

# The PTM profiling of CTCF reveals the regulation of 3D chromatin structure by O-GlcNAcylation

Xiuxiao Tang<sup>1,2,3,4,5,14</sup>, Pengguihang Zeng<sup>1,2,3,4,14</sup>, Kezhi Liu<sup>1,2,3,4</sup>, Li Qing<sup>1,2,3,4</sup>, Yifei Sun<sup>1,2,3,4</sup>, Xinyi Liu<sup>1,2,3,4</sup>, Lizi Lu<sup>1,2,3,4</sup>, Chao Wei<sup>1,2,3,4</sup>, Jia Wang<sup>6</sup>, Shaoshuai Jiang<sup>1,2,3,4</sup>, Jun Sun<sup>7,8</sup>, Wakam Chang<sup>9</sup>, Haopeng Yu<sup>7,8</sup>, Hebing Chen<sup>10</sup>, Jiaguo Zhou<sup>5</sup>, Chengfang Xu<sup>11,\*</sup>, Lili Fan<sup>12,\*</sup>, Yi-Liang Miao<sup>13,\*</sup>, Junjun Ding<sup>1,2,3,4,7,\*</sup>

<sup>1</sup> RNA Biomedical Institute, Sun Yat-Sen Memorial Hospital, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China

<sup>2</sup> Department of Rehabilitation Medicine, The Seventh Affiliated Hospital, Sun Yat-Sen University, Shenzhen, Guangdong, 518107, China

<sup>3</sup> Advanced Medical Technology Center, The First Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, Guangdong, China

<sup>4</sup> Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-Sen University, Guangzhou 510080, China

<sup>5</sup> Department of Pharmacology and Cardiac & Cerebral Vascular Research Center, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China

<sup>6</sup> GMU-GIBH Joint School of Life Sciences, Guangzhou Medical University, Guangzhou, 511436, China

<sup>7</sup> West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu 610041, CHINA

<sup>8</sup> Med-X Center for Informatics, Sichuan University, Chengdu 610041, China

<sup>9</sup> Department of Biomedical Sciences, Faculty of Health Sciences, University of Macau, Taipa, Macau, China

<sup>10</sup> Institute of Health Service and Transfusion Medicine, Beijing 100850, China

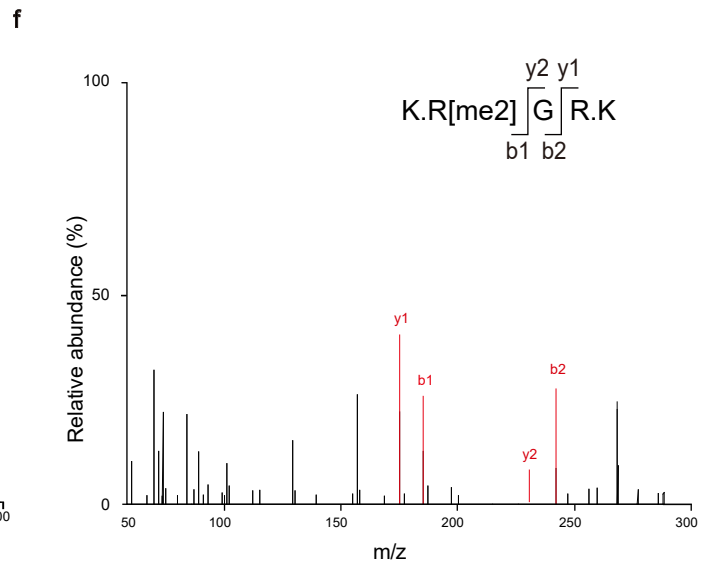
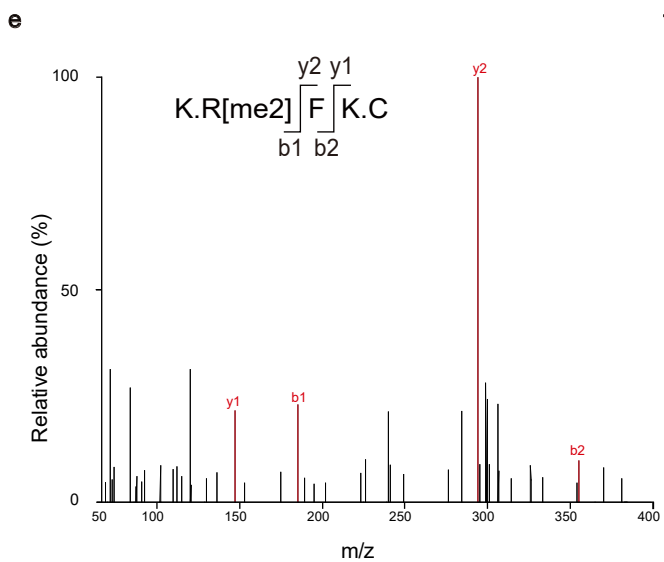
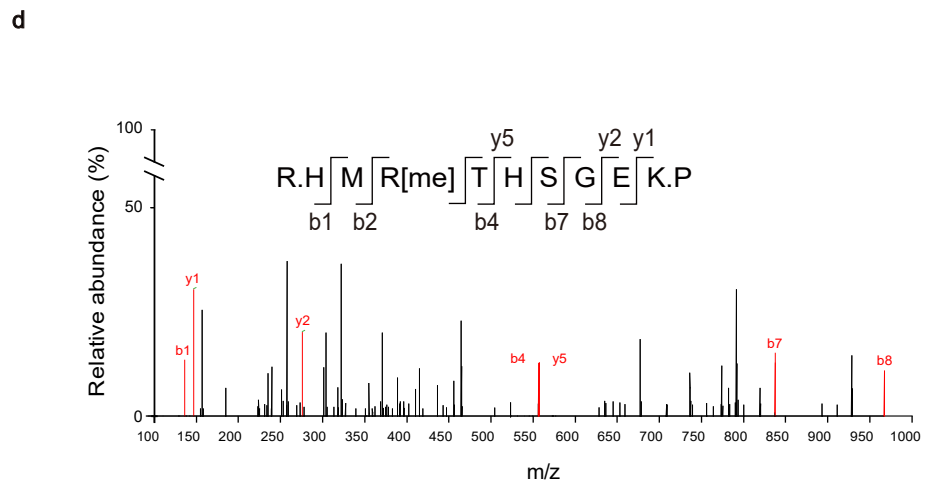
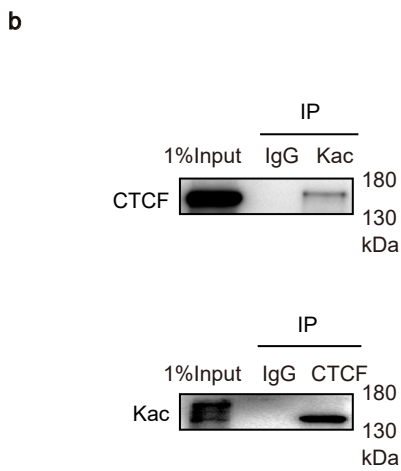
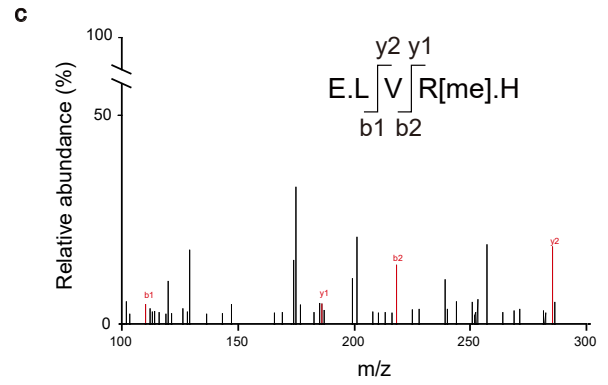
<sup>11</sup> The obstetric and gynecology Department of The third affiliated hospital of Sun Yat-Sen University

<sup>12</sup> Guangzhou Key Laboratory of Formula-Pattern of Traditional Chinese Medicine, School of Traditional Chinese Medicine, Jinan University, Guangzhou, Guangdong, China

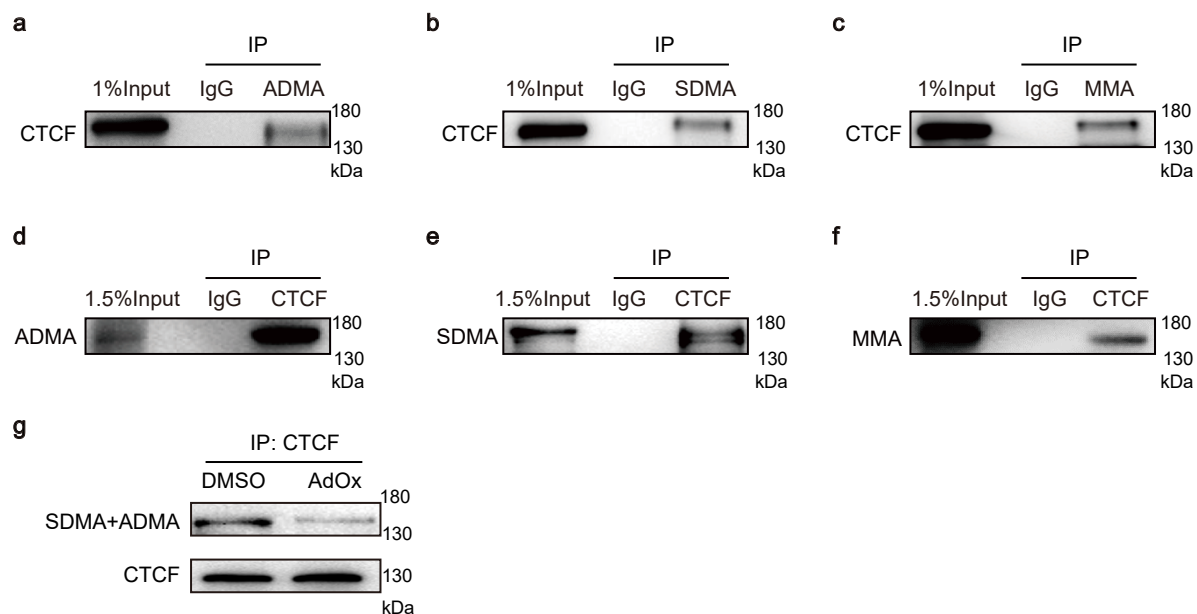
<sup>13</sup> Institute of Stem Cell and Regenerative Biology, College of Animal Science and Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei, China

<sup>14</sup> These authors contributed equally to the work

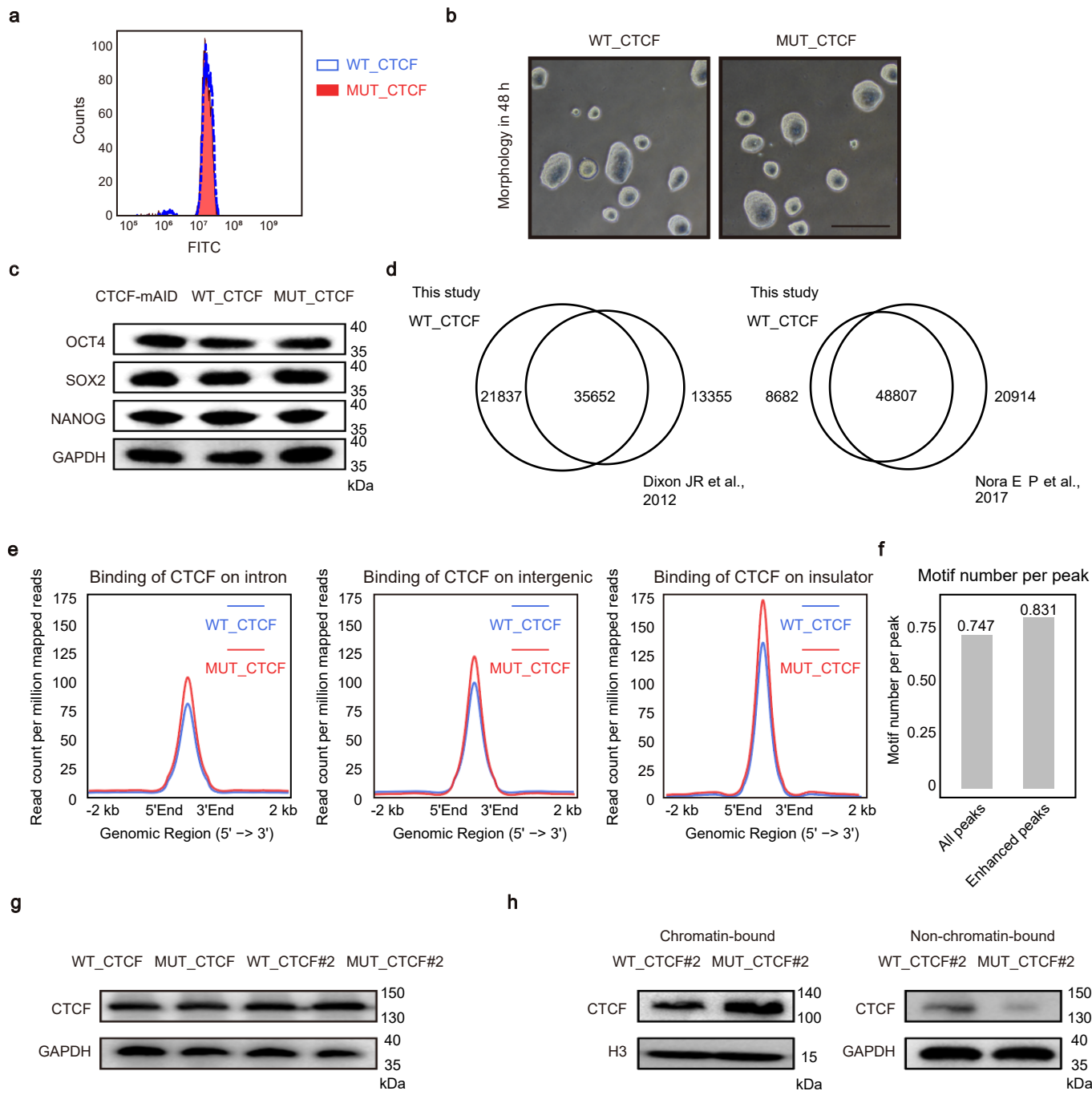
\* Correspondence: [xuchengf@mail.sysu.edu.cn](mailto:xuchengf@mail.sysu.edu.cn); [fanlili@jnu.edu.cn](mailto:fanlili@jnu.edu.cn); [miaoyl@mail.hzau.edu.cn](mailto:miaoyl@mail.hzau.edu.cn); [dingjunj@mail.sysu.edu.cn](mailto:dingjunj@mail.sysu.edu.cn)



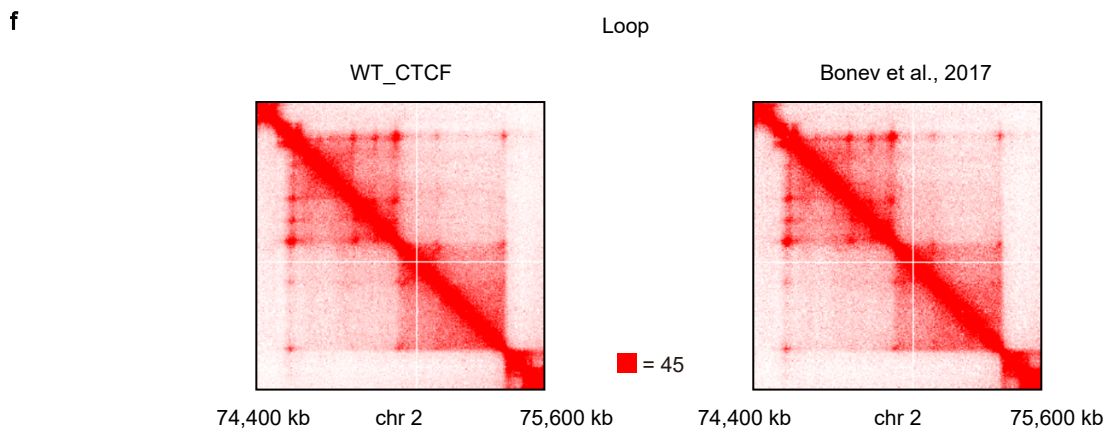
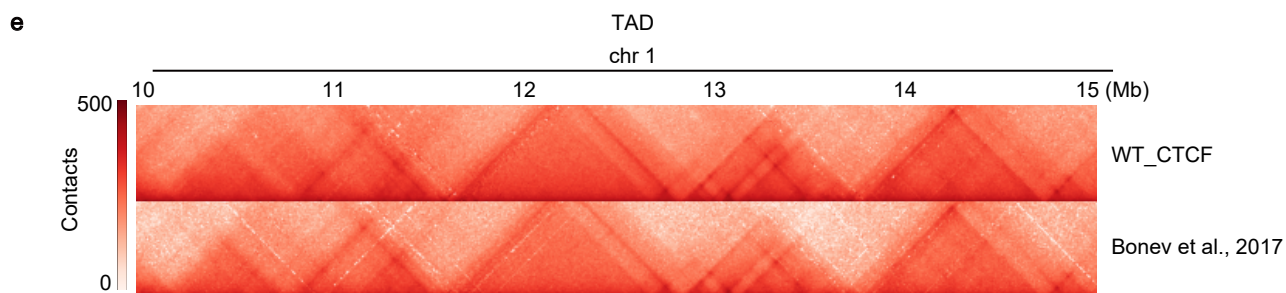
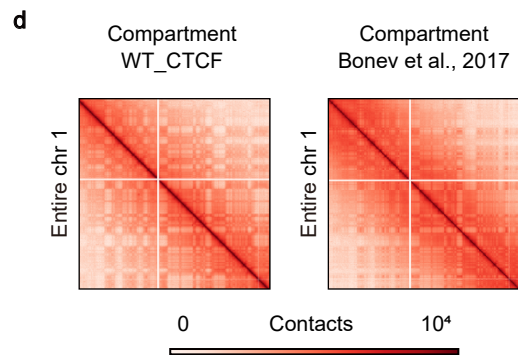
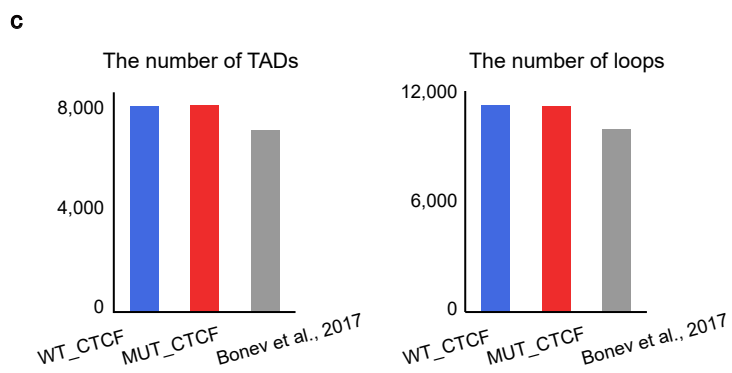
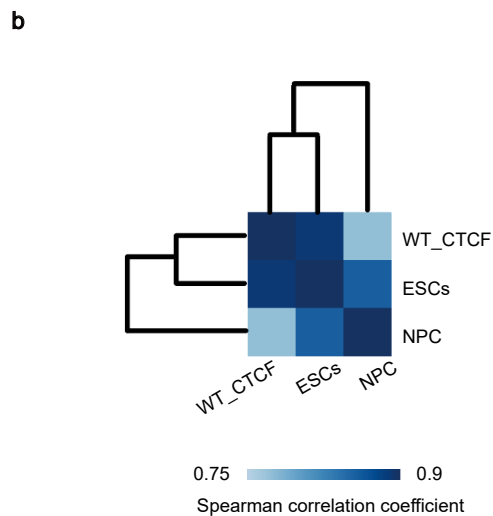
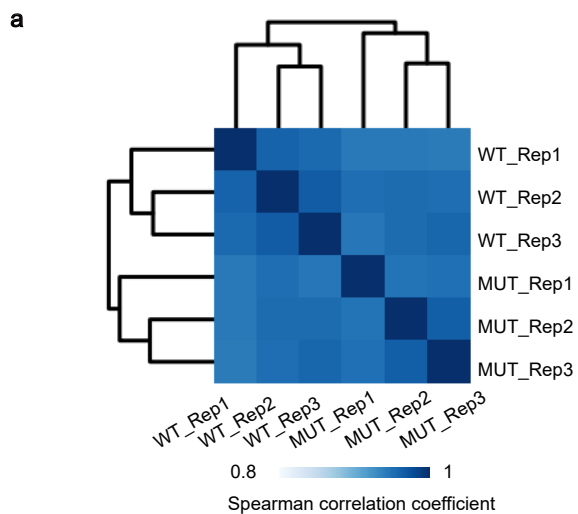
**Supplementary Fig. 1 The purification of CTCF, Related to Fig. 1.** a, Coomassie blue staining for purification of CTCF. This experiment was performed three times, with similar results. b, Whole-cell lysates were denatured and incubated with Kac-conjugated agarose beads to enrich acetylated proteins and bound proteins were analyzed by western blot (top) and whole-cell lysates were denatured and immunoprecipitated with antibodies for CTCF. Acetylation of proteins was detected by western blot with Kac (bottom). This experiment was performed three times, with similar results. c-f, LC-MS/MS spectra of methylated peptides of endogenous CTCF derived from mESCs and CTCF was purified by immunoprecipitation with antibody. The matched fragment ions were labeled in red. Source data are provided as a Source Data file.



**Supplementary Fig. 2 Validation of arginine methylation, Related to Fig. 2.** a-c, Whole-cell lysates were denatured and incubated with ADMA (a), SDMA (b) and MMA (c) -conjugated agarose beads to enrich methylated proteins respectively, and bound proteins were analyzed by western blot. This experiment was performed three times, with similar results. d-f, Whole-cell lysates were denatured and immunoprecipitated with antibody for CTCF. Methylation of proteins was detected by western blot with ADMA (d), SDMA (e) and MMA (f), respectively. This experiment was performed three times, with similar results. g, The methylation signal of CTCF decreased when treated with an arginine methylation inhibitor, periodate oxidized adenosine (AdOx). This experiment was performed three times, with similar results. ADMA: asymmetrical dimethylarginine; SDMA: symmetrical dimethylarginine; MMA: mono-methylarginine. Source data are provided as a Source Data file.

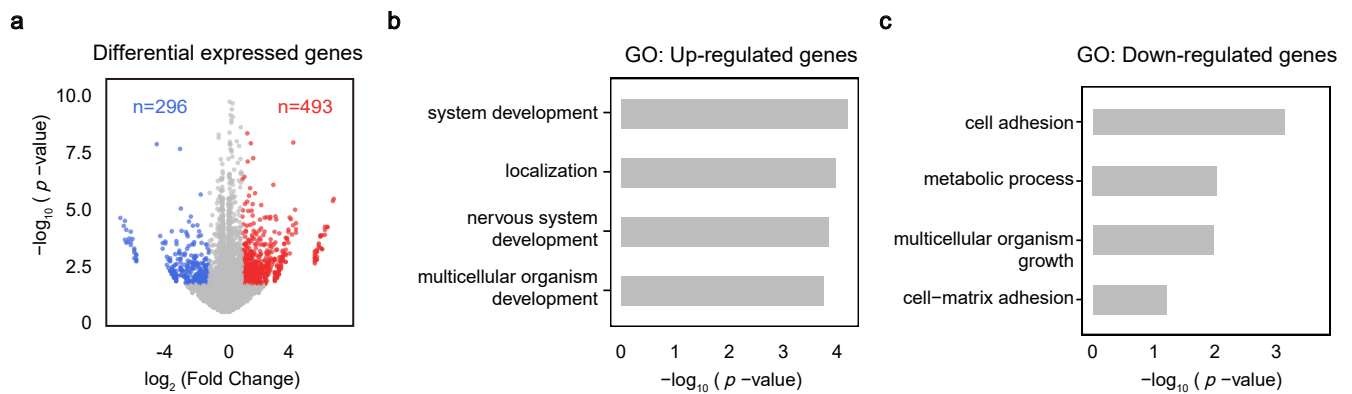


**Supplementary Fig. 3 Validation of cell lines and bioinformatics analysis of CTCF ChIP-seq, Related to Fig. 3.** a, FACS analysis showing the consistent expression of CTCF in WT\_CTCF and MUT\_CTCF mESCs. b, The cell morphology of WT\_CTCF and MUT\_CTCF mESCs in 48 h. Scale bar was 500  $\mu$ m. c, Western blot of total cell lysates from CTCF-mAID mESCs, WT\_CTCF cells and MUT\_CTCF cells to detect the expression of OCT4, SOX2 and NANOG after 48 h auxin treatment. GAPDH was shown as a loading control. d, Venn diagram showing the overlap between CTCF ChIP-seq peaks of WT\_CTCF mESCs in this study and two published CTCF ChIP-seq datasets. e, Profiling of CTCF binding indicating the overall binding of O-GlcNAc-deficient CTCF on intron (left), intergenic (middle) and insulator (right) were enhanced. f, Boxplot showing there were more CTCF motifs on the enhanced peaks. g, Western blot of total cell lysates from WT\_CTCF cells, MUT\_CTCF cells, WT\_CTCF clone#2 and MUT\_CTCF clone#2 to detect the expression of CTCF. GAPDH was showing as a loading control. h, Western blot of chromatin-bound CTCF (left) and non-chromatin-bound CTCF (right) in WT\_CTCF clone#2 and MUT\_CTCF clone#2. H3 and GAPDH were showing as loading control. All western blot experiments were performed three times, with similar results. Source data are provided as a Source Data file.

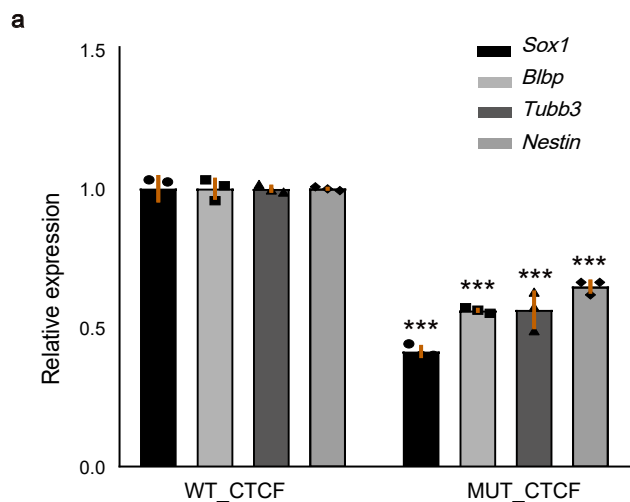


**Supplementary Fig. 4 Quality control metrics of Hi-C data, Related to Fig 4.** a, Spearman correlation coefficients among and between the WT\_CTCF and MUT\_CTCF conditions. b, Spearman correlation coefficients between the WT\_CTCF in this study, ESCs and neural progenitor cells (NPCs) in Bonev et al., 2017. c, The number of TADs and loops identified in this study and in Bonev et al., 2017. d-f, Hi-C contact heatmaps showing compartments (d), TADs (e), loops (f) in this study and in Bonev et al., 2017. Source data are provided as a Source Data file.

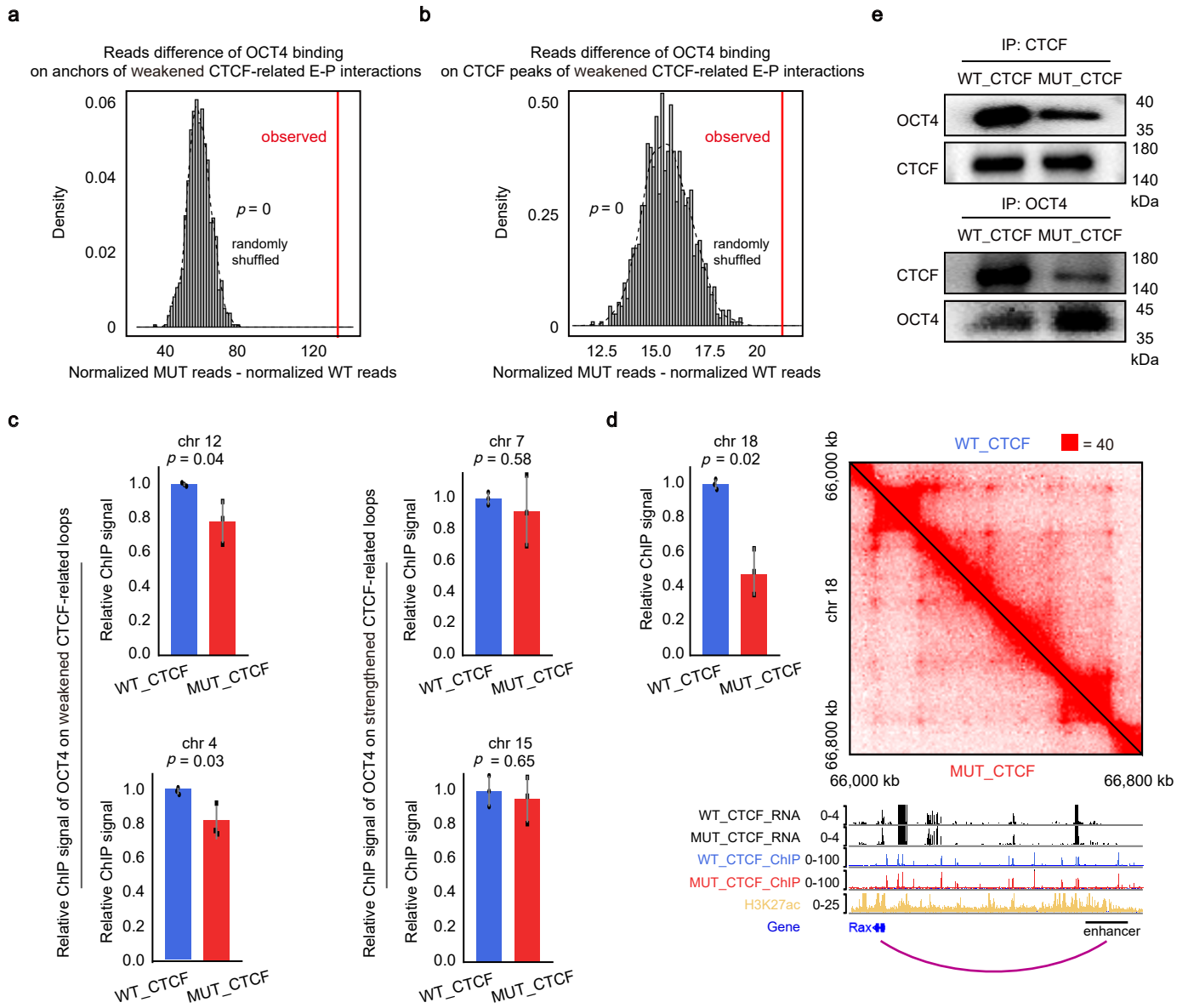




**Supplementary Fig. 5 Deficiency of CTCF O-GlcNAcylation affects gene expression, Related to Fig 6.** a, Volcano plot comparing gene expression in MUT\_CTCF mESCs against WT\_CTCF mESCs. Gray dots, genes with no significant change; blue dots, down-regulated genes; red dots, up-regulated genes, p-values by exact test based on over-dispersed Poisson regression, performed by R package edgeR. b, GO analysis revealed the up-regulated genes were associated with developmental processes, p-values by Fisher's Exact test. c, GO analysis revealed the down-regulated genes were associated with cell adhesion and metabolic process, p-values by Fisher's Exact test. Source data are provided as a Source Data file.

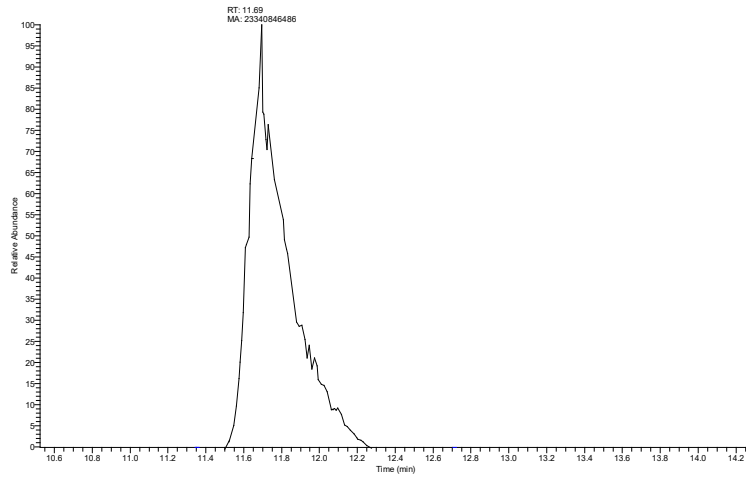


**Supplementary Fig. 6 Mutation of CTCF O-GlcNAcylation inhibits the differentiation into NPCs, Related to Fig 7. a,** Quantitative PCR analysis of NPCs markers. All statistical tests were two-sided Student t-test. Significance levels were indicated by asterisks: \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ . All data were presented as mean  $\pm$  SD ( $n = 3$  biologically independent samples). Source data are provided as a Source Data file.

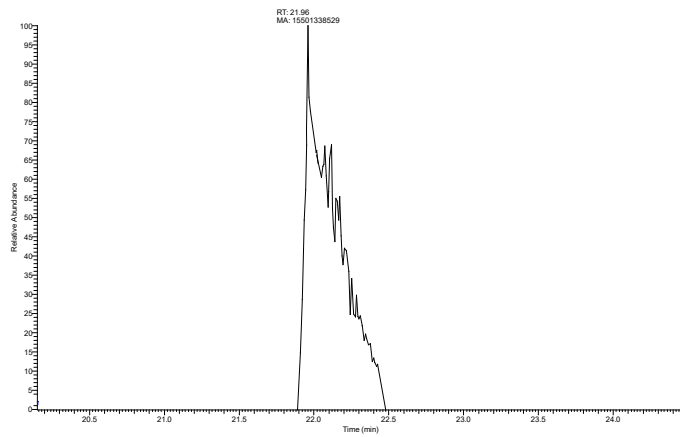


**Supplementary Fig. 7 Weakened CTCF-related E-P interactions exhibit a higher level of OCT4 signals.** a, Permutation test showing OCT4 had higher level of binding on anchors of weakened CTCF-related E-P interactions than randomly shuffled set of loops. p-value by one-tailed test based on the assumption of normal distribution,  $p=0$ . b, Permutation test showing OCT4 had higher level of binding on CTCF peaks of weakened CTCF-related E-P interactions than randomly shuffled set of loops. p-value by one-tailed test based on the assumption of normal distribution,  $p=0$ . c, CHIP-qPCR showing the enrichment of OCT4 on the anchors of weakened (left) or strengthened (right) CTCF-related loops. All statistical tests were two-sided Student t-test. All data were presented as mean  $\pm$  SD ( $n = 3$  biologically independent samples). d, Representative genomic loci showing weakened loop targeted genes and CHIP-qPCR showing the enrichment of OCT4 on the anchors of weakened loops. Data were presented as mean  $\pm$  SD ( $n = 3$  biologically independent samples). e, Western blot showing that mutation of CTCF O-GlcNAcylation weakened the interaction between CTCF and OCT4. Source data are provided as a Source Data file.

**a**



**b**



**Supplementary Fig. 8. Extracted ion chromatograms of peptide.** a, Extracted ion chromatograms of the O-GlcNAcylated peptide. b, Extracted ion chromatograms of the unmodified peptide.

Supplementary Table 1. ChIP-seq statistics

Sample	Total	Unique	Multiple	Unmapped
WT_CTCF.R1	74480336	53126196	15070262	6283878
WT_CTCF.R2	67958130	48348223	13757748	5852159
MUT_CTCF.R1	71157312	49830929	17666060	3660323
MUT_CTCF.R2	73646740	51575000	18231272	3840468
WT_CTCF_input.R1	63120404	38638698	20748354	3733352
MUT_CTCF_input.R1	57369566	21053135	31171469	5144962

Supplementary Table 2. Oligonucleotides

Primers used in this paper	
<i>Fgf5</i> realtime-F	TGTGTCTCAGGGGATTGTAGG
<i>Fgf5</i> realtime-R	AGCTGTTTTCTTGGAATCTCTCC
<i>Pax3</i> realtime-F	GCAGCGCAGGAGCAGAACCA
<i>Pax3</i> realtime-R	GCACTCGGGCCTCGGTAAGC
<i>Mixl1</i> realtime-F	ATCCGCCCGGACCCTCCAAA
<i>Mixl1</i> realtime-R	TCGGTTCTGGAACCACACCTGGA
<i>Gata4</i> realtime-F	CCCTACCCAGCCTACATGG
<i>Gata4</i> realtime-R	ACATATCGAGATTGGGGTGTCT
<i>Gata6</i> realtime-F	TTGCTCCGGTAACAGCAGTG
<i>Gata6</i> realtime-R	GTGGTCGCTTGTGTAGAAGGA
<i>Cdx2</i> realtime-F	CAAGGACGTGAGCATGTATCC
<i>Cdx2</i> realtime-R	GTAACCACCGTAGTCCGGGTA
<i>Oct3/4</i> realtime-F	GTGGAGGAAGCCGACAACAATGA
<i>Oct3/4</i> realtime-R	CAAGCTGATTGGCGATGTGAG
<i>Nanog</i> realtime-F	TGGTCCCCACAGTTTGCCTAGTTC
<i>Nanog</i> realtime-R	CAGGTCTTCAGAGGAAGGGCGA
<i>Sox2</i> realtime 1-F	CTGGACTGCGAACTGGAGAAG
<i>Sox2</i> realtime 1-R	AATTTGGATGGGATTGGTGGT
<i>sox1</i> realtime-F	GTGACATCTGCCCCCATC
<i>sox1</i> realtime-R	GAGGCCAGTCTGGTGTGTCAG
<i>Nestin</i> realtime-F	CTGCAGGCCACTGAAAAGTT
<i>Nestin</i> realtime-R	TTCAGCCAAGGAGACAGTCA
<i>Blbp</i> realtime-F	GGCAAGATGGTCGTGACTCT
<i>Blbp</i> realtime-R	ACCTCCACACCGAAGACAAA
<i>Tubb3</i> realtime-F	AACCAGATAGGGGCCAAGTT
<i>Tubb3</i> realtime-R	GGCCTGAATAGGTGTCCAAA

Supplementary Table 3. RNA-seq statistics

Sample	Total	Unique	Multiple
MUT.RNA.R1	25141990	23201868	1192166
MUT.RNA.R2	24755332	22870517	1142240
WT.RNA.R1	25185745	23211826	1197890
WT.RNA.R2	25253995	23210290	1197976



Supplementary Table 4 The quality control steps of Hi-C

Supplementary Table 4-1 Step 1 of Hi-C data quality control

Sample Name	Total Pairs	Reported Pairs	Reported Ratio
WT_Rep1	891882490	532314732	59.68%
WT_Rep2	1541879172	887438717	57.56%
WT_Rep3	1698451224	1034842720	60.93%
MUT_Rep1	960728847	560114338	58.30%
MUT_Rep2	1318284107	747740724	56.72%
MUT_Rep3	1500762042	888468710	59.20%

Reads were aligned to mm9 reference genome using bowtie2, reads with mapping quality > 10 were assigned to Mbol restriction fragments and interaction pairs were reconstructed as total pairs. Singleton or multi-hits pairs are filtered out, remaining pairs were denoted as reported pairs.

Supplementary Table 4-2. Step 2 of Hi-C data quality control

Sample Name	Valid Interaction Pairs	FF	RR	RF	FR	Ligation Efficiency
WT_Rep1	508314911	126956223	126873742	123851812	130633134	95.49%
WT_Rep2	846400045	211389337	211263652	206277665	217469391	95.38%
WT_Rep3	991209046	247498126	247423168	241458757	254828995	95.78%
MUT_Rep1	527405953	131857704	131755235	128608628	135184386	94.16%
MUT_Rep2	704704231	176138645	176029929	171798445	180737212	94.24%
MUT_Rep3	839842049	209886003	209778087	204703111	215474848	94.53%

Among reported pairs, failed ligation products (dangling end pairs, re-ligation pairs, self-cycle pairs) and pairs not able to reconstruct the ligation product were discarded, remaining pairs were denoted as valid interaction pairs. Valid interaction pairs were divided into 4 classes by the fragment directions (F for forward, R for reverse). As the ligation is a random process, about 25% of each valid ligation class is expected.

Supplementary Table 4-3. Step 3 of Hi-C data quality control

Sample Name	Valid RMDUP Pairs	PCR Duplication Ratio	Long Intra Interaction	Long Intra Ratio	Library Efficiency
WT_Rep1	490790030	3.45%	313371562	63.85%	55.03%
WT_Rep2	798693864	5.64%	509567229	63.80%	51.80%
WT_Rep3	932545214	5.92%	595005195	63.80%	54.91%

MUT_Rep1	501713985	4.87%	289967099	57.80%	52.22%
MUT_Rep2	668322376	5.16%	386579144	57.84%	50.70%
MUT_Rep3	792218069	5.67%	458197529	57.84%	52.79%

---

Among valid interaction pairs, PCR duplicates were discarded, remaining pairs were denoted as valid remove duplication interaction pairs and used for further analysis. The fractions of intra and inter-chromosomal interactions as well as long range (>20kb) versus short range (<20kb) intra-chromosomal interactions were calculated. Finally, the overall library efficiency was calculated as (valid remove duplication interaction pairs) / (total pairs).