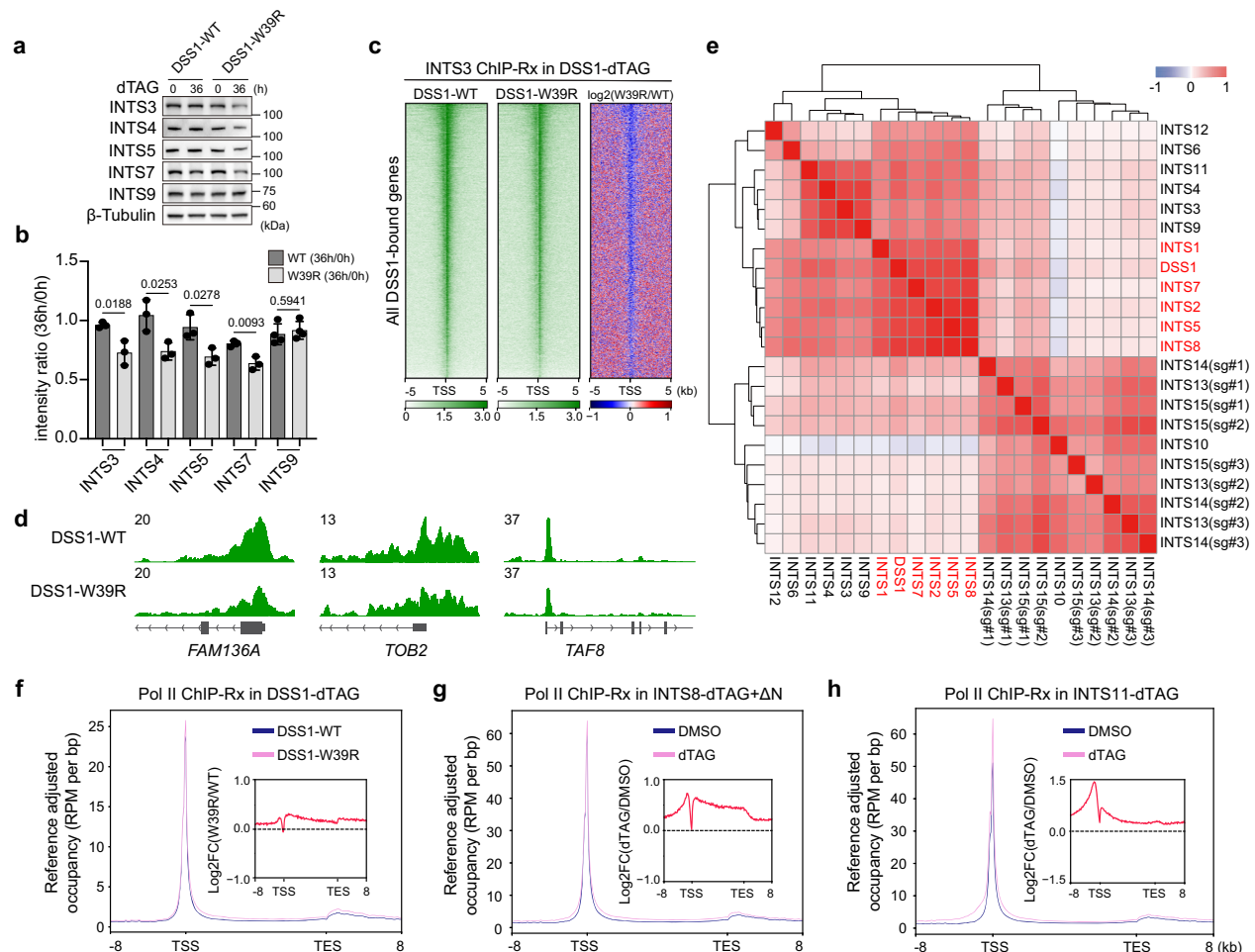


Fig. 5 | DSS1 regulates INTAC integrity and its function in transcription.

a Western blotting of INTAC subunits in DSS1-WT or DSS1-W39R expressing DSS1-Flag-dTAG cells with dTAG treatment. β -tubulin serves as loading control. Samples derive from the same experiment and that blots were processed in parallel to be suitable for subsequent quantification in **(b)**. Data represent three biological independent experiments with similar results. Source data are provided as a Source Data File. **b** Quantification of the biological independent western blotting result in **(a)**. Data are presented as mean values with SD. (INTS3, INTS4, INTS5, INTS7: $n = 3$, INTS9: $n = 4$). P values were calculated using two-tailed unpaired t -tests. Source data are provided as a Source Data file. **c** Heatmaps

showing INTS3 ChIP-Rx signals over 10 kb regions centered on the TSS of target genes after endogenous DSS1 degradation in DSS1-Flag-dTAG cells with overexpression of DSS1-WT or DSS1-W39R. **d** Representative track examples of INTS3 occupancy in dTAG treated DSS1-Flag-dTAG cells with overexpression of DSS1-WT or DSS1-W39R. **e** Analysis of Perturb-seq showing the correlation of gene expression changes in cells lacking INTAC subunits. **f** Metagenes showing Pol II ChIP-Rx signals in DSS1-Flag-dTAG cells overexpressing DSS1-WT or DSS1-W39R. **g, h** Metagenes showing Pol II ChIP-Rx signals in INTS8-dTAG + ΔN (**g**) and INTS11-dTAG cells (**h**) treated with DMSO or dTAG.

significantly with those affected by depletion of these INTAC modules (Supplementary Fig. 5d, e). These results strongly suggest that the gene expression alterations observed upon DSS1 depletion primarily stem from its role in INTAC rather than other DSS1-containing complexes.

We next performed Pol II ChIP-Rx in dTAG-treated DSS1-Flag-dTAG cells expressing either DSS1-WT or DSS1-W39R. Metagene and heatmap analysis showed increased Pol II occupancy across gene bodies in cells expressing the DSS1-W39R mutant compared to those with DSS1-WT (Fig. 5f and Supplementary Fig. 5f). We next analyzed the changes in Pol II occupancy in cells in response to rapid disruption of the INTAC shoulder and endonuclease modules (Fig. 5g, h and Supplementary Fig. 5g, h)²⁶. This analysis shows that all three perturbations result in increased Pol II occupancy across gene bodies, with the alterations induced by DSS1 mutation being more similar to the changes seen in cells expressing the mutant INTS8, a scaffolding subunit of INTAC. Collectively, these findings underscore DSS1's role in supporting INTAC's multifaceted regulation of transcriptional.

Discussion

In this study, we establish that human DSS1 is a bona fide component of the INTAC backbone module. Through our structural analysis of DSS1-INTAC, both alone and as part of a complex with the paused Pol II-DSIF-NELF elongation complex (PEC), we identify a close association between DSS1 and the INTS7 subunit of the INTAC backbone module. Notably, this interface is positioned adjacent to the INTS2-INTS7 dimer interface. We therefore propose that DSS1 stabilizes the INTAC backbone module by binding in this region. Disruption of DSS1's interactions with INTS7 leads to reduced DSS1 enrichment on chromatin and impairs the function of INTAC in transcriptional regulation.

The detection of DSS1 in association with paused Pol II suggests its continuous engagement with INTAC throughout the premature termination process. A comparison of the DSS1-INTAC structures, with and without PEC, reveals that despite a reconfiguration in the overall structure of INTAC including activation of the RNA endonuclease module upon PEC interaction, DSS1's positioning in relation to the backbone module remains unchanged in the presence of PEC, underscoring a strict scaffolding role for DSS1. Supporting this notion, a

density corresponding to DSS1 can also be observed in previously reported structural maps of INTAC and INTAC-PEC^{25,46,49}, which are all consistent with endogenous DSS1 stably associating with INTAC.

Importantly, we identified a specific point mutation in DSS1 that disrupts its interaction with INTAC, without affecting its association with the 26S proteasome, and this point mutant significantly impairs the chromatin enrichment of DSS1. This suggests that chromatin-bound DSS1 primarily exists within INTAC. Expression of this mutant leads to a modest reduction in levels of several INTAC subunits across different modules, which could result from secondary effects of disrupting the complex's stability or assembly. Moreover, disrupting the DSS1-INTAC interaction triggers transcriptional alterations akin to those seen in INTAC deficient cells, affirming DSS1's function in facilitating INTAC's role in transcriptional regulation.

Among the four established DSS1-associated complexes—26S proteasome, TREX-2, BRCA2, and INTAC—the exact role of DSS1, whether acting independently within each complex or contributing to a shared biological outcome, remains an open question. The possibility that the regulation of DSS1 levels could enable a coordinated cellular response across these complexes is an intriguing hypothesis. Supporting this idea, each of these complexes plays a role, either directly or indirectly, in genome stability. For example, BRCA2 plays a critical role in DSB repair through homologous recombination and helps mitigate the accumulation of R-loops, which can induce replication stress and genomic instability due to the susceptibility of ssDNA to damage and the obstruction of replication fork progress by DNA:RNA hybrids. BRCA2's involvement in R-loop mitigation is linked to its interaction with the TREX-2 subunit PCID2, with both being direct partners of DSS1^{54–56}.

While it is improbable for a single DSS1 molecule to simultaneously engage with BRCA2 and TREX-2 due to the spatial constraints of their binding sites, the potential role of DSS1 in bridging BRCA2 and TREX-2's activities against R-loops have yet to be clarified. Similarly, INTAC was recently found to recognize and attenuate R-loops that arise at sites of paused Pol II, thus preventing R-loop-related genomic instability^{27,57}. Moreover, evidence suggests that DSS1 is recruited along with the 26S proteasome to DSBs to facilitate their repair^{58,59}. These observations suggest that DSS1's involvement across these complexes might reflect multifaceted yet interconnected biological functions. Exploring how these complexes might collaborate, especially through the regulation of DSS1's availability or stability, presents an intriguing area for future research.

Methods

Reagents, materials, and cell culture

Detailed information for sequences of nucleotides, reagents, and materials, including antibodies and cell lines, used in this study is provided in Supplementary Data 3.

Human DLD-1 cells were cultured in McCoy's 5A medium (BasalMedia) supplemented with 10% fetal bovine serum (FBS; Yeasen) and 1× penicillin-streptomycin (Meilunbio). HEK293T cells were maintained in Dulbecco's modified Eagle medium (DMEM, BasalMedia) supplemented with 10% FBS and 1× penicillin-streptomycin.

HEK 293 Flp-In T-Rex cells were cultured in high glucose (4.5 mg/ml) DMEM medium containing 10% FBS and 1× penicillin/streptomycin. Additionally, 100 µg/ml zeocin and 15 µg/ml blasticidin were supplied in DMEM medium to maintain wild-type 293 Flp-In T-Rex cells.

Molecular cloning

The human DSS1 gene was amplified from a cDNA library reverse transcribed from SK-HEP1 cells. The DSS1 gene was subcloned into a modified pcDNA5/FRT/TO plasmid (Invitrogen), resulting in an N-terminal 2× Strep and 3× FLAG tag, yielding the pcDNA5-StFLAG-

DSS1 plasmid. To produce lentivirus for gene overexpression assays, the DSS1 gene and its mutant (W39R) were also subcloned into pLVX plasmid, resulting in pLVX-DSS1-WT or -Mutant.

Genome editing for dTAG endogenous knock-in

The procedures for generating the dTAG endogenous knock-in cell line were conducted following established criteria^{60,61}. DLD-1 cell suspension (1×10^6 cells) was combined with PITCh plasmids, including PX459-sgRNA for targeting specific genomic sites within target genes, pCRISPR-PITChv2 containing the dTAG microhomology repair template, and sg-PITCh for cutting and fetching out the repair template. Electroporation was then employed, followed by a 2-day recovery period in regular culture medium. After the recovery phase, cells underwent gradient dilution and were cultured with 1–2 µg/ml puromycin (Meilunbio) for 10–14 days. Surviving single colonies were selected, expanded, and subjected to genotype confirmation. The effectiveness of protein degradation in positive clones was validated through dTAG treatment, followed by western blots analysis.

Generation of stable overexpression cell lines

To produce lentivirus for gene overexpression assays, HEK293T cells underwent co-transfection with plasmids pLVX-DSS1-WT or Mutant (or the empty pLVX vector plasmid as the control), which harbored the blasticidin resistance gene, along with psPAX2 and pMD2.G. The transfection plasmids were mixed in a 4:2:1 ratio in Opti-MEM medium utilizing the polycationic agent polyethylenimine (Sigma-Aldrich). Following transfection, the HEK293T culture supernatant, containing viral particles, was harvested at 48 h and 72 h, subsequently filtered through a 0.45 µm filter.

DLD-1 Cell infection with lentivirus was conducted in the presence of 10 mg/ml Polybrene (YEASEN) for 24 h. Infected cells underwent treatment with 10 µg/ml blasticidin for 7–10 days before collection. Evaluation of overexpression efficiency was performed through Western blots.

Co-immunoprecipitation assays

For co-immunoprecipitation (Co-IP) assays, DLD-1 cells ($1-2 \times 10^7$) were harvested by scraping and subsequently washed twice with ice-cold phosphate-buffered saline (PBS). In experiments involving the comparison of multiple cell lines, normalization by weight was implemented to block variations in cellular mass. The cell pellet was resuspended in 900 µl of ice-cold lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.2–0.5% NP40, 10% glycerol, 1× protease inhibitor) and gently rotated at 4 °C for 1 h.

The resulting lysate was clarified by centrifugation at 20,000 ×g for 20 min at 4 °C. The supernatant was then incubated with 2–5 µg of the relevant antibody for each immunoprecipitation reaction, followed by 9.5 h of rotation at 4 °C. rProtein A/G MagPoly Beads (Smart-Lifesciences), pre-blocked with 1 mg/ml BSA for 1 h, were introduced to the samples, and the mixture was rotated for an additional 3 h at 4 °C. Following incubation, samples with beads underwent collection using a magnetic rack, and the beads were subjected to four washes with lysis buffer. Finally, the samples were eluted by adding 50–100 µl of 1× SDS loading buffer, followed by subsequent western blot analysis.

Immunofluorescence analysis

DLD-1 cells were cultured on coverslips for a minimum of 24 h preceding the experimental procedures. Following a wash with PBS, cells underwent fixation with 4% paraformaldehyde for a duration of 10 min. After three consecutive PBS washes, cell permeabilization was achieved by treatment with 0.5% Triton X-100 in PBS for 10 min, followed by blocking with 4% bovine serum albumin (BSA) in PBS for 30 min.

Primary antibodies, appropriately diluted in ice-cold 4% BSA according to the recommended ratios by the manufacturers, were applied to the cells for an overnight incubation at 4 °C. Subsequent to three PBS washes, cells underwent a 1-h incubation with the respective secondary antibodies. Following this, cells were mounted in ProLong Gold Antifade Mountant with DAPI (Invitrogen) before imaging.

Image acquisition was performed using the Leica TCS SP8 laser-scanning confocal microscopy. Unless explicitly stated otherwise, all procedures were executed at room temperature. The division of nuclei and cytoplasmic regions, as well as the quantification of immunofluorescence intensity, were executed using the Fiji built on ImageJ2 2.14.0/1.54f, complemented by statistical analysis conducted through GraphPad 9.5.1 Prism.

Western blotting

The samples underwent lysis using 1× SDS loading buffer, followed by heating at 99 °C for 10 min. Subsequently, the lysates were loaded onto a FuturePAGE™ 4–20% 15 Wells SDS page (ACE Biotechnology) and transferred to an Immobilon®-NC Transfer Membrane (Millipore). Blots were then blocked using Protein Free Rapid Blocking Buffer (Epizyme Biotech) for 20 min. After that they were subjected to overnight incubation at 4 °C with appropriately diluted primary antibodies. Following three washes with TBS containing Tween-20 (TBST), blots were incubated for 1 h at room temperature with HRP-conjugated anti-rabbit- or anti-mouse-IgG (Abclonal), diluted at 1:5000 in TBST, followed by three additional washes in TBST. For the full blots data, see the Source Data file.

Signal detection was achieved using Immobilon Western Chemiluminescent HRP Substrate (Millipore), and images were captured using a Tanon Imaging system. Subsequent quantification of band intensity was performed using Fiji built on ImageJ2 2.14.0/1.54f.

ChIP-Rx

ChIP-Rx was performed following established protocols⁶². Each immunoprecipitation (IP) utilized 1×10^7 cells, which were crosslinked with 1% formaldehyde (Sigma) for 10 min at room temperature. The crosslinking reaction was terminated by the addition of glycine (final concentration 125 mM) and subsequent incubation for 5 min. Cells were then rinsed twice with ice-cold PBS, scraped, and collected by centrifugation.

The cell pellet was resuspended in Lysis buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1× protease inhibitor (Roche)) and subjected to sonication (Qsonica) to fragment chromatin to an appropriate size range. The soluble fraction was collected by centrifugation and mixed with 25% of lysate from mouse embryonic stem cells processed identically as a spike-in for normalization.

The mixed lysate was then incubated with antibodies overnight and BSA-blocked magnetic Protein A/G beads for 2 h, followed by three washes with High-salt Wash buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 1 mM EDTA, 1.0% NP-40, 0.25% sodium deoxycholate), two washes with Low-salt Wash buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.1% sodium deoxycholate), and a final wash with TE buffer containing 50 mM NaCl.

Beads were then eluted with Elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1.0% SDS). The eluted samples were treated with Protease K, and DNA was purified using the Phenol/Chloroform/Isoamyl Alcohol extraction method. Subsequently, libraries for sequencing were prepared using the VAHTS Universal Plus DNA Library Prep Kit for Illumina (Vazymes). The libraries were sequenced using NovaSeq 6000 platform (Mingma Technologies).

CUT&Tag

The CUT&Tag procedure was optimized in accordance with a previously published protocol⁶³. DLD-1 cells were harvested using

Accutase (Thermo Fisher Scientific) to prevent overdigestion. For each CUT&Tag reaction, 5×10^5 cells were typically utilized to ensure sufficient DNA yield for subsequent library construction. Following centrifugation (600 ×g, 3 min) at room temperature, cells were washed twice with 800 μl of wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× protease inhibitor) and subsequently resuspended in 100 μl of wash buffer within low-retention PCR tubes.

The concanavalin-A-coated magnetic beads (Smart-Lifesciences) were pre-activated and resuspended in an equal volume of binding buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM CaCl₂, 1 mM MnCl₂). Subsequently, 10 μl of activated concanavalin A beads were added to 5×10^5 cells, followed by a 10-min incubation period under gentle rotation. The bead-bound cells were magnetized to remove excess liquid, then resuspended in 50 μl of antibody buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× protease inhibitor, 0.05% digitonin, 0.01% NP-40, 2 mM EDTA).

Next, 1 μg of either Flag (Abclonal) or HA (Proteintech) antibody was added to facilitate DSS1 binding, and the mixture was rotated overnight at 4 °C. For IgG control experiments, mouse or rabbit IgG was substituted accordingly. Following overnight incubation, the bead-bound cells were subjected to successive incubations with rabbit anti-mouse IgG (Solarbio, diluted 1:100) and mouse anti-rabbit IgG (Solarbio, diluted 1:100) in 100 μl of antibody buffer for 1 h at room temperature. Subsequently, the cells were washed four times with dig-wash buffer (antibody buffer without 2 mM EDTA) to remove unbound antibodies.

The pAG-Tn5 adapter complex was prepared in dig-300 buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1× protease inhibitors, 0.01% digitonin, 0.01% NP-40) to a final concentration of 0.2 μM. The bead-bound cells were resuspended in 100 μl of pAG-Tn5 mix and incubated at room temperature for 1 h before supernatant removal. Following thorough washing, the tagmentation reaction was initiated in 40 μl of tagmentation buffer (10 mM TAPS-KOH pH 8.3, 10 mM MgCl₂, 1% DMF) at 37 °C for 1 h.

To halt the reaction, 1.5 μl of 0.5 M EDTA, 0.5 μl of 10% SDS, and 1 μl of 20 mg/ml protease K were added. After a 1-h incubation at 55 °C, DNA purification was performed using VAHTS DNA Clean Beads (Vazyme), with elution in 10 μl of 0.1% Tween-20. The eluent was subsequently mixed with 10 U of Bst 2.0 WarmStart DNA polymerase (NEB) and 1× Q5 polymerase reaction buffer (NEB) in a 20 μl reaction system.

The reaction was carried out at 65 °C for 30 min, followed by an additional 20 min at 80 °C to inactivate the Bst 2.0 WarmStart DNA polymerase. The purified DNA was then subjected to amplification using Q5 high-fidelity DNA polymerase (NEB) with a universal i5 primer and a uniquely barcoded i7 primer. The exact number of PCR cycles was determined by qPCR prior to amplification, with 11–12 cycles typically sufficient to yield an adequate library for sequencing.

Following library size-selection with 0.56–0.85 VAHTS DNA Clean Beads to achieve library sizes ranging from 200 to 700 bp, the products were sequenced on the NovaSeq 6000 platform (Mingma Technologies).

Native DSS1 sample purification

To establish stable DSS1 overexpression HEK 293/Flp-In/T-Rex cell lines, cells were pre-cultured in a 10-cm dish for one day. When reaching 50% confluency, cells were co-transfected with 0.5 μg pcDNA5-StFLAG-DSS1 and 4.5 μg recombinase plasmid pOG44 (Invitrogen) for 48 h. Subsequently, cells were passaged and subjected to a 14-day selection with 200 μg/ml hygromycin B. Western blot (WB) was employed to validate the resulting cell lines.

The purification process was adapted from previous studies^{64,65}. In brief, a total of 50 dishes (15 cm) of DSS1-overexpressed HEK 293/Flp-In/T-Rex cells were used for each purification. When the cells reached 70% confluency, DSS1 expression was induced with 1 μg/ml

tetracycline for 24 h. Cells were collected with a cell scraper, washed twice with cold 1× PBS buffer, and resuspended in lysis buffer (20 mM HEPES pH 7.4, 100 mM KOAc, 5 mM MgCl₂, 1 mM DTT, 0.5 mM NaF, 1 mM Na₃V₃O₄, 0.4% NP-40, 1× protease inhibitor mix). Cells were lysed by 10–15 strokes using a 15 ml dounce homogenizer. The cell lysate was clarified by centrifugation at 10,000 × *g* at 4 °C for 15 min, and the supernatant was incubated with 200 μl Flag beads (A2220, Sigma) in a 50 ml centrifuge tube for 2 h. The beads were then transferred to a small chromatography column (Cytiva, 27-5140-01), washed once with lysis buffer and three times with wash buffer (20 mM HEPES pH 7.4, 100 mM KCl, 5 mM MgCl₂, 0.04% NP-40). Finally, the protein complex was eluted by incubating with 500 μl elution buffer (20 mM HEPES pH 7.4, 100 mM KCl, 5 mM MgCl₂, 0.4 mg/ml 3× Flag peptide) for 45 min.

The eluate was concentrated using a 100 kDa cut-off concentrator (Millipore). The purified sample was checked by SDS-PAGE gel electrophoresis and Coomassie Brilliant Blue staining. Subsequently, protein quantification and identification were performed through label-free protein quantification using LC-MS.

Cellular fractionation

DLD-1 Cells were cultured in 10 cm dishes until 70–90% confluency. After washing twice with ice-cold PBS, they were centrifuged at 500 × *g* for 5 min at 4 °C. The pellet was resuspended in Buffer A (10 mM HEPES pH 7.4, 250 mM sucrose, 5 mM MgCl₂) with protease inhibitors and incubated on ice for 5 min. NP-40 (0.1%) was added, mixed gently, and incubated for another 5 min. After centrifugation at 1000 × *g* for 5 min at 4 °C, the supernatant (cytosolic fraction) was collected. The nuclear pellet was washed, centrifuged, and resuspended in Buffer B (20 mM HEPES pH 7.4, 0.2 M NaCl, 1 mM EDTA, 10% glycerol). After gentle rocking at 4 °C for 30 min and centrifugation at 10,000 × *g* for 10 min, the supernatant (nucleoplasmic fraction) was collected. The remaining chromatin pellet was washed, centrifuged, resuspended in Buffer B, homogenized, and sonicated. After centrifugation at 21,000 × *g* for 30 min, the supernatant (chromatin-associated fraction) was collected. All fractions underwent Flag (*n* = 3) or IgG (*n* = 1) co-immunoprecipitation, followed by SDS-PAGE, trypsin digestion, and LC-MS analysis at Tsinghua University.

LC-mass spectrometry

Protein concentration of samples from immunoprecipitation were determined using the Pierce BCA Protein Assay Kit. For protein digestion, trypsin was added at a 1:50 (enzyme-to-substrate, w/w) ratio, and the mixture was incubated overnight at 37 °C. After digestion, trifluoroacetic acid was added to terminate the reaction. The sample was then centrifuged, and the supernatant was desalted using a C18 column. Finally, the peptides were lyophilized to dryness and stored at –20 °C for further analysis.

Data were analyzed using PEAKS Online (version 12 build 2.1.2024-11-20_160513). The following parameters were applied: MS1 mass tolerance was set to 10 ppm, and MS2 mass tolerance was set to 0.02 Da. The UniProt human protein database (20,375 entries) was used for database searching. A false discovery rate of 1% was used as the cutoff for peptide, protein, and peptide-spectrum match (PSM) identification.

Protein expression and purification

Protein expression and purification were performed essentially as previously described²⁵. In brief, the 14 full-length open reading frames of human INTAC subunits (INTS1 to INTS14) were separately cloned into a modified pCAG vector. INTS2, INTS3, INTS4, and INTS10 were tagged with an N-terminal Protein A (ProA) tag. To purify DSS1 containing INTAC complex, pLVX-DSS1-WT plasmid was co-expressed together with all 14 INTAC subunits plasmids in Expi293F cells and purified by immunoglobulin G (IgG) affinity chromatography and on-

column digestion. The eluate was concentrated for further purification by glycerol density gradient sedimentation. The fractions, 200 μl each, were collected manually from the top of the gradient and analyzed by SDS-PAGE. Peak fractions were pooled and concentrated to 1 to 2 mg/ml accompanied with the removal of glycerol. The purified complex was used for cryo-EM sample preparation.

Pol II was isolated from *S. scrofa* thymus and purified following the reported protocol⁷. Four residue substitutions (G882S of RPB2, T75I of RPB3, S140N of RPB3, and S126T of RPB6) exist between *S. scrofa* and *H. sapiens* Pol II. Human NELF and DSIF were separately expressed and purified as previously described⁴⁶.

Cryo-EM sample preparation

To prepare the DSS1-INTAC complex for cryo-EM, the purified DSS1-INTAC sample was further cross-linked and purified using gradient fixation (Grafix) as described^{25,66}. The fractions were collected manually and the homogeneity of peak fractions was assessed by negative-stain EM. Fractions of interest were concentrated, followed by cryo-EM grid preparation.

The DSS1-INTAC-PEC complex for cryo-EM was prepared essentially as previously described⁴⁶. In brief, the purified *S. scrofa* Pol II (275 pmol) and the RNA-template DNA hybrid (550 pmol) were incubated for 15 min at 30 °C, shaking at 300 rpm, followed by the addition of non-template DNA with a twofold molar excess relative to template DNA and further incubation for 15 min at 30 °C. The purified DSIF and NELF were added in a twofold molar excess relative to Pol II for the PEC reconstitution. The sample was incubated for 1 h at 4 °C, followed by the addition of the purified DSS1-INTAC (230 pmol) and incubation for another 2 h at 4 °C. The resulting sample was subjected to GraFix⁶⁶. The fractions were collected manually and the homogeneity of peak fractions was assessed by negative-stain EM. Finally, fractions of interest were concentrated, followed by cryo-EM grid preparation.

For negative-stain EM, 5 μl of freshly purified protein sample was applied onto a glow-discharged copper grid supported by a thin layer of carbon film for 1 min before negative staining by 2% (w/v) uranyl formate at room temperature. The negatively stained grid was loaded onto a FEI Talos L120C microscope operated at 120 kV, equipped with a Ceta CCD camera.

For cryo-EM grid preparation, 4 μl of protein sample (about 0.7 mg/ml for DSS1-INTAC complex, about 0.9 mg/ml for DSS1-INTAC-PEC complex) was applied onto a glow-discharged holey carbon grid (Quantifoil Au, R1.2/1.3 or R2/2, 300 mesh). After blotting for 3 to 6 s, the grid was vitrified by plunging it into liquid ethane using a VitrobotMark IV (FEI) operated at 4 °C and 100% humidity.

Cryo-EM data collection and image processing

Cryo-EM data were acquired using a Titan Krios electron microscope (FEI) operating at 300 kV at the Cryo-EM platform of Peking University. The microscope was equipped with a K2 direct detector (Gatan) and a GIF quantum energy filter (Gatan) set to a slit width of 20 eV. Automated data acquisition was performed using Serial EM software in super-resolution mode at a nominal magnification of 130,000 × *g*, corresponding to a calibrated pixel size of 1.055 Å and 1.054 Å for the DSS1-INTAC and DSS1-INTAC-PEC sample, respectively. The defocus range was set from –1.5 to –2.5 μm. Each image stack was dose fractionated into 32 frames with a total exposure dose of about 50 e[–]/Å² and an exposure time of 6.96 s and 6.72 s, respectively.

All the image stacks were motion-corrected and dose-weighted using MotionCorr⁶⁷. The contrast transfer function (CTF) parameters were estimated by GCTF⁶⁸ from non-dose weighted micrographs. Micrographs with an estimated resolution of less than 5 Å and astigmatism of less than 5% were manually inspected for contamination or carbon breakage.

For the DSS1-INTAC dataset, 587,397 particles were picked from 3965 good micrographs using Gautomatch (<https://www2.mrc-lmb>).

[cam.ac.uk/download/gautomatch-053/](https://www2.mrc-lmb.cam.ac.uk/download/gautomatch-053/)) with the published INTAC structure (PDB: 7CUN)²⁵ as a reference. These particles were extracted in Relion⁶⁹ and then imported into cryoSPARC⁷⁰. Heterogeneous refinement in cryoSPARC⁷⁰ using the published INTAC structure as references²⁵ yielded 95,815 particles representing the INTAC structure. After 3D classification in Relion⁶⁹, 43,790 particles showing the intact INTAC structure (termed DSSI-INTAC) were identified. Final 3D refinement and CTF refinement resulted in a reconstruction at 4.1 Å resolution, based on the gold-standard Fourier shell correlation 0.143 criterion. To improve local resolution, the INTAC map was divided into three bodies: the phosphatase and shoulder part, the endonuclease module, and the backbone of the INTAC, followed by multibody refinement in Relion⁶⁹. The detailed sorting scheme is shown in Supplementary Fig. 1.

For the DSSI-INTAC-PEC dataset, an identical procedure was followed. Initially, 843,523 particles were picked from 7585 good micrographs using Gautomatch (<https://www2.mrc-lmb.cam.ac.uk/download/gautomatch-053/>) with the published INTAC-PEC structure (PDB: 7YCX)⁴⁶ as a reference. After processing in cryoSPARC⁷⁰ and Relion⁶⁹, 26,822 particles showing the INTAC associated with the paused elongation complex (DSSI-INTAC-PEC) were identified. Final 3D refinement and CTF refinement yielded a reconstruction at 4.6 Å resolution. Due to the high flexibility of the paused elongation complex (PEC), the DSSI-INTAC-PEC structure was divided into four bodies: the phosphatase and shoulder part, the endonuclease module, the backbone module, and the PEC, followed by multibody refinement in Relion⁶⁹. The data processing steps are summarized in Supplementary Fig. 1.

All the visualization and evaluation of 3D density maps were performed with UCSF Chimera⁷¹ or UCSF ChimeraX⁷². All final maps were post-processed and local resolution filtered using masks automatically generated in Relion⁶⁹. Simultaneously, the final maps were sharpened using DeepEMhancer⁷³.

Model building and structure refinement

The structural models of DSSI-INTAC and DSSI-INTAC-PEC were built based on their cryo-EM maps and corresponding focused refined maps. Human INTAC (PDB: 7CUN)²⁵ and INTAC-PEC (PDB: 7YCX)⁴⁶ structures were initially docked into the cryo-EM maps of DSSI-INTAC and DSSI-INTAC-PEC using rigid body fitting in UCSF Chimera⁷¹, followed by manual adjustments in COOT⁷⁴. The INTAC and PEC components in both models remained unchanged, while the DSSI component was manually built in COOT⁷⁴ according to its corresponding density maps.

Due to the low overall resolution, the rigid body fitted model for DSSI-INTAC-PEC was used as the final model without any further refinement. However, the structural model of the DSSI-INTAC complex was refined against the 4.1 Å overall map in real space with PHENIX⁷⁵ and was subsequently validated using MolProbity⁷⁶. The statistics of the map reconstruction and model refinement are summarized in Supplementary Data 4. Each focused refined maps from multibody refinement were used to create the composite map using PHENIX⁷⁵, which was then used to create the figures using UCSF ChimeraX⁷².

CUT&Tag data analysis

Raw CUT&Tag reads were trimmed using Trim Galore v.0.6.6 (Babraham Institute). Trimmed reads were aligned to the human hg19 genome using Bowtie v.2.4.4 with the following parameters: “-N 1 -L 25 -X 700 -no-mixed -no-discordant”⁷⁷. Duplicate reads were removed with Picard Tools v2.26.6 (Broad Institute). The reads were then shifted to compensate for the offset in tagmentation site relative to the Tn5 binding site using the alignmentSieve function of deepTools v.3.5.1 with the ‘-ATACshift’ option⁷⁸. Peak calling was performed by macs2 v2.2.6 with a q-value threshold of 0.05⁷⁹.

ChIP-Rx data analysis

The raw ChIP-Rx reads were trimmed by Trim Galore v0.6.6 (Babraham Institute) and aligned to the human hg19 and mouse mm10 assemblies using Bowtie v2.3.5.1 with default parameters⁷⁷. Low mapping quality reads (MAPQ < 30) and PCR duplicates were removed using SAMtools v1.9⁸⁰ and Picard v2.26.6 (Babraham Institute). The read number for each of the ChIP-Rx sample were then collected using SAMtools v1.9⁸⁰, and the normalization factor was calculated as $1e6/\text{spike-in_count}$. Normalized bigwig files were generated by deeptools v3.5.0⁷⁸.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The CUT&Tag and ChIP-seq data were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession codes [GSE297000](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE297000) and [GSE297001](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE297001). Published ChIP-seq data for INTS3, INTS5, and SSB1 in DLD-1 cells under GEO Series accession code [GSE223997](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE223997) and Pol II in INTS11-dTAG and INTS8-dTAG DLD-1 cell lines under GEO Series accession code [GSE223265](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE223265). All cryo-EM maps and molecular models generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) and in the Protein Data Bank (PDB) with accession codes [EMD-39338](https://www.rcsb.org/entry/EMD-39338) and [8YJB](https://www.rcsb.org/entry/8YJB) for DSSI-INTAC; [EMD-39339](https://www.rcsb.org/entry/EMD-39339) and [9VD9](https://www.rcsb.org/entry/9VD9) for DSSI-INTAC-PEC. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD063968](https://www.ebi.ac.uk/pride/archive/study/PXD063968). Requests for materials should be addressed to the corresponding authors. Source data are provided with this paper.

References

- Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
- Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).
- Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
- Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
- Roeder, R. G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol.* **26**, 783–791 (2019).
- Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
- Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, eaba8490 (2021).
- Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
- Chen, F., Gao, X. & Shilatifard, A. Stably paused genes revealed through inhibition of transcription initiation by the TFIIB inhibitor triptolide. *Genes Dev.* **29**, 39–47 (2015).
- Henriques, T. et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* **52**, 517–528 (2013).
- Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407 (2014).
- Missra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc. Natl. Acad. Sci. USA* **107**, 11301–11306 (2010).

13. Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. & Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Mol. Cell Biol.* **22**, 2918–2927 (2002).
14. Farnung, L., Vos, S. M. & Cramer, P. Structure of transcribing RNA polymerase II-nucleosome complex. *Nat. Commun.* **9**, 5432 (2018).
15. Chiu, A. C. et al. Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. *Mol. Cell* **69**, 648–663.e647 (2018).
16. Aoi, Y. et al. NELF regulates a promoter-proximal step distinct from RNA Pol II Pause-release. *Mol. Cell* **78**, 261–274.e265 (2020).
17. Yamamoto, J. et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5**, 4263 (2014).
18. Song, A. & Chen, F. X. The pleiotropic roles of SPT5 in transcription. *Transcription* **13**, 53–69 (2022).
19. Hu, S. et al. SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Mol. Cell* **81**, 4425–4439.e4426 (2021).
20. Aoi, Y. et al. SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol. Cell* **81**, 4413–4424.e4415 (2021).
21. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335–12338 (1995).
22. Sanso, M. et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev.* **30**, 117–131 (2016).
23. Huang, K. L. et al. Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Mol. Cell* **80**, 345–358.e349 (2020).
24. Vervoort, S. J. et al. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell* **184**, 3143–3162.e3132 (2021).
25. Zheng, H. et al. Identification of integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* **370**, eabb5872 (2020).
26. Hu, S. et al. INTAC endonuclease and phosphatase modules differentially regulate transcription by RNA polymerase II. *Mol. Cell* **83**, 1588–1604.e1585 (2023).
27. Xu, C. et al. R-loop-dependent promoter-proximal termination ensures genome stability. *Nature* **621**, 610–619 (2023).
28. Fan, P. et al. Catalytic-independent functions of the Integrator-PP2A complex (INTAC) confer sensitivity to BET inhibition. *Nat. Chem. Biol.* <https://doi.org/10.1038/s41589-024-01807-x> (2025).
29. Sabath, K. et al. Basis of gene-specific transcription regulation by the Integrator complex. *Mol. Cell* **84**, 2525–2541.e2512 (2024).
30. Razew, M., Fraudeau, A., Pfeleiderer, M. M., Linares, R. & Galej, W. P. Structural basis of the Integrator complex assembly and association with transcription factors. *Mol. Cell* **84**, 2542–2552.e2545 (2024).
31. Zhang, Y. et al. The enhancer module of integrator controls cell identity and early neural fate commitment. *Nat. Cell Biol.* **27**, 103–117 (2025).
32. Barbieri, E. et al. Targeted enhancer activation by a subunit of the integrator complex. *Mol. Cell* **71**, 103–116.e107 (2018).
33. Baillat, D. et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
34. Jantti, J., Lahdenranta, J., Olkkonen, V. M., Soderlund, H. & Keranen, S. SEM1, a homologue of the split hand/split foot malformation candidate gene Dss1, regulates exocytosis and pseudohyphal differentiation in yeast. *Proc. Natl. Acad. Sci. USA* **96**, 909–914 (1999).
35. Kragelund, B. B., Schenstrom, S. M., Rebula, C. A., Panse, V. G. & Hartmann-Petersen, R. DSS1/Sem1, a multifunctional and intrinsically disordered protein. *Trends Biochem. Sci.* **41**, 446–459 (2016).
36. Sone, T., Saeki, Y., Toh-e, A. & Yokosawa, H. Sem1p is a novel subunit of the 26 S proteasome from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 28807–28816 (2004).
37. Funakoshi, M., Li, X., Velichutina, I., Hochstrasser, M. & Kobayashi, H. Sem1, the yeast ortholog of a human BRCA2-binding protein, is a component of the proteasome regulatory particle that enhances proteasome stability. *J. Cell Sci.* **117**, 6447–6454 (2004).
38. Faza, M. B. et al. Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *J. Cell Biol.* **184**, 833–846 (2009).
39. Wilmes, G. M. et al. A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol. Cell* **32**, 735–746 (2008).
40. Marston, N. J. et al. Interaction between the product of the breast cancer susceptibility gene BRCA2 and DSS1, a protein functionally conserved from yeast to mammals. *Mol. Cell Biol.* **19**, 4633–4642 (1999).
41. Paraskevopoulos, K. et al. Dss1 is a 26S proteasome ubiquitin receptor. *Mol. Cell* **56**, 453–461 (2014).
42. Tomko, R. J. Jr. & Hochstrasser, M. The intrinsically disordered Sem1 protein functions as a molecular tether during proteasome lid biogenesis. *Mol. Cell* **53**, 433–443 (2014).
43. Ellisdon, A. M., Dimitrova, L., Hurt, E. & Stewart, M. Structural basis for the assembly and nucleic acid binding of the TREX-2 transcription-export complex. *Nat. Struct. Mol. Biol.* **19**, 328–336 (2012).
44. Li, J. et al. DSS1 is required for the stability of BRCA2. *Oncogene* **25**, 1186–1194 (2006).
45. Yang, H. et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297**, 1837–1848 (2002).
46. Zheng, H. et al. Structural basis of INTAC-regulated transcription. *Protein Cell* **14**, 698–702 (2023).
47. Fianu, I. et al. Structural basis of integrator-mediated transcription regulation. *Science* **374**, 883–887 (2021).
48. Pfeleiderer, M. M. & Galej, W. P. Structure of the catalytic core of the Integrator complex. *Mol. Cell* **81**, 1246–1259.e1248 (2021).
49. Fianu, I. et al. Structural basis of Integrator-dependent RNA polymerase II termination. *Nature* **629**, 219–227 (2024).
50. Nabet, B. et al. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat. Commun.* **11**, 4687 (2020).
51. Stein, C. B. et al. Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Mol. Cell* **82**, 4232–4245.e4211 (2022).
52. Li, J. et al. Proteome-wide mapping of short-lived proteins in human cells. *Mol. Cell* **81**, 4722–4735.e4725 (2021).
53. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e2528 (2022).
54. Shivji, M. K. K., Renaudin, X., Williams, C. H. & Venkitesan, A. R. BRCA2 regulates transcription elongation by RNA polymerase II to prevent R-Loop accumulation. *Cell Rep.* **22**, 1031–1039 (2018).
55. Tan, S. L. W. et al. A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability. *Cell* **169**, 1105–1118.e1115 (2017).
56. Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
57. Wang, J., Xu, C., Zhang, Z. & Chen, F. X. Mechanisms of SOSS-Integrator-PP2A complex in attenuating R-loops and promoting genome stability. *Clin. Transl. Med.* **14**, e1519 (2024).

58. Gudmundsdottir, K., Lord, C. J. & Ashworth, A. The proteasome is involved in determining differential utilization of double-strand break repair pathways. *Oncogene* **26**, 7601–7606 (2007).
59. Krogan, N. J. et al. Proteasome involvement in the repair of DNA double-strand breaks. *Mol. Cell* **16**, 1027–1034 (2004).
60. Nabet, B. et al. The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
61. Ji, Y. X., Hu, S. & Chen, F. X. Protocol for rapidly inducing genome-wide RNA Pol II hyperphosphorylation by selectively disrupting INTAC phosphatase activity. *STAR Protoc.* **4**, 102640 (2023).
62. Orlando, D. A. et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).
63. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
64. Ameismeier, M. et al. Structural basis for the final steps of human 40S ribosome maturation. *Nature* **587**, 683–687 (2020).
65. Zemp, I. et al. Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *J. Cell Biol.* **185**, 1167–1180 (2009).
66. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
67. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
68. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
69. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021).
70. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
71. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
72. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
73. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
74. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **60**, 2126–2132 (2004).
75. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **66**, 213–221 (2010).
76. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **66**, 12–21 (2010).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
79. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
80. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Acknowledgements

This work was supported by grants from the National Key R&D Program of China (2024YFA1803600, to X.C., 2021YFA1300100, to X.Y. and C.F.X., 2021YFA1301700, to C.F.X., 2023YFC2413200, to C.J.), the National Natural Science Foundation of China (32422014, to X.C., 32371350, to C.J., 32271308, to Z.H.), Shanghai Rising-Star program to X.C., the Shanghai Natural Science Foundation (22410712400, 22ZR1413600, to C.J.), and the Shanghai Oriental Talent Plan-Youth Program (to Z.H.). C.F.X. is a SANS Exploration Scholar.

Author contributions

C.F.X., C.J., X.Y., and X.C. conceptualized the project, X.C., Z.Q., and Z.H. conducted the experiments with help from Z.W., X.T., X.Y., Z.Y., H.Z., and S.A. performed the bioinformatics analyses. C.F.X. and X.C. supervised the project, C.F.X., C.J., X.Y., and X.C. writing the original draft, and review & editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-61257-4>.

Correspondence and requests for materials should be addressed to Congling Xu, Yanhui Xu, Jingdong Cheng or Fei Xavier Chen.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025