

Structural insights into a highly flexible zinc finger module unravel INSM1 function in transcription regulation

Received: 30 September 2024

Accepted: 24 February 2025

Published online: 04 March 2025

Check for updates

Heng Zhou^{1,2,3}, Xiaoling He^{1,3}, Yue Xiong^{1,2,3}, Yixuan Gong^{1,2}, Yuanyuan Zhang^{1,2}, Shuangli Li¹, Rui Hu^{1,2}, Ying Li^{1,2}, Xu Zhang^{1,2}, Xin Zhou^{1,2}, Jiang Zhu^{1,2} , Yunhuang Yang^{1,2} & Maili Liu^{1,2}

Orderly development of neuroendocrine and nervous system of mammals requires INSM1, a key regulator for cell differentiation. Ectopic expression of INSM1 is closely correlated with human neuroendocrine tumorigenesis, which makes INSM1 a reliable diagnostic biomarker and potential therapeutic target. To date, INSM1 is known as a transcription repressor binding to GGGG-contained DNA element and TEAD1 using its five zinc fingers (ZFs), while the binding mechanism remains unknown. Here, we reveal highly variable conformations of the whole structure of the five ZFs, among which ZF1 adopts an unusual CCHC-fold. ZF1 binds to the TEAD domain of TEAD1 through hydrophobic interactions, and forms a ternary complex with TEAD1 and TEAD1-targeted DNA. Based on this, INSM1 cooperates with TEAD1 to repress the transcription of TEAD1-targeted genes. ZF2 and ZF3 of INSM1 can bind to DNA but have no specificity to the GGGG-contained element due to long flexible interdomain linker. Instead, INSM1 collaborates with CTCF to target genome loci having the GGGG-contained element and regulate the expression of adjacent genes. This study defines a functional mode of INSM1 by cooperating with diverse DNA-binding proteins for targeting specific genome loci in transcription regulation, and provides structural information for designing INSM1-related therapeutic drugs and diagnostic probes.

The spatial-temporal specific transcription of genes is crucial for cell proliferation and differentiation and the development of eukaryotes, and is controlled by numerous regulatory proteins. Abnormal change of functions of the transcriptional regulators often causes developmental defect of eukaryotes and human diseases^{1,2}. One of the most well-known groups of transcription factors (TFs) in eukaryotes is the C2H2-type zinc finger (ZF)-containing TFs that are primarily recognized by their function of binding to DNA *cis*-elements within gene promoters in a sequence-specific manner^{3–5}. C2H2 ZFs typically adopt

a $\beta\beta\alpha$ fold that is stabilized through the coordination of a zinc ion by two Cys and two His residues and interactions within several hydrophobic residues. Two or more ZFs are usually employed by one TF to constitute a modular ZF array for synergistically binding to DNA in high affinity, with the specific residues in the α -helix being responsible for DNA sequence selectivity⁶.

Insulinoma-associated protein 1 (INSM1) is a conserved C2H2 ZF-containing transcriptional regulator broadly expressed in animals^{7–11}. In mammals, INSM1 plays an essential role in the regulation of cell

¹State Key Laboratory of Magnetic Resonance Spectroscopy and Imaging, Key Laboratory of Magnetic Resonance in Biological Systems, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences – Wuhan National Laboratory for Optoelectronics, Wuhan, China. ²University of Chinese Academy of Sciences, Beijing, China. ³These authors contributed equally: Heng Zhou, Xiaoling He, Yue Xiong. e-mail: jiangzhu@apm.ac.cn; yang_yh@apm.ac.cn

differentiation in neuroendocrine and nervous system, and loss of function of INSM1 causes abnormal development of cerebral neocortex^{12–14}, sympathetic adrenal system¹⁵, pituitary¹⁶, and pancreas^{17–19}, and thymus^{20–22}. In recent years, the key role of INSM1 in regulating the differentiation of cochlear outer hair cells which have critical function in hearing was revealed^{23,24}. On the other hand, the abnormal expression of INSM1 in mature tissues is closely correlated with neuroendocrine tumorigenesis. A series of studies identified INSM1 as a reliable biomarker for a variety of neuroendocrine tumors, such as pancreatic tumor, small cell lung cancer, prostate cancer, etc.^{25–30}. In 2022, the World Health Organization (WHO) recommended INSM1 as one of the conventional tools available to pathologists for the accurate classification of neuroendocrine neoplasms³¹. The diagnostic and therapeutic strategies for neuroendocrine tumors employing INSM1 are under development^{32–35}, which demands more understandings on the mechanism of INSM1 function including the structure-function relationship.

Human INSM1 can be divided into two parts structurally: the N-terminal (residue 1–260) part contains a Snail/Ggi-1 (SNAG) motif and two proline-rich (PR) regions, and the C-terminal (residue 261–510) part harbors five ZFs⁷. Sequence analysis identified ZF2/3/4/5 as typical C2H2 ZF, while ZF1 is lack of the last His residue for Zn coordination. NMR structure determination confirmed the typical C2H2 ZF fold of ZF4 and ZF5³⁶. Structures of ZF1/ZF2/ZF3 have not been reported. While the SNAG motif contributes to the transcription repression by INSM1 through mediating the interactions with histone-modifying proteins including KDM1A (also named LSD1), HDAC1/2, and RCOR1/2/3¹⁶, the INSM1 ZFs were suggested to be responsible for binding to DNA *cis*-elements in gene promoters and targeted regulation of gene expression^{37–40}. The consensus sequences bound by INSM1 were explored both *in vitro* and *in vivo*^{41,42}, and the two sequences showed moderate similarity containing a GGGG element. The mechanism by which the INSM1 ZFs recognize and bind to the DNA elements is still unclear. On the other hand, INSM1 was proposed to cooperate with FOXA2 and NEUROD1 for targeted regulation of gene expression in pancreatic β -cell. Collaboration between INSM1 and TEAD1 for transcription regulation was also implied, as the homologs of the two, Nerfin1 and Sd in *Drosophila* can interact with each other to repress the transcriptional output of Hippo signaling^{43,44}. The ZF1 of Nerfin1 was manifested to bind with the transcriptional enhanced associate domain (TEAD) of Sd, but the binding mechanism remains to be revealed.

In this study, the solution structure of INSM1 ZFs was comprehensively investigated, based on which the interaction mechanisms of INSM1 ZFs with DNA and TEAD1-TEAD were elucidated. The results indicate that INSM1 binds to TEAD1-TEAD with ZF1 in a previously undiscovered binding mode, but binds to DNA through ZF2 and ZF3 with an affinity at micromolar level and does not have sequence preference to the GGGG-contained element. We further reveal that INSM1 can collaborate with TEAD1 to repress the expression of TEAD1-targeted genes and cell proliferation. Moreover, INSM1 can interact with CTCF to regulate the expression of CTCF-targeted genes cooperatively. Combined with the previous study, we propose a functional mechanism of INSM1 distinct from the canonical C2H2-ZF proteins, in which INSM1 regulates gene transcription by cooperating with other DNA-binding TFs to indirectly bind to gene promoters.

Results

INSM1 ZF1 adopts an unusual CCHC-fold

The disordered probability prediction using PrDOS⁴⁵ indicated that the N-terminal half of INSM1 containing SNAG and PRs is largely disordered, while the C-terminal half harbors five folded ZFs but the long linkers connecting ZF2, ZF3 and ZF4 are also disordered (Fig. 1A). In order to understand the structure-function relationship of the ZFs, we investigated their solution structures, and according to the folded/

disordered probability, the ZFs were constructed into different truncations to facilitate structural study. The structure of a truncated INSM1 containing ZF4 and ZF5 (termed ZF4-ZF5 hereafter) had been solved previously³⁶. Here, three INSM1 truncations including ZF1-ZF5, ZF1-ZF2, and ZF3 were constructed for structure study (Supplementary Fig. 1A). Inductively coupled plasma mass spectrometry (ICP-MS) analysis manifested that ZF1-ZF5, ZF1-ZF2 and ZF3 have approximate 5, 2, and 1 zinc ions respectively (Supplementary Fig. 1B), consistent with the ZF numbers. Static light scattering-size exclusion chromatography (SLS-SEC) indicated that ZF1-ZF5, ZF1-ZF2, and ZF3 all exist as monomer in solution (Supplementary Fig. 1C).

The ¹H–¹⁵N HSQC spectra of ZF1-ZF2 and ZF3 showed well dispersed resonances in both ¹H and ¹⁵N dimensions, suggesting that they are well folded (Supplementary Fig. 2A, B). The backbone and side chain resonances of ZF1-ZF2 and ZF3 were assigned, and NMR structures were solved subsequently (Supplementary Fig. 2C–E, Supplementary Tables 1 and 2). Similar to ZF4 and ZF5, ZF2 and ZF3 are typical C2H2 ZFs (Fig. 1B and C). The last His residue typical for Zn coordination in C2H2 ZF is missing in ZF1, and a Cys residue (C287) is alternatively involved in Zn coordination, forming an atypical CCHC-fold. Mutation of C287 but not R289 to Ala caused broad chemical shift perturbation (CSP), confirming the importance of C287 in ZF1 folding (Supplementary Fig. 3A, B). All five ZFs have one α -helix and two β -strands. The α -helix of ZF1 is short, due to the less interval residues between Zn-coordinated H285 and C287. Structure-based sequence alignment showed that the five ZFs have very low sequence similarity, and only the residues coordinated with Zn ion and three hydrophobic residues potential for stabilizing the structure are conserved (Fig. 1B).

In established C2H2 ZF-DNA binding model, the residues at the positions of –1, +2, +3, and +6 relative to the beginning of the α -helix are involved in specific DNA binding, which prefer charged or polar residues⁶. ZF2 and ZF3 of INSM1 basically obey this principle, while the residues at +2 and +3 positions of the α -helices from ZF1, ZF4 and ZF5 have hydrophobic sidechains or no sidechain (Fig. 1D). Electrostatic potential analysis showed that the surface of ZF1 is mainly negatively charged, while positive and negative charges are evenly distributed on the surface of ZF2/3/4/5 (Fig. 1E). These results implied that ZF1 hardly has any DNA-binding ability, and ZF2/3/4/5 may be able to bind with DNA.

INSM1 ZF1-ZF5 assumes a collapsed and dynamic interdomain conformation

Although the linkers connecting tandem ZFs vary in length and amino acid type, canonical linker lengths of 4–6 amino acids were well-acknowledged for the C2H2-ZF proteins that bind to DNA in sequence-specific manner. As the linker restricts the spatial distance and relative motion of the connected ZFs, the functions of the ZFs, especially DNA binding, are remarkably affected by the linker⁶. The linkers connecting INSM1 ZFs vary a lot in length, thus the interdomain conformation should be investigated to understand their DNA-binding function. A 4-residue rigid linker connects ZF4 and ZF5 of INSM1, prompting the two ZFs to assume a relatively fixed interdomain conformation³⁶. Unlike the case of ZF4-ZF5, a 7-residue disordered linker connects ZF1 and ZF2, and relative distance and orientation between ZF1 and ZF2 could not be determined by NOE restraints. To assess the possible interdomain contact, an INSM1 truncation containing only ZF1 was constructed (Supplementary Fig. 1A). The NMR spectrum of ZF1 was well overlapped with that of ZF1-ZF2, and only slight CSPs for a small number of residues were found, although V291 and R292 at the C-terminus displayed large change of chemical shift and were not assigned in the ZF1 truncation (Supplementary Fig. 3C, D). This result indicated that interdomain contact between ZF1 and ZF2 is weak. Small angle X-ray scattering (SAXS) analysis determined an R_g of 21.41 Å and a D_{max} of 73.24 Å of ZF1-ZF2 (Fig. 1F and G, and Supplementary Table 3). The experimental curve was similar to the theoretical scattering curve

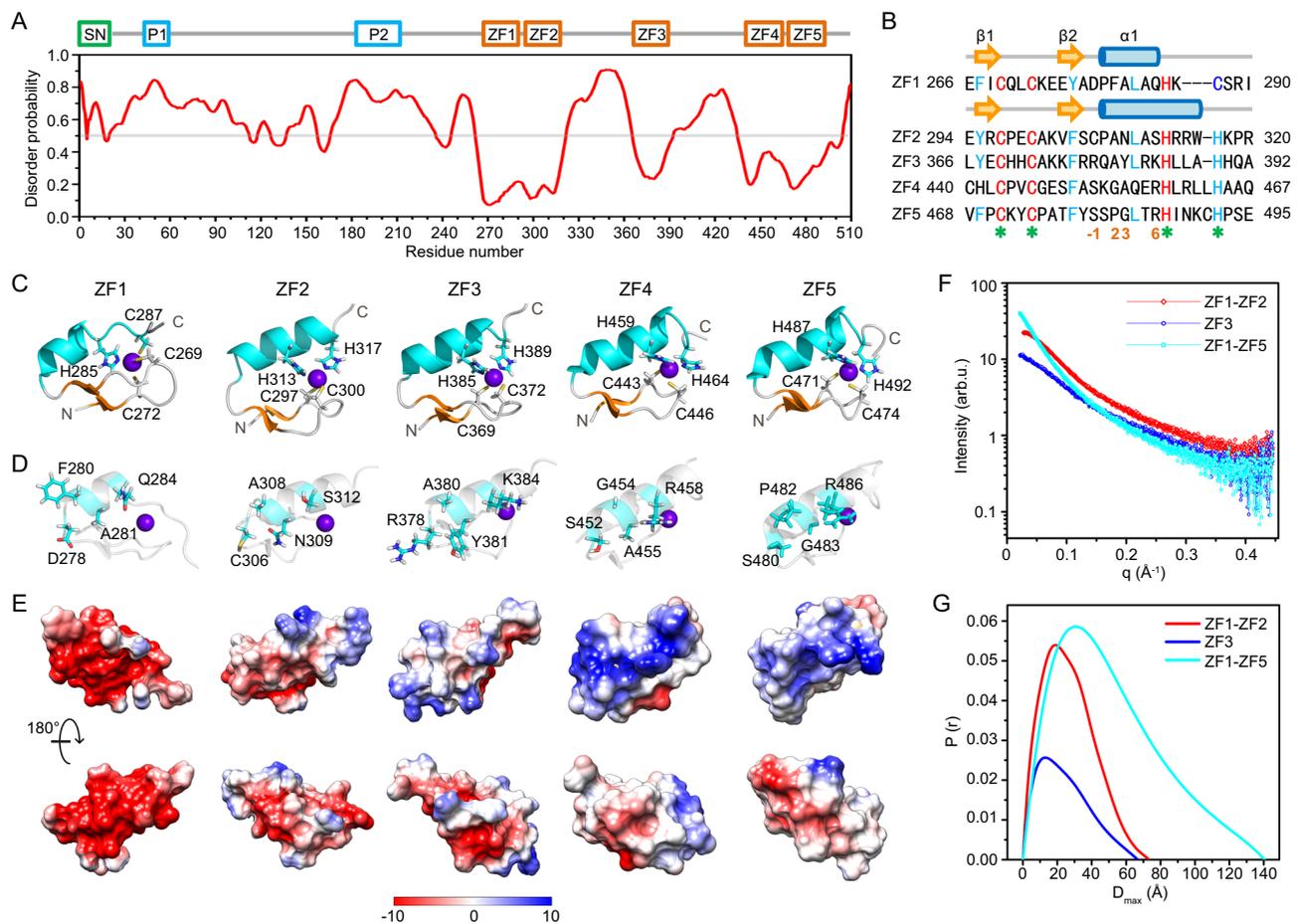


Fig. 1 | Solution structures of INSM1 ZFs. **A** Schematic domain structure of human INSM1 with an analysis of structurally disordered probability. SN, SNAG motif; P1 and P2, proline-rich region 1 and 2; ZF, zinc finger. The regions showing a probability over the gray line (>0.5) are considered to be disordered. Source data are provided as a Source Data file. **B** Sequence alignment of INSM1 ZFs. Identical residues are colored in red, and similar residues are in cyan. The secondary structure elements of INSM1 ZFs are shown on top. The residues at -1, +2, +3, and +6 positions of the α -helix potentially involved in DNA recognition are labeled at the

bottom. The residues involved in Zn coordination are marked with green asterisks. **C** NMR structures of INSM1 ZFs presented in cartoon. The residues involved in Zn coordination are marked and shown as sticks. Zn ions are shown as purple spheres. **D** The residues at the -1, +2, +3, and +6 positions of the α -helix shown as sticks. **E** Surface views of the structures of INSM1 ZFs colored by electrostatic potential. **F** Experimental SAXS data of INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5. **G** Particle distance distribution curves transformed from the SAXS data in (F). Source data are provided as a Source Data file.

of conformer 17 among the ensemble of ZF1-ZF2 NMR structures (Supplementary Fig. 4A), suggesting an extended conformation with hardly any interdomain interaction. Likewise, the experimental SAXS curve of ZF3 truncation was similar to the theoretical scattering curve of conformer 6 among the ensemble of ZF3 structures (Supplementary Fig. 4B), suggesting that the N- and C-terminal disordered regions adjacent to ZF3 also mainly adopt an extended conformation.

There are linkers with up to 49 and 50 residues connecting ZF2 and ZF3, and ZF3 and ZF4, respectively. For ZF1-ZF5 containing all five ZFs, SAXS analysis determined an R_g of 40.94 Å and a D_{max} of 140.39 Å (Fig. 1F and G, and Supplementary Table 3), only about two folds of the values of ZF1-ZF2, suggesting a collapsed conformation of ZF1-ZF5. EOM analysis based on the SAXS data showed an ensemble of varied interdomain conformations of ZF1-ZF5, and all the conformations adopt collapsed states with spatial approaching of different ZFs (Supplementary Fig. 4C, D). We tried to assign the ^1H - ^{15}N HSQC spectrum of ZF1-ZF5, which showed obviously lower quality compared to those of ZF1-ZF2, ZF3 and ZF4-ZF5. Many peaks disappeared in the spectrum of ZF1-ZF5, likely due to the molecular weight over 25 kDa and unfavorable conformational dynamics. The peaks in the central region of the spectrum are very crowded, which are mainly from the

residues in the two long disordered linkers connecting ZF2, ZF3 and ZF4 (Supplementary Fig. 5A). Finally, 67% non-proline backbone chemical shifts were assigned. Based on the assignments, it was found that many disappeared peaks belong to ZF1, ZF2 and ZF3, implying that the three ZFs undergo conformation exchange in ZF1-ZF5 at intermediate regime of NMR time scale (Supplementary Fig. 5B, C). Collectively, the two long disordered linkers connecting ZF2, ZF3, and ZF4 make ZF1-ZF5 assume a collapsed state with dynamic interdomain conformation, which will markedly affect the DNA binding specificity of INSM1.

INSM1 ZF2 and ZF3 bind to DNA with an affinity at micromolar level

Based on the structure data, we subsequently investigated the DNA-binding mechanism of INSM1 ZFs. Previous studies showed that the INSM1 ZFs can bind with two GGGG-contained DNA motifs (M1 and M2) with 8-bp similarity (Fig. 2A)^{41,42}. We firstly studied the binding of INSM1 ZFs with a 17-bp DNA containing the M2 motif using NMR titration. The DNA was respectively titrated into the ^{15}N -labeled ZF1-ZF2, ZF3, and ZF4-ZF5 proteins. Among the three truncated INSM1 proteins, ZF3 showed broad changes of chemical shifts in its ^1H - ^{15}N HSQC spectrum, while only a few peaks were remarkably affected in the

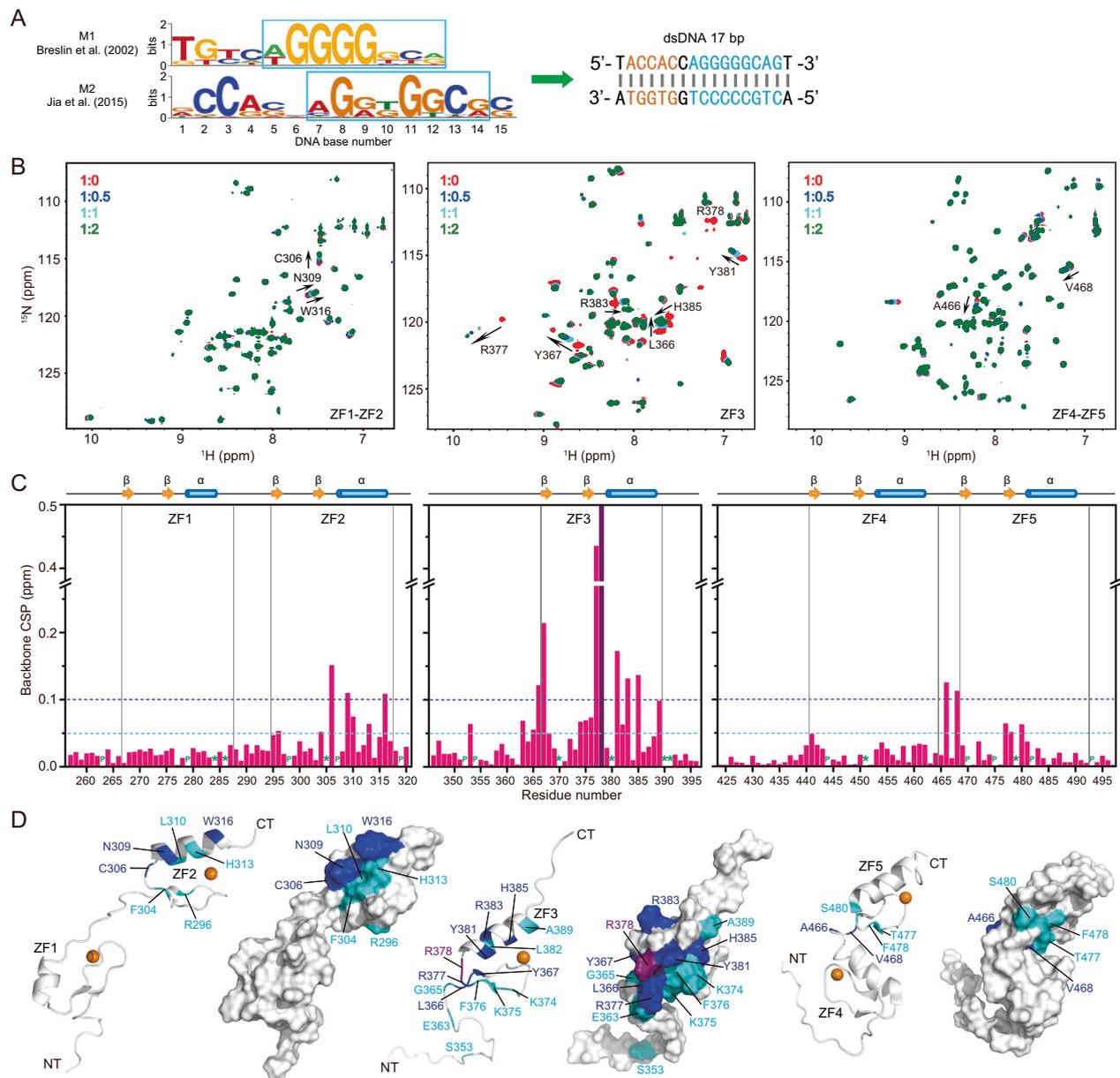


Fig. 2 | NMR titration of INSM1 ZF1-ZF2, ZF3, and ZF4-ZF5 with DNA.

A Consensus motifs (M1 and M2) in DNA bound by INSM1 as determined by selected and amplified binding site selection (SAAB) (Breslin, M.B. et al., 2002) and ChIP-seq (Jia, S. et al. *Insm1* cooperates with NeuroD1 and Foxa2 to maintain mature pancreatic beta-cell function. *EMBO J*, 2015, 34, 1417–1433). The DNA sequence (17-bp) derived from M2 was used in following NMR titration. The region with similarity to both M1 and M2 are colored in cyan, while the region similar only to M2 is in orange. **B** Overlay of a series of ^1H - ^{15}N HSQC spectra of ^{15}N -labeled INSM1 ZF1-ZF2, ZF3, and ZF4-ZF5 in the presence of M2 DNA at different protein:DNA molar ratios, which are

colored differently as indicated. **C** Chemical shift perturbation (CSP) for each residue is calculated and illustrated with the secondary structure elements of each ZF shown on top. CSP values of 0.05 and 0.1 were selected as two significant levels for global comparison of the five ZFs. The residues with CSP over 0.1 were labelled in **(B)**. Unassigned residues are marked with asterisks. P, proline. Source data are provided as a Source Data file. **D** The residues with CSP values over 0.1 (blue) and ranging 0.05–0.1 (cyan) are shown on the structures of INSM1 ZF1-ZF2, ZF3, and ZF4-ZF5 in cartoon and surface views, respectively. R378 with disappeared peak in the spectra during titration is marked in purple.

spectra of ZF1-ZF2 and ZF4-ZF5 (Fig. 2B). The CSP values were calculated for each residue according to the NMR titrations and mapped onto the structures of ZF1-ZF2, ZF3, and ZF4-ZF5, respectively (Fig. 2C and D). In ZF3, six and five residues respectively showed CSPs over 0.1 ppm and ranging in 0.05–0.1 ppm. The peak for R378 disappeared during NMR titration. The affected residues clearly define a DNA-binding interface of ZF3 mainly including the α -helix. The residues at -1 (R378) and $+3$ (Y381) positions of the α -helix exhibited remarkable chemical shift changes. Although the overall CSP of ZF2 was lower than ZF3, the three and four residues respectively showing CSPs over 0.1

ppm and ranging in 0.05–0.1 ppm also defined a DNA-binding interface including the α -helix of ZF2. Similarly, the residues at -1 (C306) and $+3$ (N309) positions of the α -helix exhibited remarkable chemical shift changes. No residue in ZF1 and ZF4 exhibited CSP greater than 0.05 ppm. Two residues from the linker between ZF4 and ZF5, and three residues of ZF5 close to ZF4 showed CSPs over 0.1 and 0.05 ppm, respectively, none of which is located in the α -helix. The NMR titration data of ZF4-ZF5 are consistent with the previous study using a 12-bp DNA containing the M1 motif, which suggested that ZF4 and ZF5 have very weak DNA binding ability³⁶. These results manifested that ZF2 and

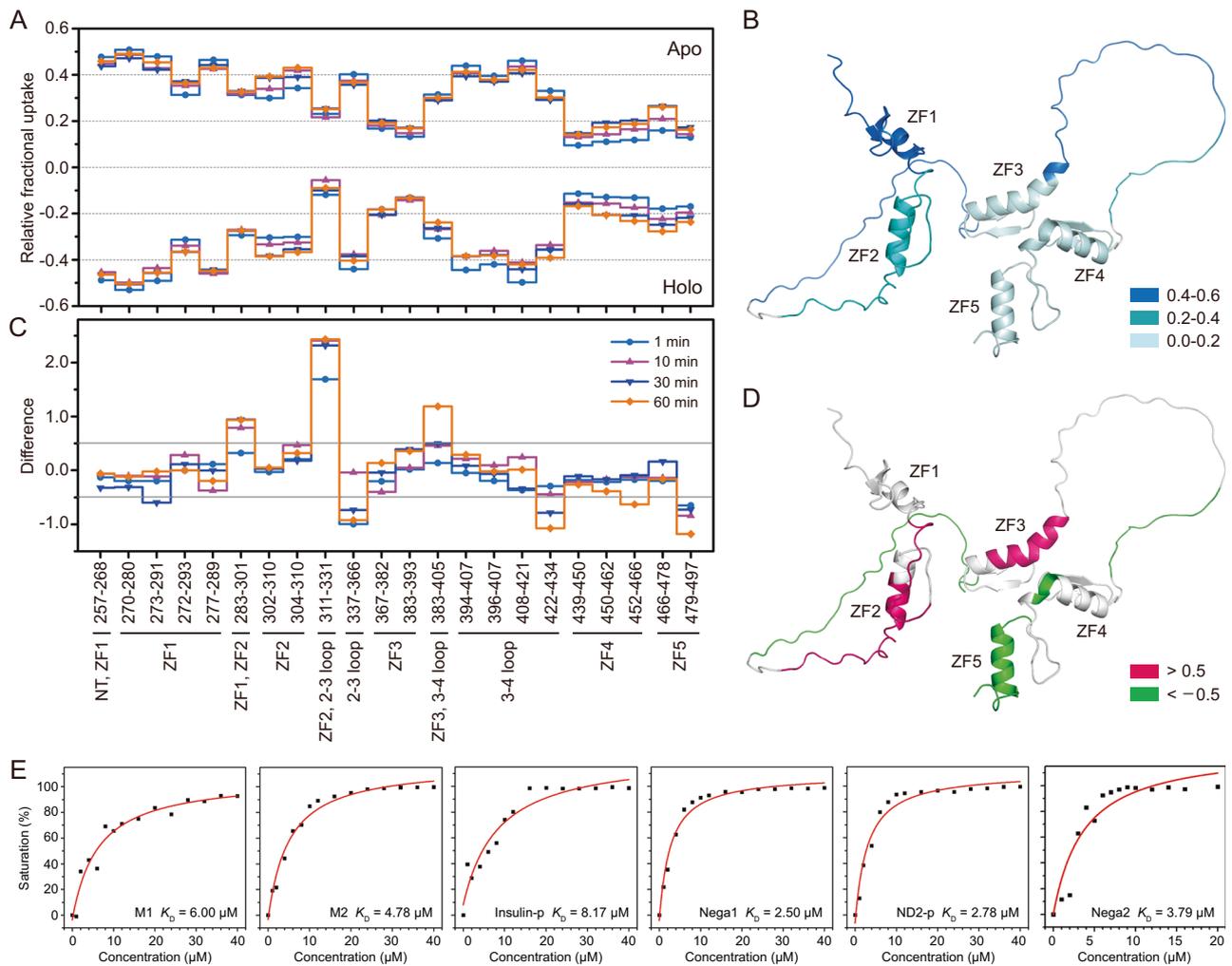


Fig. 3 | Interaction of INSM1 ZF1-ZF5 with DNA. **A** A mirror plot of HDX-MS data comparing the relative deuterium uptake for each peptide detected from the N to C terminus of free INSM1 ZF1-ZF5 (apo) and INSM1 ZF1-ZF5 bound by M2 DNA (holo) at 1 min, 10 min, 30 min, and 60 min. **B** The relative deuterium uptake for each peptide of apo INSM1 ZF1-ZF5 at 1 min was mapped on the structure of INSM1 ZF1-ZF5 modelled by AlphaFold2, colored with marine blue, cyan, and light cyan for different levels of uptake. **C** A difference plot of HDX-MS data showing the changes

in relative deuterium uptake between apo and holo INSM1 ZF1-ZF5 at indicated time points. Source data are provided as a Source Data file. **D** The segments of INSM1 ZF1-ZF5 with significant changes (>0.5 or <-0.5 in **C**) between apo and holo states at 60 min were marked in pink or green on the structure of INSM1 ZF1-ZF5 modelled by AlphaFold2. **E** Curves and K_D values fitted from the EMSA data of INSM1 ZF1-ZF5 with different DNA fragments. Source data are provided as a Source Data file.

ZF3 of INSM1 can bind with DNA, and ZF3 has higher binding ability than ZF2 in the separated state.

To test whether the DNA binding properties of INSM1 ZFs in the separated and integrated states are consistent, the NMR titration of ZF1-ZF5 protein with M2 DNA was carried out. Likely due to the overall big size of the complex of ZF1-ZF5 and M2 DNA, most affected peaks showed decrease of signal intensity rather than change of peak position (Supplementary Fig. 6A). Thus, instead of CSP, change of signal intensity was calculated for each assigned residue to evaluate the DNA-binding sites (Supplementary Fig. 6B). The results revealed that the residues with signal intensity at the bound state lower than 0.2 of that at the free state are mostly from ZF2 and ZF3. These residues defined the DNA-binding interfaces of ZF2 and ZF3 mainly formed by the α -helices (Supplementary Fig. 6C), which include the -1, +3, and +6 residues of ZF2 α -helix (C306, N309, and S312), and +3 and +6 residues of ZF3 α -helix (Y381 and K384). Taken together, no matter in separated or integrated state, ZF2 and ZF3 are the major DNA-binding modules of INSM1.

Considering that the residues in the two long linkers were not assigned in ZF1-ZF5 spectrum, hydrogen-deuterium exchange mass spectrometry (HDX-MS) experiments that can provide high coverage were carried out to study the DNA-binding of ZF1-ZF5. Deuterium uptake data was collected at four different time points (1 min, 10 min, 30 min, and 60 min) for both apo and holo ZF1-ZF5 (bound by M2 DNA) with coverage of 96%. At apo state, ZF3, ZF4, and ZF5 showed low deuterium uptake rate, while the two long linkers exhibited high deuterium uptake rate (Fig. 3A and B), consistent with their disordered structure. ZF1 and ZF2 showed a moderate deuterium uptake rate, higher than ZF3/4/5, suggesting a more flexible fold of ZF1 and ZF2. Upon binding to DNA, ZF2 and ZF3 showed marked decrease of deuterium uptake, further demonstrating their involvement in DNA binding (Fig. 3C and D). Interestingly, three regions in ZF2-ZF3 linker, ZF3-ZF4 linker, and ZF5 displayed decrease deuterium uptake, indicating that they adopt more open conformation in presence of DNA.

In established C2H2 ZF-DNA recognition theory, recognition of sequence-specific DNA requires more than two C2H2 ZFs tandemly

connected by linkers with favorable length for synergistic binding. There are only two DNA-binding ZFs in INSM1, and the two ZFs, ZF2 and ZF3 is connected by a long disordered linker with 49 residues, which is very flexible. Cooperative recognition of sequence-specific DNA may be hardly able to be achieved by ZF2 and ZF3. Thus, we tested the binding properties of ZF1-ZF5 for a series of DNA fragments including the DNA fragments containing M1 and M2 motifs, DNA fragments from insulin and NeuroD2 promoters^{37,38} that show high similarity to M1 and M2 motifs, and two negative control DNA fragments in different length without the two motifs (Supplementary Fig. 7A) using electrophoretic mobility shift assay (EMSA). The results manifested that ZF1-ZF5 binds to all six DNA fragments in low affinity (Supplementary Fig. 7B). No band for specific protein-DNA complex was found for M1, M2 and NeuroD2-p, similar to the negative controls. A weak band for protein-DNA complex was observed for insulin-p. The K_D values of ZF1-ZF5 for the six DNA fragments were calculated by fitting the changes of band intensity of the free DNA (Fig. 3E), which did not show marked difference and were all in the range of 1–10 μ M. These results indicated that INSM1 ZF1-ZF5 binds to DNA in an affinity at micromolar level, and has no preference of binding to the GGGG-contained M1 and M2 motifs. Similarly, ZF1-ZF2 and ZF3 truncations also bind to M2 DNA in an affinity (K_D) of micromolar level (Supplementary Fig. 7C, D). Thus, we suppose that INSM1 is not a typical C2H2 ZF-containing protein, which generally can bind with DNA in high affinity and sequence-dependent manner.

INSM1 ZF1 interacts with the TEAD of TEAD1 through hydrophobic interactions

Transcription factor TEAD1 in Hippo signaling was evidenced to interact with INSM1 ZFs through co-immunoprecipitation (Co-IP) in 293 T cell with ectopic expression of INSM1 previously⁴⁴. We also observed the binding of the two proteins in HeLa cell with ectopic expression of INSM1 (Fig. 4A). As INSM1 is not expressed in 293 T and HeLa cells, we further test their interaction with small cell lung cancer cell NCI-H69. H69 cells have high endogenous expression of both TEAD1 and INSM1, and similarly their interaction was found using Co-IP (Fig. 4A). Bimolecular fluorescence complementation (BiFC) experiments in HeLa cell revealed the interaction between TEAD1 and INSM1 occurs in nucleus, mostly in nucleus speckles (Fig. 4B). The binding of INSM1 to TEAD1 is dependent on ZF1, as C287 mutation that unfolds ZF1 abolished the binding, while C297A and C369A unfolding ZF2 and ZF3 respectively did not (Fig. 4B). Yeast two hybrid (Y2H) experiments further confirmed ZF1 and TEAD are sufficient for the binding between INSM1 and TEAD1 (Fig. 4C), similar to the binding between Nerfin1 and Sd in *D. melanogaster*.

We then prepared the recombinant protein of the TEAD of TEAD1 along with the INSM1 ZF1 to study the binding mechanism between TEAD1 and INSM1. The binding affinity was evaluated using fluorescence polarization (FP) experiment. The K_D between TEAD1-TEAD and INSM1-ZF1 was determined to be 39.5 nM, indicating a high-affinity binding (Fig. 4D). Analytical ultracentrifugation (AUC) experiment indicated that ZF1 and TEAD bind with each other at a molar ratio of 1:1 (Fig. 4E). We further study their binding interface using NMR titration. When ¹⁵N-labeled ZF1 was titrated with unlabeled TEAD, remarkable changes of chemical shifts were found. Peaks from eight residues of ZF1 disappeared at the TEAD-bound state, and three residues showed CSPs over 0.1 ppm (Fig. 4F and H). On the other hand, titration of ¹⁵N-labeled TEAD with unlabeled ZF1 caused disappearance of TEAD peaks from seven residues and CSPs of seven residues greater than 0.1 ppm (Fig. 4G and I). These residues are located in β 1-strand and the α -helix of ZF1, and α 1 and α 2 helices of TEAD, which defined the binding interface between the two proteins (Fig. 4J and K). Most residues in the binding interface disappeared at the bound state suggested that they underwent conformation exchange at intermediate regime of NMR time scale.

As the disappearance of peaks for most interface residues, de novo calculation of ZF1-TEAD complex structure was not possible. Recently, the AlphaFold3 (AF3) server efficiently predicting protein complex structure was online⁴⁶. We tried to predict the ZF1-TEAD complex structure using AF3. The resulted model has a pTM of 0.78 and an ipTM of 0.72, which suggested a moderate confidence (Supplementary Fig. 8A). The resulted complex structure models showed high similarity at the bound interface (Supplementary Fig. 8B), which basically resembles that derived from the NMR titration data (Fig. 5A and 5B), indicating that the complex structure model predicted by AF3 probably reflects the conformation of the complex. In the structure model, the side chains of ten residues of TEAD were embedded into the shallow pits on ZF1 surface mainly through hydrophobic interactions. Two hydrogen bonds are formed between backbone O atom of E266 (ZF1) and the guanidinium group of R80 (TEAD), and between the backbone NH atom of Q270 (ZF1) and O atom of L79 (TEAD) (Supplementary Fig. 8C). SAXS analysis determined an R_g of 22.07 Å and a D_{max} of 72.76 Å for the complex of ZF1 and TEAD (Supplementary Fig. 8D, E, and Supplementary Table 3). Because the long flexible loop between α 1 and α 2 helices of TEAD, and the disordered region N-terminal to ZF1 may contributed a lot to the SAXS signals, the SAXS data can hardly be considered in complex structure modeling.

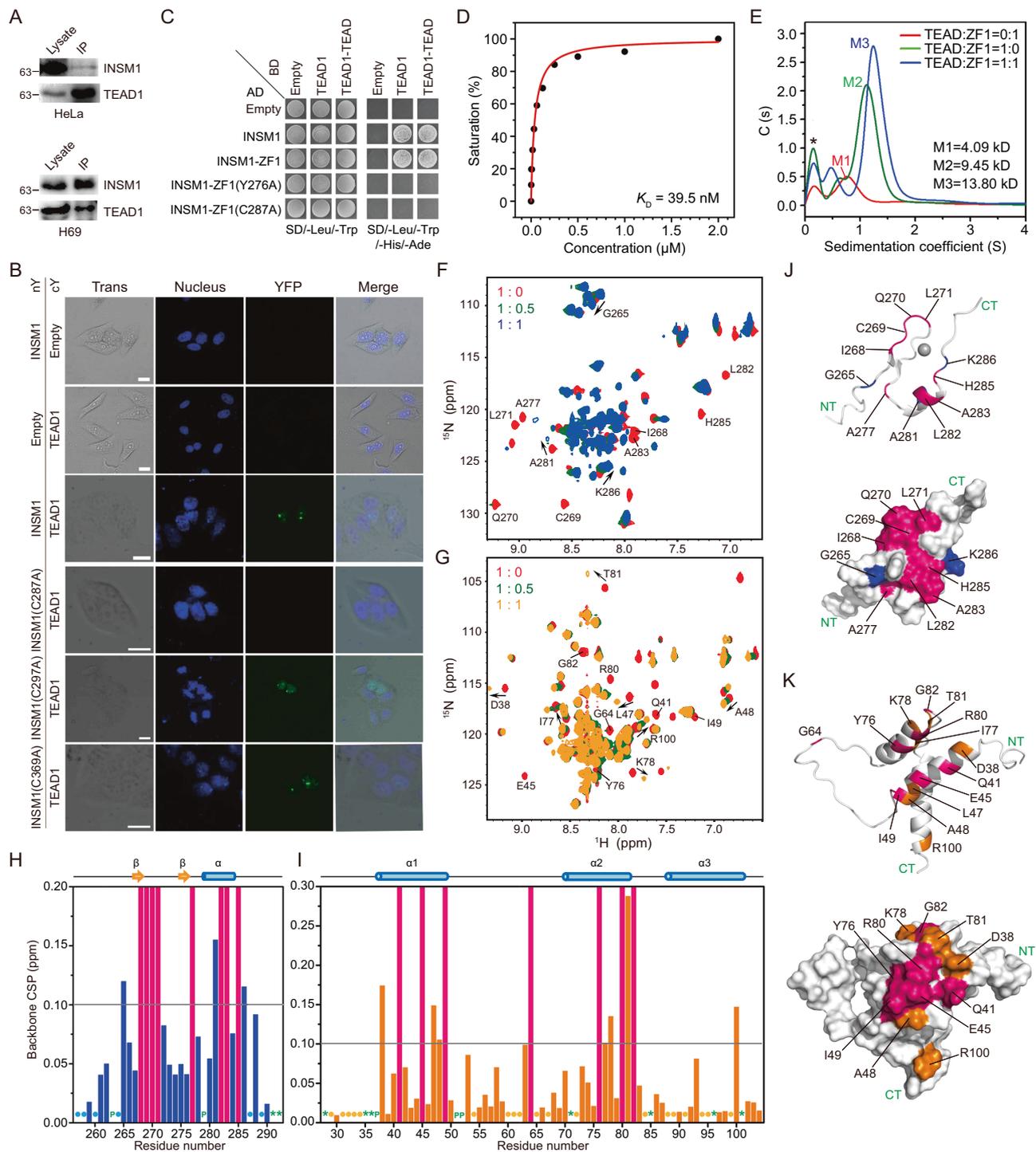
The importance of the residues located in the binding interface was further assessed through Y2H experiments with the mutants of each residue of TEAD1-TEAD and INSM1-ZF1. The mutated residues were picked according to the NMR titration results and complex structure model predicted by AF3. Sixteen residues of ZF1 and thirteen residues of TEAD were tested. The results showed that I268, L271, P279, F280, and A283 of ZF1, and Q41, E45, I49, Y50, R75, Y76, L79, and R80 of TEAD are crucial for the binding (Fig. 5C). F267 and L282 of ZF1 are highly conserved residues for ZF structure forming, and thus were not mutated for the Y2H test.

INSM1-ZF1, TEAD1-TEAD, and MCAT DNA form a ternary complex

The initially identified function of the TEAD1-TEAD is to recognize the MCAT DNA element (5'-CATTCC-3') in gene promoters⁴⁷. The binding mechanism of TEAD1-TEAD with MCAT DNA had been studied by NMR⁴⁸, and the complex structure had been solved by X-ray (PDB ID: 5NXX). It was found that the α 3 is the main DNA recognition helix which inserts into the major groove of DNA, and some residues from N-terminal loop, α 1- α 2 loop, and α 2 helix also contact with DNA (Fig. 5D–F). Four residues in α 3 including R87, K88, S92 and Q95 contribute to the specific binding to the core bases (5'-ATTCC-3') in the MCAT element through forming a hydrogen-bond network (Fig. 5D).

The DNA-binding interface of TEAD1-TEAD has almost no overlapped region with its interface for binding to INSM1-ZF1, and a superposition of the complex structures of TEAD1-TEAD:MCAT and TEAD1-TEAD:INSM1-ZF1 indicated no conflict for the simultaneous binding of TEAD1-TEAD with MCAT and INSM1-ZF1 (Fig. 5G). This inspired us to test whether a ternary complex can be formed. AUC assay discovered a complex with 20.6 kDa which is approximately equal to the sum of the MWs of TEAD1-TEAD, MCAT, and INSM1-ZF1, indicating the successful assembly of the ternary complex (Fig. 5H). SAXS analysis determined an R_g of 28.00 Å and a D_{max} of 99.04 Å for the ternary complex (Supplementary Fig. 8D, E, and Supplementary Table 3). Similar to the case of ZF1-TEAD complex, probably due to the SAXS signals contributed by the long flexible α 1- α 2 loop of TEAD, and the disordered region N-terminal to ZF1, the structure model of the ternary complex could not be well refined by the SAXS data.

Mutations of three residues of TEAD1-TEAD including R87, K88 and Q95 markedly impaired the binding with MCAT DNA (Fig. 5I), but did not affect the interaction between TEAD1-TEAD



and INSM1-ZF1 in Y2H experiment (Fig. 5C). On the contrary, mutations of R75, L79, and R80 which abolished the binding of TEAD1-TEAD with INSM1-ZF1 in Y2H experiment, did not affect the binding of TEAD1-TEAD with MCAT DNA. Taken together, a ternary complex containing TEAD1-TEAD, MCAT, and INSM1-ZF1 can be stably formed in vitro. We further tested the interactions between full-length INSM1 and TEAD1 variants containing R75A, L79A, R80A, R87A, K88A, and Q95A mutations, respectively via BiFC assays in HeLa cell. Similar to the Y2H results, R75A, L79A, and R80A mutants of TEAD1 failed to interact with INSM1, while R87A, K88A, and Q95A mutants can bind to INSM1 (Fig. 5J), which supported the proposal that INSM1, TEAD1, and MCAT element can form a ternary complex in cell.

INSM1 cooperates with TEAD1 to repress gene expression and cell proliferation

Given that INSM1, TEAD1 and MCAT element can form ternary complex, we checked whether INSM1 can bind to the MCAT elements in the promoters of TEAD1-targeted genes including *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, and *WTIP*⁴⁹, through chromatin-immunoprecipitation (ChIP) experiment. The results showed that INSM1 can bind to the promoters of these genes but not to *FAT3* and *HBB* genomic regions without MCAT element, similar to TEAD1 (Supplementary Fig. 9A). C287 mutation of INSM1 which destabilizes the structure of ZF1 and abolishes the interaction with TEAD1 but not affects INSM1 expression (Supplementary Fig. 9B), significantly decreased the binding of INSM1 with the promoters of these genes

Fig. 4 | Interaction of INSM1 with TEAD1. **A** Western blot images for the Co-IP experiments between INSM1 and TEAD1 in HeLa and H69 cells. IP was carried out using anti-Flag antibody for Flag-fused TEAD1 in HeLa cell, and using anti-INSM1 antibody in H69 cell. The experiment was repeated independently for three times with similar results. **B** BiFC between cYFP-fused TEAD1 and different variants of INSM1 fused with nYFP in HeLa cell. Bar, 20 μm . The experiment was repeated independently for three times with similar results. **C** Y2H assay between different variants of INSM1 and TEAD1. Transformants harboring both AD- and BD-derived constructs were grown on SD2 (-Trp/-Leu) medium for growth control and SD4 (-Trp/-Leu/-His/-Ade) medium for the interaction test. **D** Binding curve and K_D value between INSM1-ZF1 and TEAD1-TEAD determined by FP assay. Source data are provided as a Source Data file. **E** The fitted AUC data of INSM1-ZF1, TEAD1-TEAD, and their complex in sedimentation velocity experiments. The determined molecular weight for each peak is shown as inset. Asterisk marks the unspecific peak. Source data are provided as a Source Data file. **F** Overlay of a series of ^1H - ^{15}N HSQC

spectra of ^{15}N -labeled INSM1-ZF1 in the presence of unlabeled TEAD1-TEAD at different molar ratios, which are colored differently as indicated. **G** Overlay of a series of ^1H - ^{15}N HSQC spectra of ^{15}N -labeled TEAD1-TEAD in the presence of unlabeled INSM1-ZF1 at different molar ratios, which are colored differently as indicated. Chemical shift perturbation (CSP) for each residue of INSM1-ZF1 (**H**) and TEAD1-TEAD (**I**) during NMR titration is calculated and illustrated with the secondary structure elements shown on top. CSP value of 0.1 was selected as the significant level for globally comparison. Pink-filled columns mark the residues with disappeared signals during titration. The residues with CSP over 0.1 and those with disappeared signals were labelled in (**F**, **G**). Unassigned residues are marked with asterisks. P, proline. The residues with hardly any CSP are marked with cycles. Source data are provided as a Source Data file. The residues with CSP values over 0.1 are shown on the structures of INSM1-ZF1 (**J**, blue) and TEAD1-TEAD (**K**, orange) in cartoon and surface views, respectively. Residues with disappeared peaks in the spectra are shown and marked in pink.

(Fig. 6A), while TEAD1 can bind to these sites in absence of INSM1 (Supplementary Fig. 9C). Considering that TEAD1-TEAD exhibited a high affinity with K_D value of approximate 0.09 μM , while INSM1 ZF1-ZF5 showed a weak binding with K_D of 8.24 μM to MCAT DNA (Supplementary Fig. 10A), it can be suggested that INSM1 binds to the promoters of these genes indirectly through interacting with TEAD1. In addition, INSM1 ZF1 did not significantly affect the affinity of TEAD1-TEAD to MCAT DNA (Supplementary Fig. 10B).

The transcriptional regulation of INSM1 to *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, and *WTIP* was subsequently explored by comparing their expression levels with and without INSM1 overexpression in HeLa cell (Fig. 6B). The results indicated that INSM1 repressed the expression of these genes but not *FAT3* and *SLIT2* which are not TEAD1-targeted genes. The repressive role was enhanced when co-overexpressing with TEAD1. The C287A mutant of INSM1 did not show an enhanced repressive role when co-overexpressing with TEAD1, supporting the cooperative repression of gene expression by INSM1 and TEAD1. However, C287A mutant of INSM1 also exhibited a repressive effect to these genes when overexpressing alone, which may imply that INSM1 can also regulate the transcription of these genes in a manner independent on the interaction with TEAD1. Likewise, mutants of TEAD1 that impairs the binding of INSM1 did not show enhanced repressive effect on the expression of the six genes when co-overexpressing with INSM1, but had similar level of repressive effect as the wild type of TEAD1 when overexpressing alone (Fig. 6C). On the other hand, mutants of TEAD1 that impairs the binding to MCAT exhibited decreased repressive role to the expression of the six genes not only when overexpressing alone but also when co-overexpressing with INSM1, conforming to the prerequisite of promoter binding during transcription repression by TEAD1. Similar repressive role of INSM1 on *CTGF* promoter activity depending on TEAD1 was also verified in dual luciferase reporter assay (Supplementary Fig. 10C). In H69 cell with TEAD1 knockdown, the binding to INSM1 to promoters of *AJUBA*, *ANKRD1*, *CTGF*, and *WTIP* significantly decreased (Supplementary Fig. 11A, B), further supporting that the binding is dependent on TEAD1. Knockdown of INSM1 and TEAD1 both increased the expression of *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, and *WTIP* but not *FAT3* and *SLIT2* (Supplementary Fig. 11C), consistent with their repressive role in transcription regulation.

As both INSM1 and TEAD1 had been evidenced to play inhibitive roles in cell proliferation^{50,51}, we further examined whether they can regulate cell proliferation cooperatively. As expected, INSM1 and TEAD1 markedly inhibited the proliferation of HeLa and H69 cells when overexpressing separately (Fig. 6D). When co-overexpressing, the inhibitive effect was further increased. Mutations of ZF1 (C287A), ZF2 (C297A) and ZF3 (C369A) did not impair the inhibition role of INSM1 for cell proliferation when overexpressing alone. Nevertheless, mutation of ZF1 but not ZF2 and ZF3 abolished the increased inhibitive

effect in cell proliferation by co-overexpression of INSM1 and TEAD1, confirming that INSM1 and TEAD1 cooperatively inhibit the proliferation of HeLa and H69 cells through interaction. Likewise, the mutants of TEAD1 impairing the binding of INSM1 and those impairing the binding of MCAT did not show increased inhibitive effect on cell proliferation when co-overexpressing with INSM1, and the mutants impairing the MCAT binding exhibited less inhibitive effect on cell proliferation than the wild type when overexpressing alone (Fig. 6E). On the other hand, H69 cells with knockdown of TEAD1 showed increased proliferation (Supplementary Fig. 11D), consistent with the inhibitive role of TEAD1. Unexpectedly, H69 cells with knockdown of INSM1 displayed decreased proliferation, suggesting that INSM1 regulates cell proliferation through complicated mechanism and is partly associated with TEAD1.

Taken together, INSM1 and TEAD1 can interact with each other through ZF1 and TEAD and cooperatively repress gene expression and cell proliferation. Analysis of co-expression pattern of *INSM1* and *TEAD1* using Bgee database and ChIPBase revealed that the two genes are co-expressed in many human organs and cancers (Supplementary Tables 4–6), implying that the cooperation of INSM1 and TEAD1 has broad regulatory roles.

INSM1 cooperates with CTCF to target GGGG-contained genome loci

A previous study discovered three DNA motifs that are bound by INSM1 in cell using ChIP-seq, among which two are specifically recognized by FOXA2 and NEUROD1, respectively⁴². The third one that corresponds to the M2 motif in this study has been proven not to be recognized by INSM1 in the EMSA experiments. Thus, we hypothesized that there is a protein recognizing the M2 motif and recruiting INSM1 to the M2 element through interacting with INSM1, similar to the cases of TEAD1, FOXA2 and NEUROD1. In light of this, CTCF which recognizes a consensus sequence highly similar to the M2 element attracted our attention. CTCF has eleven tandem C2H2 ZFs (Supplementary Fig. 12A), and the consensus DNA bound by CTCF had been well characterized^{52,53}, which has only 1-bp mismatching with the M2 element (Fig. 7A). Mechanically, five ZFs of CTCF including ZF3/4/5/6/7 recognize the 15-bp core consensus sequence obeying the established C2H2 ZF-DNA recognition theory^{54,55}. Expectedly, the recombinant protein of CTCF ZF3-ZF7 exhibited high binding affinity to the M2 DNA used in this study, with a K_D of 98.4 nM (Supplementary Fig. 12B, C) in EMSA, which means about 50-fold higher affinity than that of INSM1 ZF1-ZF5 for the M2 DNA. This high affinity of CTCF to M2 DNA is consistent with the results of previous studies for its consensus DNA⁵⁴. In addition, the sequences of INSM1 ZF1/2/3/4/5 showed almost no similarity to those of CTCF ZF3/4/5/6/7 (Supplementary Fig. 12D), further supporting that INSM1 ZFs can't recognize the M2 element.

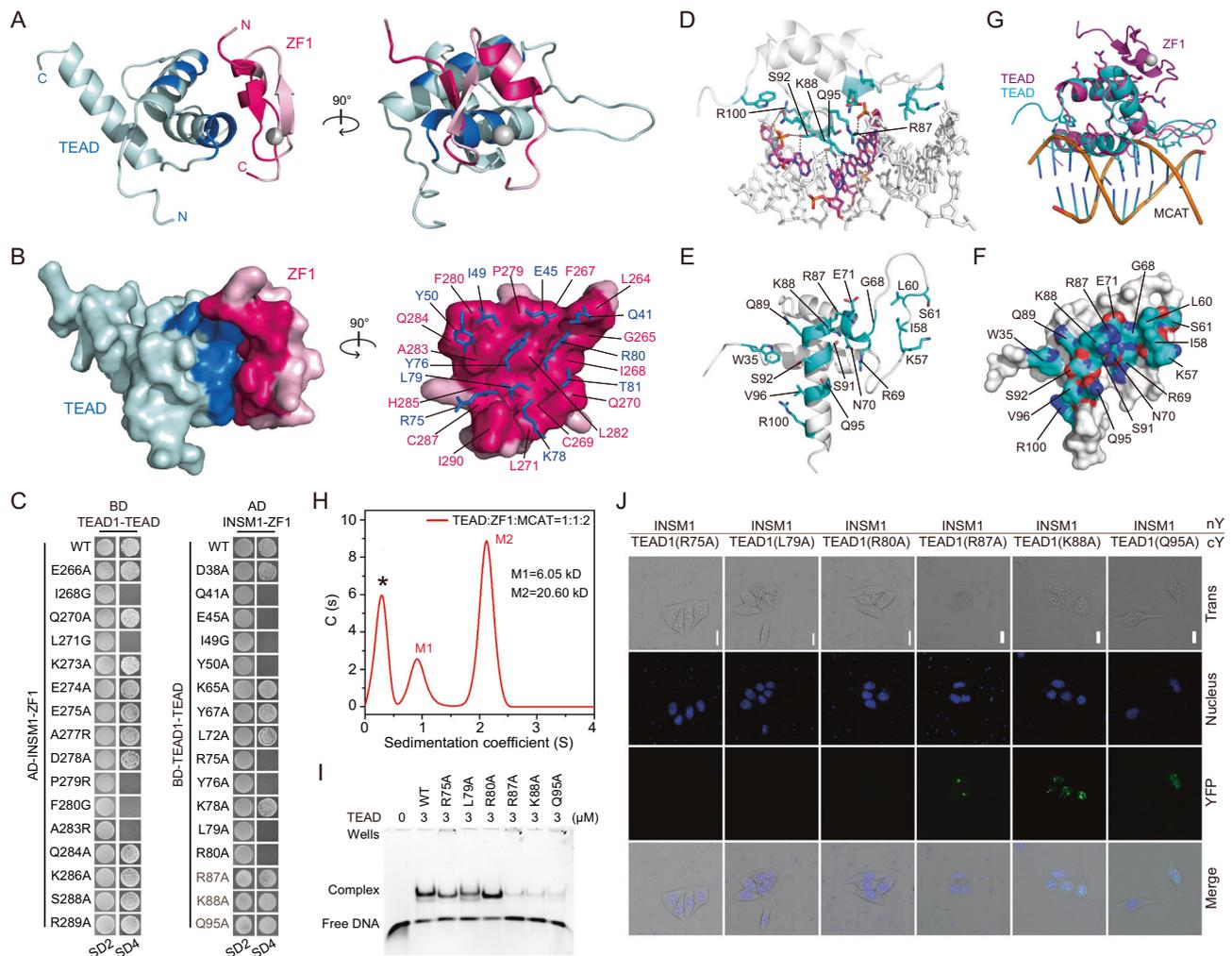


Fig. 5 | The interaction of TEAD1-TEAD with INSM1-ZF1 and MCAT DNA.

A Structural model of the complex of TEAD1-TEAD and INSM1-ZF1 modelled by AlphaFold3. The interfaces of TEAD1-TEAD and INSM1-ZF1 are colored in marine blue and pink, respectively. **B** Structural model of the complex of TEAD1-TEAD and INSM1-ZF1 presented in surface view (left). The residues in the interfaces of TEAD1-TEAD and INSM1-ZF1 are respectively labeled in marine blue and pink, with those residues from TEAD1-TEAD shown in sticks (right). **C** Y2H assay between different mutants of INSM1-ZF1 and TEAD1-TEAD. Transformants harboring both AD- and BD-derived constructs were grown on SD2 (-Trp/-Leu) medium for growth control and SD4 (-Trp/-Leu/-His/-Ade) medium for the interaction test. **D** Crystal structure of the complex of TEAD1-TEAD (cartoon) and MCAT DNA (sticks), PDB ID: 5NNX. The residues of TEAD1-TEAD forming hydrogen bonds (black dashed lines) with DNA are marked and shown as sticks. The residues on the interface of TEAD1-

TEAD for binding to MCAT DNA are shown in cartoon (**E**) and surface (**F**) views, respectively. **G** Superposition of the complex structure of TEAD1-TEAD and INSM1-ZF1 (purple) with that of TEAD1-TEAD and MCAT DNA (cyan). **H** The fitted AUC data of the mixture of INSM1-ZF1, TEAD1-TEAD, and MCAT DNA at indicated molar ratio in sedimentation velocity experiment. The determined molecular weight for each peak is shown as inset. Asterisk marks the unspecific peak. Source data are provided as a Source Data file. **I** EMSA of wild-type TEAD1-TEAD and its mutants with FAM-labeled MCAT DNA. The concentrations of TEAD1-TEAD variants are indicated on top. The experiment was repeated independently for three times with similar results. **J** BiFC between cYFP-fused TEAD1 mutants and INSM1 fused with nYFP in HeLa cell. Bar, 20 μ m. The experiment was repeated independently for three times with similar results.

The interaction between CTCF and INSM1 was subsequently examined through CoIP and BiFC experiments. CoIP data evidenced that INSM1 and CTCF can form complex not only when expressing ectopically in HeLa cell but also when expressing endogenously in H69 cell (Fig. 7B). BiFC data manifested that INSM1 interacts with CTCF in nucleus (Fig. 7C). Unlike the case of TEAD1, the C-terminal part of INSM1 harboring the five ZFs doesn't interact with CTCF. Instead, the N-terminal intrinsically disordered part of INSM1 contributes to the interaction. The region for interacting with CTCF was further mapped to residues 221-256 of INSM1 (Fig. 7D). The region of CTCF binding with INSM1 was mapped to the C-terminal part of CTCF, which is also disordered (Fig. 7C and Supplementary Fig. 12A), implying that the two interact with each other in a manner different from that between INSM1 and TEAD1.

We further checked whether CTCF can recruit INSM1 to its binding sites in genome. The binding of INSM1 to twelve sites that had been evidenced to be bound by CTCF in previous studies were determined by ChIP-PCR. These sites are adjacent to the genes including *CPM*, *EGRI*, *IFI6*, *MDM2*, *MYC*, *PUMA*, *KLK5*, *KLK8*, and *KLK9*, respectively⁵⁶⁻⁶⁰. The results indicated that INSM1 can bind to all tested sites, although the enrichments by INSM1 were lower than those by CTCF (Fig. 7E and Supplementary Fig. 13). The binding of INSM1 to these sites decreased in cell with CTCF knockdown (Supplementary Fig. 14A, B), suggesting that the binding is dependent on CTCF. The regulatory role of INSM1 on the expression of these nine genes was further assessed by comparing their expression levels with and without INSM1 overexpression. The results manifested that INSM1 can repress the expression of all nine genes (Fig. 7F). On the other hand, H69 cells with knockdown of INSM1 displayed higher expression of these nine genes

(Supplementary Fig. 14C). Collectively, these data demonstrated that INSM1 can bind to the gene promoters harboring the M2 elements and repress transcription cooperating with CTCF. Analysis of co-expression pattern of *INSM1* and *CTCF* using Bgee database and

ChIPBase revealed that the two genes are co-expressed in many human organs and cancers (Supplementary Table 7-9), implying that the cooperative transcription regulation by *INSM1* and *CTCF* may broadly exist.



Fig. 6 | Cooperative inhibition of gene expression and cell proliferation by INSM1 and TEAD1. **A** ChIP-qPCR analysis of INSM1 and TEAD1 binding to the MCAT elements in the promoters of *AJUBA*, *ANKRD1*, *CTGF*, and *CYR61* in HeLa cells co-transfected with TEAD1 and wild-type INSM1 or its C287A mutant. Genomic regions from *FAT3* and *HBB* genes without MCAT element were used as negative controls. Data are presented as mean values \pm SD of three independent experiments ($n = 3$). Two-sided Student's *t*-test was used to show the statistical difference. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The exact *P*-values are provided in Source Data. Source data are provided as a Source Data file. **B** Relative expression of *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, *WTIP*, *FAT3*, and *SLIT2* in HeLa cells co-transfected with TEAD1 and wild-type INSM1 or its C287A mutant. **C** Relative expression of *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, *WTIP*, *FAT3*, and *SLIT2* in HeLa cells co-transfected with INSM1 and wild-type TEAD1 or its mutants. The expression level of the

indicated gene in HeLa cells without transfection is set to 1 in **(B, C)**. Data are presented as mean values \pm SD of three independent experiments ($n = 3$). The differences between columns annotated with different letters are significant ($P < 0.05$ by one-way ANOVA and Tukey's test with adjustment, two-sided). The exact *P*-values were provided in Source Data. Source data are provided as a Source Data file. **D** Cell proliferation assays of HeLa and H69 cells co-transfected with TEAD1 and wild-type INSM1 or its mutants. **E** Cell proliferation assays of HeLa and H69 cells co-transfected with INSM1 and wild-type TEAD1 or its mutants. The absorbance of HeLa or H69 cells without transfection is set to 1 in **(D, E)**. Data are presented as mean values \pm SD of three independent experiments ($n = 3$). The differences between columns annotated with different letters are significant ($P < 0.05$ by one-way ANOVA and Tukey's test with adjustment, two-sided). The exact *P*-values were provided in Source Data. Source data are provided as a Source Data file.

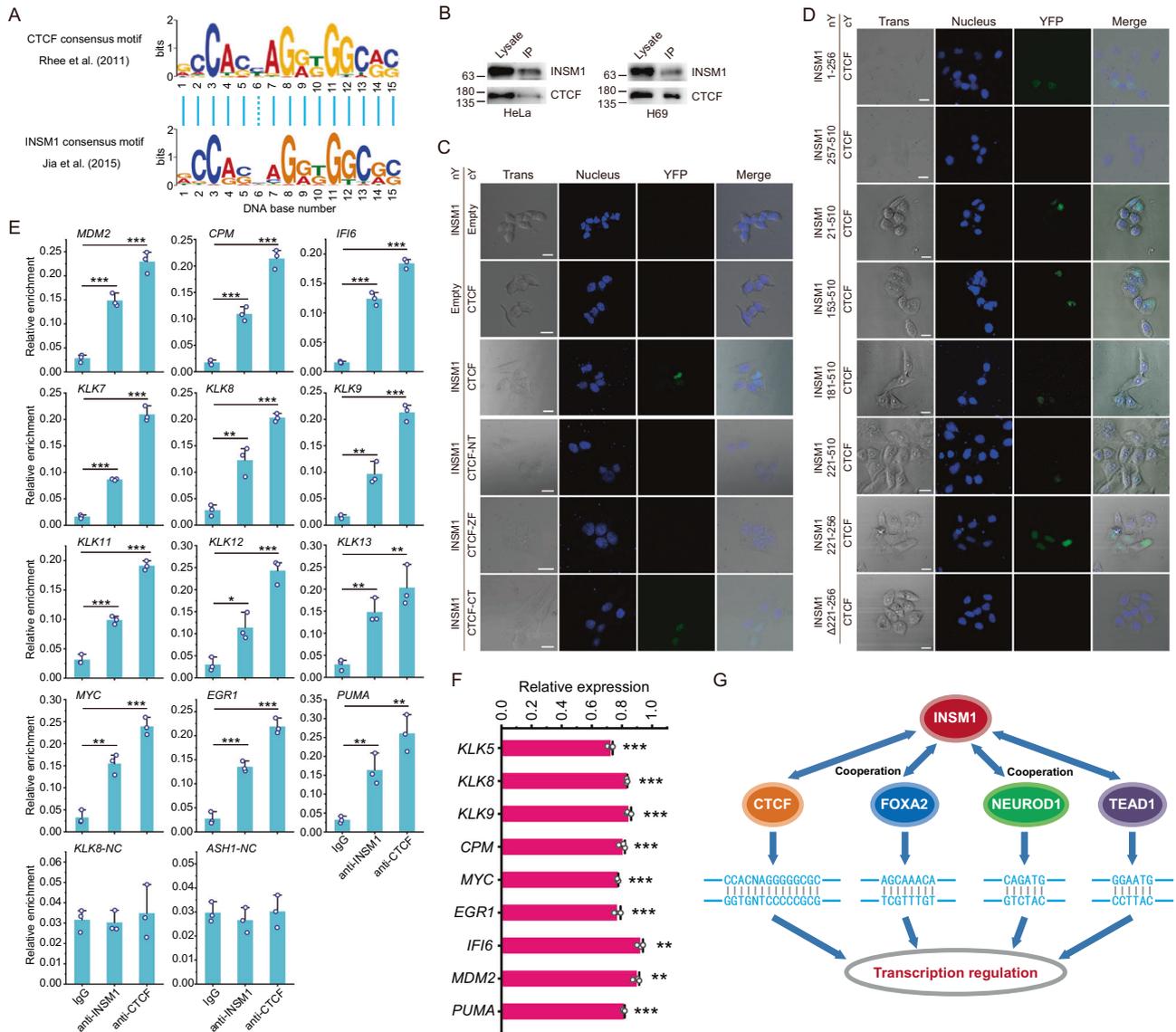


Fig. 7 | Cooperative repression of gene expression by INSM1 and CTCF.

A Consensus motifs in DNA bound by CTCF and INSM1 as determined by ChIP-seq (Reprinted from Cell, 147, Rhee, H.S. & Pugh, B.F., Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution, 1408–1419, Copyright (2011), with permission from Elsevier; Jia, S. et al. Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic beta-cell function. EMBO J, 2015, 34, 1417-1433). The similar nucleotides between the two motifs are connected by cyan lines. **B** Western blot images for the Co-IP experiments between INSM1 and CTCF in HeLa and H69 cells. IP was carried out using anti-INSM1 antibody. The experiment was repeated independently for three times with similar results. **C** BiFC between cYFP-fused CTCF variants and INSM1 fused with nYFP in HeLa cell. **D** BiFC between cYFP-fused CTCF and INSM1 variants fused with nYFP in HeLa cell. Bar, 20 μm. The experiment was repeated independently for three times with similar results. **E** CHIP-qPCR analysis of INSM1 and CTCF binding to the CTCF-targeted

genome loci in HeLa cells co-transfected with INSM1 and CTCF. Genomic regions from *KLK8* and *ASH1* genes without CTCF-binding site were used as negative controls (NC). Data are presented as mean values ± SD of three independent experiments (n = 3). Two-sided Student's t-test was used to show the statistical difference. *P < 0.05, **P < 0.01, ***P < 0.001. The exact P-values were provided in Source Data. Source data are provided as a Source Data file. **F** Relative expression of CTCF-targeted genes in HeLa cells transfected with INSM1. The expression level of CTCF-targeted genes in HeLa cells transfected with empty vector is set to 1. Data are presented as mean values ± SD of three independent experiments (n = 3). Two-sided Student's t-test was used to show the statistic difference. **P < 0.01, ***P < 0.001. The P-values were provided in Source Data. Source data are provided as a Source Data file. **G** A functional mode of INSM1 by cooperating with CTCF, FOXA2, NEUROD1, and TEAD1 to indirectly bind to different DNA elements in genome loci for regulating gene transcription.

We also mapped the region of INSM1 interacting with FOXA2 and NEUROD1 using BiFC experiments. Similar to the case of CTCF, INSM1 binds to FOXA2 and NEUROD1 through a disordered region containing residues 221-256 (Supplementary Fig. 15). This is consistent with a previous study which indicated that SNAG domain is not involved in the interaction of INSM1 with FOXA2 and NEUROD1¹⁹. Finally, to sum up the previous study and this study, we conclude that INSM1 itself may be not a TF in the usual sense that can bind to sequence-specific

DNA. It collaborates with other transcription regulators such as TEAD1, CTCF, NEUROD1 and FOXA2 to bind to sequence-specific DNA and regulate downstream gene expression (Fig. 7G).

Discussion

INSM1 is a ZF-containing protein playing a crucial role in the development of neuroendocrine and nervous system in mammals and is associated with many human neuroendocrine cancers^{7,31}. For more

than two decades, it was thought that INSM1 is a transcription repressor that binds to specific DNA elements through its ZFs⁴¹. At the same time, many proteins that interact with INSM1 have been identified, among which FOXA2, NEUROD1, and TEAD1 are TFs that bind to specific DNA elements, and a cooperation of INSM1 with FOXA2 and NEUROD1 in transcription regulation was demonstrated^{42,44}. Here, we comprehensively studied the solution structure of INSM1 ZFs, and investigated their binding with DNA and TEAD1. Based on these results and the discovery of CTCF as a new interacting partner of INSM1, a functional mode of INSM1 by cooperating with diverse DNA-binding TFs to indirectly bind to gene promoters for transcription regulation is proposed, which is distinct from canonical C2H2-ZF proteins.

Previous studies have shown that tandem ZFs that can bind to specific DNA sequences have the following characteristics⁶. The residues at positions -1, +2, +3, and +6 of the α -helix tend to be polar or positively charged. Residues with negative charge can also exist, but rarely. The linker between the ZFs generally contains 4 to 6 residues, which confers a certain degree of freedom but also conformation constraints. The tandem ZFs can synergistically bind to the large grooves of DNA to recognize continuous 3 + 3 bases of DNA motifs. By analyzing the structure of INSM1 ZFs, it can be found that only ZF3 has the most characteristics of typical DNA-binding ZF. ZF2 also meets some of the requirements. But it lacks the positively charged residues in the α -helix. Therefore, the CSP of ZF2 from the ZF1-ZF2 truncation was weak in the NMR titration with DNA, but was significantly enhanced in the ZF1-ZF5 truncation, which may be with the help of ZF3. However, the long linker connecting ZF2 and ZF3 lead to a highly flexible interdomain conformation, thereby eliminating the conformation constraints between ZF2 and ZF3 that are required for the collaborative recognition of sequence-specific DNA by tandem C2H2 ZFs. The surface of ZF1 is negatively charged, which basically prevents its binding to DNA. ZF4 and ZF5 are also positively charged, but the positive charges are not sufficiently located on the α -helix, and the residues at the +2 and +3 positions of α -helix are either glycine without side chains or alanine and proline with hydrophobic side chains in ZF4 and ZF5. These sites make it likely that ZF4 and ZF5 have very weak ability to recognize double-stranded DNA. These structure features provide reasonable explains for the low DNA affinity and incapable recognition of the M2 element of INSM1 ZF1-ZF5. In contrast, CTCF ZF3-ZF7 perfectly matched the characteristics of ZFs binding to DNA, and its binding affinity to the M2 element is far higher than that of INSM1 ZF1-ZF5.

INSM1 ZF1 adopts an unusual CCHC-type Zn coordination with short interval between the His and the last Cys residues, which prompts the forming of a shorter helix than typical C2H2 ZF. INSM1 ZF1 binds to TEAD1-TEAD with its β 1-strand and α -helix through hydrophobic interactions, representing a previously undiscovered binding mode of C2H2 ZF to other bio-macromolecules. Both INSM1 ZF1 and TEAD1-TEAD are common in animals, and their sequences are conserved in the evolution of species (Supplementary Fig. 16). Although INSM1 ZF1 is not as conserved as TEAD1-TEAD, the residues at the binding interface for TEAD1-TEAD are basically unchanged during evolution, indicating that their orthologs in these animals also interact with each other. The interaction of INSM1 with TEAD1 can be described as a very ancient transcriptional regulatory complex. This study provides a primary insight into the mechanisms of action of this partner in transcriptional regulation. A recent study has shown that TEAD1 regulates pancreatic β cell development in a way that is independent of its YAP-binding domain, called pocket-independent transcriptional inhibition⁵¹. We suppose that the partner of TEAD1 in this pocket-independent inhibition is likely INSM1, as INSM1 has a similar regulatory effect on the development of pancreatic β cells.

Searching of C2H2 ZF in INSM1 orthologs from different animals found that the number of ZFs varied a lot during evolution (Supplementary Fig. 17). ZF1 and ZF2 are the most conserved ZFs in INSM1,

while ZF3 is missing in several lower animals, such as *Caenorhabditis elegans* and *Helobdella robusta*. ZF4 and ZF5 appear to emerge later in evolution than ZF3. Interestingly, INSM1 orthologs from some lower animals in sea have a sixth ZF (ZF6, named after the ZFs of INSM1 in human) with sequence significant different from the ZFs in human INSM1. The linker connecting ZF2 and ZF3 also showed great changes in length, but the shortest one, which is in the INSM1 ortholog from *Tigriopus californicus* is still 27-residue long, far beyond the favorite length of 4–6 residues for canonical DNA-binding C2H2 ZFs. While the function of ZF1 in the cooperative transcription repression with TEAD1 is well elucidated, the meanings of the DNA binding of ZF2 and ZF3 remain unclear. We suspect that ZF2 and ZF3 may help INSM1 get close to and move along chromatin DNA to search for its interacting TFs such as TEAD1, NEUROD1 and FOXA2. It is also possible that ZF2 and ZF3 of INSM1 are involved in the shaping of chromatin 3D structure by CTCF. The specific functions of ZF4 and ZF5, and the uncommon ZF6, are not yet known and need to be explored in future studies.

A recent study reported that mouse *Insm1* regulates gene expression in medullary thymic epithelial cells by binding to chromatin, and the majority of the *Insm1* binding sites are co-occupied by *Aire*²¹, although the interaction between *Insm1* and *Aire* was not evidenced. This is in line with the functional mode of INSM1 claimed by us that INSM1 collaborates with other TFs to bind to sequence-specific DNA and regulate downstream gene expression. Different from TEAD1, INSM1 interacts with CTCF, NEUROD1, and FOXA2 through its N-terminal part with disordered structure. The region in CTCF for interacting with INSM1 was also predicted to be disordered. Study on the interaction mechanism between two regions with disordered structure is a big challenge. Whether these proteins fold into new structures when they interact with each other is not yet known. It is also possible that INSM1 interacts with CTCF indirectly by binding to a common third protein. It was shown that the SNAG motif in the N-terminal part of INSM1 can interact with histone methylase¹⁶, and a proline-rich region can interact with cyclin⁵⁰. In addition, Cbl-associated protein (CAP) was shown to interact with the proline-rich region of INSM1⁶¹. Subsequent studies are required for further exploring the interaction mechanism between these proteins and the N-terminal part of INSM1.

In the development process of mammals, INSM1 mainly plays a role in promoting the differentiation of nerve cells and neuroendocrine cells. The expression of INSM1 itself is highly spatio-temporal specific during development, and so its targeted genes. In adults, ectopic expression of INSM1 is related to diseases. The expression of INSM1 causes cancer cells to transform to gain neuroendocrine properties, which is conducive to the survival of cancer cells and attenuates the treatment of cancer. Therefore, it is necessary to target INSM1 protein for therapeutic intervention at specific cancer stages. At the same time, probes targeting INSM1 can also be used to detect whether cancer cells are undergoing neuroendocrine transformation. The interface of INSM1 ZF1 with TEAD1-TEAD may be a workable site for drug designing. Recently, TEAD1 and its homologs become hot targets for drug designing. Nevertheless, the current targeted site is focused on the pocket in YBD for binding to YAP and VGLL4^{62–64}. In light of our study, some attentions of drug designing should be paid on the interface between INSM1 ZF1 with TEAD1-TEAD for some neuroendocrine cancers like small cell lung cancer. Our structural data provide reference information for the design of drugs and probes.

The role of INSM in cell proliferation seems to be complicated and dependent on the cell type and context. A previous study revealed that overexpression of INSM1 caused cell cycle arrest and inhibited Panc-1 cell proliferation through affecting Cyclin D1 and CDK4⁵⁰. INSM1 ectopic expression inhibited H1975 cell proliferation, while H69 and H889 cell with knockdown of INSM1 showed reduced cell number⁶⁵. INSM1 silencing decreased viability of H510A, H1417, and H526 cell, which may be related to its interaction with LSD1³⁴. In this study,

overexpression INSM1 in HeLa and H69 cell inhibit proliferation partly through the interaction with TEAD1, while H69 cells with knockdown of INSM1 also showed reduced cell proliferation. A similar nonlinear relationship between expression level and phenotype was also observed for many genes, such as *SRPK1*⁶⁶, *MiR-125*⁶⁷, and *Psn*⁶⁸. As INSM1 interacts with many proteins, it possibly regulates cell proliferation through multiple pathways. The effect of change of INSM1 expression on cell proliferation may be related to the dominant pathway in specific cell type and the concentration of INSM1 in cell, which should be cautiously considered during drug test.

In summary, we analyzed the structure of the five ZFs of human INSM1. It is proved that ZF1 is an atypical ZF with CCHC type of Zn coordination, while ZF2/3/4/5 are typical C2H2 ZFs. The five ZFs adopt collapsed conformations with highly dynamic change of interdomain distance. ZF2 and ZF3 can interact with double-stranded DNA, but the affinity is at the micromolar level and do not have specificity to previous identified DNA elements. ZF1 can interact with the TEAD1-TEAD in a previously undiscovered mode and then bind to the MCAT DNA element indirectly to form a ternary complex. Mediated by ZF1, INSM1 can cooperate with TEAD1 to bind to the promoters of many genes, regulate their transcriptions, and then inhibit cell proliferation. On the other hand, INSM1 can use a region adjacent to the N-terminal of ZF1 to interact with CTCF, and work together with CTCF to bind to genome loci having GGGG-contained elements and inhibit the expression of genes nearby. Overall, INSM1 is possibly not a typical C2H2-ZF protein that can bind to sequentially specific DNA. It is more like a general transcriptional co-repressor, which can interact with at least four DNA-binding proteins, CTCF, TEAD1, NEUROD1, and FOXA2, and synergistically regulate the expression of downstream genes with them. This study significantly updates the understanding of the functional mechanism of INSM1, and will help in the design of drugs targeting INSM1 and TEAD1.

Methods

Production of recombinant proteins

The DNA fragments encoding ZF1 (residue 257-292), ZF1-ZF2 (residue 257-320), ZF3 (residue 346-396) and ZF1-ZF5 (residue 257-497) of INSM1 were cloned into a modified pET32 vector (pET32m) which allows the expression with an N-terminal 6×His-tag (MHHHHHHSSGLVPRGS). The DNA fragment encoding TEAD (residue 28-104) of TEAD1 was cloned into the pET32m to be expressed with C-terminal 6×His-tag (LVPRGSLHHHHHH). The DNA fragment encoding ZF3-ZF7 of CTCF (residue 317-467) was cloned into the pGEX4T vector to be expressed with an N-terminal GST-tag. The above plasmids were transformed into *Escherichia coli* strain BL21 (DE3) for protein expression, respectively.

The expression and purification of the truncated proteins of INSM1 including ZF1, ZF1-ZF2, ZF3, and ZF1-ZF5, were carried out following the same method as that for ZF4-ZF5 described previously³⁶. The expression and purification of TEAD1-TEAD and its mutants basically followed the method for the truncated proteins of INSM1, except that ZnCl₂ was not added to the LB or M9 medium. The 6×His-tag was removed through incubation with thrombin (Biosharp) at 25 °C for 4 h. For GST-fused CTCF ZF3-ZF7 protein, the *E. coli* cells were cultured in LB medium at 37°C until the OD₆₀₀ reached 0.6–0.8. Then, 0.3 mM IPTG was added to induce protein expression at 25°C for 10 h. The culture was harvested by centrifugation and lysed using the high-pressure homogenizer (ATS Engineering). The clarified supernatant was first purified by the ÄKTAexpress™ chromatography system (GE Healthcare) with a GSTrap™ FF column (5 mL), and then purified using the NGC™ chromatography system (Bio-rad) equipped with the HiLoad 26/60 Superdex™ 75 pg column.

NMR chemical shift assignment and structure calculation

The ¹⁵N,¹³C-labeled INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5 were concentrated to 0.6 mM in a buffer solution containing 10% D₂O(v/v),

20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl₂, and 0.02% NaN₃ at pH 6.5 for NMR data collection. A Bruker Avance III 600 MHz spectrometer and a Bruker Avance III 850 MHz spectrometer with a 5-mm HCN cryogenic probe were used to record the NMR data at 298 K. Backbone and side chain resonance assignments and structure calculation of INSM1 ZF1-ZF2 and ZF3 were carried out following the method described previously³⁶. The backbone and side chain resonances were 95.7% and 66.1% assigned for INSM1 ZF1-ZF2, and 95.4% and 61.2% assigned for INSM1 ZF3. The ¹H, ¹³C and ¹⁵N assignments for INSM1 ZF1-ZF2 and ZF3 have been deposited to the BioMagResBank (BMRB accession number 36573 and 36586). The final NMR ensembles of 20 structures of ZF1-ZF2 and ZF3 were deposited in the Protein Data Bank (PDB, ID: 8JPY and 8K81). Structural statistics and global structure quality factors are shown in Supplementary Tables 1 and 2). The backbone chemical shifts of INSM1 ZF1-ZF5 were assigned according to the assignments of ZF1-ZF2, ZF3, and ZF4-ZF5, and from the NMR data including HNCA and HN(CO)CA. Finally, 67% non-proline backbone chemical shifts were assigned. The backbone chemical shifts of INSM1 ZF1 were assigned according the assignments of ZF1-ZF2. The backbone chemical shifts of TEAD1-TEAD were assigned from the NMR data including HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH, and 90.7% non-proline backbone chemical shifts were assigned.

Small angle X-ray scattering (SAXS) data collection and analysis

SAXS data were collected in Shanghai Synchrotron Radiation Facility, at the BL19U2 BioSAXS beamline with a wavelength of 1.033 Å and a sample-to-detector distance of 2.6 m. The range of momentum transfer covered was 0.007 < q < 0.445 Å⁻¹ (q = 4π sinθ/λ, where 2θ is the scattering angle, λ is the wavelength). The data for INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5 were collected at two different concentrations (2 mg/mL and 4 mg/mL) and analyzed using program PRIMUS from the ASTAS package⁶⁹. No concentration-dependent effect was found after comparing the two I(q) curves of different concentrations. Data collected at high concentration (4 mg/mL) were used for further analysis. The data for ZF1-TEAD and ZF1-TEAD-MCAT complex were collected at a concentration of 300 μM and molar ratios of 1:1 and 1:1:1, respectively. The theoretical I(q) scattering curves for the NMR structures of INSM1 ZF1-ZF2 and ZF3 were calculated using CRY SOL⁷⁰. The particle distance distribution function P(r) and the maximum dimension (D_{max}) of the scattering particle were obtained for each sample using GNOM⁷¹. Ensemble optimization method (EOM)⁷² was used to generate an ensemble of various conformations for evaluating the interdomain flexibility of INSM1 ZF1-ZF5. The interdomain linkers (residues 289-293, 319-365, and 391-439 of INSM1) and N-terminal residues (257-265 of INSM1) were allowed to be flexible, while ZF1, ZF2, ZF3, and ZF4-ZF5 were treated as rigid bodies during EOM analysis. Sample details, data collection parameters, data analysis software, and structural parameters derived from SAXS data are summarized in Supplementary Table 3.

NMR titration

Single-stranded DNA (ssDNA) containing M2 element (5'-TACCAC-CAGGGGGCAGT-3') and its reverse-complement DNA were synthesized by Sangon Biotech Co., and were dissolved with Milli-Q water and mixed at equimolar concentration. The M2 double-stranded DNA (dsDNA) was obtained through annealing the two ssDNA in a buffer solution containing 10 mM Tris, 50 mM NaCl, 1 mM EDTA, at pH 7.5. Annealing was conducted by heating the mixed DNA to 95 °C in a water bath for 5 min and then slowly cooling it down to room temperature. After desalting, the DNA was lyophilized and re-suspended with the same buffer as that for INSM1 ZFs to a final concentration of 4 mM. NMR titration experiments were performed at 298 K using on a Bruker Avance III 600 MHz instrument. The M2 dsDNA was added to the sample solution containing 0.2 mM ¹⁵N-labeled ZF1-ZF2, ZF3, ZF4-ZF5,

and ZF1-ZF5, respectively, to the indicated molar ratios. The mixed samples were allowed to equilibrate for over 1 h, and subsequently used for ^1H - ^{15}N HSQC spectrum collection.

For NMR titration between INSM1 ZF1 and TEAD1-TEAD, unlabeled TEAD1-TEAD or INSM1 ZF1 was added to the sample solution containing 0.2 mM ^{15}N -labeled INSM1 ZF1 or TEAD1-TEAD, respectively, to the indicated molar ratios. The mixed samples were allowed to equilibrate for over 1 h before ^1H - ^{15}N HSQC spectrum collection. The equation used for calculating chemical shift perturbations (CSPs) was the same as described in the previous study⁷³.

Hydrogen–deuterium exchange mass spectrometry (HDX-MS)

INSM1 ZF1-ZF5 samples at a concentration of 100 μM and molar ratios of 1:0 (apo) and 1:2 (holo) with M2 dsDNA were prepared and allowed to equilibrate for 1 h. Peptide-level HDX-MS was performed as follows. Both apo- and holo- INSM1 ZF1-ZF5 samples were diluted with equilibrium buffer (100 mM sodium phosphate, pH 7.0) to a concentration of 20 μM . Briefly, at the start of an HDX reaction, at time zero ($t = 0$), protein solution at 20 μM was constituted with labeling buffer (100 mM sodium phosphate, D_2O , pD 6.6) at an approximate 20-fold dilution. The labeling mixtures were incubated at 20 $^\circ\text{C}$, and deuterium exchange was quenched at the following time points: 1 min, 10 min, 30 min, and 60 min. Deuterium labelling was quenched by making a 1:1 dilution with chilled quenching buffer (100 mM sodium phosphate, pH 2.0). Quenched samples were digested, desalted, and separated online using an Acquity UPLC M-Class system with HDX-2 automation coupled with Synapt XS HDMS. The online digestion was performed using an immobilized pepsin column, 2.1 mm \times 30 mm (Enzymate Pepsin Column, Waters Corp., Milford, MA, USA) for 5 min in 0.1% formic acid, with H_2O and ACN at a flow rate of 40 $\mu\text{L}/\text{min}$. The entire digestion was held at 20 $^\circ\text{C}$ within the column compartment of the HDX manager. The peptides were collected and desalted online by a trap column (ACQUITY UPLC BEH C18 VanGuard pre-column, 130 Å , 1.7 μm , 2.1 mm \times 5 mm, Waters), and subsequently separated with an Acquity UPLC BEH C18, 130 Å , 1.7 μm , 1 mm \times 100 mm column (Waters) held at 0 $^\circ\text{C}$. After 7 min of linear elution, eluent was directed into a Synapt XS HDMS and lockmass corrected. Mass spectra were acquired in MSE mode over the mass range of 100 to 2000. Blank injections were inserted after each sample injection to eliminate the effect of protein carryover. Peptic peptides were identified utilizing dual analytical approaches: exact mass determination coupled with MSE technology through ProteinLynx Global Server 3.0.3 (Waters Corp., Milford, MA, USA) with self-conducted databank referred to the amino acid sequence of INSM1 and default search parameters as following: primary digest reagent, non-specific; missed cleavages, 1; false discovery rate, 4. Quantitative deuterium incorporation analysis was performed at both peptide and residue resolution using Waters DynamX 3.0 software, with isotopic pattern centroid comparison between deuterated and native peptide ions. To ensure data reliability, rigorous MSE filtering criteria were implemented including a minimum sequence length of 5, a maximum peptide length of 20 residues, minimum products per amino acid of 0.3, and a maximum MH⁺ error of 20 ppm. A summary of HDX-MS experiment and data is shown in Supplementary Table 10.

Electrophoretic mobility shift assay (EMSA)

EMSA experiments were carried out using dsDNA probes with FAM label at the 5' end of one strand. The sequences of used DNA fragments are listed in Supplementary Table 11. Single-stranded DNAs were synthesized by Sangon Biotech Co., and then annealed to obtain dsDNAs as described above. For each sample in the binding assay, a final concentration of 0.01 μM dsDNA probe was mixed with the tested protein

at indicated concentration up to 10 μL . The mixed samples were allowed to equilibrate at 4 $^\circ\text{C}$ for 30 min, and then loaded into a 10% native PAGE gel in 0.5 \times TBE buffer. After separation by electrophoresis in ice bath, the gel was subjected to fluorescence detection by ChemiDoc MP instrument (Bio-rad).

Co-immunoprecipitation (Co-IP)

The DNA fragment encoding INSM1 (residue 1-510) was cloned into the pCDNA3.1-HA-N vector, while the DNA fragments encoding TEAD1 (residue 1-426) and CTCF (residue 1-727) were cloned into the pECMV-3 \times FLAG-N vector. Plasmid transfection of HeLa S3 cell was carried out using Hieff TransTM Liposomal Transfection Reagent (Yeasen, 40802ES02) following the manuals. Approximate 10^7 HeLa S3 cells (48 h after transfection) or NCI-H69 cells (without transfection) were lysed with RIPA buffer solution (50 mM Tris, pH7.4, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 1 mM EDTA, 1 \times protein inhibitor mixture (Solarbio, P6730), 1 mM PMSF) on ice for 20 min. Subsequently, 500 μL supernatant of the cell lysate with protein concentration of 1 mg/mL was incubated with anti-INSM1 (A-8) antibody conjugated to agarose (Santa Cruz, sc-271408 AC) or anti-Flag rabbit polyclonal antibody (BBI, D110005) bound by protein A/G plus-agarose (Santa Cruz, sc-2003) at 4 $^\circ\text{C}$ with gentle shaking overnight. After washing with RIPA buffer solution for 10 times, the agarose was recovered, and the immunocomplex was eluted by adding 4 \times SDS loading buffer followed by 5 min of boiling. The eluent was analyzed by western blot using the antibodies including anti-INSM1 (A-8) (Santa Cruz, sc-271408), anti-TEAD1 (E-5) (Santa Cruz, Sc-393976), anti-CTCF (G-8) (Santa Cruz, Sc-271474), and anti-Flag/OctA-Probe (H-5) (Santa Cruz, Sc-166355). The bands were visualized through chemiluminescence using the BeyoECL Moon kit (Beyotime, P0018FS-2), and imaged by ChemiDoc MP instrument (Bio-rad). Raw western blot images are shown in Source Data file.

Bimolecular fluorescence complementation (BiFC)

The DNA fragments coding INSM1 and its variants (Supplementary Fig. 1A) were inserted into the pCDNA3.1-HA-N vector between BamHI and EcoRI sites, while the DNA fragment coding nYFP was inserted into the XhoI site, to express the proteins fused with nYFP at the C-terminal. The DNA fragments encoding TEAD1 (residue 1–426), CTCF (residue 1–727), NEUROD1 (residue 1–356), FOXA2 (residue 1–463), and their variants were cloned into the pECMV-3 \times FLAG-N vector between the BamHI and EcoRI sites, while the DNA fragment coding cYFP was inserted into the XhoI site, to express the proteins fused with cYFP at the C-terminal. Different pairs of plasmids with nYFP and cYFP were co-transfected into HeLa S3 cell, respectively. 24 h after transfection, cells were incubated with NucBlueTM Live ReadyProbesTM Reagent (Invitrogen, R37605) at 37 $^\circ\text{C}$ for 10 min for nuclear staining, and then the fluorescence images were taken with a laser scanning confocal microscope (Nikon AI Ti). The correct expression of different INSM1 variants fused with nYFP was checked through western blot analysis with anti-HA antibody (F7) (Santa Cruz, Sc-7392) (Supplementary Fig. 1D).

Yeast two-hybrid (Y2H)

Yeast two-hybrid assays were performed using the Matchmaker Yeast Transformation System (Clontech, Palo Alto, CA, USA). The DNA fragments coding INSM1 and TEAD1 and their variants were inserted into pGADT7 and pGBKT7 vectors, respectively. Mutants of specific residues were obtained through PCR following the QuikChange site-directed mutagenesis method. Yeast AH109 cells were co-transformed with different pairs of pGADT7 and pGBKT7 constructs as indicated, according to the manuals. All yeast transformants were grown on SD2 (–Trp/–Leu) medium for validating successful transformation and SD4 (–Trp/–Leu/–His/–Ade) medium for testing the prey–bait interaction.

Fluorescence polarization (FP)

The recombinant INSM1 ZF1 protein was labeled with FITC with the FITC rapid labeling kit (Frdbio, ARL0021K) following the manuals. Free FITC was removed through ultrafiltration. The label-free TEAD1-TEAD and FITC-labeled INSM1 ZF1 were kept in the same buffer solution (100 mM NaH₂PO₄, 100 mM NaCl, pH 5.85). A series of binding mixtures were constructed for FP determination, in which the concentration of FITC-labeled INSM1 ZF1 was kept at 3 nM, while TEAD1-TEAD was continuously 2-fold diluted for 10 times with the maximum final concentration of 2 μM in the mixtures. By recording the excitation at 485 nm and the emission at 528 nm, the FP value of FITC-labeled INSM1 ZF1 bound by TEAD1-TEAD was measured using a SpectraMax i3x microplate reader (Molecular Devices). Each binding reaction was repeated three times, and the polarization value was averaged (in mP). The averaged polarization value of each titration point was subtracted with that of the free INSM1 ZF1 to obtain the final value. The dissociation constants (K_D) were calculated using the final polarization value of each titration point as previously described⁷⁴.

Analytical ultracentrifugation (AUC)

Sedimentation velocity experiments were performed at 20 °C on Proteomelab XL-I analytical ultracentrifuge instrument (Beckman Coulter, Brea, US). The samples of TEAD1-TEAD and INSM1 ZF1 at free state were at a concentration of 10 μM in a buffer solution containing 20 mM MES and 100 mM NaCl at pH 6.5. In the samples of ZF1-TEAD and ZF1-TEAD-MCAT mixture, the concentrations of TEAD1-TEAD and INSM1 ZF1 are both 10 μM, while the MCAT DNA was at a concentration of 20 μM. Data were collected every minute at 50,000 rpm and a wavelength of 280 nm. Sedimentation coefficient distribution, $c(s)$, was calculated from the sedimentation velocity data using SEDFIT software.

Chromatin immunoprecipitation (ChIP)

ChIP experiments were carried out using the BeyoChip™ enzymatic ChIP assay kit (Beyotime, P2083S) following the manuals. Approximately 10⁷ HeLa S3 cells (48 h after transfection) were used for each ChIP sample. Anti-INSM1 (A-8) (Santa Cruz, sc-271408), anti-TEAD1 (E-5) (Santa Cruz, Sc-393976), and anti-CTCF (G-8) (Santa Cruz, Sc-271474) antibodies were used for enriching the corresponding protein-DNA complex, and mouse IgG (BBI, D110503) was used as a negative control. The isolated chromatin sample before immunoprecipitation was used as the input control. The enrichment of DNA fragments was identified by traditional PCR, and quantified by quantitative real-time PCR (qPCR) with TB Green Premix Ex Taq™ II (TaKaRa, RR820A) reagent on a QuantStudio™ 3 Real-Time PCR Instrument (Thermo Fisher Scientific, A28132). The relative enrichment was calculated by normalizing the amount of a target DNA fragment after ChIP against that before ChIP (Input). Primers for specific genome loci are listed in Supplementary Table 12. The experiments were performed in biological and technical triplicates.

Gene expression analysis

RNA extraction was performed with approximately 10⁷ HeLa S3 cells (48 h after transfection) for each sample, using the RNAiso Plus reagent (TaKaRa, 108-95-2). Residual trace amounts of DNA were digested by DNase I (TaKaRa, 2270 A). Reverse transcription-PCR (RT-PCR) was carried out with 2 μg RNA using the SMART MMLV Reverse Transcriptase (Clontech, 639522). The expression of TEAD1 and CTCF-targeted genes was analyzed by qPCR with TB Green Premix Ex Taq™ II reagent on a QuantStudio™ 3 Real-Time PCR Instrument (Thermo Fisher Scientific, A28132). *GAPDH* was selected as an internal control. Gene-specific primers used in qPCR are listed in Supplementary

Table 13. The experiments were performed in biological and technical triplicates.

Dual luciferase reporter assay

A 770-bp DNA fragment for *CTGF* promoter region was amplified by PCR with primers 5'-ACGCGTCGACTTAGGGTTTACAGGTAGGC-3' and 5'-ACGCGTCGACGGTCATGGTTGGCACTGC-3', and inserted into the *Sall* site of p2LUC vector. The resulted plasmid was co-transfected with INSM1 variants and TEAD1 variants into HeLa cells. Firefly luciferase and Renilla luciferase activity were determined using a Dual Luciferase Reporter Assay Kit (Yeasen, 11402ES60) 48 h later according to manufacturer's protocol. Renilla luciferase activity was used as an internal control for normalization.

Gene knockdown

DNA fragments containing the target sequences of *INSM1*, *TEAD1*, and *CTCF* were synthesized by Sangon Biotech Co., annealed to obtain dsDNAs, and then inserted into the *BbsI* site of pSpCas9(BB)-2A-GFP (PX458) vector. The resulted plasmids were transfected into H69 cells. After 48 h, the expressions of INSM1, TEAD1, and CTCF were analyzed using western blot with specific antibodies, respectively. The expression of GAPDH was analyzed with anti-GAPDH antibody (Proteintech, 60004-1-Ig) as a loading control. Target sequences of *INSM1*, *TEAD1*, and *CTCF* are listed in Supplementary Table 14. Raw western blot images are shown in the Source Data file.

Cell proliferation assay

Cell proliferation was determined using MTS assay. Briefly, after removing the culture medium, the HeLa S3 and NCI-H69 cells (48 h after transfection) were rinsed with PBS solution once, and then 100 μL of fresh culture medium and 20 μL of the CellTiter 96 Aqueous One Solution reagent (Promega, G3580) were added for incubation for 2 h. Subsequently, the absorbance at 490 nm was recorded for each sample using a SpectraMax i3x microplate reader (Molecular Devices). Three parallel experiments were conducted each time.

Reporting summary

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

Data availability

Source data are provided with this paper. The Structure coordinate data for ZF1-ZF2 and ZF3 of INSM1 generated in this study have been deposited in the Protein Data Bank database with ID [8JPY](#) and [8K81](#). The crystal structure of the complex of TEAD1-TEAD with MCAT DNA can be found in the Protein Data Bank under ID [5NXX](#). The NMR chemical shift assignment data are deposited in the Biological Magnetic Resonance Data Bank with ID [36573](#) and [36586](#). Source data are provided with this paper.

References

1. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
2. Clark, K. L., Yutzey, K. E. & Benson, D. W. Transcription factors and congenital heart defects. *Annu Rev. Physiol.* **68**, 97–121 (2006).
3. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
4. Najafabadi, H. S. et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015).
5. Persikov, A. V. et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res* **43**, 1965–1984 (2015).

6. Zhang, X., Blumenthal, R. M. & Cheng, X. Updated understanding of the protein-DNA recognition code used by C2H2 zinc finger proteins. *Curr. Opin. Struct. Biol.* **87**, 102836 (2024).
7. Lan, M. S. & Breslin, M. B. Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation. *FASEB J.* **23**, 2024–2033 (2009).
8. Tourniere, O., Gahan, J. M., Busengdal, H., Bartsch, N. & Rentsch, F. Insm1-expressing neurons and secretory cells develop from a common pool of progenitors in the sea anemone *Nematostella vectensis*. *Sci. Adv.* **8**, eabi7109 (2022).
9. Wu, J., Duggan, A. & Chalfie, M. Inhibition of touch cell fate by egl-44 and egl-46 in *C. elegans*. *Genes Dev.* **15**, 789–802 (2001).
10. Forbes-Osborne, M. A., Wilson, S. G. & Morris, A. C. Insulinoma-associated 1a (Insm1a) is required for photoreceptor differentiation in the zebrafish retina. *Dev. Biol.* **380**, 157–171 (2013).
11. Frolidi, F. et al. The transcription factor Nerfin-1 prevents reversion of neurons into neural stem cells. *Genes Dev.* **29**, 129–143 (2015).
12. Tavano, S. et al. Insm1 induces neural progenitor delamination in developing neocortex via downregulation of the adherens junction belt-specific protein *Plekha7*. *Neuron* **97**, 1299–1314 e8 (2018).
13. Farkas, L. M. et al. Insulinoma-associated 1 has a panneurogenic role and promotes the generation and expansion of basal progenitors in the developing mouse neocortex. *Neuron* **60**, 40–55 (2008).
14. Monaghan, C. E. et al. REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation-differentiation balance in the developing brain. *Proc. Natl Acad. Sci. USA* **114**, E406–E415 (2017).
15. Wildner, H., Gierl, M. S., Strehle, M., Pla, P. & Birchmeier, C. Insm1 (IA-1) is a crucial component of the transcriptional network that controls differentiation of the sympatho-adrenal lineage. *Development* **135**, 473–481 (2008).
16. Welcker, J. E. et al. Insm1 controls development of pituitary endocrine cells and requires a SNAG domain for function and for recruitment of histone-modifying factors. *Development* **140**, 4947–4958 (2013).
17. Gierl, M. S., Karoulias, N., Wende, H., Strehle, M. & Birchmeier, C. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev.* **20**, 2465–2478 (2006).
18. Mellitzer, G. et al. IA1 is NGN3-dependent and essential for differentiation of the endocrine pancreas. *EMBO J.* **25**, 1344–1352 (2006).
19. Liang, X. et al. The SNAG domain of Insm1 regulates pancreatic endocrine cell differentiation and represses beta- to delta-cell transdifferentiation. *Diabetes* **70**, 1084–1097 (2021).
20. Givony, T. et al. Thymic mimetic cells function beyond self-tolerance. *Nature* **622**, 164–172 (2023).
21. Tao, W. et al. Insm1 regulates mTEC development and immune tolerance. *Cell Mol. Immunol.* **20**, 1472–1486 (2023).
22. James, K. D. & Cowan, J. E. Insm1: orchestrating cellular mimicry in the thymus medulla. *Cell Mol. Immunol.* **21**, 416–418 (2024).
23. Wiwatpanit, T. et al. Trans-differentiation of outer hair cells into inner hair cells in the absence of INSM1. *Nature* **563**, 691–695 (2018).
24. Li, S., He, S., Lu, Y., Jia, S. & Liu, Z. Epistatic genetic interactions between Insm1 and *Ikzf2* during cochlear outer hair cell development. *Cell Rep.* **42**, 112504 (2023).
25. Rosenbaum, J. N. et al. INSM1: A novel immunohistochemical and molecular marker for neuroendocrine and neuroepithelial neoplasms. *Am. J. Clin. Pathol.* **144**, 579–591 (2015).
26. Chen, C., Notkins, A. L. & Lan, M. S. Insulinoma-Associated-1: From neuroendocrine tumor marker to cancer therapeutics. *Mol. Cancer Res.* **17**, 1597–1604 (2019).
27. Kobayashi, S. et al. Alleles of Insm1 determine whether RIP1-Tag2 mice produce insulinomas or nonfunctioning pancreatic neuroendocrine tumors. *Oncogenesis* **8**, 16 (2019).
28. Mukhopadhyay, S., Dermawan, J. K., Lanigan, C. P. & Farver, C. F. Insulinoma-associated protein 1 (INSM1) is a sensitive and highly specific marker of neuroendocrine differentiation in primary lung neoplasms: an immunohistochemical study of 345 cases, including 292 whole-tissue sections. *Mod. Pathol.* **32**, 100–109 (2019).
29. Tsai, H. K., Hornick, J. L. & Vivero, M. INSM1 expression in a subset of thoracic malignancies and small round cell tumors: rare potential pitfalls for small cell carcinoma. *Mod. Pathol.* **33**, 1571–1580 (2020).
30. Asrani, K. et al. Reciprocal YAP1 loss and INSM1 expression in neuroendocrine prostate cancer. *J. Pathol.* **255**, 425–437 (2021).
31. Rindi, G. et al. Overview of the 2022 WHO classification of neuroendocrine neoplasms. *Endocr. Pathol.* **33**, 115–154 (2022).
32. Akerstrom, V., Chen, C., Lan, M. S. & Breslin, M. B. Modifications to the INSM1 promoter to preserve specificity and activity for use in adenoviral gene therapy of neuroendocrine carcinomas. *Cancer Gene Ther.* **19**, 828–838 (2012).
33. Tseng, A. W. S., Chen, C. C., Breslin, M. B. & Lan, M. S. Tumor-specific promoter-driven adenoviral therapy for insulinoma. *Cell. Oncol.* **39**, 279–286 (2016).
34. Takagi, S. et al. LSD1 inhibitor T-3775440 inhibits SCLC cell proliferation by disrupting LSD1 interactions with SNAG domain proteins INSM1 and GF11B. *Cancer Res.* **77**, 4652–4662 (2017).
35. Lan, M. S. & Chen, C. C. Small molecules targeting INSM1 for the treatment of high-risk neuroblastoma. *Biology* **12**, 1134 (2023).
36. Zhu, J. et al. Solution NMR structure of zinc finger 4 and 5 from human INSM1, an essential regulator of neuroendocrine differentiation. *Proteins* **85**, 957–962 (2017).
37. Liu, W. D., Wang, H. W., Muguira, M., Breslin, M. B. & Lan, M. S. INSM1 functions as a transcriptional repressor of the neuroD/beta2 gene through the recruitment of cyclin D1 and histone deacetylases. *Biochem. J.* **397**, 169–177 (2006).
38. Wang, H. W. et al. Identification of an INSM1-binding site in the insulin promoter: negative regulation of the insulin gene transcription. *J. Endocrinol.* **198**, 29–39 (2008).
39. Osipovich, A. B. et al. Insm1 promotes endocrine cell differentiation by modulating the expression of a network of genes that includes *Neurog3* and *Ripply3*. *Development* **141**, 2939–2949 (2014).
40. Jia, S., Wildner, H. & Birchmeier, C. Insm1 controls the differentiation of pulmonary neuroendocrine cells by repressing *Hes1*. *Dev. Biol.* **408**, 90–98 (2015).
41. Breslin, M. B., Zhu, M., Notkins, A. L. & Lan, M. S. Neuroendocrine differentiation factor, IA-1, is a transcriptional repressor and contains a specific DNA-binding domain: identification of consensus IA-1 binding sequence. *Nucleic Acids Res.* **30**, 1038–1045 (2002).
42. Jia, S. et al. Insm1 cooperates with *Neurod1* and *Foxa2* to maintain mature pancreatic beta-cell function. *EMBO J.* **34**, 1417–1433 (2015).
43. Vissers, J. H. A. et al. The Scalloped and Nerfin-1 transcription factors cooperate to maintain neuronal cell fate. *Cell Rep.* **25**, 1561–1576 e7 (2018).
44. Guo, P. et al. Nerfin-1 represses transcriptional output of Hippo signaling in cell competition. *Elife* **8**, e38843 (2019).
45. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464 (2007).
46. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
47. Farrance, I. K., Mar, J. H. & Ordahl, C. P. M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. *J. Biol. Chem.* **267**, 17234–17240 (1992).
48. Anbanandam, A. et al. Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proc. Natl Acad. Sci. USA* **103**, 17225–17230 (2006).

49. Zanonato, F. et al. Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* **17**, 1218–1227 (2015).
50. Zhang, T., Liu, W. D., Saunee, N. A., Breslin, M. B. & Lan, M. S. Zinc finger transcription factor INSM1 interrupts cyclin D1 and CDK4 binding and induces cell cycle arrest. *J. Biol. Chem.* **284**, 5574–5581 (2009).
51. Li, F. et al. TEAD1 regulates cell proliferation through a pocket-independent transcription repression mechanism. *Nucleic Acids Res.* **50**, 12723–12738 (2022).
52. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–5231 (2008).
53. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
54. Hashimoto, H. et al. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* **66**, 711–720 e3 (2017).
55. Yin, M. et al. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.* **27**, 1365–1377 (2017).
56. Li, J. et al. An alternative CTCF isoform antagonizes canonical CTCF occupancy and changes chromatin architecture to promote apoptosis. *Nat. Commun.* **10**, 1535 (2019).
57. Khoury, A. et al. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.* **11**, 54 (2020).
58. Gombert, W. M. & Krumm, A. Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. *PLoS One* **4**, e6109 (2009).
59. Gomes, N. P. & Espinosa, J. M. Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. *Genes Dev.* **24**, 1022–1034 (2010).
60. Sekiya, T., Kato, K., Kawaguchi, A. & Nagata, K. Involvement of CTCF in transcription regulation of EGR1 at early G1 phase as an architecture factor. *Sci. Rep.* **9**, 329 (2019).
61. Xie, J., Cai, T., Zhang, H., Lan, M. S. & Notkins, A. L. The zinc-finger transcription factor INSM1 is expressed during embryo development and interacts with the Cbl-associated protein. *Genomics* **80**, 54–61 (2002).
62. Gibault, F., Sturbaut, M., Bailly, F., Melnyk, P. & Cotellet, P. Targeting Transcriptional Enhanced Associate Domains (TEADs). *J. Med. Chem.* **61**, 5057–5072 (2018).
63. Bum-Erdene, K. et al. Small-molecule cyanamide Pan-TEAD.YAP1 covalent antagonists. *J. Med. Chem.* **66**, 266–284 (2023).
64. Lu, W. et al. Structure-based design of Y-shaped covalent TEAD inhibitors. *J. Med. Chem.* **66**, 4617–4632 (2023).
65. Fujino, K. et al. Insulinoma-associated Protein 1 is a crucial regulator of neuroendocrine differentiation in lung cancer. *Am. J. Pathol.* **185**, 3164–3177 (2015).
66. Wang, P. et al. Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of Akt. *Mol. Cell* **54**, 378–391 (2014).
67. Li, Z. et al. Overexpression and knockout of miR-126 both promote leukemogenesis. *Blood* **126**, 2005–2015 (2015).
68. Gao, Y. H., Wen, D. T., Du, Z. R., Wang, J. F. & Wang, S. J. Muscle Psn gene combined with exercise contribute to healthy aging of skeletal muscle and lifespan by adaptively regulating Sirt1/PGC-1 α and arm pathway. *PLoS One* **19**, e0300787 (2024).
69. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
70. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLO - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
71. Semenyuk, A. V. & Svergun, D. I. Gnom - a program package for small-angle scattering data-processing. *J. Appl. Crystallogr.* **24**, 537–540 (1991).
72. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr J* **2**, 207–217 (2015).
73. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
74. Wang, Z. et al. Solution structure of SHIP2 SH2 domain and its interaction with a phosphotyrosine peptide from c-MET. *Arch. Biochem. Biophys.* **656**, 31–37 (2018).

Acknowledgements

This work was supported by funds from the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant number: XDB0540000 to Y. Y.), the National Natural Science Foundation of China (grant numbers: 22327901 to M. L., 21703283 to J. Z., 21991080 to M. L., 22074152 to Y. Y., 22374155 to R. H., and 21921004 to X. Zhou.), and National Key R&D Program of China (Grant number: 2022YFF0707000 to Y. L.). We thank Prof. Liang Chen at Wuhan Botanical Garden, Chinese Academy of Sciences for help in determining the zinc ion content of INSM1 protein sample through ICP-MS, and Dr. Yiwen Li, Dr. Guangfeng Liu, Qipan Song, and the staff of BL19U2 BioSAXS beamline at Shanghai Synchrotron Radiation Facility for assistance during SAXS data collection.

Author contributions

J. Z. and Y. Y. conceived the study and designed the experiments. H. Z. and Y. G. produced the protein samples. Y. X. and S. L. performed the NMR measurements. X. H. and S. L. assigned the chemical shifts and performed the structure calculation. H. Z. conducted the biochemical and cell experiments. X. H. performed SAXS experiments. Y. Z. carried out the HDX-MS experiments. R. H., Y. L., and J. Z. analyzed the data. X. Zhou., X. Zhang., and M. L. contributed reagents/materials/analysis tools. H. Z., X. H., and J. Z. designed the figures and wrote the manuscript with the contributions of all authors. All authors approved 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/s41467-025-57478-2>.

Correspondence and requests for materials should be addressed to Jiang Zhu or Yunhuang Yang.

Peer review information *Nature Communications* thanks Shiqi Jia, Fudong Li, and the other, 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