Supplementary Information

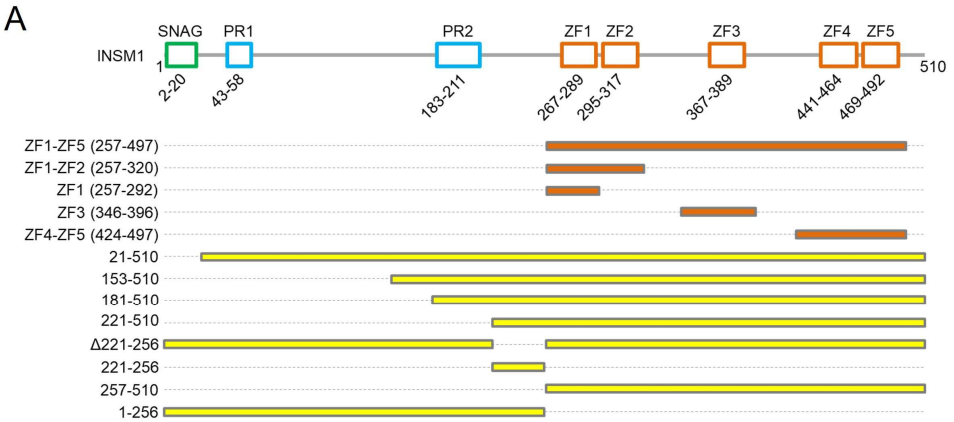
Structural insights into a highly flexible zinc finger module unravel INSM1 function in transcription regulation

Heng Zhou[1,2,#], Xiaoling He[1,#], Yue Xiong[1,2,#], Yixuan Gong[1,2], Yuanyuan Zhang[1,2], Shuangli Li[1], Rui Hu[1,2], Ying Li[1,2], Xu Zhang[1,2], Xin Zhou[1,2], Jiang Zhu[1,2,*], Yunhuang Yang[1,2,*] and Maili Liu[1,2]

[1] State 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, 430071, China
[2] University of Chinese Academy of Sciences, Beijing, 10049, China

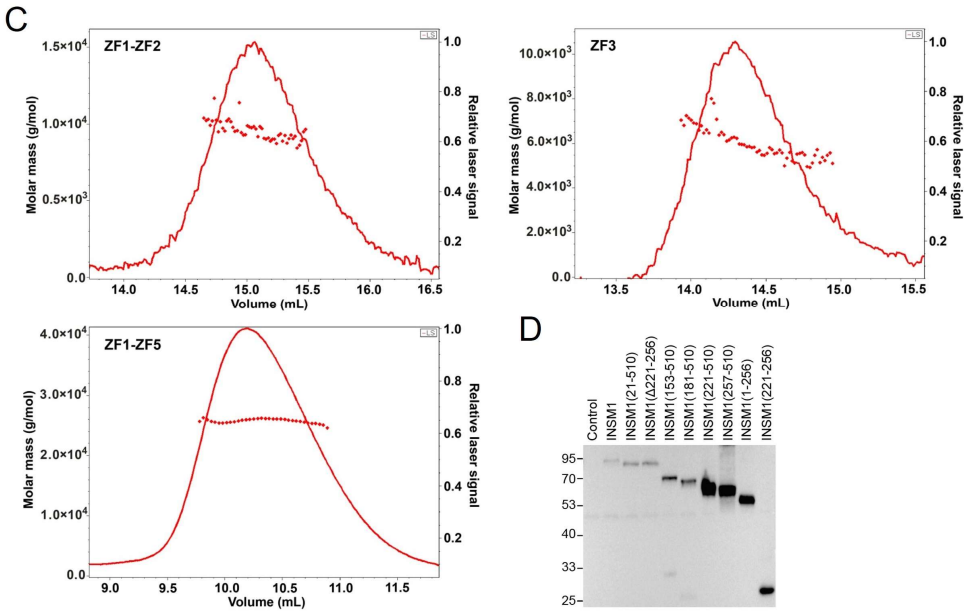
[*]Corresponding author: Jiang Zhu, Email: jiangzhu@apm.ac.cn; Yunhuang Yang, Email: yang_yh@apm.ac.cn.

[#]The authors contributed equally to this work.

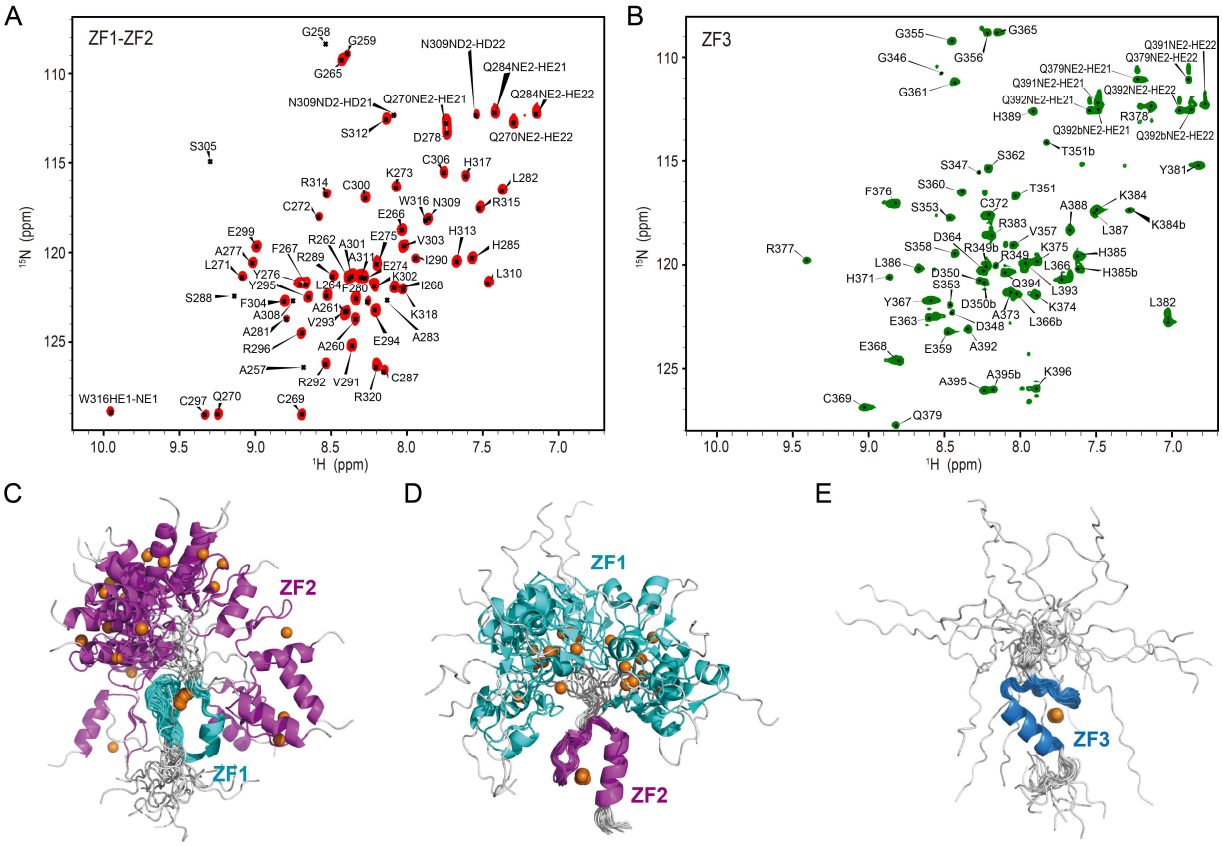


B

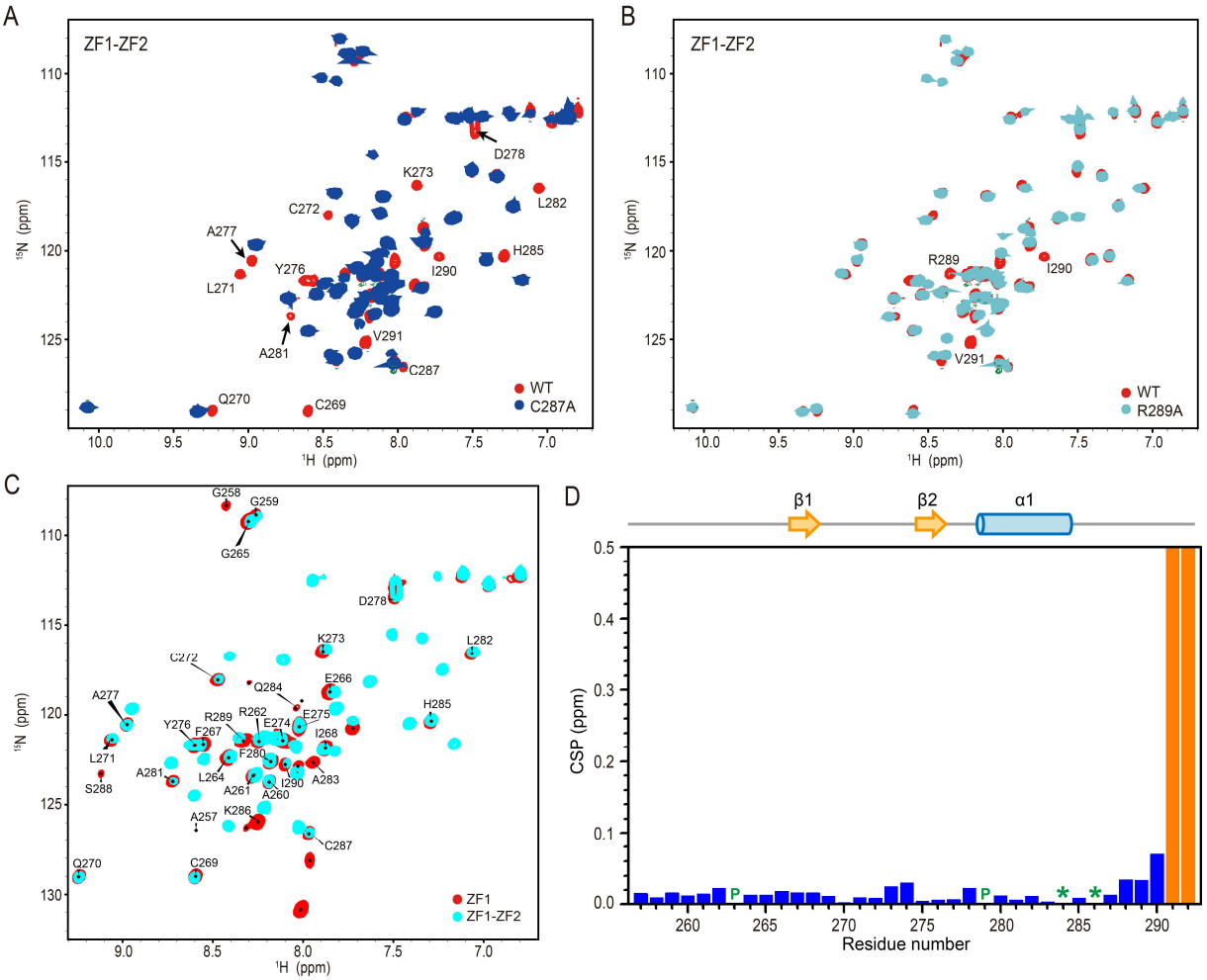
Sample	Molecular weight (g/mol)	Zn^{2+} concentration (mg/g)	Zn^{2+} mass (g/mol)	Zn^{2+} number
ZF1-ZF5	25976	12.08	313.78	4.83
ZF1-ZF2	7417	18.73	138.92	2.14
ZF3	5674	13.46	76.37	1.17



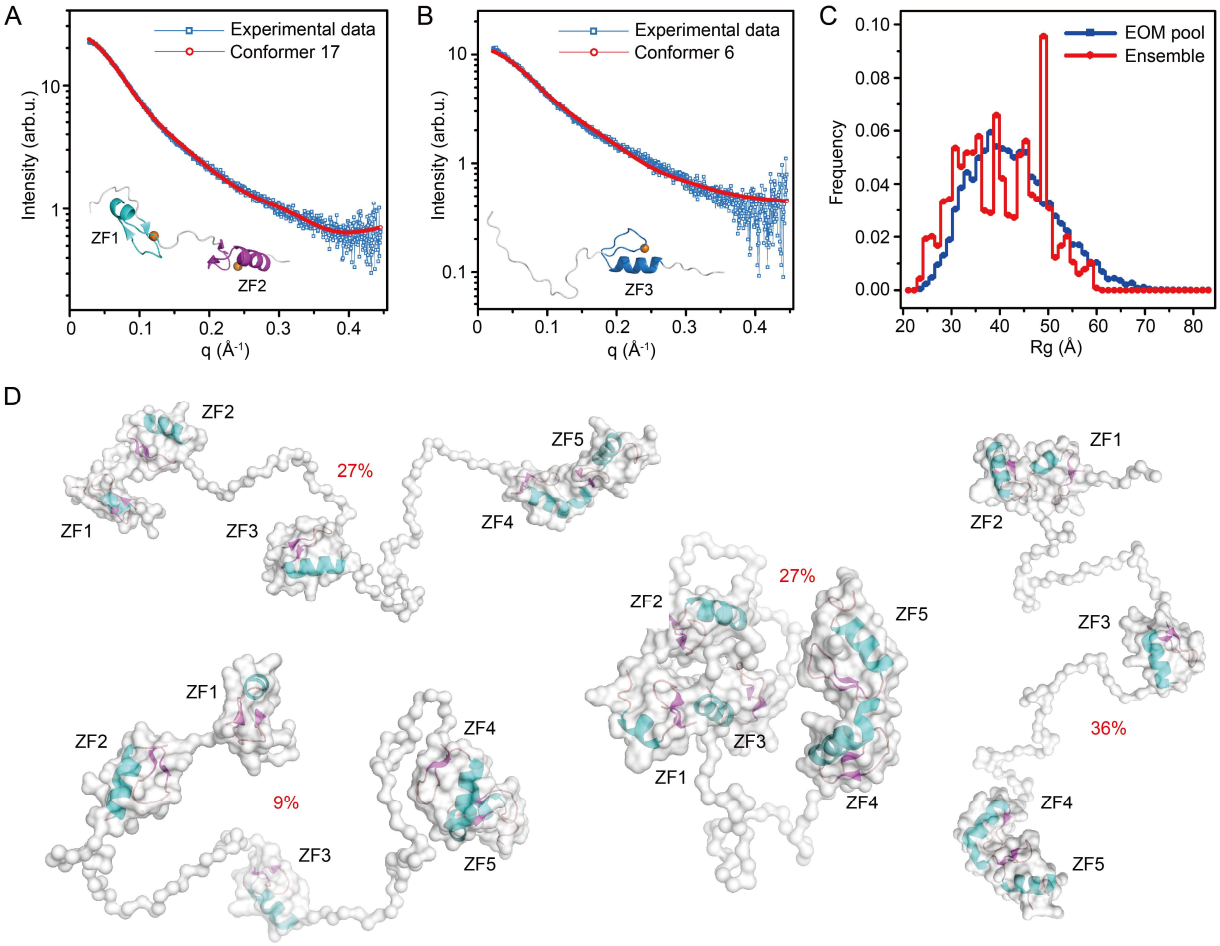
Supplementary Figure 1. (A) Diagram showing the different truncations of INSM1 used in this study. The truncations used for protein production in *E.coli* are colored in orange, while those used in BiFC are in yellow. (B) ICP-MS data for determining the number of Zn ion in INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5. (C) Multi-angle static light scattering assay coupled with size exclusion chromatography (SEC-MALS) of INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5. The light scattering curves are shown in a relative scale of each sample, and the determined molar masses are shown with dots, which indicate weight-average molar masses of 9.5 ± 0.3 kDa, 6.1 ± 0.2 kDa, and 25.8 ± 0.1 kDa for INSM1 ZF1-ZF2, ZF3, and ZF1-ZF5, respectively. Source data are provided as a Source Data file. (D) Western blot analysis of the correct expression of different INSM1 variants fused with nYFP with anti-HA antibody in HeLa cell after transfection. Control, HeLa cell without transfection. The experiment was repeated independently for three times with similar results.



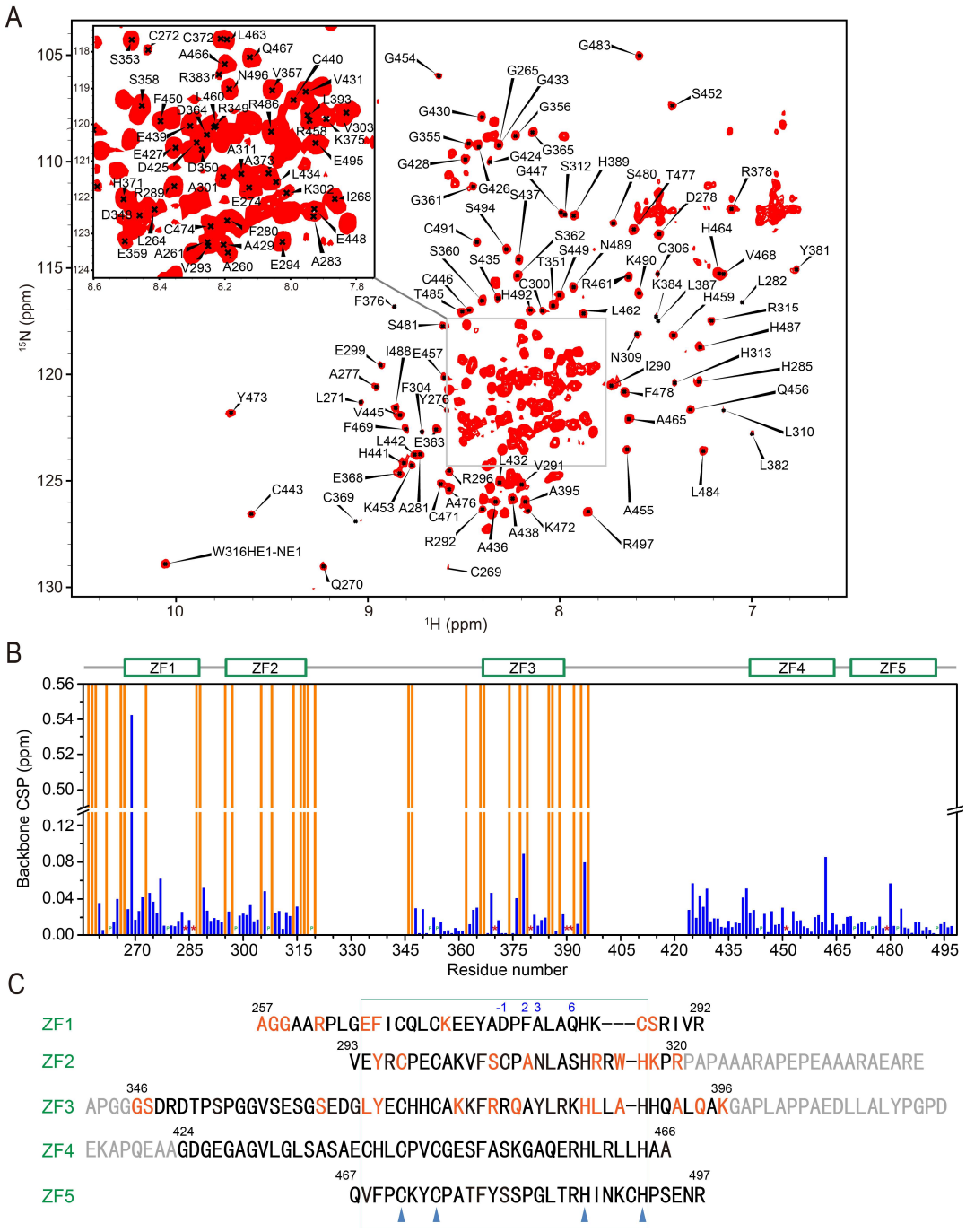
Supplementary Figure 2. (A, B) ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled INSM1 ZF1-ZF2 and ZF3 with chemical shift assignments. (C, D) Structures of 20 lowest energy conformers of INSM1 ZF1-ZF2 determined by NMR spectroscopy, aligned with ZF1 (C) and ZF2 (D). ZF1 and ZF2 are colored in cyan and purple, respectively. Loops are colored in gray. (E) Structures of 20 lowest energy conformers of INSM1 ZF3 determined by NMR spectroscopy, aligned with ZF3. ZF3 is colored in marine blue, and loops are colored in gray. Zn ions are shown as orange spheres in (C-E).



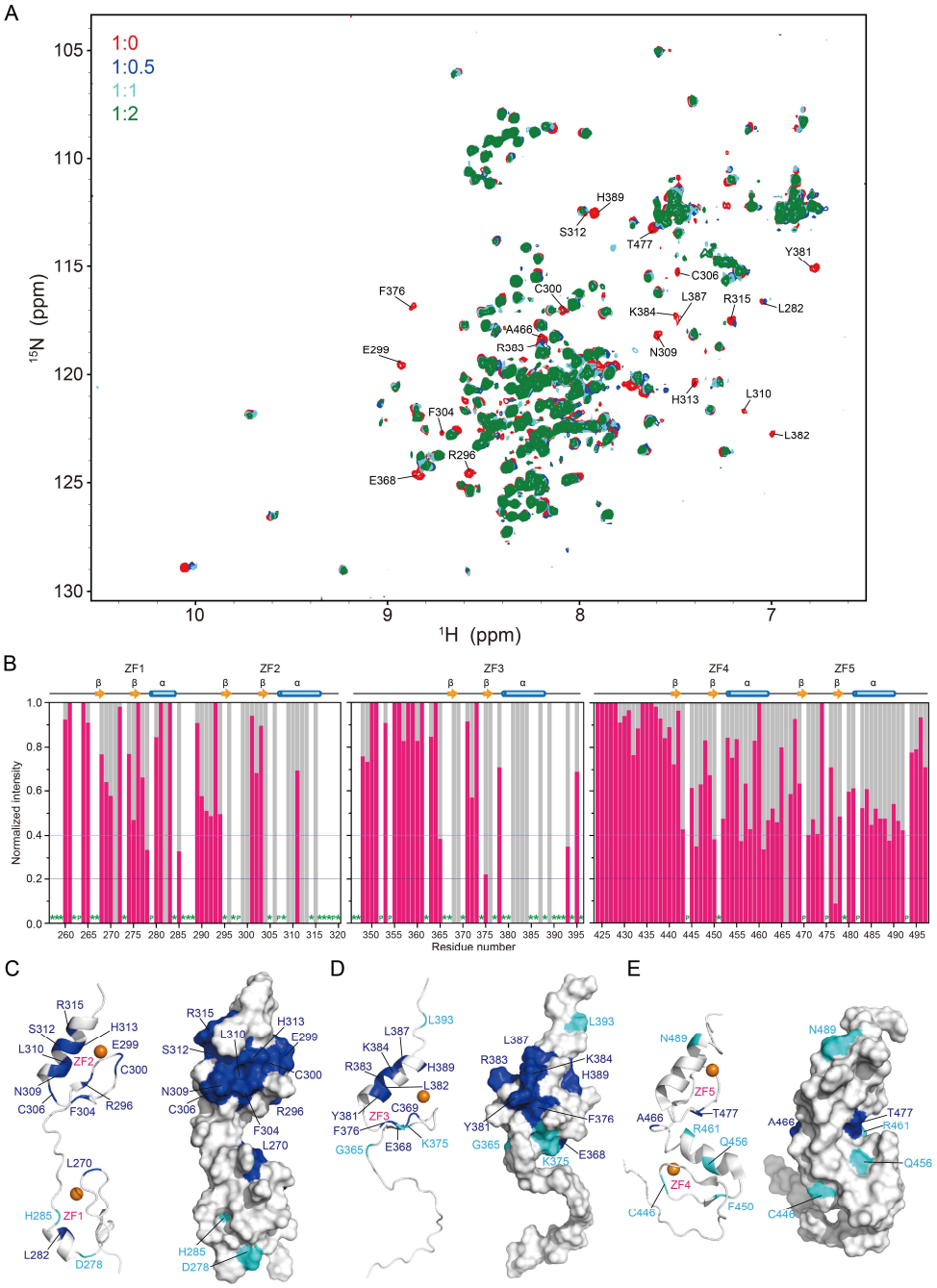
Supplementary Figure 3. (A, B) Overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled INSM1 ZF1-ZF2 wild type and C287A mutant (A) or R289A mutant (B). Residues with remarkable changes of chemical shifts are marked. (C) Overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled INSM1 ZF1-ZF2 and ZF1 truncations, with chemical shifts for the residues in ZF1 truncation assigned. (D) Chemical shift differences for the residues in ZF1 between INSM1 ZF1-ZF2 and ZF1 truncations are calculated and shown. Unassigned residues in both ZF1-ZF2 and ZF1 truncations are marked with asterisks. Orange-filled columns mark the residues unassigned only in ZF1 truncation. P, proline. Source data are provided as a Source Data file.



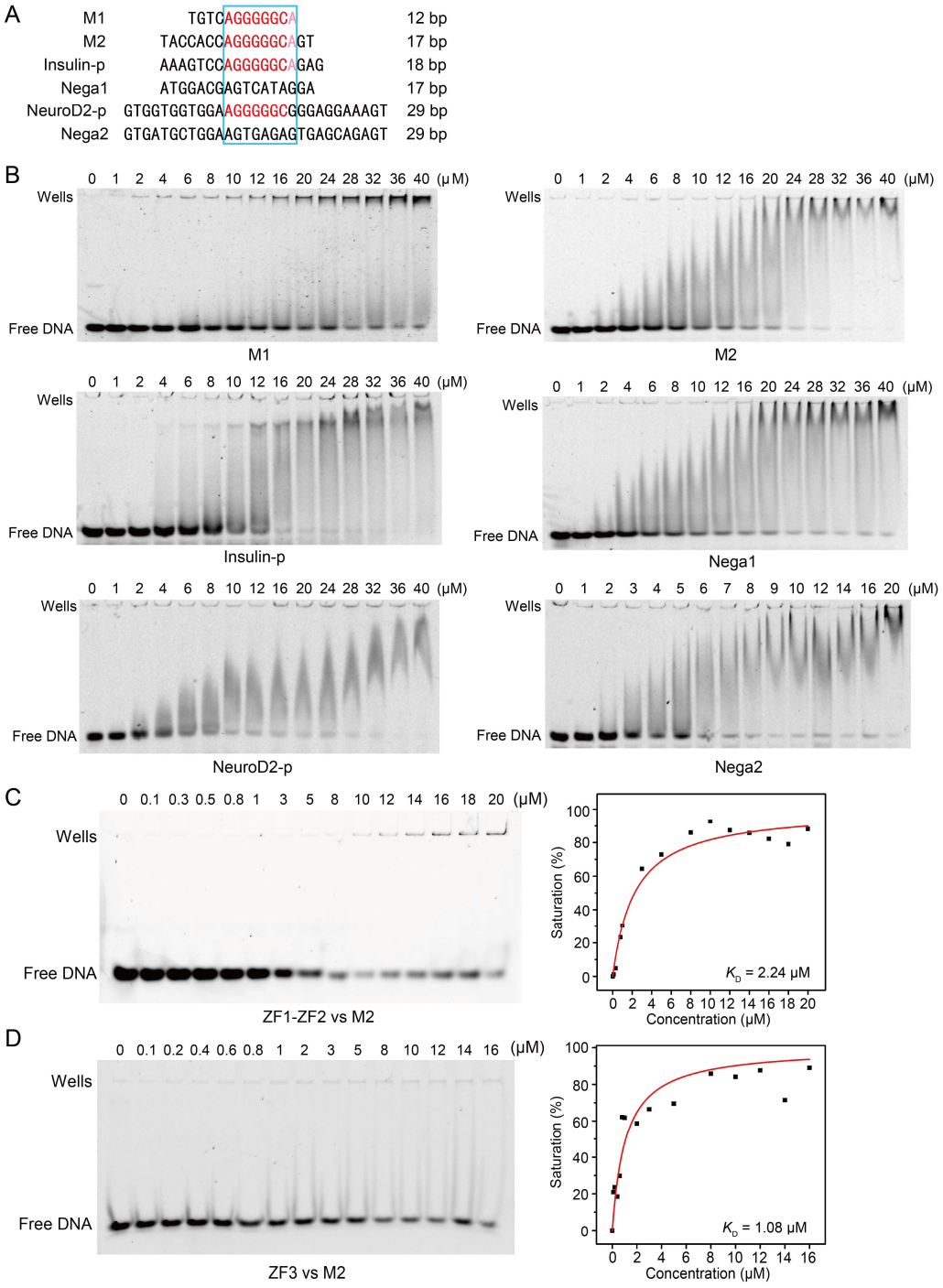
Supplementary Figure 4. (A) Experimental scattering data of INSM1 ZF1-ZF2 and theoretical scattering curve of conformer 17 in the ensemble of INSM1 ZF1-ZF2 structure. The structure of conformer 17 is shown as inset. (B) Experimental scattering data of INSM1 ZF3 and theoretical scattering curve of conformer 6 in the ensemble of INSM1 ZF3 structure. The structure of conformer 6 is shown as inset. Source data are provided as a Source Data file. (C) The line charts of the gyration radius (Rg) for the pool of 10,000 structures generated by EOM and for the selected ensemble of INSM1 ZF1-ZF5. Source data are provided as a Source Data file. (D) Four models of INSM1 ZF1-ZF5 obtained from EOM modelling. Frequency of each model in the pool is depicted in red font.



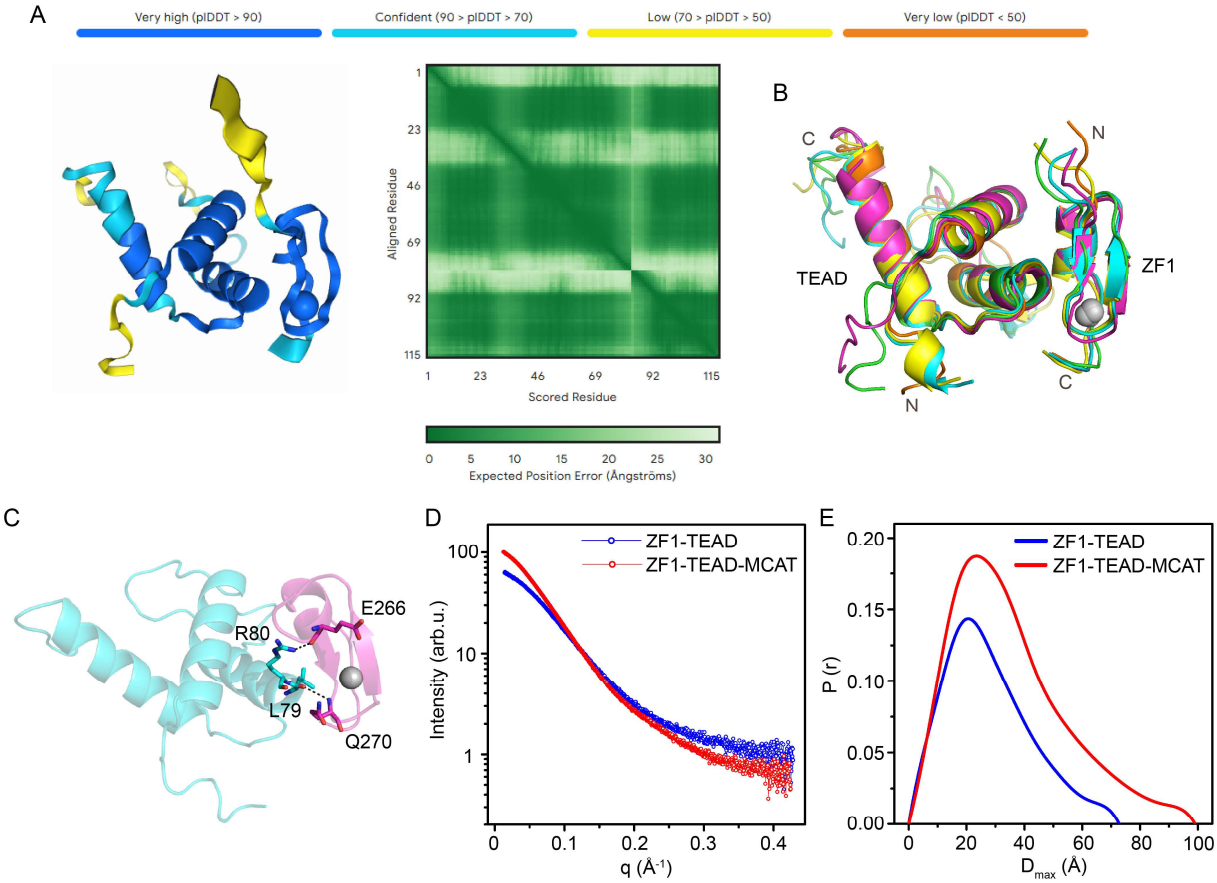
Supplementary Figure 5. (A) ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled INSM1 ZF1-ZF5 with backbone chemical shift assignments. (B) Chemical shift differences for each residue of INSM1 ZF1-ZF5 compared to ZF1-ZF2, ZF3, and ZF4-ZF5. Unassigned residues in all truncations are marked with asterisks. Orange-filled columns mark the residues unassigned in ZF1-ZF5. P, proline. Source data are provided as a Source Data file. (C) Sequence of INSM1 ZF1-ZF5 aligned with the ZFs (green box). The residues assigned in ZF1-ZF2, ZF3, or ZF4-ZF5, but not in ZF1-ZF5 are colored in orange. The residues that are not included in ZF1-ZF2, ZF3, and ZF4-ZF5 truncations are colored in gray.



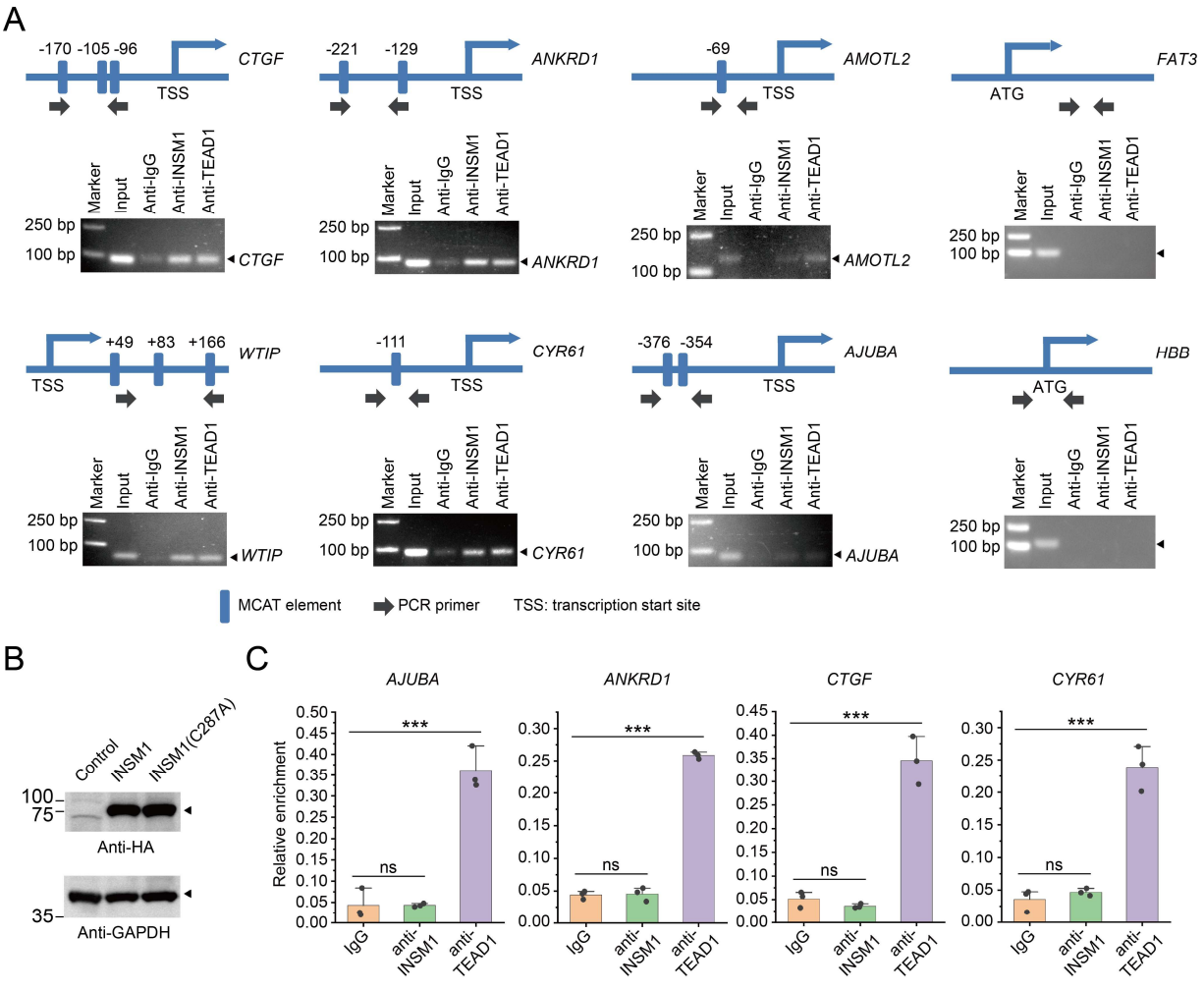
Supplementary Figure 6. (A) Overlay of a series of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled INSM1 ZF1-ZF5 in the presence of DNA at different molar ratios of protein:DNA, which are colored differently as indicated. (B) Chemical shift intensity (CSI) for each residue of INSM1 ZF1-ZF5 at protein:DNA ratio of 1:2 (pink) is calculated and normalized to that at protein:DNA ratio of 1:0 (gray). CSI values of 0.2 and 0.4 were selected as two significant levels for globally comparison of the five ZFs. The residues with CSI lower than 0.2 were labelled in (A). Unassigned residues are marked with asterisks. P, proline. Source data are provided as a Source Data file. (C-E) The residues with CSI values lower than 0.2 (blue) and ranging 0.2-0.4 (cyan) are shown on the structures of INSM1 ZF1-ZF2 (C), ZF3 (D), and ZF4-ZF5 (E) in cartoon and surface views, respectively.



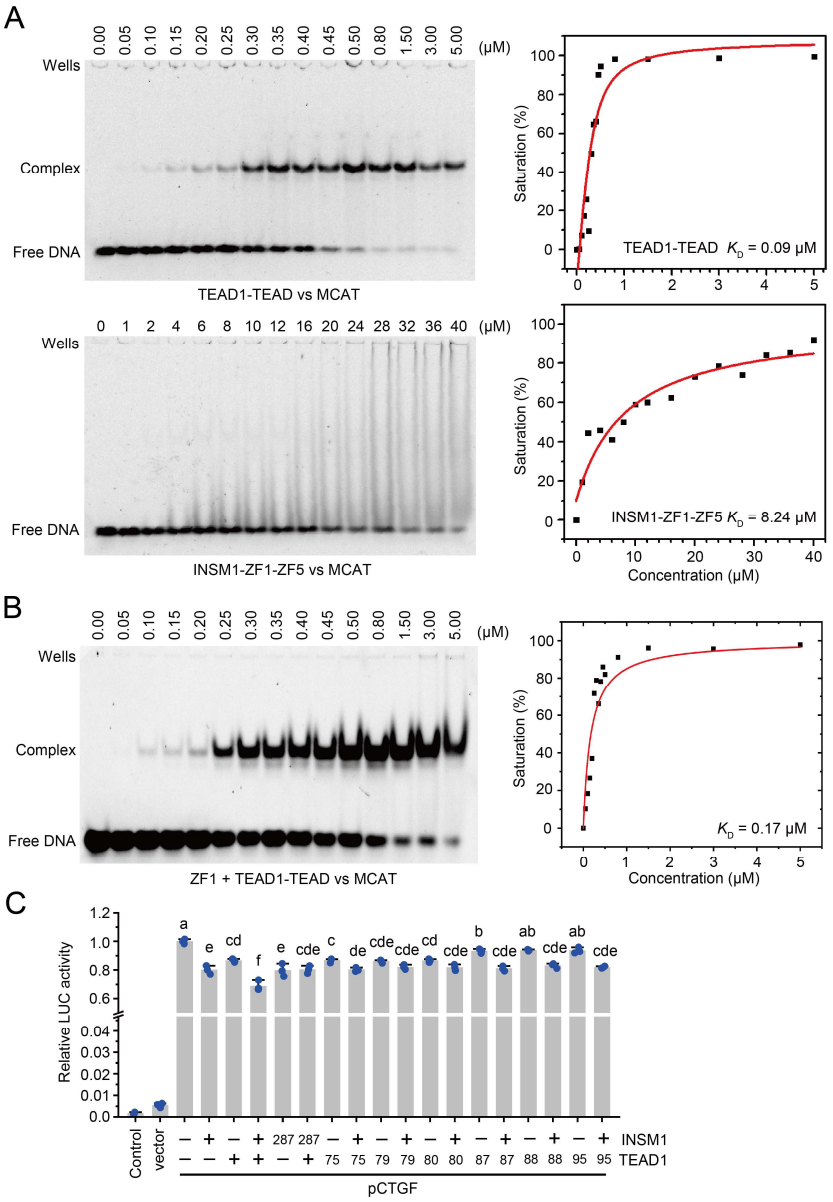
Supplementary Figure 7. (A) Sequences of the DNA fragments used in EMSA with INSM1 ZF1-ZF5. Similar nucleotides are colored in red. (B) EMSA of INSM1 ZF1-ZF5 with different FAM-labeled DNA fragments. The concentrations of INSM1 ZF1-ZF5 are indicated on top of each gel image. (C) and (D) EMSA of INSM1 ZF1-ZF2 (C) and ZF3 (D) with FAM-labeled M2 DNA. The concentrations of proteins are indicated on top of gel images. Binding curves and K_D values fitted from band intensities in the EMSA gel images are shown on the right, respectively. Source data are provided as a Source Data file. The experiment was repeated independently for three times with similar results.



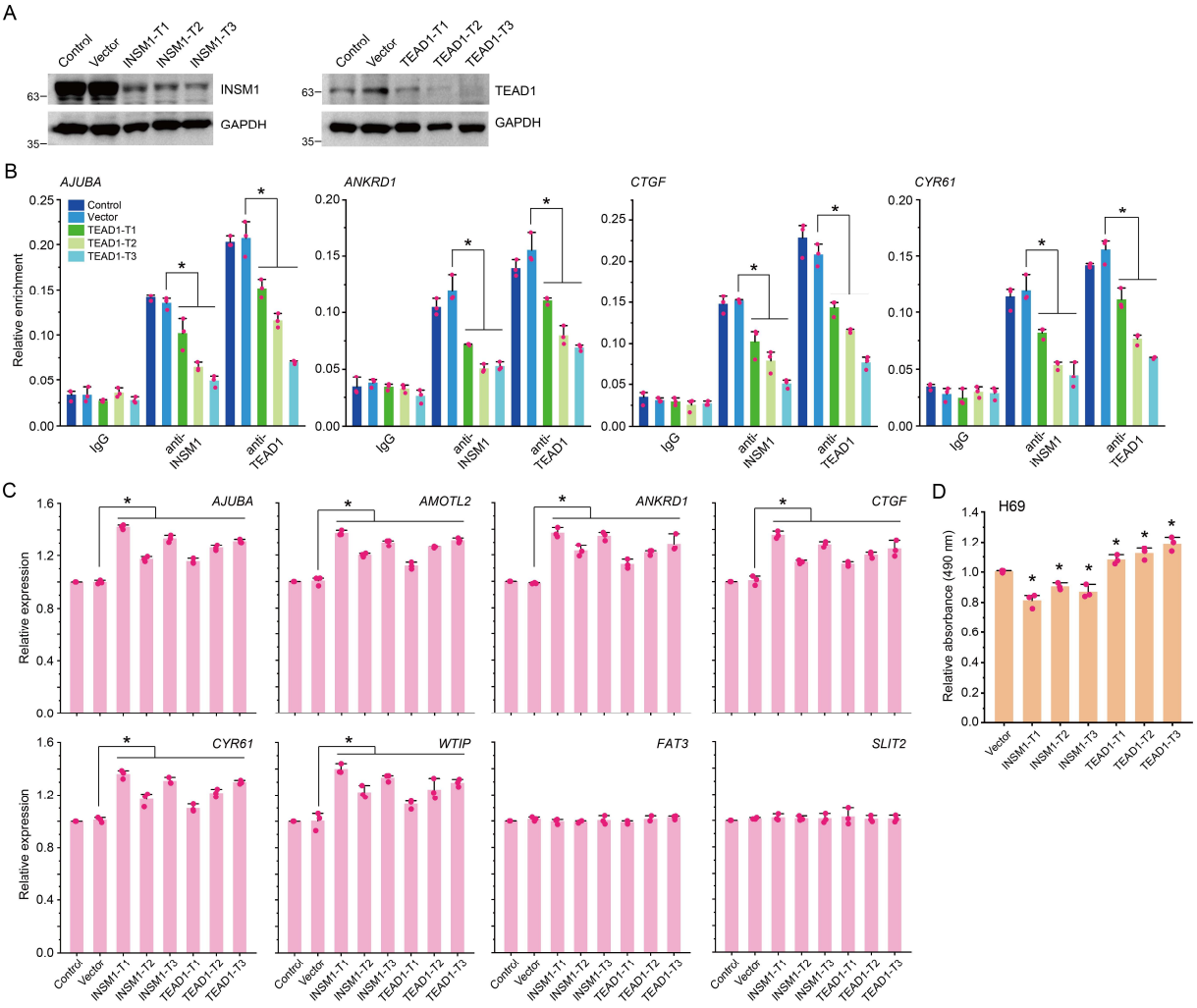
Supplementary Figure 8. (A) Quality assessment of structural models of INSM1-ZF1 and TEAD1-TEAD complex modelled by AlphaFold3. (B) Superposition of the five complex structures of INSM1-ZF1 and TEAD1-TEAD modelled by AlphaFold3. (C) The hydrogen bonds formed between INSM1-ZF1 and TEAD1-TEAD in the structure model 1. The residues forming hydrogen bonds (black dashed lines) are marked and shown as sticks. (D) Experimental SAXS data of the complex of INSM1-ZF1 and TEAD1-TEAD, and the complex of INSM1-ZF1, TEAD1-TEAD, and MCAT DNA. (E) Particle distance distribution curves transformed from the SAXS data in (D). Source data are provided as a Source Data file.



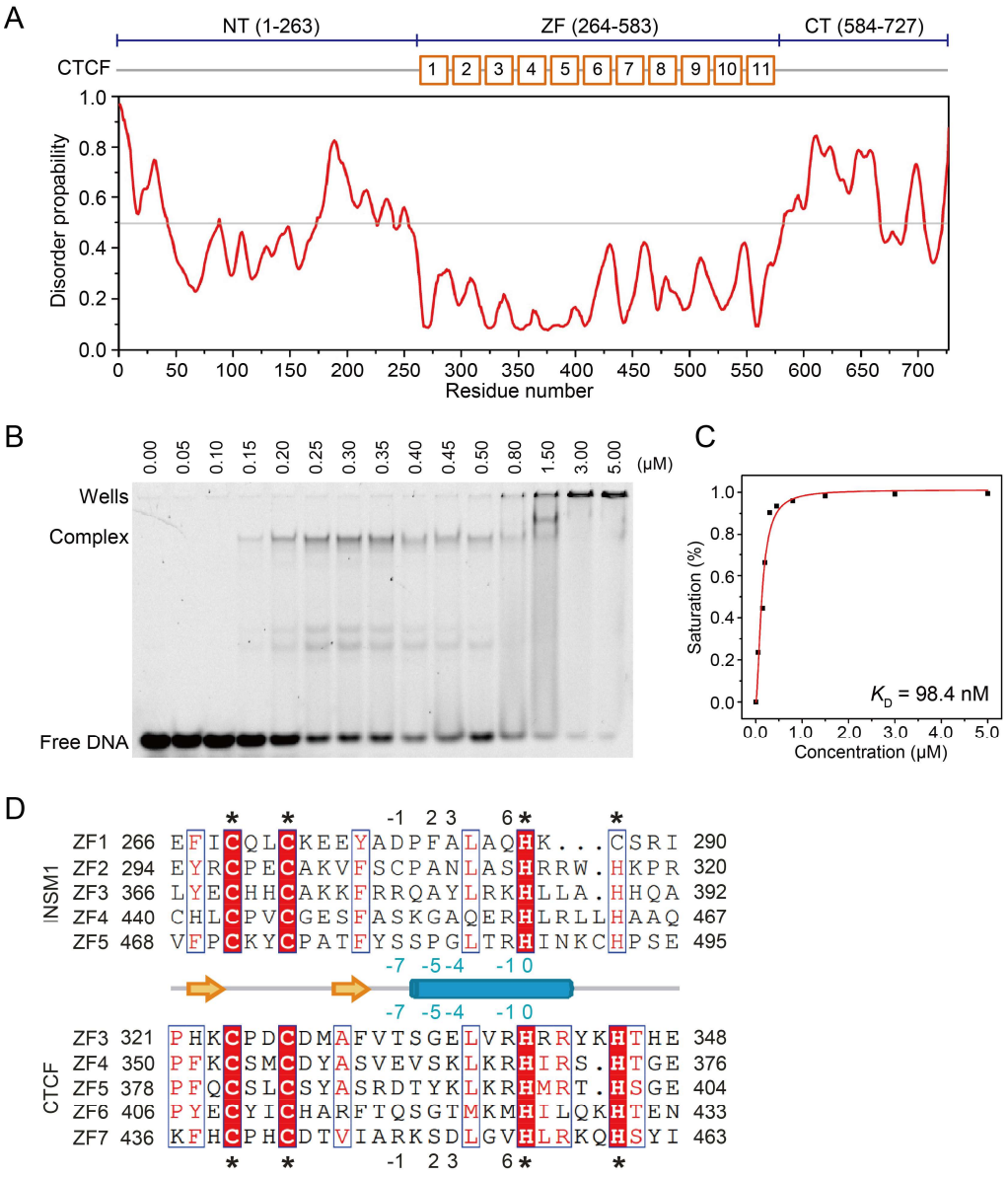
Supplementary Figure 9. (A) Gel images of ChIP-PCR of INSM1 and TEAD1 binding to the TEAD1-targeted genome loci in HeLa cells co-transfected with INSM1 and TEAD1. The targeted genome loci adjacent to *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, and *WTIP* genes are illustrated on top of each gel image. Genomic regions from *FAT3* and *HBB* genes without MCAT element were used as negative controls. (B) Western blot analysis of the correct expression of INSM1 C287A mutant fused with nYFP with anti-HA antibody in HeLa cell after transfection. Control, HeLa cell without transfection. The experiment was repeated independently for three times with similar results. (C) ChIP-qPCR analysis of INSM1 and TEAD1 binding to the MCAT elements in the promoters of *AJUBA*, *ANKRD1*, *CTGF* and *CYR61* in HeLa cells without endogenous expression and transfection of INSM1. 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. *** $P < 0.001$. NS, not significant. The exact P-values were provided in Source Data. Source data are provided as a Source Data file.



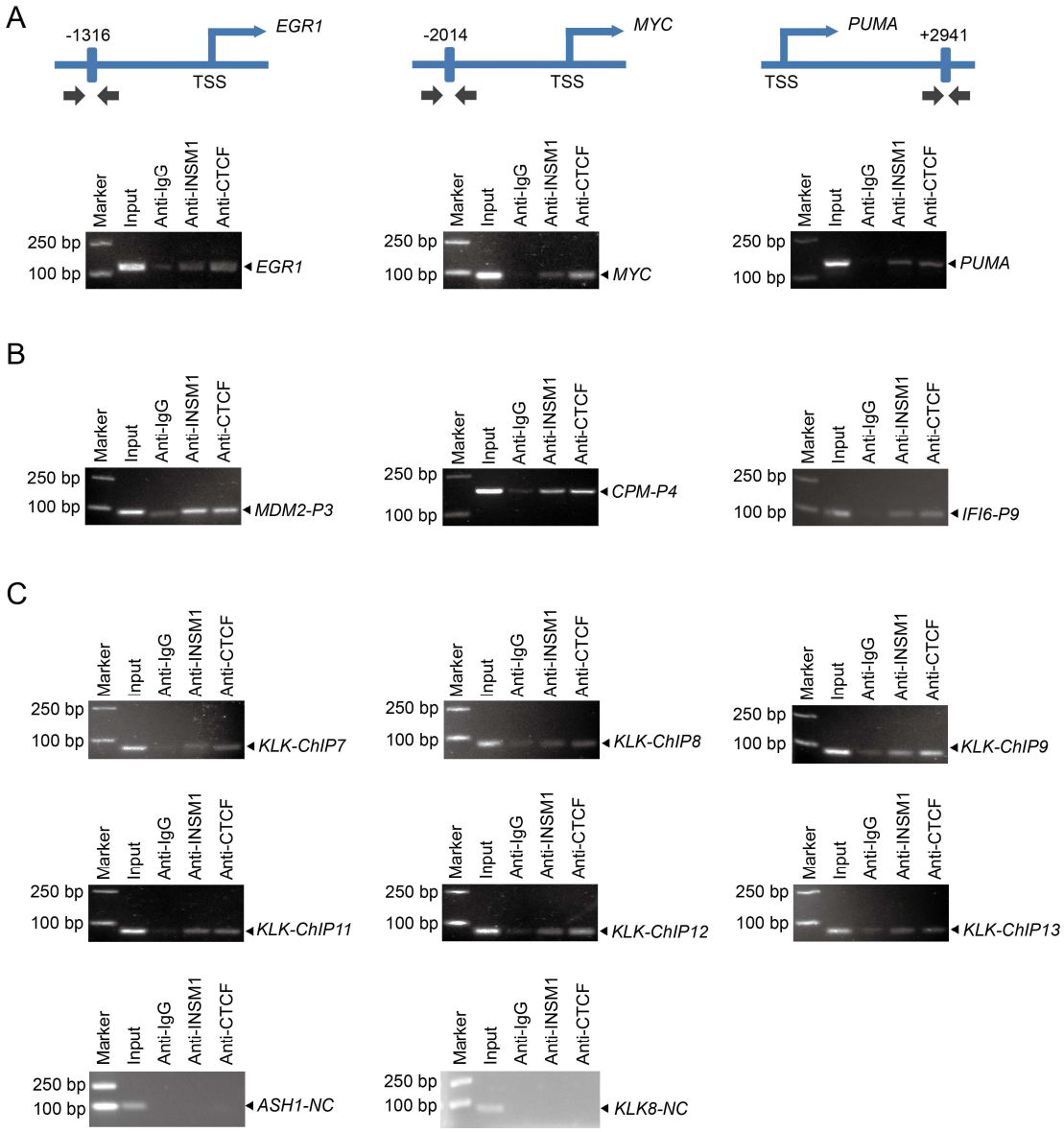
Supplementary Figure 10. (A) EMSA of TEAD1-TEAD and INSM1 ZF1-ZF5 with FAM-labeled MCAT DNA. The protein concentrations are indicated on top of gel images. Binding curves and K_D values fitted from band intensities in gel images of TEAD1-TEAD and INSM1 ZF1-ZF5 are shown on the right, respectively. (B) EMSA of TEAD1-TEAD in presence of INSM1 ZF1 with FAM-labeled MCAT DNA. The concentrations of proteins are indicated on top of gel images. TEAD1-TEAD and INSM1 ZF1 was premixed at molar ratio of 1:1. Binding curve and K_D value fitted from band intensities in the EMSA gel image are shown on the right. Source data are provided as a Source Data file. The experiment was repeated independently for three times with similar results. (C) Dual-luciferase reporter assay in HeLa cells co-transfected with different combinations of TEAD1 variants and INSM1 variants. The relative luciferase activity driven by CTGF promoter without transfection of TEAD1 and INSM1 was set to 1. Data are presented as mean values ± 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 P-values were provided in Source Data. Source data are provided as a Source Data file.



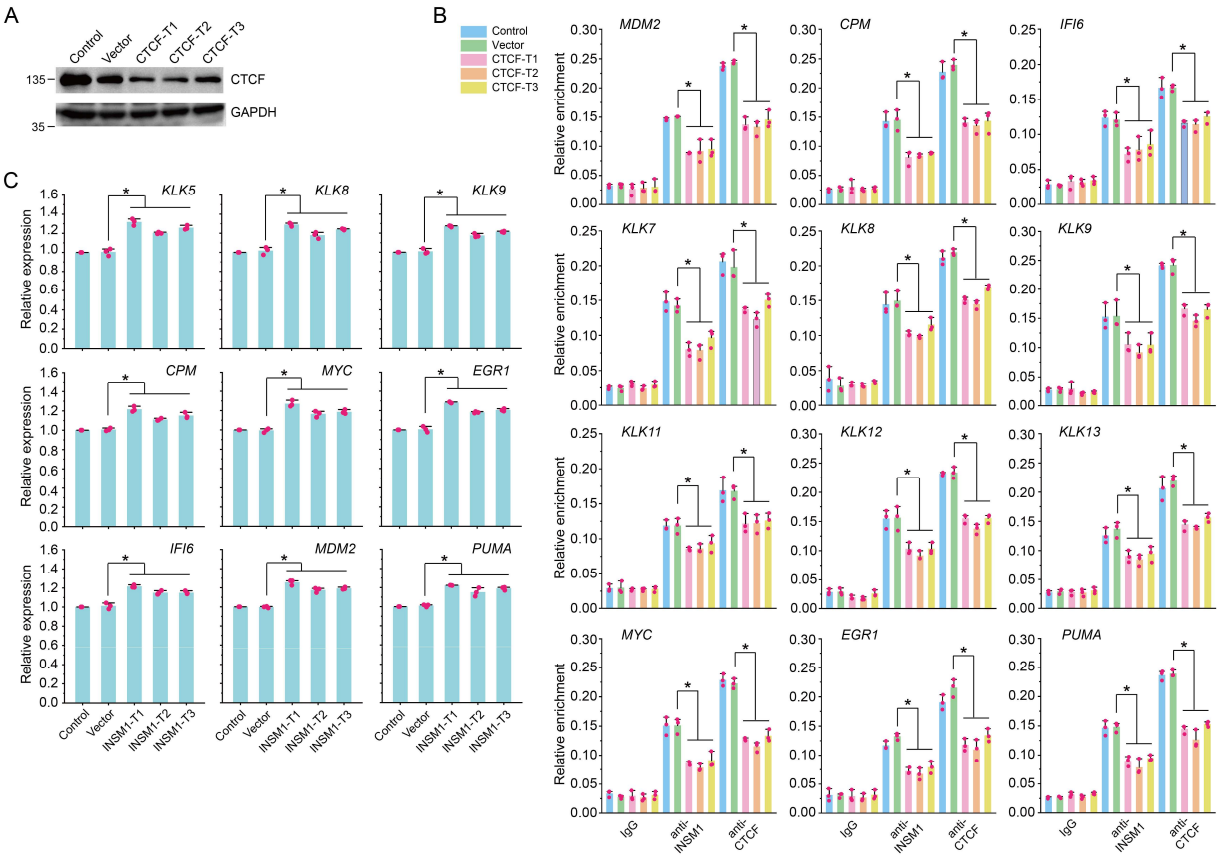
Supplementary Figure 11. (A) Western blot analysis of the expressions of INSM1 (left) or TEAD1 (right) in H69 cell expressing CRISPR/Cas9 and INSM1-targeted (left) or TEAD1-targeted (right) guide RNAs. GAPDH was used as loading control. The experiment was repeated independently for three times with similar results. (B) ChIP-qPCR analysis of INSM1 and TEAD1 binding to the MCAT elements in the promoters of *AJUBA*, *ANKRD1*, *CTGF* and *CYR61* in H69 cells used in (A, right). 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. * $P < 0.05$. The exact P-values were provided in Source Data. Source data are provided as a Source Data file. (C) Relative expression of *AJUBA*, *AMOTL2*, *ANKRD1*, *CTGF*, *CYR61*, *WTIP*, *FAT3*, and *SLIT2* in H69 cells used in (A). The expression level of the indicated gene in control H69 cells 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. * $P < 0.05$. The exact P-values were provided in Source Data. Source data are provided as a Source Data file. (D) Cell proliferation assays of H69 cells used in (A). The absorbance of control H69 cells (column not shown) is set to 1. Data are presented as mean values ± SD of three independent experiments (n=3). The differences between vector column and columns annotated with asterisks are significant (* $P < 0.05$, by two-sided Student's t-test). The exact P-values were provided in Source Data. Source data are provided as a Source Data file.



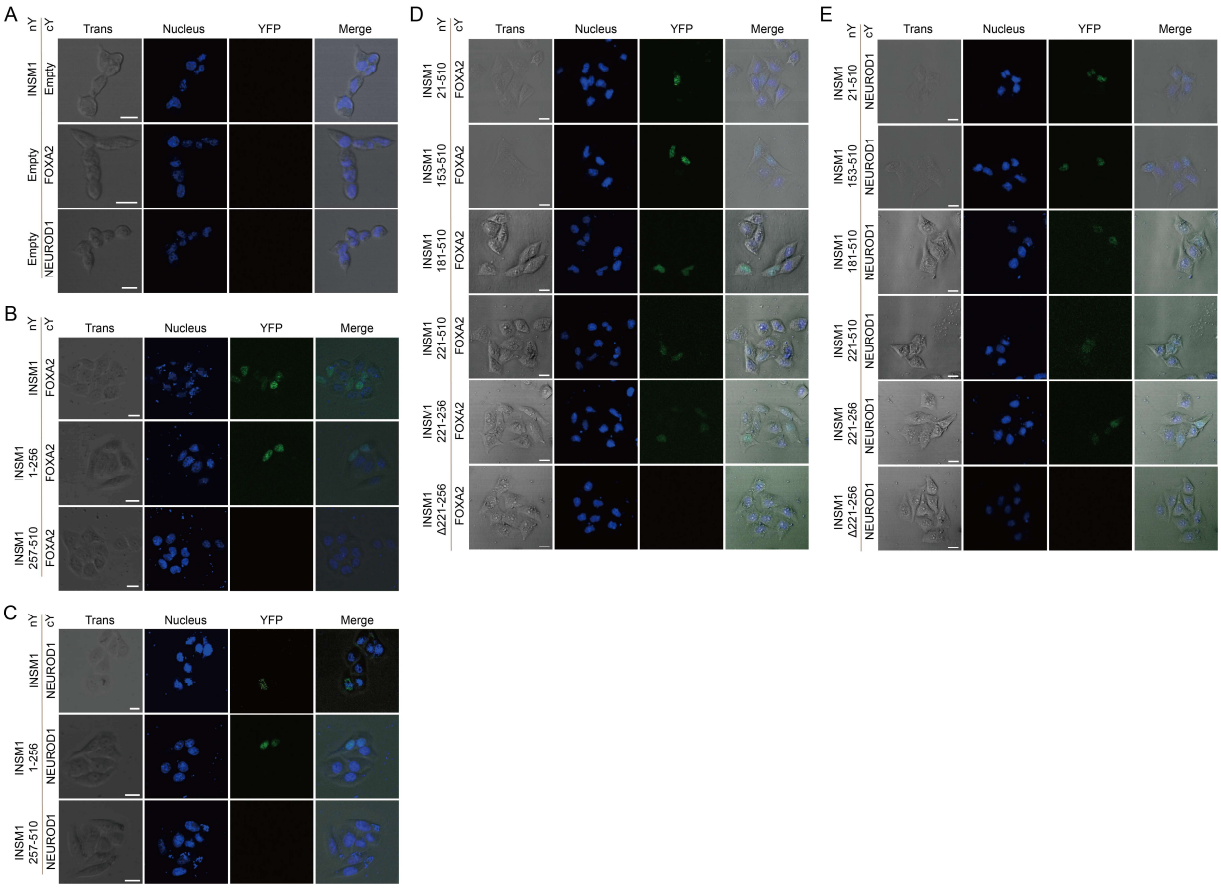
Supplementary Figure 12. (A) Schematic domain structure of human CTCF with an analysis of structurally disordered probability. NT, N-terminal part; CT, C-terminal part; ZF, zinc finger. The regions showing a probability over the gray line (> 0.5) are considered to be disordered. Source data are provided as a Source Data file. (B) EMSA of CTCF ZF3-ZF7 with FAM-labeled M2 DNA. The concentrations of CTCF ZF3-ZF7 are indicated on top. The experiment was repeated independently for three times with similar results. (C) Curve and K_D value fitted from the EMSA data of CTCF ZF3-ZF7 with M2 DNA. Source data are provided as a Source Data file. (D) Sequence alignment of INSM1 ZF1/2/3/4/5 and CTCF ZF3/4/5/6/7, respectively. Identical (white letters filled with red color) and similar (red letters with blue box) amino acids are denoted. The secondary structure elements of INSM1 ZF5 are shown in the middle. The residues at -1, +2, +3, and +6 positions of the α-helix, that are -7, -5, -4, and -1 positions from the first His residue coordinated with Zn ion, are indicated. The residues involved in Zn coordination are marked with asterisks.



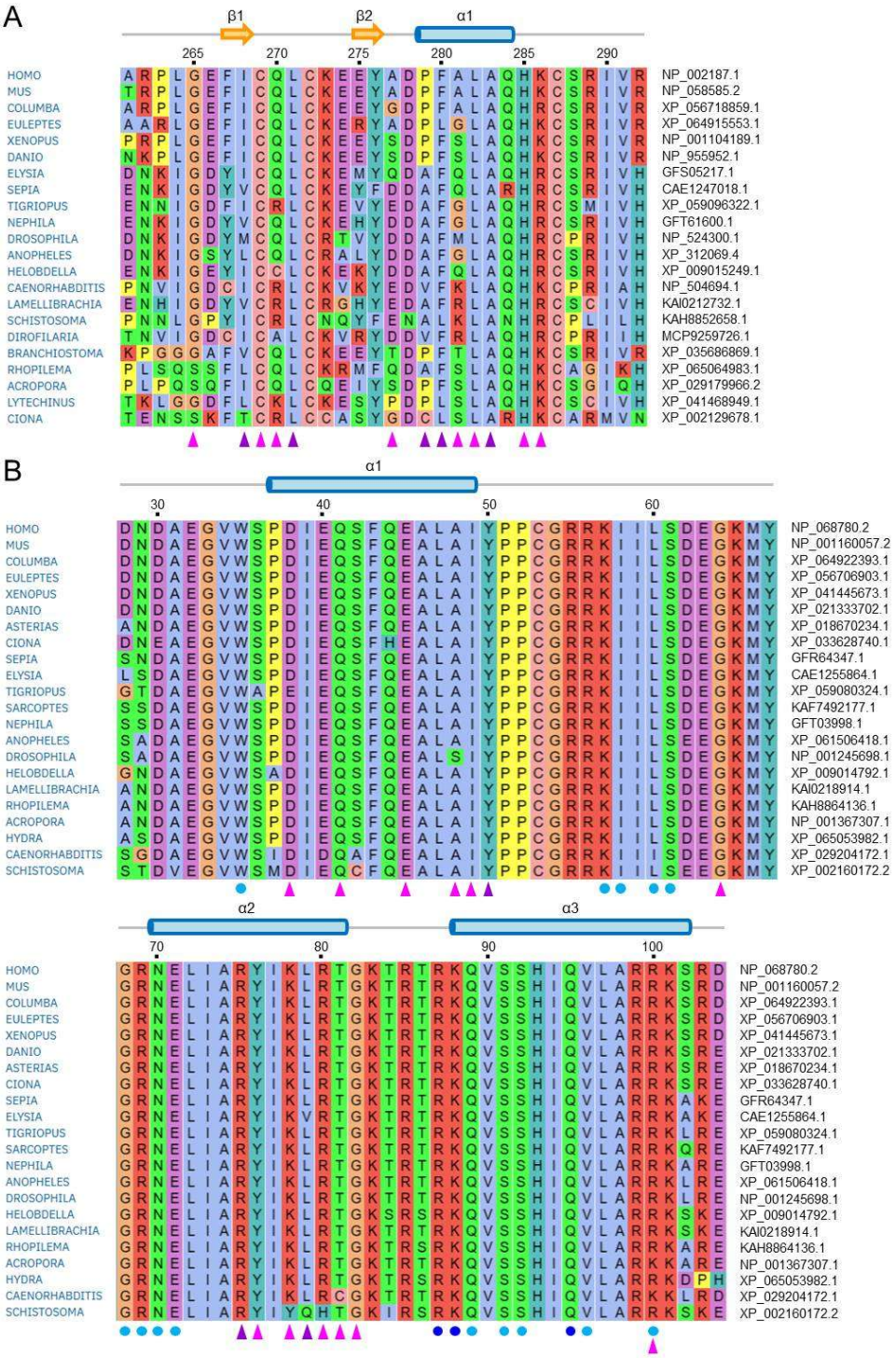
Supplementary Figure 13. (A-C) Gel images of ChIP-PCR of INSM1 and CTCF binding to the CTCF-targeted genome loci in HeLa cells co-transfected with INSM1 and CTCF. The targeted genome loci adjacent to *EGR1*, *MYC*, and *PUMA* genes are illustrated on top in (A). The targeted genome loci in (B) and (C) were described in previous studies (Li et al., 2019; Khoury et al., 2020). Genomic regions from *ASH1* and *KLK8* genes without CTCF-binding site were used as negative controls (NC).



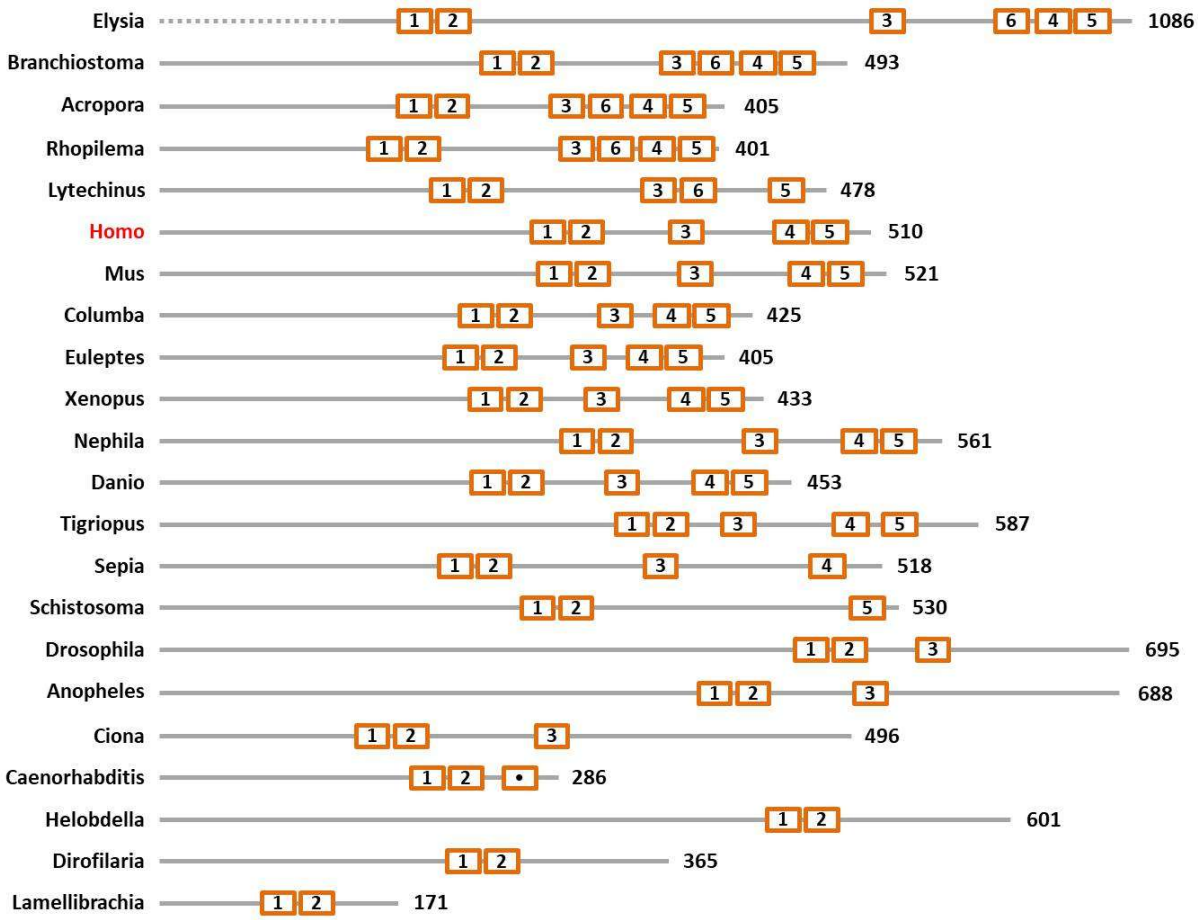
Supplementary Figure 14. (A) Western blot analysis of the expression of CTCF in H69 cell expressing CRISPR/Cas9 and CTCF-targeted guide RNAs. GAPDH was used as loading control. The experiment was repeated independently for three times with similar results. (B) ChIP-qPCR analysis of INSM1 and CTCF binding to the CTCF-targeted genome loci in H69 cells used in (A). 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. * $P < 0.05$. The exact P-values were provided in Source Data. Source data are provided as a Source Data file. (C) Relative expression of CTCF-targeted genes in INSM1-knockdown H69 cells used in (Figure S11A, left). The expression level of the indicated gene in control H69 cells 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. * $P < 0.05$. The exact P-values were provided in Source Data. Source data are provided as a Source Data file.



Supplementary Figure 15. (A-E) BiFC between cYFP-fused FOXA2 or NEUROD1 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.



Supplementary Figure 16. (A) Sequence alignment of ZF1 in INSM1 orthologs from different animals. The residues in the binding interface for TEAD1 are marked with triangles at bottom. (B) Sequence alignment of TEAD in TEAD1 orthologs from different animals. The residues in the binding interface for INSM1 are marked with triangles at bottom, while the residues in the binding interface for MCAT DNA are marked with cycles. The NCBI accession number for each protein is listed on the right.



Supplementary Figure 17. Diagram showing the number and position of zinc fingers in INSM1 orthologs from different animals. The total residue numbers of different INSM1 orthologs are shown on the right. The dot-marked zinc finger in *Caenorhabditis elegans* shows low sequence similarity to other ZFs found in the listed animals. The NCBI accession numbers for the proteins are the same to those in Figure S16A.

Supplementary Table 1. Structural statistics for INSM1 ZF1-ZF2 (residues 257-320).

Conformationally-restricting constraints [a]		
NOE-based distance constraints		
Total		559
Intra-residue (i=j)		187
Sequential (|i-j|=1)		194
Medium-range (1<|i–j|<5)		104
Long-range (|i–j|≥5)		74
NOE constraints per restrained residue [b]		9.4
Hydrogen bond constraints		
Long-range (|i–j|≥5)/total		4/20
Dihedral angle constraints		82
Total number of restricting constraints [b]		661
Total number of restricting constraints per restrained residue [b]		10.8
Restricting long-range constraints per restrained residue [b]		1.3
Number of structures used		20
Residue constraint violations [a, c]		
Average number of distance violations per structure		
0.1-0.2Å		5.45
0.2-0.5 Å		2
>0.5Å		0
Average RMS distance violation/constraint (Å)		0.13
Average number of dihedral angle violations per structure		
1-10º		6.8
>10º		0
Average RMS dihedral angle violation/constraint (degree)		4.11
RMSD Values (pairwise) [e,f]	ZF1	ZF2
Backbone /Heavy atoms (Å)	0.6±0.09 /1.1±0.06	0.5±0.06/1.2±0.03
Ramachandran plot statistics	ZF1	ZF2
Most favored regions (%)	97.4	90
Allowed regions (%)	2.6	10
Disallowed regions (%)	0	0
Global quality scores(raw/Z-score) [g]		
Verify3D		0.03/-6.9
Prosall		-3.4/-4.09
Procheck(phi-psi) [d]		-0.22/-0.55
Procheck(all) [d]		-0.28/-1.66
Molprobity clash		8.37/0.09

[a] Generated using PSVS 1.5. Analyzed for residues 1 to 64.

[b] There are 61 residues with conformationally restricting constraints.

[c] Calculated for all constraints for the given residues, using sum over r^{-6}.

[d] Largest constraint violation among all the reported structures.

[e] Ordered residue ranges (with sum of phi and psi order parameters > 1.8): 10A-17A,19A-31A for ZF1, and 38A-63A for ZF2.

[f] Residues were selected based on dihedral angle order parameter with S(phi)+S(psi) ≥1.8. Selected residue ranges: 10A-17A,19A-31A for ZF1, and 38A-63A for ZF2.

[g] With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution ≤ 1.80 Å, R-factor ≤ 0.25 and R-free ≤ 0.28; a positive value indicates a 'better' score.

Supplementary Table 2. Structural statistics for INSM1 ZF3 (residues 346-396).

Conformationally-restricting constraints [a]	
NOE-based distance constraints	
Total	288
Intra-residue (i=j)	102
Sequential (|i-j|=1)	1.3
Medium-range (1<|i–j|<5)	47
Long-range (|i–j|≥5)	36
NOE constraints per restrained residue [b]	5.8
Hydrogen bond constraints	
Long-range (|i–j|≥5)/total	0/14
Dihedral angle constraints	50
Total number of restricting constraints [b]	352
Total number of restricting constraints per restrained residue [b]	7.0
Restricting long-range constraints per restrained residue [b]	0.7
Number of structures used	20
Residue constraint violations [a, c]	
Average number of distance violations per structure	
0.1-0.2Å	3.1
0.2-0.5 Å	0.8
>0.5Å	0
Average RMS distance violation/constraint (Å)	0.03
Maximum distance violation (Å) [d]	0.33
Average number of dihedral angle violations per structure	
1-10º	3.45
>10º	0
Average RMS dihedral angle violation/constraint (degree)	0.82
Maximum dihedral angle violation (degree) [d]	9.1
RMSD Values (pairwise) [e,f]	
Backbone /Heavy atoms (Å)	0.6±0.03 /1.3±0.06
Ramachandran plot statistics	
Most favored regions (%)	98
Allowed regions (%)	2
Disallowed regions (%)	0
Global quality scores(raw/Z-score) [g]	
Verify3D	0.17/-4.65
Prosall	-0.01/-2.65
Procheck(phi-psi) [d]	0.07/0.59
Procheck(all) [d]	-0.00/-0.00
Molprobity clash	3.48/0.93

[a] Generated using PSVS 1.5. Analyzed for residues 1 to 51.

[b] There are 50 residues with conformationally restricting constraints.

[c] Calculated for all constraints for the given residues, using sum over r^{-6}.

[d] Largest constraint violation among all the reported structures.

[e] Ordered residue ranges (with sum of phi and psi order parameters > 1.8): 20A-24A,27A-46A.

[f] Residues were selected based on dihedral angle order parameter with S(phi)+S(psi) ≥1.8. Selected residue ranges: 20A-24A,27A-46A.

[g] With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution ≤ 1.80 Å, R-factor ≤ 0.25 and R-free ≤ 0.28; a positive value indicates a 'better' score.

Supplementary Table 3. SAXS sample details, data collection parameters, data analysis software, and structural parameters.

(*a*) Sample details					
Sample	ZF1-ZF2	ZF3	ZF1-ZF5	ZF1/TEAD	ZF1/TEAD/MCAT
Organism	human				
Source	Recombinant expression and purification from *E.coli*				
Sequence	GSAGGAARPLGEFICQLCKEEYADPFALAQHKCSRIVRVEYRCPECAKVFSCPANLASHRRWHKPR	GSDRDTPSPGGVSESGSEDGLYECHHCAKKFRRQAYLRKHLLAHHQALQAK	GSAGGAARPLGEFICQLCKEEYADPFALAQHKCSRIVRVEYRCPECAKVFSCPANLASHRRWHKPRPAPAAARAPEPEAAARAEAREAPGGGSDRDTPSPGGVSESGSEDGLYECHHCAKKFRRQAYLRKHLLAHHQALQAKGAPLAPPAEDLLALYPGPDEKAPQEAAGDGEGAGVLGLSASAECHLCPVCGESFASKGAQERHLRLLHAAQVFPCKYCPATFYSSPGLTRHINKCHPSENR	ZF1: GSAGGAARPLGEFICQLCKEEYADPFALAQHKCSRIVR TEAD: MDNDAEGVWSPDIEQSFQEALSIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKSRDLVPR	ZF1: GSAGGAARPLGEFICQLCKEEYADPFALAQHKCSRIVR TEAD: MDNDAEGVWSPDIEQSFQEALSIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKSRDLVPR MCAT-F (DNA): AGTGGAATGTGC MCAT-R (DNA): GCACATTCCACT
Extinction coefficient ε (280 nm and M^{-1} cm^{-1})	8480	2980	15930	ZF1: 1490, TEAD: 9970	ZF1: 1490, TEAD: 9970, MCAT: 196999 (260 nm)
Molecular mass *M* (Da)	7417.6	5674.3	25976.3	ZF1: 4093.7, TEAD: 9531.0	ZF1: 4093.7, TEAD: 9531.0, MCAT: 7290.8
Loading volume/concentration (µL / mg mL^{-1})	60 / 4.0	60 / 4.0	60 / 4.0	60 / ZF1: 1.3, TEAD: 2.9	60 / ZF1: 1.3, TEAD: 2.9, MCAT: 2.2
Solvent composition	50 mM NaH_2PO_4, 150 mM NaCl, pH 5.8				

(*b*) SAS data collection parameters					
Source, instrument			Shanghai Synchrotron Radiation Facility, at the BL19U2 BioSAXS beamline		
Wavelength (Å)			1.033		
Beam geometry (size, sample-to-detector distance)			1475 × 1679, 2.6 m		
q-measurement range ($Å^{-1}$)			0.007 < q < 0.445		
Exposure time, number of exposures			1 s, 20		
Sample temperature (ºC)			25		
(*c*) Software employed for SAS data reduction, analysis and interpretation					
SAS data reduction to sample–solvent scattering, and extrapolation			RAW 2.2.1		
Calculation of ε from sequence			Expasy - ProtParam		
Basic analyses: Guinier, *P*(*r*), scattering particle volume (*e.g.* Porod volume V_P or volume of correlation V_c)			PRIMUS		
Atomic structure modelling (ensemble)			EOM		
Molecular graphics			PyMOL		
(*d*) Structural parameters					
Guinier Analysis	ZF1-ZF2	ZF3	ZF1-ZF5	ZF1/TEAD	ZF1/TEAD/MCAT
I(0)	26.20	11.63	51.66	61.81	100.70
R_g (Å)	21.41	19.81	40.94	22.07	28.00
q-range ($Å^{-1}$)	0.02768-0.44573	0.0267-0.43644	0.02574-0.44302	0.02368-0.4256	0.02777-0.42192
P(*r*) analysis	ZF1-ZF2	ZF3	ZF1-ZF5	ZF1/TEAD	ZF1/TEAD/MCAT
I(0)	26.20	11.63	51.66	61.81	100.70
R_g (Å)	21.50	19.92	41.26	22.16	28.21
D_{max} (Å)	73.24	66.76	140.39	72.76	99.04
D_{max} / R_g	3.421	3.370	3.429	3.297	3.537
q-range ($Å^{-1}$)	0.02768-0.44573	0.0267-0.43644	0.02574-0.44302	0.02368-0.4256	0.02777-0.42192
Volume (V_c)	159	120	373	196	280
MW_{Vc} [kDa][a]	9.992	6.321	31.023	14.025	23.831

[a] MW_{Vc}, molecular weight calculated using volume of correlation method (Vc)

Supplementary Table 4. Coexpression analysis of INSM1 and TEAD1 in human tissues and organs using Bgee database. Source data are provided as a Source Data file.

Anatomical entities	Max expression score	Genes with presence of expression	Anatomical entity IDs
embryo	81.28	INSM1, TEAD1	UBERON:0000922
left testis	67.74	INSM1, TEAD1	UBERON:0004533
subcutaneous adipose tissue	89.31	INSM1, TEAD1	UBERON:0002190
small intestine Peyer's patch	80.22	INSM1, TEAD1	UBERON:0003454
vagina	83.43	INSM1, TEAD1	UBERON:0000996
kidney	80.27	INSM1, TEAD1	UBERON:0002113
minor salivary gland	79.88	INSM1, TEAD1	UBERON:0001830
sigmoid colon	90.54	INSM1, TEAD1	UBERON:0001159
neocortex	74.6	INSM1, TEAD1	UBERON:0001950
cerebellar hemisphere	80.2	INSM1, TEAD1	UBERON:0002245
central nervous system	76.38	INSM1, TEAD1	UBERON:0001017
right frontal lobe	73.96	INSM1, TEAD1	UBERON:0002810
colonic mucosa	85.91	INSM1, TEAD1	UBERON:0000317
ventral tegmental area	84.48	INSM1, TEAD1	UBERON:0002691
body of pancreas	80.52	INSM1, TEAD1	UBERON:0001150
muscle organ	90.9	INSM1, TEAD1	UBERON:0001630
nucleus accumbens	76.82	INSM1, TEAD1	UBERON:0001882
pylorus	93.69	INSM1, TEAD1	UBERON:0001166
mucosa of sigmoid colon	87.6	INSM1, TEAD1	UBERON:0004993
fundus of stomach	87.5	INSM1, TEAD1	UBERON:0001160
saliva-secreting gland	78.01	INSM1, TEAD1	UBERON:0001044
middle temporal gyrus	83.84	INSM1, TEAD1	UBERON:0002771
substantia nigra pars reticulata	80.74	INSM1, TEAD1	UBERON:0001966
brain	76.36	INSM1, TEAD1	UBERON:0000955
transverse colon	83.22	INSM1, TEAD1	UBERON:0001157
jejunum	90.08	INSM1, TEAD1	UBERON:0002115
small intestine	80.98	INSM1, TEAD1	UBERON:0002108
pleura	91.2	INSM1, TEAD1	UBERON:0000977
primary visual cortex	78.35	INSM1, TEAD1	UBERON:0002436
Brodmann (1909) area 23	82.19	INSM1, TEAD1	UBERON:0013554
islet of Langerhans	88.63	INSM1, TEAD1	UBERON:0000006
testis	71.68	INSM1, TEAD1	UBERON:0000473
urinary bladder	90.71	INSM1, TEAD1	UBERON:0001255
occipital lobe	79.04	INSM1, TEAD1	UBERON:0002021
intestine	85.28	INSM1, TEAD1	UBERON:0000160
tendon	89.78	INSM1, TEAD1	UBERON:0000043
calcaneal tendon	90.89	INSM1, TEAD1	UBERON:0003701
female reproductive system	87.28	INSM1, TEAD1	UBERON:0000474
lung	82	INSM1, TEAD1	UBERON:0002048
cerebellum	80.92	INSM1, TEAD1	UBERON:0002037
postcentral gyrus	85.3	INSM1, TEAD1	UBERON:0002581
corpus callosum	85.12	INSM1, TEAD1	UBERON:0002336
large intestine	86.58	INSM1, TEAD1	UBERON:0000059

right adrenal gland cortex	77.41	INSM1, TEAD1	UBERON:0035827
connective tissue	90.11	INSM1, TEAD1	UBERON:0002384
dorsolateral prefrontal cortex	73.11	INSM1, TEAD1	UBERON:0009834
muscle of leg	92.76	INSM1, TEAD1	UBERON:0001383
caecum	78.28	INSM1, TEAD1	UBERON:0001153
hypothalamus	75.19	INSM1, TEAD1	UBERON:0001898
lymph node	75.48	INSM1, TEAD1	UBERON:0000029
spleen	70.65	INSM1, TEAD1	UBERON:0002106
caudate nucleus	77.73	INSM1, TEAD1	UBERON:0001873
ascending aorta	82.77	INSM1, TEAD1	UBERON:0001496
esophagus mucosa	84.28	INSM1, TEAD1	UBERON:0002469
aorta	83.62	INSM1, TEAD1	UBERON:0000947
putamen	76.28	INSM1, TEAD1	UBERON:0001874
substantia nigra	74.79	INSM1, TEAD1	UBERON:0002038
tibial artery	84.19	INSM1, TEAD1	UBERON:0007610
skin of leg	82.36	INSM1, TEAD1	UBERON:0001511
mucosa of transverse colon	69.87	INSM1, TEAD1	UBERON:0004991
anterior cingulate cortex	73.39	INSM1, TEAD1	UBERON:0009835
metanephros	73.97	INSM1, TEAD1	UBERON:0000081
material anatomical entity	80.06	INSM1, TEAD1	UBERON:0000465
Brodmann (1909) area 9	72.33	INSM1, TEAD1	UBERON:0013540
epithelial cell of pancreas	67.07	INSM1, TEAD1	CL:0000083
ventricular zone	90.78	INSM1, TEAD1	UBERON:0003053
zone of skin	82.57	INSM1, TEAD1	UBERON:0000014
cerebral cortex	75.68	INSM1, TEAD1	UBERON:0000956
thoracic aorta	82.72	INSM1, TEAD1	UBERON:0001515
adult organism	88.18	INSM1, TEAD1	UBERON:0007023
amygdala	78.67	INSM1, TEAD1	UBERON:0001876
prefrontal cortex	73.38	INSM1, TEAD1	UBERON:0000451
right testis	68.52	INSM1, TEAD1	UBERON:0004534
telencephalon	76.61	INSM1, TEAD1	UBERON:0001893
skin of abdomen	81.58	INSM1, TEAD1	UBERON:0001416
cerebellar vermis	82.11	INSM1, TEAD1	UBERON:0004720
entorhinal cortex	85.19	INSM1, TEAD1	UBERON:0002728
muscle layer of sigmoid colon	91.41	INSM1, TEAD1	UBERON:0035805
superior frontal gyrus	81.96	INSM1, TEAD1	UBERON:0002661
anatomical system	79.75	INSM1, TEAD1	UBERON:0000467
adrenal tissue	83.48	INSM1, TEAD1	UBERON:0018303
left lobe of thyroid gland	81.19	INSM1, TEAD1	UBERON:0001120
body of stomach	82.08	INSM1, TEAD1	UBERON:0001161
right adrenal gland	77.16	INSM1, TEAD1	UBERON:0001233
pituitary gland	86.7	INSM1, TEAD1	UBERON:0000007
duodenum	80.96	INSM1, TEAD1	UBERON:0002114
epithelium of nasopharynx	87.17	INSM1, TEAD1	UBERON:0001951
cortical plate	74.26	INSM1, TEAD1	UBERON:0005343
esophagus	86.39	INSM1, TEAD1	UBERON:0001043
right hemisphere of cerebellum	79.93	INSM1, TEAD1	UBERON:0014890

upper leg skin	95.09	INSM1, TEAD1	UBERON:0004262
thymus	63.87	INSM1, TEAD1	UBERON:0002370
adrenal cortex	79.35	INSM1, TEAD1	UBERON:0001235
thyroid gland	81.71	INSM1, TEAD1	UBERON:0002046
visceral pleura	92.94	INSM1, TEAD1	UBERON:0002401
midbrain	75.43	INSM1, TEAD1	UBERON:0001891
stomach	83.67	INSM1, TEAD1	UBERON:0000945
prostate gland	82.4	INSM1, TEAD1	UBERON:0002367
pancreas	82.92	INSM1, TEAD1	UBERON:0001264
parietal lobe	84.61	INSM1, TEAD1	UBERON:0001872
colon	86.5	INSM1, TEAD1	UBERON:0001155
temporal lobe	80.92	INSM1, TEAD1	UBERON:0001871
Ammon's horn	80.01	INSM1, TEAD1	UBERON:0001954
mouth mucosa	80.93	INSM1, TEAD1	UBERON:0003729
frontal cortex	74.4	INSM1, TEAD1	UBERON:0001870
vermiform appendix	74.08	INSM1, TEAD1	UBERON:0001154
adenohypophysis	85.79	INSM1, TEAD1	UBERON:0002196
Brodmann (1909) area 46	64.82	INSM1, TEAD1	UBERON:0006483
ovary	85.95	INSM1, TEAD1	UBERON:0000992
popliteal artery	84.2	INSM1, TEAD1	UBERON:0002250
adipose tissue	90.3	INSM1, TEAD1	UBERON:0001013
rectum	83.92	INSM1, TEAD1	UBERON:0001052
cingulate cortex	73.69	INSM1, TEAD1	UBERON:0003027
substantia nigra pars compacta	79.92	INSM1, TEAD1	UBERON:0001965
blood	45.68	INSM1, TEAD1	UBERON:0000178
jejunal mucosa	84.46	INSM1, TEAD1	UBERON:0000399
ganglionic eminence	97.51	INSM1, TEAD1	UBERON:0004023
adrenal gland	79.74	INSM1, TEAD1	UBERON:0002369
forebrain	77.18	INSM1, TEAD1	UBERON:0001890
spinal cord	77.1	INSM1, TEAD1	UBERON:0002240
left adrenal gland	78.65	INSM1, TEAD1	UBERON:0001234
left adrenal gland cortex	78.78	INSM1, TEAD1	UBERON:0035825
cerebellar cortex	80.26	INSM1, TEAD1	UBERON:0002129
tonsil	82.1	INSM1, TEAD1	UBERON:0002372
oocyte	53.66	INSM1, TEAD1	CL:0000023

Supplementary Table 5. Overview of co-expression pattern of INSM1 and TEAD1 in human organs analyzed using ChIPBase v3.0 with Genotype-Tissue Expression (GTEx) project.

Organs	Sample number	Pearson coefficient *r*	p-value(two tail, t-test)
Pancreas	203	0.9393	2.66E-95
Pituitary	126	0.9348	1.32E-57
Testis	208	0.9156	1.93E-83
Brain	1426	0.8725	0
Small Intestine	106	0.6972	1.03E-16
Colon	384	0.656	1.29E-48
Prostate	122	0.6099	8.84E-14
Spleen	121	0.5391	1.78E-10
Stomach	209	0.5001	1.26E-14
Adrenal Gland	161	0.4242	2.06E-08
Salivary Gland	71	0.4179	0.000287
Lung	381	0.4001	4.45E-16
Adipose Tissue	621	0.3832	3.75E-23
Breast	221	0.3683	1.67E-08
Thyroid	366	0.3664	4.52E-13
Blood	595	0.3536	5.79E-19
Kidney	38	0.3161	0.0532
Esophagus	805	0.2829	2.77E-16
Muscle	478	0.2825	3.2E-10
Vagina	99	0.2729	0.00628
Skin	977	0.2642	4.56E-17
Ovary	112	0.2538	0.00694
Heart	493	0.2397	7.17E-08
Nerve	335	0.2322	0.0000177
Blood Vessel	753	0.226	3.54E-10
Liver	141	0.2203	0.00866
Bone Marrow	102	0.2103	0.0339
Bladder	13	0.184	0.547
Uterus	93	0.1681	0.107
Cervix Uteri	11	0.0906	0.791
Fallopian Tube	7	0	1

Supplementary Table 6. Overview of co-expression pattern of INSM1 and TEAD1 in human cancers analyzed using ChIPBase v3.0 with TCGA Pan-Cancer (PANCAN).

Diseases	Sample number	Pearson coefficient *r*	p-value (two tail, t-test)
glioblastoma multiforme	1228	0.9632	0
colon adenocarcinoma	1096	0.6826	3.36E-151
stomach adenocarcinoma	1162	0.6362	7.88E-133
uterine corpus endometrioid carcinoma	1190	0.5741	2.77E-105
lung adenocarcinoma	1410	0.4329	1.73E-65
kidney clear cell carcinoma	1888	0.3772	6.96E-65
rectum adenocarcinoma	368	0.728	5.79E-62
brain lower grade glioma	1060	0.4761	4.45E-61
ovarian serous cystadenocarcinoma	1254	0.4168	7.2E-54
lung squamous cell carcinoma	1252	0.3643	1.39E-40
prostate adenocarcinoma	1132	0.3547	6.77E-35
pancreatic adenocarcinoma	392	0.5229	6.87E-29
pheochromocytoma & paraganglioma	374	0.5119	2.27E-26
head & neck squamous cell carcinoma	1208	0.2873	2.22E-24
breast invasive carcinoma	2482	0.1279	1.59E-10
sarcoma	542	0.2286	7.44E-08
kidney papillary cell carcinoma	704	0.187	0.00000058
bladder urothelial carcinoma	872	0.1597	0.00000214
esophageal carcinoma	408	0.2205	0.00000696
thyroid carcinoma	1160	0.1272	0.000014
diffuse large B-cell lymphoma	96	0.3708	0.0002
adrenocortical cancer	184	0.2416	0.000953
liver hepatocellular carcinoma	876	0.1048	0.00191
cervical & endocervical cancer	626	0.1233	0.002
acute myeloid leukemia	400	0.1538	0.00204
skin cutaneous melanoma	960	0.0955	0.00306
testicular germ cell tumor	312	0.1434	0.0112
kidney chromophobe	182	0.1042	0.161
mesothelioma	174	-0.091	0.233
uterine carcinosarcoma	114	0.1017	0.282
thymoma	252	0.0594	0.347
uveal melanoma	160	0.0669	0.401
cholangiocarcinoma	90	-0.0845	0.428

Supplementary Table 7. Coexpression analysis of INSM1 and CTCF in human tissues and organs using Bgee database. Source data are provided as a Source Data file.

Anatomical entities	Max expression score	Genes with presence of expression	Anatomical entity IDs
embryo	95.6	INSM1, CTCF	UBERON:0000922
left testis	80.38	INSM1, CTCF	UBERON:0004533
subcutaneous adipose tissue	87.39	INSM1, CTCF	UBERON:0002190
small intestine Peyer's patch	87.86	INSM1, CTCF	UBERON:0003454
vagina	89.52	INSM1, CTCF	UBERON:0000996
kidney	86.81	INSM1, CTCF	UBERON:0002113
minor salivary gland	87.22	INSM1, CTCF	UBERON:0001830
sigmoid colon	88.8	INSM1, CTCF	UBERON:0001159
neocortex	87.41	INSM1, CTCF	UBERON:0001950
cerebellar hemisphere	89.56	INSM1, CTCF	UBERON:0002245
central nervous system	87.21	INSM1, CTCF	UBERON:0001017
right frontal lobe	86.6	INSM1, CTCF	UBERON:0002810
colonic mucosa	91.11	INSM1, CTCF	UBERON:0000317
ventral tegmental area	85.79	INSM1, CTCF	UBERON:0002691
body of pancreas	87.94	INSM1, CTCF	UBERON:0001150
muscle organ	90.14	INSM1, CTCF	UBERON:0001630
nucleus accumbens	85.96	INSM1, CTCF	UBERON:0001882
pylorus	89.37	INSM1, CTCF	UBERON:0001166
mucosa of sigmoid colon	91.34	INSM1, CTCF	UBERON:0004993
fundus of stomach	87.89	INSM1, CTCF	UBERON:0001160
saliva-secreting gland	87.26	INSM1, CTCF	UBERON:0001044
middle temporal gyrus	88.21	INSM1, CTCF	UBERON:0002771
brain	87.2	INSM1, CTCF	UBERON:0000955
transverse colon	87.97	INSM1, CTCF	UBERON:0001157
jejunum	89.51	INSM1, CTCF	UBERON:0002115
small intestine	87.92	INSM1, CTCF	UBERON:0002108
pleura	83.23	INSM1, CTCF	UBERON:0000977
primary visual cortex	87.6	INSM1, CTCF	UBERON:0002436
Brodmann (1909) area 23	86.43	INSM1, CTCF	UBERON:0013554
islet of Langerhans	91.25	INSM1, CTCF	UBERON:0000006
testis	82.28	INSM1, CTCF	UBERON:0000473
urinary bladder	89.19	INSM1, CTCF	UBERON:0001255
occipital lobe	87.51	INSM1, CTCF	UBERON:0002021
intestine	88.45	INSM1, CTCF	UBERON:0000160
tendon	87.21	INSM1, CTCF	UBERON:0000043
calcaneal tendon	86.5	INSM1, CTCF	UBERON:0003701
female reproductive system	88.96	INSM1, CTCF	UBERON:0000474
lung	86.46	INSM1, CTCF	UBERON:0002048
cerebellum	89.75	INSM1, CTCF	UBERON:0002037
type B pancreatic cell	97.57	INSM1, CTCF	CL:0000169
postcentral gyrus	84.25	INSM1, CTCF	UBERON:0002581
corpus callosum	86.59	INSM1, CTCF	UBERON:0002336
large intestine	88.6	INSM1, CTCF	UBERON:0000059

right adrenal gland cortex	86.32	INSM1, CTCF	UBERON:0035827
connective tissue	86.45	INSM1, CTCF	UBERON:0002384
dorsolateral prefrontal cortex	87.35	INSM1, CTCF	UBERON:0009834
muscle of leg	89.95	INSM1, CTCF	UBERON:0001383
caecum	89.89	INSM1, CTCF	UBERON:0001153
hypothalamus	86.22	INSM1, CTCF	UBERON:0001898
lymph node	92.48	INSM1, CTCF	UBERON:0000029
spleen	89.31	INSM1, CTCF	UBERON:0002106
paraflocculus	92.94	INSM1, CTCF	UBERON:0005351
caudate nucleus	86.61	INSM1, CTCF	UBERON:0001873
CA1 field of hippocampus	84.33	INSM1, CTCF	UBERON:0003881
ascending aorta	87.74	INSM1, CTCF	UBERON:0001496
esophagus mucosa	88.07	INSM1, CTCF	UBERON:0002469
aorta	88.93	INSM1, CTCF	UBERON:0000947
putamen	86.93	INSM1, CTCF	UBERON:0001874
substantia nigra	86.48	INSM1, CTCF	UBERON:0002038
tibial artery	89.73	INSM1, CTCF	UBERON:0007610
skin of leg	89.32	INSM1, CTCF	UBERON:0001511
mucosa of transverse colon	89.2	INSM1, CTCF	UBERON:0004991
anterior cingulate cortex	86.3	INSM1, CTCF	UBERON:0009835
metanephros	86.14	INSM1, CTCF	UBERON:0000081
material anatomical entity	87.91	INSM1, CTCF	UBERON:0000465
Brodmann (1909) area 9	87.61	INSM1, CTCF	UBERON:0013540
epithelial cell of pancreas	86.66	INSM1, CTCF	CL:0000083
ventricular zone	97.27	INSM1, CTCF	UBERON:0003053
nasopharynx	88.57	INSM1, CTCF	UBERON:0001728
zone of skin	89.36	INSM1, CTCF	UBERON:0000014
cerebral cortex	87.23	INSM1, CTCF	UBERON:0000956
thoracic aorta	87.83	INSM1, CTCF	UBERON:0001515
adult organism	88.46	INSM1, CTCF	UBERON:0007023
amygdala	87.01	INSM1, CTCF	UBERON:0001876
prefrontal cortex	88.72	INSM1, CTCF	UBERON:0000451
right testis	80.99	INSM1, CTCF	UBERON:0004534
telencephalon	86.98	INSM1, CTCF	UBERON:0001893
skin of abdomen	89.36	INSM1, CTCF	UBERON:0001416
cerebellar vermis	93.66	INSM1, CTCF	UBERON:0004720
entorhinal cortex	82.37	INSM1, CTCF	UBERON:0002728
muscle layer of sigmoid colon	88.21	INSM1, CTCF	UBERON:0035805
superior frontal gyrus	84.73	INSM1, CTCF	UBERON:0002661
anatomical system	87.99	INSM1, CTCF	UBERON:0000467
adrenal tissue	85.64	INSM1, CTCF	UBERON:0018303
left lobe of thyroid gland	86.03	INSM1, CTCF	UBERON:0001120
body of stomach	86.69	INSM1, CTCF	UBERON:0001161
right adrenal gland	86.31	INSM1, CTCF	UBERON:0001233
pituitary gland	86.7	INSM1, CTCF	UBERON:0000007
duodenum	86.82	INSM1, CTCF	UBERON:0002114
epithelium of nasopharynx	88.59	INSM1, CTCF	UBERON:0001951

cortical plate	96.45	INSM1, CTCF	UBERON:0005343
esophagus	87.7	INSM1, CTCF	UBERON:0001043
right hemisphere of cerebellum	88.86	INSM1, CTCF	UBERON:0014890
upper leg skin	91.53	INSM1, CTCF	UBERON:0004262
thymus	96.39	INSM1, CTCF	UBERON:0002370
adrenal cortex	85.99	INSM1, CTCF	UBERON:0001235
thyroid gland	86.4	INSM1, CTCF	UBERON:0002046
visceral pleura	82.52	INSM1, CTCF	UBERON:0002401
secondary oocyte	81.11	INSM1, CTCF	CL:0000655
orbitofrontal cortex	85.02	INSM1, CTCF	UBERON:0004167
midbrain	86.45	INSM1, CTCF	UBERON:0001891
choroid plexus epithelium	81.13	INSM1, CTCF	UBERON:0003911
stomach	87.05	INSM1, CTCF	UBERON:0000945
prostate gland	88.14	INSM1, CTCF	UBERON:0002367
pancreas	89.01	INSM1, CTCF	UBERON:0001264
parietal lobe	84.46	INSM1, CTCF	UBERON:0001872
colon	88.56	INSM1, CTCF	UBERON:0001155
temporal lobe	85.54	INSM1, CTCF	UBERON:0001871
Ammon's horn	87.24	INSM1, CTCF	UBERON:0001954
mouth mucosa	87.36	INSM1, CTCF	UBERON:0003729
frontal cortex	87.28	INSM1, CTCF	UBERON:0001870
vermiform appendix	89.61	INSM1, CTCF	UBERON:0001154
adenohypophysis	85.79	INSM1, CTCF	UBERON:0002196
Brodmann (1909) area 46	86.68	INSM1, CTCF	UBERON:0006483
ovary	88.6	INSM1, CTCF	UBERON:0000992
popliteal artery	89.74	INSM1, CTCF	UBERON:0002250
adipose tissue	86.31	INSM1, CTCF	UBERON:0001013
rectum	89.14	INSM1, CTCF	UBERON:0001052
cingulate cortex	86.41	INSM1, CTCF	UBERON:0003027
substantia nigra pars compacta	86.38	INSM1, CTCF	UBERON:0001965
blood	90.64	INSM1, CTCF	UBERON:0000178
jejunal mucosa	88.39	INSM1, CTCF	UBERON:0000399
ganglionic eminence	97.51	INSM1, CTCF	UBERON:0004023
adrenal gland	86.5	INSM1, CTCF	UBERON:0002369
forebrain	86.78	INSM1, CTCF	UBERON:0001890
spinal cord	87.45	INSM1, CTCF	UBERON:0002240
left adrenal gland	86.1	INSM1, CTCF	UBERON:0001234
left adrenal gland cortex	85.83	INSM1, CTCF	UBERON:0035825
cerebellar cortex	89.64	INSM1, CTCF	UBERON:0002129
multicellular organism	87.91	INSM1, CTCF	UBERON:0000468
skeletal muscle organ	90.14	INSM1, CTCF	UBERON:0014892
tonsil	91.01	INSM1, CTCF	UBERON:0002372
oocyte	76.4	INSM1, CTCF	CL:0000023

Supplementary Table 8. Overview of co-expression of INSM1 and CTCF in human organs analyzed using ChIPBase v3.0 with Genotype-Tissue Expression (GTEx) project.

Organs	Sample number	Pearson coefficient *r*	p-value (two tail, t-test)
Pancreas	203	0.9352	1.34E-92
Pituitary	126	0.9337	3.66E-57
Testis	208	0.9198	1.28E-85
Brain	1426	0.9087	0
Small Intestine	106	0.8372	5.15E-29
Colon	384	0.7626	2.89E-74
Stomach	209	0.6414	1.3E-25
Prostate	122	0.6364	3.35E-15
Spleen	121	0.5814	2.72E-12
Skin	977	0.4357	1.54E-46
Lung	381	0.4142	3.14E-17
Salivary Gland	71	0.4009	0.000531
Adrenal Gland	161	0.3976	0.000000177
Adipose Tissue	621	0.3911	3.91E-24
Thyroid	366	0.3754	1.08E-13
Breast	221	0.3743	9.34E-09
Blood	595	0.3652	3.29E-20
Kidney	38	0.3451	0.0339
Esophagus	805	0.3095	2.5E-19
Vagina	99	0.2943	0.00311
Muscle	478	0.2791	5.3E-10
Heart	493	0.2546	9.89E-09
Ovary	112	0.2511	0.00756
Nerve	335	0.2503	0.00000351
Bladder	13	0.2475	0.415
Blood Vessel	753	0.2403	2.37E-11
Liver	141	0.228	0.00654
Bone Marrow	102	0.2097	0.0344
Uterus	93	0.1716	0.1
Cervix Uteri	11	0.1122	0.743
Fallopian Tube	7	0	1

Supplementary Table 9. Overview of co-expression of INSM1 and CTCF in human cancers analyzed using ChIPBase v3.0 with TCGA Pan-Cancer (PANCAN).

Diseases	Sample number	Pearson coefficient *r*	p-value (two tail, t-test)
glioblastoma multiforme	1228	0.9672	0
rectum adenocarcinoma	368	0.7145	9.87E-59
colon adenocarcinoma	1096	0.6945	1.28E-158
pheochromocytoma & paraganglioma	374	0.6417	8.9E-45
stomach adenocarcinoma	1162	0.6398	8.78E-135
brain lower grade glioma	1060	0.6086	2.13E-108
uterine corpus endometrioid carcinoma	1190	0.5735	5.46E-105
pancreatic adenocarcinoma	392	0.5413	3.24E-31
prostate adenocarcinoma	1132	0.4663	3.58E-62
lung adenocarcinoma	1410	0.4382	3.24E-67
ovarian serous cystadenocarcinoma	1254	0.4264	1.48E-56
kidney clear cell carcinoma	1888	0.3827	6.55E-67
lung squamous cell carcinoma	1252	0.3607	9.29E-40
diffuse large B-cell lymphoma	96	0.3572	0.000353
thymoma	252	0.3194	0.00000022
head & neck squamous cell carcinoma	1208	0.3032	4.19E-27
cholangiocarcinoma	90	0.2886	0.00581
adrenocortical cancer	184	0.2569	0.000431
testicular germ cell tumor	312	0.2456	0.0000115
kidney papillary cell carcinoma	704	0.2099	0.000000019
breast invasive carcinoma	2482	0.2051	5.64E-25
esophageal carcinoma	408	0.2019	0.0000398
acute myeloid leukemia	400	0.1988	0.0000626
skin cutaneous melanoma	960	0.1743	5.49E-08
bladder urothelial carcinoma	872	0.1695	0.000000479
mesothelioma	174	0.1671	0.0275
liver hepatocellular carcinoma	876	0.153	0.0000054
sarcoma	542	0.1516	0.000398
cervical & endocervical cancer	626	0.1427	0.000342
kidney chromophobe	182	0.1237	0.0963
thyroid carcinoma	1160	0.1186	0.0000517
uterine carcinosarcoma	114	0.0969	0.305
uveal melanoma	160	0.0283	0.723

Supplementary Table 10. Summary of HDX-MS experiment and data.

Data Set	ZF1-ZF5	ZF1-ZF5 /M2 DNA
HDX reaction details	100 mM sodium phosphate, pD 6.6, 20 °C	100 mM sodium phosphate, pD 6.6, 20 °C
HDX time course (min)	1, 10, 30, 60	1, 10, 30, 60
HDX control samples	Maximally-labeled ZF1-ZF5	Maximally-labeled ZF1-ZF5
Back-exchange (mean / IQR)	44.51%/35.75%	
# of Peptides	31	31
Sequence coverage	96%	96%
Average peptide length / Redundancy	15.83/2.31	15.83/2.31
Replicates (biological or technical)	3 (technical)	3 (technical)
Repeatability	0.370 (average standard deviation)	0.476 (average standard deviation)
Significant differences in HDX (delta HDX > X D)	0.5 D (99%CI)	

Supplementary Table 11. DNA fragments used in EMSA.

Name	Sequence	Length (bp)
M1	FAM-5′-TGTCAGGGGGCA-3′ 5′-TGCCCCCTGACA -3′	12
M2	FAM-5′-TACCACCAGGGGGCAGT-3′ 5′-ACTGCCCCCTGGTGGTA -3′	17
NeuroD2-p	FAM-5′-GTGGTGGTGGAAGGGGGCGGGAGGAAAGT-3′ 5′-ACTTTCCTCCCGCCCCCTTCCACCACCAC -3′	29
Insulin-p	FAM-5′-AAAGTCCAGGGGGCAGAG-3′ 5′-CTCTGCCCCCTGGACTTT -3′	18
Nega1	FAM-5′-ATGGACGAGTCATAGGA-3′ 5′-TCCTATGACTCGTCCAT -3′	17
Nega2	FAM-5′-GTGATGCTGGAAGTGAGAGTGAGCAGAGT-3′ 5′-ACTCTGCTCACTCTCACTTCCAGCATCAC -3′	29
MCAT	FAM-5′-AGTGGAATGTGC-3′ 5′-GCACATTCCACT -3′	12

Supplementary Table 12. Primers used in ChIP-PCR and qPCR.

Gene	Forward primer	Reverse primer
CTGF	5′-AATGCGAGGAATGTCCCTGTT-3′	5′-GCGGCCCGAGGCTTTTATAC-3′
ANKRD1	5′-CAGCTGTCCCCTGACTCTTG-3′	5′-AGAGGGCCATTCCTTTGGTG-3′
AMOTL2	5′-TCGGCGGCGAAGATGTGT-3′	5′-TTTGTGGTTCGCTTTGATCCC-3′
WTIP	5′-CTCCACGCCCCTCCGA-3′	5′-GAATGCGAGGAAGGAGACGG-3′
CYR61	5′-CAAAGGTGCAATGGAGCCAG-3′	5′-CGGAGCCCGCCTTTTATACG-3′
AJUBA	5′-AGAACATGCCTCCTGTCGTC-3′	5′-TTGTCTGCAGTTGGAATGCG-3′
EGR1	5′-CAGGCACCTCTTAATGCTTGTC-3′	5′-GCCAGGGAAAGTTTGTGCTG-3′
PUMA	5′-ATGCTGCGCCTGCTAAATG-3′	5′-CAGACGCTGATGCCAGACA-3′
MYC	5′-ACTTTCGCAAACCTGAACGC-3′	5′-GCAACCAATCGCTATGCTGG-3′
CPM-P4	5′-TAAAGGGGTGTTTGCTGCCA-3′	5′-TGCCCCCTTGTGGTTTAAGA-3′
IFI6-P9	5′-AGTAAACGGTTCTCCGGCTG-3′	5′-AAACTCCGACAGGGATTGGC-3′
MDM2-P3	5′-TGTATGAACGCATACCTGCC-3′	5′-CATCATGCCATCTAGCGGTC-3′
KLK-ChIP7	5′-TGCGCCTACAACTGTGCAAT-3′	5′-ACAACCTTTAGTCCTCCCCG-3′
KLK-ChIP8	5′-TGCCCCTCCCACAACTTTC-3′	5′-CATTTGCTCTATCGCGCCA-3′
KLK-ChIP9	5′-CACCGGACCCTGGCTTCT-3′	5′-TCACTCCTCTCCCGCTGAT-3′
KLK-ChIP11	5′-TGATGTCAATGACCACCGGA-3′	5′-GTGCTCTTTTCGTCTCTGGC-3′
KLK-ChIP12	5′-CACCCCCTACAGCTCATTC-3′	5′-CAACAGTGGGTCCAATTTCC-3′
KLK-ChIP13	5′-CCATCTCCCCGGGAAGTTT-3′	5′-CGGGTCTCAGGGTTGCTGT-3′
HBB-NC	5′-TGACATAACAGTGTTCACTAGC-3′	5′-CCAACTGCATCCACGTTCAC-3′
FAT3-NC	5′-GGCTTCCACTTCACACATTCC-3′	5′-TGCCCATTCTACTCTGGCTGTT-3′
ASH1-NC	5′-TGGAGAGAGGTAGGGCAG-3′	5′-AAGAGCCCCACTTTGGAAAT-3′
KLK8-NC	5′-CCGACAAACCTGGCGTCTAT-3′	5′-GCCCTTGCTGCCTATGATCT-3′

Supplementary Table 13. Primers used in qPCR for gene expression analysis.

Gene	Forward primer	Reverse primer
CTGF	5′-GTTTGGCCCAGACCCAACTA-3′	5′-GGCTCTGCTTCTCTAGCCTG-3′
ANKRD1	5′-GGTGAGGACTGGCCACTATG-3′	5′-GGTTCAGTCTCACCGCATCA-3′
AMOTL2	5′-GACATGACCAAGTGGGAGCA-3′	5′-ATGAGAGTGGTGTCACGCTG-3′
WTIP	5′-GGCATGTTACCACTGTGAGG-3′	5′-TGGCAACGACGACACAGTAG-3′
CYR61	5′-CAGGACTGTGAAGATGCGGT-3′	5′-GCCTGTAGAAGGGAAACGCT-3′
AJUBA	5′-CCAGGGAGGACTACTTCGGA-3′	5′-CTCTCCACTGCAGACTTCCA-3′
EGR1	5′-CCTTCGCTAACCCCTCTGTC-3′	5′-TGGGTTTGATGAGCTGGGAC-3′
PUMA	5′-ACCTCAACGCACAGTACGAG-3′	5′-TAAGGGCAGGAGTCCCATGA-3′
MYC	5′-CCCTCCACTCGGAAGGACTA-3′	5′-GCTGGTGCATTTTCGGTTGT-3′
CPM	5′-TCAAGGCAGCCTGAAACTGT-3′	5′-GCTGGTGCATTTTCGGTTGT-3′
IFI6	5′-GGGTGGAGGCAGGTAAGAAA-3′	5′-GTCAGGGCCTTCCAGAACC-3′
KLK-5	5′-CAAAGTGCTTGGTGTCTGGC-3′	5′-GCGCAGAACATGGTGTCATC-3′
KLK-8	5′-CCGACAAACCTGGCGTCTAT-3′	5′-GCCCTTGCTGCCTATGATCT-3′
KLK-9	5′-TGCAGTGTCTCATCTCAGGC-3′	5′-CAGGGTATGCCCAGTGACAG-3′
MDM2	5′-AGGAGATTTGTTTGGCGTGC-3′	5′-GAGTCCGATGATTCCTGCTG-3′
FAT3	5′-TGATATGGACTGGGGAGCC-3′	5′-AGCCCGTGTTGCTGTCAAT-3′
SLIT2	5′-AAGGTGTCCCGATTAGAGTG-3′	5′-GGAGCCGTCAGTGCATTCG-3′
GAPDH	5′-CTCCTGCACCACCAACTGC-3′	5′-GGCCATCCACAGTCTTCTG-3′

Supplementary Table 14. DNA fragments used for CRISPR/Cas9 plasmid construction targeting *INSM1*, *TEAD1* and *CTCF*.

Target	Forward	Reverse
INSM1-T1	5′-CACCGAGTCCACGCCCGTTTCCTAC-3′	5′-AAACGTAGGAAACGGGCGTGGACTC-3′
INSM1-T2	5′-CACCGTGCTGCTCTCGCCCAGCTGC-3′	5′-AAACGCAGCTGGGCGAGAGCAGCAC-3′
INSM1-T3	5′-CACCGCGCGCACTTCGGCAACCCCG-3′	5′-AAACCGGGGTTGCCGAAGTGCGCGC-3′
TEAD1-T1	5′-CACCGCAAGACGAGGACCAGAAAAC-3′	5′-AAACGTTTTCTGGTCCTCGTCTTGC-3′
TEAD1-T2	5′-CACCGCGACATCGAGCAAAGCTTTC-3′	5′-AAACGAAAGCTTTGCTCGATGTCGC-3′
TEAD1-T3	5′-CACCGACATTCAGGTTCTTGCCAGA-3′	5′-AAACTCTGGCAAGAACCTGAATGTC-3′
CTCF-T1	5′-CACCGTGCAGTCGAAGCCATTGTGG-3′	5′-AAACCCACAATGGCTTCGACTGCAC-3′
CTCF-T2	5′-CACCGACTTACCAGAGACGCCGGGA-3′	5′-AAACTCCCGGCGTCTCTGGTAAGTC-3′
CTCF-T3	5′-CACCGGATGGTGATGATGGAACAGC-3′	5′-AAACGCTGTTCCATCATCACCATCC-3′