A negatively-charged region within carboxy-terminal domain maintains proper CTCF DNA binding

Lian Liu, Yuanxiao Tang, Qiang Wu*

Center for Comparative Biomedicine, Ministry of Education Key Laboratory of Systems Biomedicine, State Key Laboratory of Medical Genomics, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

*Lead Contact and Correspondence: qiangwu@sjtu.edu.cn

SUMMARY

As an essential regulator of higher-order chromatin structures, CTCF is a highly-conserved DNA binding protein with a central DNA-binding domain of 11 tandem zinc fingers, which are flanked by N- and C- terminal domains of intrinsically disordered regions. The N-terminal domain interacts with cohesin complex and the central zinc finger domains recognize a large range of diverse genomic sites. However, the function of unstructured C-terminal domain remains unknown. Here we report that CRISPR deletion of the entire C-terminal domain of alternating charge blocks decreases CTCF DNA binding but deletion of C-terminal fragment of 116 amino acids results in increased CTCF DNA binding and aberrant gene regulation. Through a series of genetic targeting experiments, in conjunction with ChIP-seq and ATAC-seq, we uncovered a negatively-charged region responsible for weakening CTCF DNA binding and chromatin accessibility. Thus, the unstructured C-terminal domain plays an intricate role in maintaining proper CTCF-DNA interactions and 3D genome organization.

Key words

Alternating charge block, Chromatin accessibility, CTCF, Unstructured C-terminal domain, CRISPR screening, DNA binding, Gene regulation, Higher-order chromatin structure, Intrinsically disordered region, Protocadherin

INTRODUCTION

CCCTC-binding factor (CTCF) is a principal architectural protein for the construction of 3D genomes and is highly conserved in bilateria.[1-5] Together with the cohesin complex, CTCF mediates the formation of long-distance chromatin loops between distant sites, known as CBS (CTCF binding site), through an ATP-dependent active process known as "loop extrusion", leading to higher-order chromatin structures such as TADs.[6-13] Interestingly, CTCF/cohesin-mediated chromatin loops are preferentially formed between pairs of CBS elements in a forward-reverse convergent orientation.[14-16] In particular, topological chromatin loops are formed between tandem-arrayed CBS elements via cohesin-mediated dynamic loop extrusion, leading to balanced promoter choice.[17-19] The dynamic cohesin loop extrusion and its asymmetric blocking by oriented CTCF binding on numerous CBS elements distributed throughout mammalian genomes constitute a general principle in 3D genome organization and play an important role in gene regulation.

The clustered protocadherin (*cPcdh*) genes are an excellent model to investigate the relationships between CTCF/cohesin-mediated chromatin looping and gene expression programs. The 53 highly-similar human *cPCDH* genes are organized into three tandemly-linked clusters of *PCDH*α, *PCDH*β, and *PCDH*γ, spanning a large region of ~1M bps genomic DNA.[20] The *PCDH*α gene cluster comprises an upstream region of 15 variable exons and a downstream region with 3 constant exons. Similarly, *PCDH*γ comprises an upstream region of 22 variable exons and a downstream region of 3 constant exons. Each variable exon is separately spliced to the respective set of 3 constant exons within the *PCDH* α or γ gene cluster. By contrast, the *PCDH*β gene cluster comprises only 16 variable exons with no constant exons.

Similar to the intriguing *Dscam* gene for generating enormous diversity of cell-recognition codes in fly, the *cPCDH* genes generate an exquisite diversity for neuronal self-avoidance in vertebrates.[21,22] Different from competitive RNA pairing-mediated mutually exclusive splicing mechanism for *Dscam*, the *cPCDH* diversity is generated by a combination of *cis*-alternative splicing and promoter choice determined by CTCF-directed DNA looping.[23,24] In this complicated 3D genome configuration, CTCF directionally binds to tandem arrays of oriented CBS elements associated with *Pcdh* variable promoters and super-enhancers and CTCF-mediated chromatin loops are formed between pairs of convergent forward-reverse CBS elements. For example, in the *PCDH*α gene cluster, there are two forward CBS elements flanking each of the 13 alternate variable exons and two reverse CBS elements flanking the *HS5-1* enhancer.[25] A "double-clamping" chromatin interaction between these convergent pairs of CBS elements determines the *cPCDH* promoter choice.[24] In summary, CTCF/cohesin-mediated loop extrusion bridges remote super-enhancers in close contact with target variable promoters to form long-distance chromatin loops and this looping is essential for establishing proper expression patterns of the c*PCDH* genes in the brain.[16-18,24,26-28]

CTCF contains a central domain of 11 zinc fingers (ZFs) organized in a tandem array flanked by intrinsically disordered regions of the N-terminal domain (NTD) and C-terminal domain (CTD) (Figure 1A).[1,29] The central domain of CTCF binds to DNA directly through ZFs 3-7 and ZFs 9-11.[16,30-32] Recently, several lines of evidence suggest that ZF1 and ZF2 recognize base pairs downstream of the CTCF core motif.[33-37] Remarkably, CTCF also interacts with RNA to mediate chromatin loop formation and to regulate gene expression pattern.[38,39] Finally, the intrinsically disordered NTD, but not CTD, of CTCF interacts with cohesin complex to anchor chromatin loops between distant DNA elements.[4,12,40-43] Here by a combination of a series of genetic deletions, in conjunction with chromosome conformation capture and gene expression analyses, we found that a negatively-charged region (NCR) within the disordered CTD is important for proper CTCF DNA binding, higher-order chromatin organization, and gene regulation.

RESULTS

Deletion of CTCF CTD results in decreased DNA binding

We analyzed the amino acid (AA) composition of CTCF NTD and CTD and found that the NTD of CTCF contains all of the 20 types of amino acids while the CTD of CTCF contains only 15 AA types, suggesting that CTCF CTD has an amino acid compositional bias and is a low complexity region (Figure 1B and 1C). Computational analyses suggest that both NTD and CTD regions have a high intrinsically disordered score, especially the CTD region (Figure 1D).[44] In addition, the CTD region is highly conserved in vertebrates (Figure 1E). Owning to the lethality of CTCF deletion in mice or in cultured cells,[45,46] we tried to delete the CTCF NTD or CTD and found that deletion of NTD, but not CTD, is lethal in cultured cells. Specifically, we screened 254 single-cell clones for deletion of NTD, and could not find a single homozygous cell clone. Therefore, we focused our genetic dissection on CTCF CTD.

We screened for CTD-deletion clones by CRISPR DNA-fragment editing programed with dual sgRNAs and a donor construct containing FLAG sequences for tagging[47,48] and obtained two single-cell clones with precise deletion of the CTCF CTD (ΔCTD) (Figure S1A). We performed CTCF ChIP-seq with these two ΔCTD clones as well as with WT clones as a control (Figure S1B) and found that almost every CTCF peak within the three *PCDH* gene clusters is decreased upon CTD deletion, suggesting that CTD has an important role in CTCF binding to DNA (Figures 1F and S2A, S2B). Aggregated peak analysis showed that there is a significant decrease of CTCF binding at the *PCDH* CBS elements (Figure 1G). DNA-bound CTCF anchors cohesin complex via its NTD but not CTD.[4,12,40-43] To this end, we performed ChIP-seq experiments with a specific antibody against Rad21, a cohesin subunit, and found that cohesin is colocalized with CTD-deleted CTCF, suggesting that CTD-deleted CTCF is still able to anchor cohesin at the *cPCDH* locus (Figures 1F and S2A). However, there is a significant decrease of cohesin enrichments

upon deletion of CTCF CTD (Figures 1F, 1H and S2A, S2C). We then analyzed genome-wide CTCF and cohesin enrichments and found that both CTCF and cohesin enrichments are significantly decreased upon CTD deletion (Figure 1I and 1J). However, computational analyses revealed no alternation of CTCF motifs of all three types of CBS elements upon CTD deletion, suggesting that CTD does not alter the DNA binding specificity of the central ZF domain (Figure 1K-1M), despite the fact that CTCF enrichments are decreased for all three types of CBS elements (Figure S2D-S2F).

We next performed quantitative high-resolution circularized chromosome conformation capture experiments (QHR-4C, see methods)[18] with the *HS5-1* enhancer as an anchor and found that there is a significant decrease of long-distance chromatin interactions between the *HS5-1* enhancer and its target promoters (Figure 1N). Finally, we performed RNA-seq experiments and found that, consistent with decreased chromatin interactions between enhancers and promoters, there is a significant decrease of expression levels of members of the *PCDH*α gene cluster upon CTCF CTD deletion (Figure 1O).

Deletion of CTCF CTD affects gene regulation

We next analyzed the RNA-seq data using DESeq2 with adjusted *P* value < 0.05 and $\log_2$FC (fold change) >1 as cutoffs. We found 150 up-regulated genes (Table S1, ingenuity pathway analysis IPA, Figure S3A) with mean $\log_2$FC of 1.70 and 207 down-regulated genes (Table S2, Figure S3B) with mean $\log_2$FC of -1.84 (Figure S3C). We also found that the down-regulated genes are closer to CBS elements than the up-regulated genes (Figure S3D).

Deletion of C-terminal 116 AAs leads to increased CTCF binding

The CTD of CTCF contains an internal RNA-binding region (RBR_i) and a downstream region of 116 amino acids.[39,49] To investigate its function, we generated targeted deletion by screening single-cell CRISPR clones using CRISPR DNA-fragment editing with Cas9 programmed by dual sgRNAs.[16,47] We obtained two clones with deletion of the C-terminal 116 AAs (ΔCT116) (Figure S1C). We performed CTCF ChIP-seq experiments and found, remarkably, that there is a significant increase of CTCF enrichments in the *cPCDH* gene complex upon deletion of the C-terminal 116 AAs (Figures 2A, 2B and S4A, S4B). In addition, we performed Rad21 ChIP-seq experiments and found that there is a significant increase of cohesin enrichments at the *cPCDH* locus (Figures 2A, 2C and S4A, S4C), consistent with the model of CTCF asymmetrical blocking of cohesin “loop extrusion”.

We next performed genome-wide analyses and found similar enrichments of CTCF and cohesin upon deletion of the C-terminal 116 AAs (Figure 2D and 2E). Genome-wide analyses of CTCF motifs showed no alteration of all three types of the CBS elements (Figure S4D-S4I). We also performed QHR-4C experiments using the *HS5-1* enhancer as an anchor and found a significant increase of chromatin contacts with the target promoters of *PCDH*α*6* and

PCDHα12 (Figure 2F). Finally, RNA-seq experiments showed a significant increase of expression levels of *PCDHα6*, *PCDHα12*, and *PCDHαc2* (Figure 2G). These data demonstrated that the unstructured region of C-terminal 116 AAs inhibits CTCF binding to DNA.

Deletion of the CTCF C-terminal 116 AAs affects gene expression

RNA-seq experiments revealed 432 up-regulated genes (Table S3, Figure S3E) with mean $\log_2$FC of 2.36 and 461 down-regulated genes (Table S4, Figure S3F) with mean $\log_2$FC of -1.98 in ΔCT116 cell (Figure S3G). Interestingly, we found that the up-regulated genes are closer to increased CTCF peaks (Figure S3H).

Rescue with series of C-terminal truncated CTCFs reveals negatively-charged region (NCR)

We next generated a series of C-terminal truncated CTCF with V5 tags and transfected them into ΔCTD cells (Figure 3A). Specifically, truncated CTCF with 116 AAs deleted at the C-terminus was constructed as $CTCF_{1-611}$. In addition, $CTCF_{1-637}$ contains an additional region of 26 highly-conserved mostly negatively-charged amino acids (NCR). Finally, $CTCF_{1-663}$ contains a further downstream region of 26 amino acids with 9 proline residues and 7 positively-charged lysine or arginine residues. We generated stable cell lines by infecting with lentiviruses containing these constructs and verified their expression by Western blots (Figure 3A and 3B).

We performed ChIP-seq experiments with a specific antibody against V5 tag to investigate CTCF binding profiles and found that the binding strength of $CTCF_{1-611}$ at the *cPCDH* locus and throughout the entire genome is the highest among all of these four CTCF transgenes including full-length CTCF (Figures 3C-3E and S5A, S5B). Specifically, $CTCF_{1-611}$ has the highest affinity for all three types of genome-wide CBS elements (Figures 3F-3H and S5C). In conjunction with data of endogenous C-terminal truncation (Figure 2), we concluded that $CTCF_{1-611}$ has the highest DNA binding affinity and that the negatively charged region (NCR) of 26 AAs from 612 to 637 suppresses CTCF-DNA interactions.

NCR deletion increases CTCF binding and *cPCDH* expression

To investigate the endogenous function of NCR, we genetically deleted it by screening single-cell CRISPR clones and obtained two cell clones (ΔNCR) (Figure S1D). We performed CTCF ChIP-seq experiments with these clones and found that there is a significant increase of CTCF enrichments in the *cPCDH* locus compared with WT controls (Figures 4A, 4B and S6A, S6B). We also performed Rad21 ChIP-seq and found a similar increase of cohesin enrichments at the *cPCDH* locus (Figures 4A, 4C and S6A, S6C). In addition, genome-wide CTCF and cohesin enrichments are also significantly increased upon NCR deletion (Figure 4D and 4E). Furthermore, CTCF or cohesin enrichments are significantly increased at all three types of CBS elements

(Figure S6D-S6I). We next performed QHR-4C experiments using these single-cell clones with *HS5-1* as an anchor and found there is a significant increase of long-distance chromatin interactions with the target promoters of *PCDHα6* and *PCDHα12* (Figure 4F). Finally, we performed RNA-seq experiments and found that there is a significant increase of expression levels of *PCDHα6* and *PCDHα12* upon NCR deletion. These data suggest an important function of NCR in CTCF binding and gene regulation.

Deletion of NCR affects gene regulation

Genome-wide analyses of RNA-seq data identified 312 up-regulated genes (Table S5 and Figure S3I) with mean $\log_2$FC of 1.72 and 125 down-regulated genes (Table S6 and Figure S3J) with mean $\log_2$FC of -1.55 (Figure 4H). We also found that the up-regulated genes are closer to increased CTCF peaks (Figure 4I).

A role of NCR in rewiring chromatin accessibility

We next performed the assay for transposase-accessible chromatin with sequencing (ATAC-seq) and found that chromatin accessibilities are increased at most ATAC-seq peaks in the *cPCDH* locus upon CTCF NCR deletion (Figure S7). We also observed significantly increased ATAC-seq signals genome wide at increased CTCF peaks (Figure S8A). We next calculated the $\log_2$FC of ATAC-seq peaks and found that the majority of ATAC-seq peaks are also increased and minority peaks are decreased (Figure S8B). Specifically, we identified 7373 increased differential accessibility regions (DARs) and 1311 decreased DARs genome wide (Figure S8C). Interestingly, we found that most increased DARs are located at gene promoters and that many decreased DARs are located in the intergenic regions (Figure S8C).

We noted that ~3/4 DARs (6403) are overlapped with CBS elements and ~1/4 DARs (2281) are with no CBS (Figure S8D). Among all DARs overlapped with CBS elements, there are 5431 increased DARs and 972 decreased DARs (Figure S8E). We calculated the distance of non-CBS DARs to nearest CBS elements and found that increased DARs are located closer to CBS than decreased DARs (Figure S8F). Integrated analysis with RNA-seq showed that increased DARs correlate with enhanced levels of gene expression (Figure S8G). Together, these data suggest that CTCF NCR regulates chromatin accessibility and gene expression.

DISCUSSION

Combinatorial and patterned expression of diverse cadherin-like *cPcdh* genes in single cells in the brain enables a molecular logic of self-avoidance between neurites from the same neurons as well as a functional assembly of synaptic connectivity between neurons of the same developmental origin.[27,50-53] This complicated *cPcdh* expression program is achieved by ATP-dependent active cohesin “loop extrusion” which brings remote super-enhancer in close contacts

to target variable promoters via CTCF-mediated anchoring at tandemly-arrayed directional CBS elements.[16-18,24,25,28,54-56]. In particular, tandemly-arrayed CTCF sites function as topological chromatin insulators to balance distance- and context-dependent promoter choice to activate cell-specific gene expression in the brain.[17,18,28,54,55] Consequently, the dynamic interactions between CTCF and its recognition sites at variable promoters and super-enhancers are central for establishing *cPcdh* expression programs during brain development. In this work we systematically investigated the function of CTCF CTD and uncovered a negatively-charged region for maintaining proper CTCF binding to DNA in the large *cPCDH* gene complex and throughout the entire genome.

CTCF is a key 3D genome architectural protein that recognizes a large range of genomic sites via the central domain of 11 tandem zinc fingers and anchors loop-extruding cohesin via the YDF motif of the NTD.[4,12,40-43] In addition, other architectural protein such as ZNF143 resides at the hinge region between YDF motif and central ZFD and helps maintain proper CTCF/cohesin geometry.[56] The CTCF last ZF and CTD contain an internal RNA-binding region (RBR_i) of 38 amino acids and this region helps CTCF clustering and searching for authentic CBS elements.[39,49,57,58] Through a series genetic deletion and rescue experiments, we uncovered an important negatively-charged region (NCR) immediately downstream of the positively-charged region of RBR_i and provided strong genetic evidence that NCR plays an inhibitory role in CTCF clustering via RNA and dynamic recognition of cognate genomic sites (Figure 4J).[58,59]

NCR deletion leads to a significant increase of CTCF enrichments at all three types of CBS elements throughout the whole genome. In particular, NCR deletion results in increased CTCF binding at *cPCDH* variable promoters and super-enhancers, accompanied by increased chromatin accessibility and long-distance DNA looping. NCR either repulses DNA directly or weakens the strength of CTCF interaction with RNA, thus inhibits CTCF clustering and searching for cognate CBS elements (Figure 4J). Either way NCR appears to be important to maintain proper CTCF affinity for cognate genomic sites and specific chromatin looping at the *cPcdh* gene complex and likely throughout the entire genome. Thus, CTCF has an intricate self-adjusting mechanism to control the dynamic binding to genomic sites.

NCR contains 10 glutamate and 4 aspartate residues for a total of 14 negatively-charged amino acid residues (Figure 3A). In addition, there are four serine residues immediately upstream of NCR that may be phosphorylated by casein kinase II and thus switching to negative charges upon phosphorylation.[60] Numerous CTCF mutations are related to multiple cancers or a group of neurodevelopmental diseases known as CTCF-related disorders (CRD).[61,62] Interestingly, mutations within the CTCF NCR which alter its electronic charges are associated with several types of cancers. For example, CTCF mutations of E616K or E626K are associated with melanoma and lung cancers, respectively.[61,62] The exact pathogenetic mechanisms are not known but are

very likely related to disruptions of alternate positive-negative amino acid block patterns within CTD and selective partitioning into CTCF-specific trapping zones.[58,63]

Limitations of the study

While genetic deletion of NCR results in increased CTCF enrichments at the *cPCDH* locus and throughout the genome, how NCR biophysically suppress CTCF DNA binding is not known although it is likely involving multiple repulsive weak electric interactions between charged residues and DNA or RNA. In addition, our genetic experiments, in conjunction with chromosome conformation capture and RNA-seq, demonstrated that CTCF CTD, in particular NCR, play a crucial role in maintaining proper CTCF binding at the *cPCDH* gene complex, and subsequent *Pcdh* chromatin looping and gene regulation, whether this close correlation between chromatin looping and gene expression could be generalized to the entire genome remains to be tested. Finally, while our ATAC-seq experiments showed that enhanced CTCF DNA binding correlates mostly with higher chromatin accessibility, the exact mechanism is not known but most likely related to pioneering factors.

ACKNOWLEDGEMENTS

We are grateful for advice on bioinformatics from Dr. Jingwei Li and all members of our group for discussion. This study was supported by grants to Q.W. from the National Natural Science Foundation of China (32330016), the National Key R&D Program of China (2022YFC3400200), and the Science and Technology Commission of Shanghai Municipality (21DZ2210200).

AUTHOR CONTRIBUTIONS

Q.W. conceived the research. L.L. performed experiments. L.L. and Y.T. analyzed data. L.L. and Q.W. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V.V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell Biol. *13*, 7612-7624. https://doi.org/10.1128/mcb.13.12.7612-7624.1993.
2. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc. Natl. Acad. Sci. USA *109*, 17507-17512. https://doi.org/10.1073/pnas.1111941109.
3. Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19*, 789-800. https://doi.org/10.1038/s41576-018-0060-8.
4. Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., and Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. Nature *578*, 472-476. https://doi.org/10.1038/s41586-019-1910-z.

5. Wu, Q., Liu, P., and Wang, L. (2020). Many facades of CTCF unified by its coding for three-dimensional genome architecture. J. Genet. Genomics *47*, 407-424. https://doi.org/10.1016/j.jgg.2020.06.008.
6. Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422-433. https://doi.org/10.1016/j.cell.2008.01.011.
7. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801. https://doi.org/10.1038/nature06634.
8. Rubio, E.D., Reiss, D.J., Weicsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc. Natl. Acad. Sci. USA *105*, 8309-8314. https://doi.org/10.1073/pnas.0801273105.
9. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. USA *111*, 996-1001. https://doi.org/10.1073/pnas.1317788111.
10. Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456-E6465. https://doi.org/10.1073/pnas.1518552112.
11. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038-2049. https://doi.org/10.1016/j.celrep.2016.04.085.
12. van Ruiten, M.S., and Rowland, B.D. (2021). On the choreography of genome folding: A grand pas de deux of cohesin and CTCF. Curr. Opin. Cell Biol. *70*, 84-90. https://doi.org/10.1016/j.ceb.2020.12.001.
13. Popay, T.M., and Dixon, J.R. (2022). Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. J. Biol. Chem. *298*, 102117. https://doi.org/10.1016/j.jbc.2022.102117.
14. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680. https://doi.org/10.1016/j.cell.2014.11.021.
15. de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676-684. https://doi.org/10.1016/j.molcel.2015.09.023.
16. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910. https://doi.org/10.1016/j.cell.2015.07.038.
17. Canzio, D., Nwakeze, C.L., Horta, A., Rajkumar, S.M., Coffey, E.L., Duffy, E.E., Duffié, R., Monahan, K., O'Keeffe, S., Simon, M.D., et al. (2019). Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice. Cell *177*, 639-653. https://doi.org/10.1016/j.cell.2019.03.008.
18. Jia, Z., Li, J., Ge, X., Wu, Y., Guo, Y., and Wu, Q. (2020). Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. Genome Biol. *21*, 75. https://doi.org/10.1186/s13059-020-01984-7.
19. Liang, Z.Y., Zhao, L.J., Ye, A.Y., Lin, S.G., Zhang, Y.W., Guo, C.G., Dai, H.Q., Ba, Z.Q., and Alt, F.W. (2023). Contribution of the IGCR1 regulatory element and the 3'Igh CTCF-binding elements to regulation of Igh V(D)J recombination. Proc. Natl. Acad. Sci. USA *120*, e2306564120. https://doi.org/10.1073/pnas.2306564120.
20. Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell *97*, 779-790. https://doi.org/10.1016/S0092-

8674(00)80789-8.
21. Zipursky, S.L., and Sanes, J.R. (2010). Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly. Cell *143*, 343-353. https://doi.org/10.1016/j.cell.2010.10.009.
22. Dong, H.Y., Li, J.H., Wu, Q., and Jin, Y.F. (2023). Confluence and convergence of Dscam and Pcdh codes. Trends Biochem. Sci. *48*, 1044-1057. https://doi.org/10.1016/j.tibs.2023.09.001.
23. Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin α and -γ pre-mRNA splicing. Mol. Cell *10*, 21-33. https://doi.org/10.1016/S1097-2765(02)00578-6.
24. Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. Proc. Natl. Acad. Sci. USA *109*, 21081-21086. https://doi.org/10.1073/pnas.1219280110.
25. Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of Protocadherin-α gene expression. Proc. Natl. Acad. Sci. USA *109*, 9125-9130. https://doi.org/10.1073/pnas.1205074109.
26. Allahyar, A., Vermeulen, C., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M.J.A.M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J.H.I., et al. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. Nat. Genet. *50*, 1151-1160. https://doi.org/10.1038/s41588-018-0161-5.
27. Lv, X.H., Li, S., Li, J.W., Yu, X.Y., Ge, X., Li, B., Hu, S.H., Lin, Y., Zhang, S.B., Yang, J.J., et al. (2022). Patterned cPCDH expression regulates the fine organization of the neocortex. Nature *612*, 503-511. https://doi.org/10.1038/s41586-022-05495-2.
28. Kiefer, L., Chiosso, A., Langen, J., Buckley, A., Gaudin, S., Rajkumar, S.M., Servito, G.I.F., Cha, E.S., Vijay, A., Yeung, A., et al. (2023). WAPL functions as a rheostat of Protocadherin isoform diversity that controls neural wiring. Science *380*, eadf8440. https://doi.org/10.1126/science.adf8440.
29. Martinez, S.R., and Miranda, J.L. (2010). CTCF terminal segments are unstructured. Protein Sci. *19*, 1110-1116. https://doi.org/10.1002/pro.367.
30. Nakahashi, H., Kieffer Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678-1689. https://doi.org/10.1016/j.celrep.2013.04.024.
31. Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol. Cell *66*, 711-720.e713. https://doi.org/10.1016/j.molcel.2017.05.004.
32. Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., and Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res. *27*, 1365-1377. https://doi.org/10.1038/cr.2017.131.
33. Xu, D., Ma, R., Zhang, J., Liu, Z., Wu, B., Peng, J., Zhai, Y., Gong, Q., Shi, Y., Wu, J., et al. (2018). Dynamic nature of CTCF tandem 11 zinc fingers in multivalent recognition of DNA as revealed by NMR spectroscopy. J. Phys. Chem. Lett. *9*, 4020-4028. https://doi.org/10.1021/acs.jpclett.8b01440.
34. Soochit, W., Sleutels, F., Stik, G., Bartkuhn, M., Basu, S., Hernandez, S.C., Merzouk, S., Vidal, E., Boers, R., Boers, J., et al. (2021). CTCF chromatin residence time controls three-dimensional genome organization, gene expression and DNA methylation in pluripotent cells. Nat. Cell Biol. *23*, 881-893. https://doi.org/10.1038/s41556-021-00722-w.
35. Lebeau, B., Zhao, K., Jangal, M., Zhao, T., Guerra, M., Greenwood, C.M.T., and Witcher, M. (2022). Single base-pair resolution analysis of DNA binding motif with MoMotif reveals an oncogenic function of CTCF zinc-finger 1 mutation. Nucleic Acids Res. *50*, 8441-8458. https://doi.org/10.1093/nar/gkac658.
36. Hyle, J., Djekidel, M.N., Williams, J., Wright, S., Shao, Y., Xu, B., and Li, C. (2023). Auxin-inducible degron 2 system deciphers functions of CTCF domains in transcriptional regulation. Genome Biol. *24*, 14. https://doi.org/10.1186/s13059-022-02843-3.

37. Yang, J., Horton, J.R., Liu, B., Corces, V.G., Blumenthal, R.M., Zhang, X., and Cheng, X.D. (2023). Structures of CTCF-DNA complexes including all 11 zinc fingers. Nucleic Acids Res. *51*, 8447-8462. https://doi.org/10.1093/nar/gkad594.
38. Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., and Reinberg, D. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol. Cell *76*, 412-422.e415. https://doi.org/10.1016/j.molcel.2019.08.015.
39. Hansen, A.S., Hsieh, T.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. Mol. Cell *76*, 395-411.e313. https://doi.org/10.1016/j.molcel.2019.07.039.
40. Nishana, M., Ha, C., Rodriguez-Hernaez, J., Ranjbaran, A., Chio, E., Nora, E.P., Badri, S.B., Kloetgen, A., Bruneau, B.G., Tsirigos, A., and Skok, J.A. (2020). Defining the relative and combined contribution of CTCF and CTCFL to genomic regulation. Genome Biol. *21*, 108. https://doi.org/10.1186/s13059-020-02024-0.
41. Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A.L., Coulon, A., et al. (2020). Molecular basis of CTCF binding polarity in genome folding. Nat. Commun. *11*, 5612. https://doi.org/10.1038/s41467-020-19283-x.
42. Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., and Lobanenkov, V.V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. USA *117*, 2020-2031. https://doi.org/10.1073/pnas.1911708117.
43. Hansen, A.S. (2020). CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. Nucleus *11*, 132-148. https://doi.org/10.1080/19491034.2020.1782024.
44. Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., and Kurgan, L. (2021). flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nat. Commun. *12*, 4438. https://doi.org/10.1038/s41467-021-24773-7.
45. Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V., and Filippova, G.N. (2012). Loss of Maternal CTCF Is Associated with Peri-Implantation Lethality of Null Embryos. Plos One *7*, e34915. https://doi.org/10.1371/journal.pone.0034915.
46. González-Buendía, E., Pérez-Molina, R., Ayala-Ortega, E., Guerrero, G., and Recillas-Targa, F. (2014). Experimental Strategies to Manipulate the Cellular Levels of the Multifunctional Factor CTCF. In Cancer Cell Signaling: Methods and Protocols, M. Robles-Flores, ed. (Springer New York), pp. 53-69. https://doi.org/10.1007/978-1-4939-0856-1_5.
47. Li, J.H., Shou, J., Guo, Y., Tang, Y.X., Wu, Y.H., Jia, Z.L., Zhai, Y.N., Chen, Z.F., Xu, Q., and Wu, Q. (2015). Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. J. Mol. Cell Biol. *7*, 284-298. https://doi.org/10.1093/jmcb/mjv016.
48. Shou, J., Li, J.H., Liu, Y.B., and Wu, Q. (2018). Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. Molecular Cell *71*, 498-509.e4. https://doi.org/10.1016/j.molcel.2018.06.021.
49. Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723-734. https://doi.org/10.1101/gad.236869.113.
50. Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature *488*, 517-521.
51. Thu, C.A., Chen, W.V., Rubinstein, R., Chevee, M., Wolcott, H.N., Felsovalyi, K.O., Tapia, J.C., Shapiro, L., Honig, B., and Maniatis, T. (2014). Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. Cell *158*, 1045-1059. https://doi.org/10.1016/j.cell.2014.07.012.
52. Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, S.,

Ahlsen, G., Chevee, M., Halim, A., Clausen, H., et al. (2015). Molecular logic of neuronal self-recognition through protocadherin domain interactions. Cell *163*, 629-642. https://doi.org/10.1016/j.cell.2015.09.026.
53. Mountoufaris, G., Chen, W.V., Hirabayashi, Y., O'Keeffe, S., Chevee, M., Nwakeze, C.L., Polleux, F., and Maniatis, T. (2017). Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. Science *356*, 411-414. https://doi.org/10.1126/science.aai8801.
54. Zhou, Y.X., Xu, S.Y., Zhang, M., and Wu, Q. (2021). Systematic functional characterization of antisense eRNA of protocadherin composite enhancer. Genes Dev. *35*, 1383-1394. https://doi.org/10.1101/gad.348621.121.
55. Ge, X., Huang, H., Han, K., Xu, W., Wang, Z., and Wu, Q. (2023). Outward-oriented sites within clustered CTCF boundaries are key for intra-TAD chromatin interactions and gene regulation. Nat. Commun. *14*, 8101. https://doi.org/10.1038/s41467-023-43849-0.
56. Zhang, M., Huang, H., Li, J., and Wu, Q. (2024). ZNF143 deletion alters enhancer/promoter looping and CTCF/cohesin geometry. Cell Rep. *43*. 11363. https://doi.org/10.1016/j.celrep.2023.113663.
57. Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell *57*, 361-375. https://doi.org/10.1016/j.molcel.2014.12.006.
58. Hansen, A.S., Amitai, A., Cattoglio, C., Tjian, R., and Darzacq, X. (2020). Guided nuclear exploration increases CTCF target search efficiency. Nat. Chem. Biol. *16*, 257-266. https://doi.org/10.1038/s41589-019-0422-3.
59. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6:25776*. https://doi.org/10.7554/eLife.25776.
60. Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., León, J., et al. (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol. Cell. Biol. *21*, 2221-2234. https://doi.org/10.1128/mcb.21.6.2221-2234.2001.
61. de Morales, H.G.V., Wang, H.L.V., Garber, K., Cheng, X.D., Corces, V.G., and Li, H. (2023). Expansion of the genotypic and phenotypic spectrum of CTCF-related disorder guides clinical management: 43 new subjects and a comprehensive literature review. Am. J. Med. Genet. A *191*, 718-729. https://doi.org/10.1002/ajmg.a.63065.
62. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385.e18. https://doi.org/10.1016/j.cell.2018.02.060.
63. Lyons, H., Veettil, R.T., Pradhan, P., Fornero, C., De La Cruz, N., Ito, K., Eppert, M., Roeder, R.G., and Sabari, B.R. (2023). Functional partitioning of transcriptional regulators by patterned charge blocks. Cell *186*, 327-345.e28. https://doi.org/10.1016/j.cell.2022.12.013.
64. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359. https://doi.org/10.1038/Nmeth.1923.
65. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
66. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*, W39-W49. https://doi.org/10.1093/nar/gkv416.
67. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. https://doi.org/10.1093/bioinformatics/bts635.
68. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511-515. https://doi.org/10.1038/nbt.1621.

69. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160-W165. https://doi.org/10.1093/nar/gkw257.
70. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Proc, G.P.D. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079. https://doi.org/10.1093/bioinformatics/btp352.
71. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. https://doi.org/10.1093/bioinformatics/btq033.
72. Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. *41*, e132. https://doi.org/10.1093/nar/gkt373.
73. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
74. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389-393. https://doi.org/10.1038/nature10730.
75. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018. https://doi.org/10.1093/bioinformatics/btr064.
76. Yu, G.C., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383. https://doi.org/10.1093/bioinformatics/btv145.

STAR METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CTCF	Millipore	Cat# 07-729; RRID: AB_441965
Anti-Rad21	Abcam	Cat# AB992; RRID: AB_2176601
Anti-V5	Abcam	Cat# AB15828; RRID: AB_443253
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody	ThermoFisher Scientific	Cat# A32735; RRID: AB_2633284
Anti-GAPDH	Abmart	Cat# P30008; RRID: AB_2936506
Bacterial and Virus Strains		
Stbl3 competent cells	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
EcoRI	NEB	Cat# R0101S
BsaI	NEB	Cat# R0535S
DpnII	NEB	Cat# R0176L
BamHI	NEB	Cat# R0136S
NEB buffer 2	NEB	Cat# B7002S
T4 DNA ligase	NEB	Cat# M0202L
Formaldehyde	ThermoFisher Scientific	Cat# 28908
RNase A	ThermoFisher Scientific	Cat# EN0531
Proteinase K	NEB	Cat# P8107S
Glycogen	ThermoFisher Scientific	Cat# R0561
TRIzol Reagent	ThermoFisher Scientific	Cat# 15596026
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich	Cat# M8823
Streptavidin Dynabeads	ThermoFisher Scientific	Cat# 65001
AMPure XP beads	Beckman Coulter	Cat# A63881
Lipofectamine 3000	Invitrogen	Cat# L30001
Polybrene	Sigma-Aldrich	Cat# TR-1003-G
Puromycin dihydrochloride	Sigma-Aldrich	Cat# P8833
Critical Commercial Assays		
MinElute Gel Extraction Kit	QIAGEN	Cat# 28606
VAHTSTM Universal DNA Library	Vazyme	Cat# ND606

Prep Kit for Illumina V2		
VAHTSTM Multiplex Oligos set 4 for Illumina	Vazyme	Cat# N321
QIAquick PCR Purification Kit	QIAGEN	Cat# 28106
VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina	Vazyme	Cat# NR604
VAHTS mRNA Capture Beads	Vazyme	Cat# N401
VAHTS® RNA Multiplex Oligos Set 1 for Illumina	Vazyme	Cat# N323
Vazyme Hyperactive ATAC-Seq Library Prep Kit for Illumina	Vazyme	Cat# TD711
TruePrep Index Kit V2 for Illumina	Vazyme	Cat# TD202
pClone007 Versatile Simple Vector Kit	TsingKe	Cat# TSV-007VSm
BCA protein assay kit	Beyotime	Cat# P0009
Deposited Data		
High-throughput sequencing files (QHR-4C)	This study	
High-throughput sequencing files (ChIP-seq)	This study	
High-throughput sequencing files (RNA-seq)	This study	
High-throughput sequencing files (ATAC-seq)	This study	
Raw imaging files	This study, Mendeley data	
Experimental Models: Cell Lines		
Human: HEC-1-B	ATCC	Cat# HTB-113, RRID: CVCL_0294
Human: HEK293T	ATCC	Cat# CRL-3216, RRID: CVCL_0063
Human: HEC-1-B ΔCTD clone1	This paper	N/A
Human: HEC-1-B ΔCTD clone2	This paper	N/A
Human: HEC-1-B ΔCT116 clone1	This paper	N/A
Human: HEC-1-B ΔCT116 clone2	This paper	N/A

Human: HEC-1-B ΔNCR clone1	This paper	N/A
Human: HEC-1-B ΔNCR clone2	This paper	N/A
Human: HEC-1-B WT-FLAG clone1	This paper	N/A
Human: HEC-1-B WT-FLAG clone2	This paper	N/A
Human: HEC-1-B $CTCF_{1-611}$	This paper	N/A
Human: HEC-1-B $CTCF_{1-637}$	This paper	N/A
Human: HEC-1-B $CTCF_{1-663}$	This paper	N/A
Human: HEC-1-B CTCF-FL	This paper	N/A
Oligonucleotides		
See Supplementary Table S7	This study	NA
Recombinant DNA		
Plasmid: pMD2.G	Addgene	Cat# 12259, RRID: Addgene_12259
psPAX2	Addgene	Cat# 12260, RRID: Addgene_12260
Plasmid: pGL3-U6-sgRNA-PGK-Puro	Li et al.[47]	https://academic.oup.com/jmcb/article/7/4/284/901042
pcDNA3.1-Cas9-WT	J. Xi (Peking University)	N/A
Plasmid: pClone007-ΔCTD homology arm	This paper	N/A
Plasmid: pClone007-ΔCT116 homology arm	This paper	N/A
Plasmid: pClone007-ΔNCR homology arm	This paper	N/A
Plasmid: pClone007-WT-FLAG homology arm	This paper	N/A
Plasmid: pGL3- ΔCTD-sg1	This paper	N/A
Plasmid: pGL3- ΔCTD-sg2	This paper	N/A
Plasmid: pGL3- ΔCT116-sg1	This paper	N/A
Plasmid: pGL3- ΔNCR-sg1	This paper	N/A
Plasmid: pGL3- ΔNCR-sg2	This paper	N/A
Plasmid: pLVX-$CTCF_{1-611}$	This paper	N/A
Plasmid: pLVX-$CTCF_{1-637}$	This paper	N/A
Plasmid: pLVX-$CTCF_{1-663}$	This paper	N/A
Plasmid: pLVX-CTCF-FL	This paper	N/A
Software and Algorithms		
Bowtie2 software (v2.3.4.2)	Langmead et al.[64]	http://bowtie-bio.sourceforge.net/bowtie2/

		index.shtml
MACS2 (v2.1.2)	Zhang et al.[65]	https://github.com/taoliu/MACS
MEME v4.12.0	Bailey et al.[66]	https://meme-suite.org/meme/
STAR (v2.7.3a)	Dobin et al.[67]	https://github.com/alexdobin/STAR
Cufflinks (v2.2.1)	Trapnell et al.[68]	http://cole-trapnell-lab.github.io/cufflinks/
Deeptools (v3.5.3)	Ramírez et al.[69]	https://deeptools.readthedocs.io/en/latest/
Samtools (v1.12)	Li et al.[70]	http://www.htslib.org/doc/
Bedtools (v2.30.0)	Quinlan et al.[71]	https://bedtools.readthedocs.io/en/latest/index.html
R3Cseq (v1.38.0)	Thongjuea et al.[72]	https://bioconductor.org/packages/release/bioc/html/r3Cseq.html
DESeq2 (v1.32.0)	Love et al.[73]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DiffBind (v3.4)	Ross-Innes et al.[74]	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
Fimo (v5.5.5)	Grant et al.[75]	http://meme-suite.org/tools/fimo
ChIPseeker (v1.30.3)	Yu et al.[76]	https://guangchuangyu.github.io/software/ChIPseeker
PicardToolshttp://broadinstitute.github.io/picard/	N/A	http://broadinstitute.github.io/picard/
UCSC Genome Browser	N/A	https://genome.ucsc.edu/
ggplot2 (v3.4.2)	Open source	https://ggplot2.tidyverse.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qiang Wu (qiangwu@sjtu.edu.cn).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- High-throughput sequencing files (ChIP-seq, RNA-seq, ATAC-seq, and QHR-4C) have been deposited into the NCBI Gene Expression Omnibus (GEO) database.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cells and culture conditions

Human HEC-1-B cells were cultured as previously described in MEM medium (Hyclone) supplemented with 10% (v/v) FBS (Gibco), 2 mM GlutaMAX (Gibco), 1 mM sodium pyruvate (Sigma), and 1% penicillin-streptomycin (Gibco).[16] Briefly, HEC-1-B cells were maintained at 37 °C in a humidified incubator containing 5% CO_2. Medium of cultured cells was changed every 24 h. Cells were passed every 72 h. When cells were passaged, the medium was removed and cells were washed by PBS and then digested by trypsin (Gibco) for 5 min at 37 °C in a humidified incubator containing 5% CO_2 and quenched by 10% FBS supplemented PBS. Digested cells were collected by centrifuging at 500 rpm for 5 min at room temperature. Pelleted cells were resuspended with media and seeded to new plates.

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (Hyclone) supplemented with 10% (v/v) FBS (Gibco) and 1% penicillin-streptomycin (Gibco). Cells were maintained at 37 °C in a humidified incubator containing 5% CO_2. Medium of cultured was changed every 24 h. HEK293T cells were passed every 48 h. The passage process of HEK293T is similar to HEC-1-B except with shortened trypsin digesting time of 2 min.

METHOD DETAILS

Plasmid construction

For all CRISPR/Cas9 experiments, the sgRNA plasmids were constructed as described before.[47] Briefly, pGL3-U6 vector was linearized by BsaI (NEB) to generate the cloning backbone with 5' overhangs of 'TGGC' and 'TTTG' at the two ends. In addition, a pair of complementary oligonucleotides (Table S7)

containing the sgRNA targeting sequences with 5' overhangs of 'ACCG' or 'AAAC' was annealed and ligated to BsaI-digested linearized vector backbone using T4 DNA ligase (NEB). The complete sgRNA sequences are under the control of the U6 promoter and will be transcribed by Pol III in mammalian cells. Finally, the Cas9 plasmid was obtained as a gift from Peking University.

For donor plasmids used in establishing ΔCTD and ΔCT116 stable cell lines via homologous recombination (HR), the donor was designed such that ΔCTD or ΔCT116 were tagged with FLAG sequences (Table S7) for tracing CTCF proteins with FLAG-specific antibodies because deletions of endogenous CTCF C-terminal fragment removed the epitope (659-675 AAs) for CTCF antibodies (Millipore). In addition, we FLAG-tagged WT CTCF via HR at its C-terminus at the endogenous locus as a control. To generate donor plasmid for CRISPR screening single-cell clones through the homologous recombination (HR) pathway, two homologous arms each of ~1kb flanking the target sites were amplified from genomic DNA by PCR with primers. The amplified homologous arms were PCR-purified and a donor DNA fragment for HR was generated through overlapping-PCR. Finally, the donor DNA fragment was ligated to a T-vector through TA cloning. For ΔNCR cells, the epitope of CTCF antibodies was intact and the donor plasmid was constructed without FLAG.

All rescue CTCF constructs were V5-tagged to distinguish with the endogenous FLAG-tagged CTCF in ΔCTD cells. To construct Lentivirus pLVX vectors containing a series of truncated CTCF C-terminal domains, pLVX vector was first linearized by EcoR1 and BamH1 (NEB) and purified as the cloning backbone. $CTCF_{1-611}$, $CTCF_{1-637}$, $CTCF_{1-663}$, and full length CTCF were cloned from a cDNA library of HEC-1-B cells with the same 5' primers containing EcoRI restriction endonuclease sites and different 3' primers containing BamHI restriction endonuclease sites in conjunction with the V5 tag sequence (Table S7). Truncated CTCF fragments amplified from cDNA were digested with restriction endonucleases and ligated into pLVX linearized vector using T4 DNA ligase.

ChIP-seq

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) experiments were performed as described before.[24] Briefly, 5×10^6 cells were collected and crosslinked with 1% formaldehyde, quenched by 2 M glycine at a final concentration of 125 mM and washed by ice-cold PBS twice. Crosslinked cells were lysed twice by the ChIP buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 0.15 M NaCl, and 1×protease inhibitors, Roche) with slow rotation at 4 °C for 10 min. The lysed cells were centrifuged at 2,500 g at 4 °C for 10 min to isolate cell nuclei. The isolated nuclei were resuspended with the ChIP buffer and sonicated using a Bioruptor Sonicator (with the high energy setting at a train of 30-s sonication with 30-s interval for 30 cycles) to fragment DNA to 200-400 bp. The sonicated mixture was centrifuged at 14,000 g at 4 °C for 30 min and the

supernatant was then precleared by protein A/G magnetic beads (Thermo 26162) for 3h at 4 °C with slow rotation. Antibody (CTCF: Millipore 07-729, Rad21: Abcam ab992, V5: Abcam ab15828) or Anti-FLAG antibody conjugated Magnetic Beads (Sigma M8823) were added to the precleared solution and incubated at 4 °C overnight with slow rotation to precipitate CTCF or cohesin protein-DNA complex.

Protein A/G beads were added and incubated for 3 h at 4 °C with slow rotation to capture the antibody-protein-DNA complex. The ChIP buffer, high salt buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 0.4 M NaCl), no salt buffer (high salt buffer without NaCl), LiCl buffer (50 mM HEPES pH 7.5, 1 mM EDTA, 1% NP-40, 0.7% sodium deoxycholate, 0.5 M LiCl), and 10 mM Tris-HCl buffer (pH 7.5) were used sequentially to wash the beads. Finally, the elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) was used to elute ChIP DNA from beads at 65 °C for 1h with 1,000 rpm shaking. The eluted protein-DNA complex was reverse-crosslinked at 65 °C overnight with 1,000 rpm shaking to dissociate DNA. Finally, the proteinase K (NEB) was added and incubated for 2 h at 55 °C to digest protein and the RNase A (Thermo) was added and incubated for 2h at 37 °C to digest RNA.

The DNA was purified by adding equal volume of phenol-chloroform and mixed by vigorously shaking. The mixture was centrifuged at 4 °C at 14,000 g for 10 min to separate the proteins and DNA. The supernatant containing DNA was transferred to a new tube. 2.5-fold volume of ice-cold ethanol, 1/10 volume of 3 M NaAc (pH 5.2), and 1.5 μl of glycogen (Thermo) were added to participate DNA at -80 °C for 1 h. The sample was centrifuged at 14,000 g at 4 °C for 30 min to pellet DNA. 70% ethanol was added to wash DNA pellets and finally the sample was centrifuged at 14,000g at 4 °C for 10 min. The supernatant was removed and the pelleted DNA was air-dried for 5 min to remove residual ethanol. The DNA was then dissolved in the nuclease-free water. The concentration of DNA was measured using Qubit (Invitrogen).

To prepare DNA library for deep sequencing, we used the Universal DNA Library Prep Kit (Vazyme ND606). Briefly, DNA was first end-repaired and ligated to adapters from the Multiplex Oligos Set (Vazyme, N321). Adapter-ligated DNA was then purified using AMPure XP beads (Beckman) and the final ChIP library was amplified by PCR. The library was sequenced on an Illumina NovaSeq platform.

Western blot

1×10^6 cells were collected and lysed by the RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, and 1× protease inhibitors) on ice for 30min. The sample was sonicated and centrifuged at 14,000 g for 15 min at 4 °C to remove cell debris. The concentration of supernatant proteins mixture was measured using the BCA protein assay kit (Beyotime). The supernatant proteins were mixed with 5× SDS protein loading buffer with DTT and incubated at 95 °C for 10 min to denature

the proteins. The denatured protein mix was centrifuged at 14,000g for 15 min at room temperature and separated on SDS-PAGE.

The separated proteins in gel were transferred to the nitrocellulose membrane through electrophoreses. After transferring, the membrane was blocked by 5% non-fat milk in PBST for 2 h at room temperature and then washed 3 times with PBST. Corresponding antibodies were incubated with the membrane overnight at 4°C with slow shaking. The membrane was washed 3 times with PBST. The secondary antibody (Invitrogen) was then incubated with the membrane for 90 min at room temperature with slow shaking. Finally, the membrane was washed by PBST for 3 times and scanned by the Odyssey System (LI-COR Biosciences).

CRISPR genome editing and single-cell cloning

CRISPR genome editing was performed as described.[47,48] Cas9 and sgRNA plasmids, and donor constructs were transfected in 12-well plates at 70 % confluency using lipofectamine 3000 (Invitrogen). Medium containing lipofectamine 3000 and plasmid DNA was then replaced with fresh medium after 6 h. After transfection for 48 h, the transfected cells were incubated with medium containing 2 μg/ml puromycin and refreshed daily. After 4 days, the cells were recovered for 2 days in fresh medium without puromycin. The cells were then dissociated by trypsin and seeded to 96-wells cell plates at the concentration of about one cell per well. After incubation for about 1 week in 96-well plates, the seeded cells were checked under microscope and single-cell clones were marked manually.

PCR genotyping was performed when seeded single-cell clones reached ~80% confluence in 96-well plates. The cells were digested using trypsin and PCR-genotyped with a pair of primers matching sequences outside of the homologous arms (Table S7). The positive PCR products were purified and sequenced by Sanger sequencing for confirmation. The identified single-cell clones were incubated for another 4~5 passages and confirmed again by PCR genotyping and Sanger sequencing.

Given that the CTCF antibody used in ChIP-seq recognizes 659-675 AAs of CTCF, we added a FLAG tag in ΔCTD and ΔCT116 cell lines via homologous recombination. We also inserted a FLAG tag in WT cell as an experimental control. For ΔCTD cells, we genotyped 191 single-cell clones and found 3 single-cell clones with CTD deletion. For WT-FLAG cells, we genotyped 160 single-cell clones and found 4 of them with FLAG tag insertion. For ΔCT116 cells, we genotyped 192 single-cell clones and found 8 of them with CT116 deletion. For ΔNCR cells, we genotyped 240 single-cell clones and found 6 of them with NCR deletion. For the cells with same genotype, we used two independent single-cell clones for subsequent experiments.

Lentivirus packaging

Lentiviruses were used to build stable transgenic cell lines that expressing different CTCF rescue constructs in ΔCTD cells. pLVX plasmids containing truncated CTCF sequences and puromycin selectable marker, and helper

plasmids (psPAX2, pMD2.G, Addgene) were transfected into HEK293T cells in 6-well plate at 70% confluence using lipofectamine 3000 reagents. The medium containing lipofectamine and plasmid DNA was replaced with fresh medium after 6 h. The medium containing lentivirus was harvested after transfection for 48 h for the first time, and then replaced with fresh medium and harvested again after another 24 h. The harvested medium was then centrifuged at 1,300 rpm at 4 °C for 5 min to remove cells and then filtered by 0.45 μm filter (Merck) and the virus was collected and stored at -80 °C.

Lentivirus infection and stable cell line establishment

All lentivirus particles expressing different truncated CTCF constructs were thawed on ice and infected with ΔCTD cells by adding polybrene at 8 ug/ml. After infection for two days, the infected cells were incubated with medium containing 2 μg/ml puromycin for 4 days and replaced daily. After selection, cells were incubated with fresh medium without puromycin for 2 days for recovering. The cells were then cultured for another 12 days and harvested for Western blot and ChIP-seq experiments.

RNA-seq

The RNA-seq experiment was performed as previously described.[16] Briefly, about 1×10^6 cells in 6-well plates were collected and washed by ice-cold PBS for 3 times. One ml of TRIzol (Invitrogen) was added to the cells and mixed thoroughly. After incubation for 5 min at room temperature, 0.2 ml of chloroform was added and mixed well by hands. The sample was then incubated for 2 min at room temperature and centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatant was transferred to a new tube and mixed with equal volume of isopropanol. The sample was incubated at room temperature for 10 min and centrifuged at 12,000 rpm for 15 min at 4 °C to pellet RNA. After removing the supernatant, 1 ml of 75% ethanol was added to wash RNA. Finally, RNA pellets were dissolved in nuclease-free water. The quantity and concentration of the total RNA was measured by NanoDrop (Thermo).

The Oligo (dT) coupled magnetic beads (Vazyme) were used to purify mRNA from total RNA. The mRNA library preparation was performed using the Universal V6 RNA-seq Library Prep Kit (Vazyme) according to the manual. Briefly, mRNA was fragmented by heating at 94 °C for 8 min. Next, reverse transcription for the first-strand cDNA was performed and then the double-strand cDNA was synthesized. Adapters from the RNA Multiplex Oligos Set (Vazyme) were added to cDNA and the adapter-ligated cDNA was purified by AMPure XP beads (Beckman). Finally, Library was amplified through PCR and was sequenced on an Illumina NovaSeq platform.

ATAC-seq

ATAC-seq was performed as previous described.[55] Briefly, cells in 12-well plates at 80% confluence were digested by trypsin and were collected by centrifuging at 500 rpm at room temperature for 5 min. 50,000 cells were collected and washed twice by PBS. The cells were then washed twice by ice-cold buffer by centrifuging at 2,300 rpm at 4 °C for 5 min. The cells were then

lysed by fresh prepared lysis buffer containing NP-40, Tween-20, and digitonin on ice for 5 min and centrifuged for 10 min at 2,300 rpm at 4 °C to collect the cell nuclei.

The chromatins were fragmented with Tn5 transposase and sequence adaptor at 37 °C for 30 min. The fragmentation reaction was terminated by adding the stop buffer at room temperature for 5 min. The fragmented DNA was purified by ATAC DNA Extract Beads (Vazyme). Finally, PCR was performed to amplify the library using primers in TruePrep Index Kit V2 (Vazyme) and the library was sequenced on an Illumina NovaSeq platform.

QHR-4C

Quantitative high-resolution circularized chromosome conformation capture (QHR-4C) experiments were performed as previously described.[18] Briefly, 1×10^6 cell were collected and crosslinked by 2% formaldehyde at room temperature for 10 min with slow rotation. The crosslinking reaction was quenched by adding 2 M glycine to the final concentration of 200 mM. The crosslinked cells were washed twice by ice-cold PBS by centrifuging for 5 min at 800 g at 4 °C and then lysed twice by ice-cold lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40,1% Triton X-100, and 1× protease inhibitors) for 10 min at 4 °C with slow rotation. The cell nuclei were collected by centrifuge at 800 g at 4 °C for 5 min. The pelleted nuclei were resuspended in 73 μl of nuclease-free water, 10μl of DpnII buffer (NEB), and 2.5 μl of 10% SDS and incubated for 1 h at 37 °C with shaking at 900 rpm. 12.5 μl of 20% Triton X-100 was added for 1 h at 37 °C with shaking at 900 rpm. 2 μl of DpnII (NEB) was added to digest the chromatin overnight at 37 °C with shaking at 900 rpm and deactivated at 65 °C for 20 min.

The nuclei were then collected through centrifuging at 1,000 g for 1 min. The supernatant was carefully removed. The proximity ligation was performed with resuspended nuclei by adding 100μl of T4 ligation buffer (NEB) containing 1 μl of T4 ligase and incubated at 16 °C for 24 h. 1 μl of proteinase K was then added to the ligation mixture to digest the proteins and incubated for 4 h at 65 °C for reverse crosslinking. The DNA was purified by phenol-chloroform as described above in ChIP-seq. Finally, the purified DNA was dissolved in 50 μl nuclease-free water and sonicated with the Bioruptor Sonicator (with the low energy setting by a train of 30-s sonication with 30-s interval for 12 cycles) to fragment DNA to sizes of 200-600 bp.

The PCR amplification was performed using 5' biotin-labeled primers (Table S7) to capture DNA anchored at the *HS5-1* enhancer. To maximize the PCR product, 100 μl of reaction system and 60 cycles of PCR were used. The PCR product was incubated at 95 °C for 5 min and immediately chilled on ice to obtain single-strand DNA (ssDNA). The biotin-labeled ssDNA was collected by incubating with Streptavidin Beads (Invitrogen) for 2 h at room temperature and the beads were washed twice by the washing buffer (5 mM Tris-HCl pH7.5, 1 M NaCl, 0.5 mM EDTA).

To prepare library for deep sequencing, adapters containing sequences

that match the 3' end of the Illumina P7 sequence were ligated to ssDNA at 16 °C for 24 h. The adapters were generated through annealing of two complementary primers (Table S7) in the annealing buffer (25 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA). The beads with ssDNA-adapter were washed twice by the washing buffer. Finally, the library was amplified through PCR with two primers. The forward primer contains the Illumina P5 sequences and the sequences adjacent to the *HS5-1* anchor. The reverse primer contains the Illumina P7 sequences and indexes. The amplified library was sequenced on an Illumina NovaSeq platform.

Data analysis of ChIP-seq

Raw FASTQ files were aligned to the human reference genome (GRCh37/hg19) using Bowtie2.[64] The MarkDuplicates module of PICARD tools (http://broadinstitute.github.io/picard/) was used to remove the duplicates and Samtools[70] was used to index or sort bam files. ChIP-seq peaks were called by MACS2[65] with the default parameter. The reads counts were normalized to reads per kilobase per million mapped (RPKM) using bamCoverage module of Deeptools[69] with a bin sized of 20 bp. The plotHeatmap module of Deeptools was used to generate heatmaps. Normalized read counts were converted to bedGraph to be visualized in the UCSC genome browser. Differential binding analyses were performed using DiffBind with the default parameter.[74] Motif analyses of CTCF were performed using the MEME suite[66] and FIMO.[75]

Data analysis of RNA-seq

Raw FASTQ files were aligned to the human reference genome (GRCh37/hg19) using STAR[67] with default parameters and the RPKMs were calculated using Cufflink.[68] Differential analysis of gene expression was performed using DEseq2.[73] Volcano plot was generated by ggplot2. TSS distance to nearest CTCF peaks were calculated by Bedtools.[71]

Data analysis of ATAC-seq

Raw FASTQ files were aligned to the human reference genome (GRCh37/hg19) using Bowtie2.[64] ATAC-seq peaks were called by MACS2.[65] Differential accessibility regions (DARs) were analyzed using DESeq2.[73] ATAC-seq peak annotation was performed with ChIPseeker.[76] Bedtools was used in overlapping analyses of DAR and CBS elements and in calculating the distance of DARs to nearest CBS elements.[71]

Data analysis of QHR-4C

Raw FASTQ files were aligned to the human reference genome (GRCh37/hg19) using Bowtie2.[64] Reads were normalized by r3Cseq program (version 1.20).[72]

QUANTIFUCANTION AND STATISTICAL ANALYSIS

All statistical tests were performed with the *R* scripts. Statistical significance values were calculated using the Student's *t* test. $p < 0.05$ was shown as *, $p < 0.01$ was shown as **, $p < 0.001$ was shown as *** and $p < 0.0001$ was shown as ****.

Figure Legends

Figure 1. Deletion of CTCF C-terminal domain results in decreased CTCF binding and gene dysregulation.

(A) Schematic of the CTCF zinc finger domain (ZFD) and the flanking N-terminal domain (NTD) and C-terminal domain (CTD).

(B and C) Amino acid composition of CTCF NTD (B) and CTD (C).

(D) Disordered propensity by the flDPnn program indicates that NTD and CTD are intrinsically disorder region (IDR).

(E) Multiple sequence alignment of CTD of the vertebrate CTCF proteins. The negatively-charged region (NCR) is highlighted in a yellow background.

(F) CTCF and Rad21 ChIP-seq peaks at the human *cPCDH* gene complex. The human *cPCDH* locus comprises three tandemly-linked gene clusters: *PCDHα*, *PCDHβ*, and *PCDHγ.* The *PCDHα* and *PCDHγ* clusters each consist of a variable region with multiple highly-similar and unusually-large exons and a constant region with three small exons. Each variable exon is separately *cis*-spliced to a single set of cluster-specific constant exons. The *PCDHβ* cluster contains 16 variable exons but with no constant exon. These three clusters form a superTAD with two subTADs: *PCDHα* and *PCDHβγ*. Each subTAD has its own downstream super-enhancer. Var, variable; Con, constant; HS, hyper-sensitive; SE, super-enhancer.

(G and H) Heatmaps of CTCF (G) and Rad21 (H) normalized signals at CBS elements in the *cPCDH* locus. Student's *t* test, $^{*}p < 0.05$.

(I and J) Heatmaps of CTCF (I) and Rad21 (J) normalized signals at genome-wide CBS elements. Student's *t* test, $^{****}p < 0.0001$.

(K-M) Three types of CTCF motifs in WT and ΔCTD cells.

(N) QHR-4C interaction profiles of the *PCDHα* gene cluster using *HS5-1* as an anchor.

(O) RNA-seq showing decreased expression levels of *PCDHα6, PCDHα12,* and *PCDHαc2* upon deletion of CTD. FPKM, fragments per kilobase of exon per million fragments mapped. Data are represented as mean ± SD; Student's *t* test, $^{****}p < 0.0001$.

See also Figures S1, S2, and S3.

Figure 2. Deleting 116 amino acids from the C-terminus leads to increased CTCF binding and affects gene expression

(A) CTCF and Rad21 ChIP-seq peaks at the *cPCDH* gene complex.

(B and C) Heatmaps of CTCF (B) and Rad21 (C) normalized signals at the *cPCDH* CBS elements. Student's *t* test, $^{*}p < 0.05$.

(D and E) Heatmaps of global CTCF (D) and Rad21 (E) ChIP-seq signals. Student's *t* test, $^{****}p < 0.0001$.

(F) QHR-4C interaction profiles of the *PCDHα* gene cluster using *HS5-1* as an anchor.

(G) RNA-seq indicating increased expression levels of *PCDHα6, PCDHα12,* and *PCDHαc2.* Data are represented as mean ± SD; Student's *t* test, $^{****}p <$

0.0001.
See also Figures S1, S3, and S4.

Figure 3. Rescue with a series of C-terminal truncated CTCFs reveals a negatively-charged region (NCR)

(A) Schematic illustration of different truncated CTCFs used in constructing stable cell lines. All truncated CTCFs are tagged with V5.
(B) Western blots of stable cells lines with the V5 antibody indicating truncated CTCF proteins in ΔCTD cells.
(C) Binding profiles of different length of CTCFs at the *cPCDH* locus.
(D) Heatmaps of CTCF ChIP-seq signals at the *cPCDH* locus. Student's *t* test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
(E) Heatmaps of different truncated CTCFs, showing global binding profiles. Student's *t* test, $^{****}p < 0.0001$.
(F-H) CTCF enrichments at the three types of CTCF motifs. Student's *t* test, $^{****}p < 0.0001$.
See also Figure S5.

Figure 4. NCR deletion increases CTCF binding and affects gene regulation

(A) CTCF and Rad21 ChIP-seq signals at the *cPCDH* locus, showing increased CTCF and cohesin enrichments at CBS elements upon NCR deletion.
(B and C) Heatmaps of CTCF (B) and Rad21 (C) ChIP-seq signals. Student's *t* test, $^{*}p < 0.05$.
(D and E) Heatmaps of CTCF (D) and Rad21 (E) ChIP-seq signals, indicating genome-wide CTCF and Rad21 enrichments upon NCR deletion. Student's *t* test, $^{****}p < 0.0001$.
(F) Quantitative high-resolution 4C experiments with *HS5-1* as an anchor, indicating increased chromatin interactions of the *HS5-1* enhancer with *PCDHα6* or *PCDHα12* promoters.
(G) RNA-seq indicating increased expression levels of *PCDHα6, PCDHα12, PCDHαc2* upon NCR deletion. Data are represented as mean ± SD; Student's *t* test, $^{****}p < 0.0001$.
(H) Volcano plot of differential gene expression analyses for WT and ΔNCR cells. Red dots, fold change of gene expression upon NCR deletion ($\log_2$FC > 1 and adjusted *p* value < 0.05). Blue dots, genes only passed adjusted *P* value < 0.05. Yellow dots, gene only passed $\log_2$FC > 1. FC: fold change.
(I) TSS distance of up-, down-, or none-regulated (NC) genes to the closest increased CTCF peaks in ΔNCR cells (data are shown as a cumulative distribution function, CDF).
(J) Models of NCR function in proper CTCF binding.
See also Figures S1, S3, and S6.

Figure 1

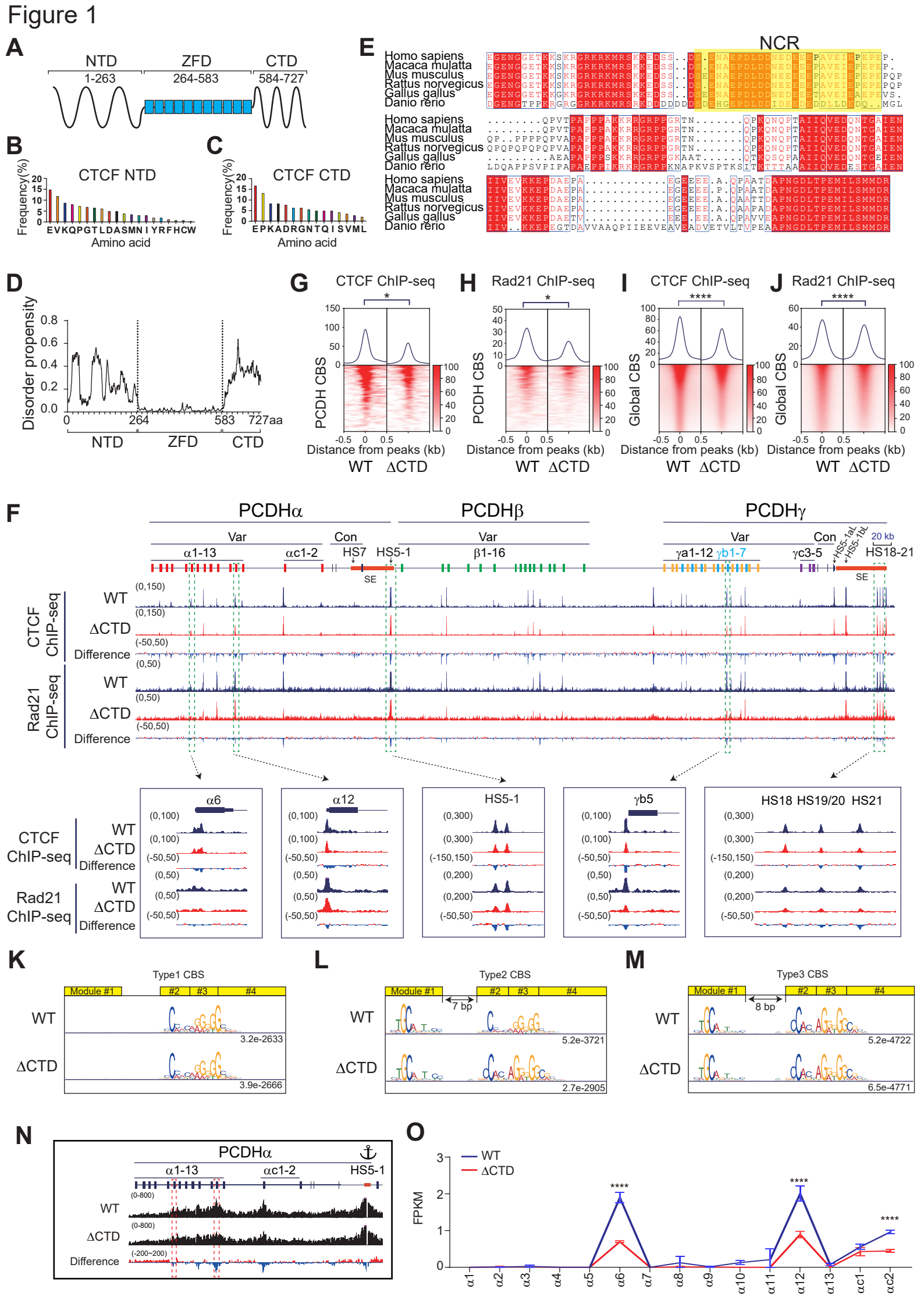

A
NTD
1-263
ZFD
264-583
CTD
584-727
B
CTCF NTD
Frequency(%)
EVKQPGTLDASMNIYRFHCW
Amino acid
C
CTCF CTD
Frequency(%)
EPKADRGNTQISVML
Amino acid
E
NCR
Homo sapiens
Macaca mulatta
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio
D
Disorder propensity
0
264
583
727aa
NTD
ZFD
CTD
G
CTCF ChIP-seq
PCDH CBS
Distance from peaks (kb)
WT
ΔCTD
H
Rad21 ChIP-seq
PCDH CBS
Distance from peaks (kb)
WT
ΔCTD
I
CTCF ChIP-seq
Global CBS
Distance from peaks (kb)
WT
ΔCTD
J
Rad21 ChIP-seq
Global CBS
Distance from peaks (kb)
WT
ΔCTD
F
PCDHα
PCDHβ
PCDHγ
Var
Con
α1-13
αc1-2
HS7
HS5-1
β1-16
γa1-12
γb1-7
γc3-5
HS5-1aL
HS5-1bL
HS18-21
20 kb
SE
CTCF ChIP-seq
Rad21 ChIP-seq
WT
ΔCTD
Difference
α6
α12
HS5-1
γb5
HS18 HS19/20 HS21
K
Type1 CBS
Module #1
#2
#3
#4
WT
ΔCTD
3.2e-2633
3.9e-2666
L
Type2 CBS
7 bp
5.2e-3721
2.7e-2905
M
Type3 CBS
8 bp
5.2e-4722
6.5e-4771
N
PCDHα
α1-13
αc1-2
HS5-1
WT
ΔCTD
Difference
O
WT
ΔCTD
FPKM
α1
α2
α3
α4
α5
α6
α7
α8
α9
α10
α11
α12
α13
αc1
αc2

Figure 2

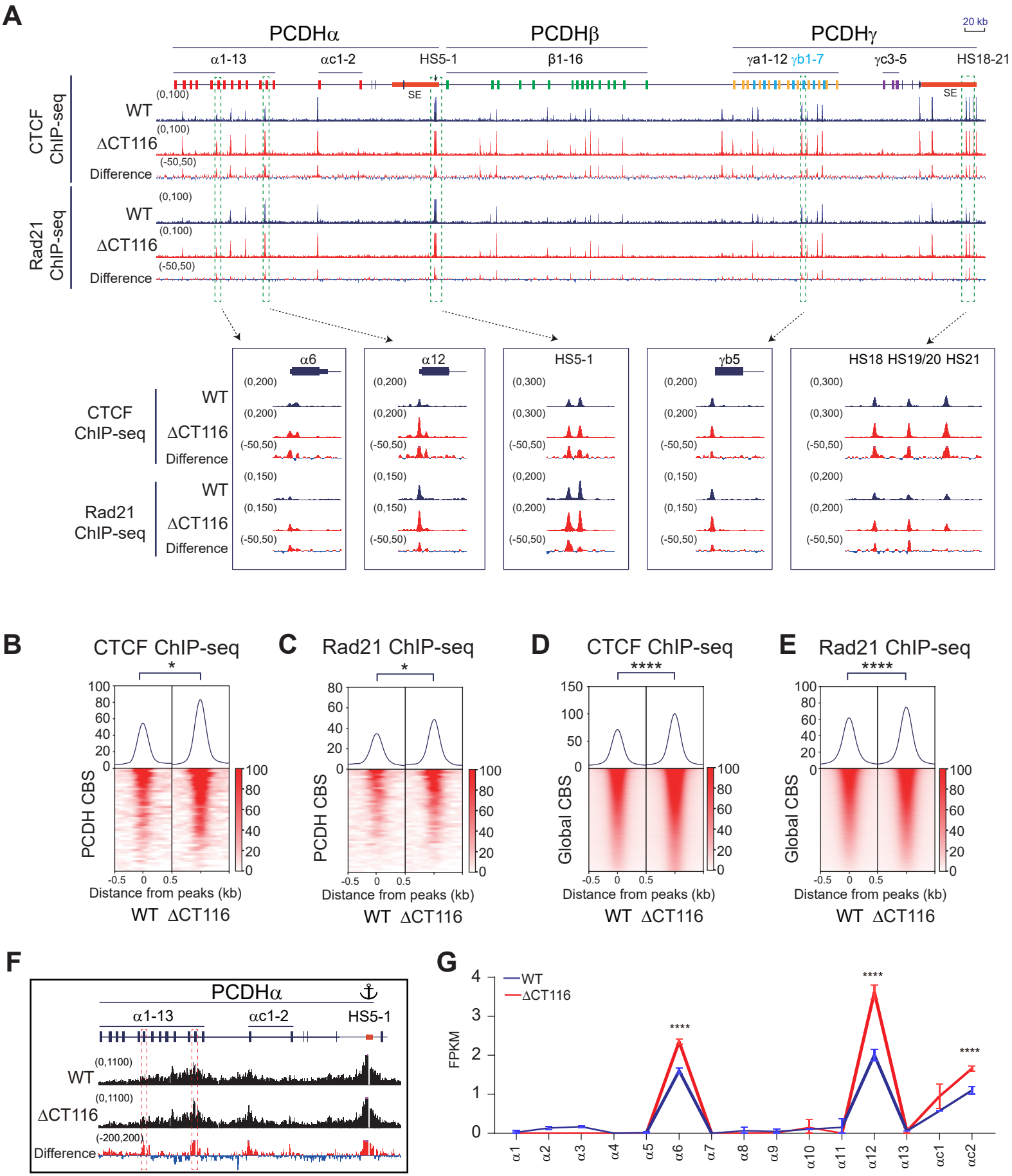

A
PCDHα
PCDHβ
PCDHγ
20 kb
α1-13
αc1-2
HS5-1
β1-16
γa1-12
γb1-7
γc3-5
HS18-21
SE
CTCF ChIP-seq
Rad21 ChIP-seq
WT
ΔCT116
Difference
α6
α12
HS5-1
γb5
HS18 HS19/20 HS21
B
CTCF ChIP-seq
C
Rad21 ChIP-seq
D
CTCF ChIP-seq
E
Rad21 ChIP-seq
PCDH CBS
Global CBS
Distance from peaks (kb)
WT ΔCT116
F
PCDHα
α1-13
αc1-2
HS5-1
WT
ΔCT116
Difference
G
WT
ΔCT116
FPKM
α1
α2
α3
α4
α5
α6
α7
α8
α9
α10
α11
α12
α13
αc1
αc2

Figure 3

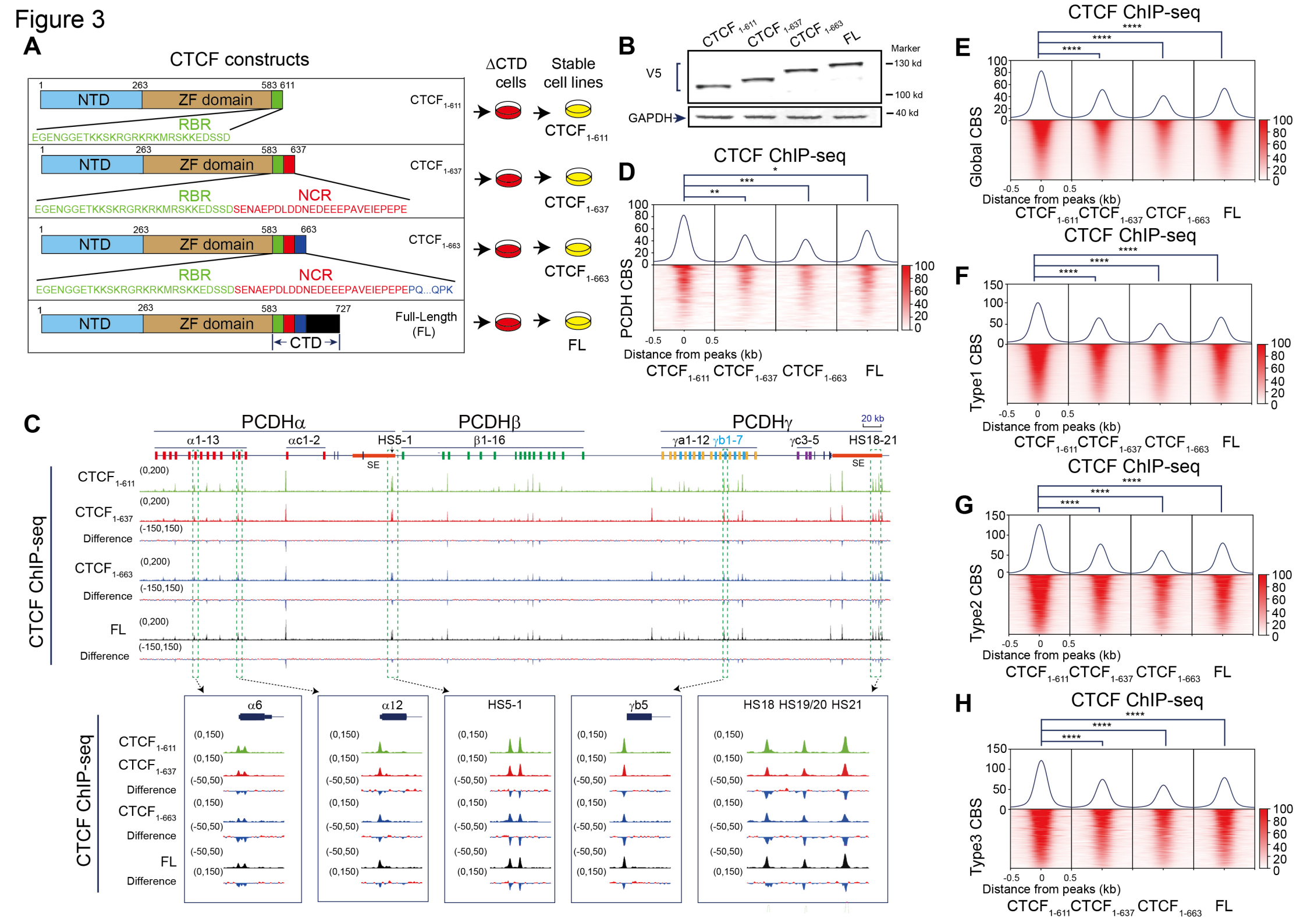
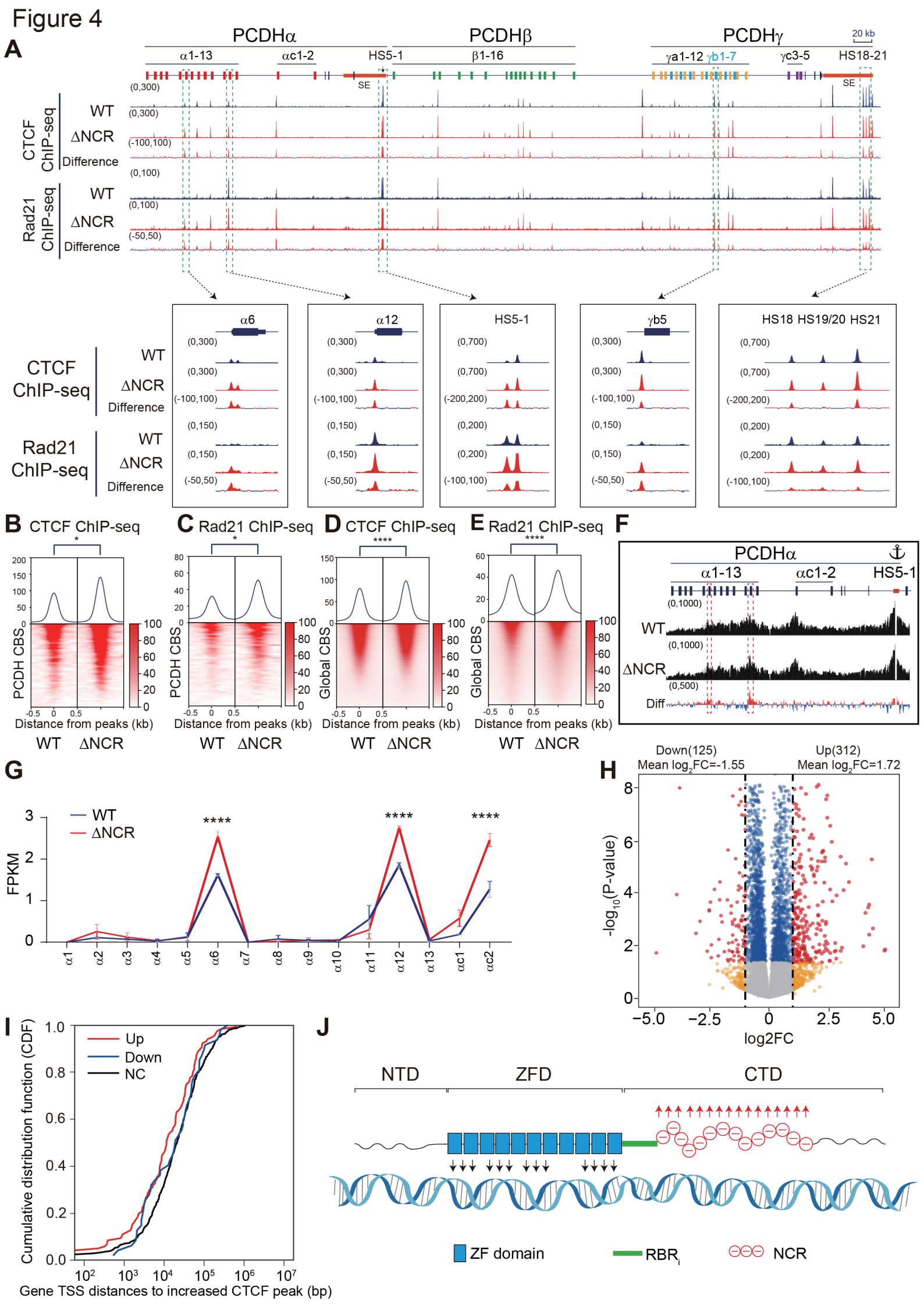

A
CTCF constructs
NTD
ZF domain
RBR
NCR
EGENGGETKKSKRGRKRKMRSKKEDSSD
SENAEPDLDDNEDEEEPAVEIEPEPE
PQ...QPK
CTD
Full-Length (FL)
ΔCTD cells
Stable cell lines
B
V5
GAPDH
Marker
130 kd
100 kd
40 kd
C
PCDHα
PCDHβ
PCDHγ
20 kb
SE
CTCF ChIP-seq
Difference
D
E
F
G
H
CTCF ChIP-seq
Global CBS
PCDH CBS
Type1 CBS
Type2 CBS
Type3 CBS
Distance from peaks (kb)
FL

Figure 4


A
PCDHα
PCDHβ
PCDHγ
20 kb
α1-13
αc1-2
HS5-1
β1-16
γa1-12
γb1-7
γc3-5
HS18-21
SE
CTCF ChIP-seq
Rad21 ChIP-seq
WT
ΔNCR
Difference
α6
α12
HS5-1
γb5
HS18 HS19/20 HS21
B CTCF ChIP-seq
C Rad21 ChIP-seq
D CTCF ChIP-seq
E Rad21 ChIP-seq
PCDH CBS
Global CBS
Distance from peaks (kb)
F
PCDHα
Diff
G
FPKM
H
Down(125)
Mean log2FC=-1.55
Up(312)
Mean log2FC=1.72
-log10(P-value)
log2FC
I
Up
Down
NC
Cumulative distribution function (CDF)
Gene TSS distances to increased CTCF peak (bp)
J
NTD
ZFD
CTD
ZF domain
RBRi
NCR

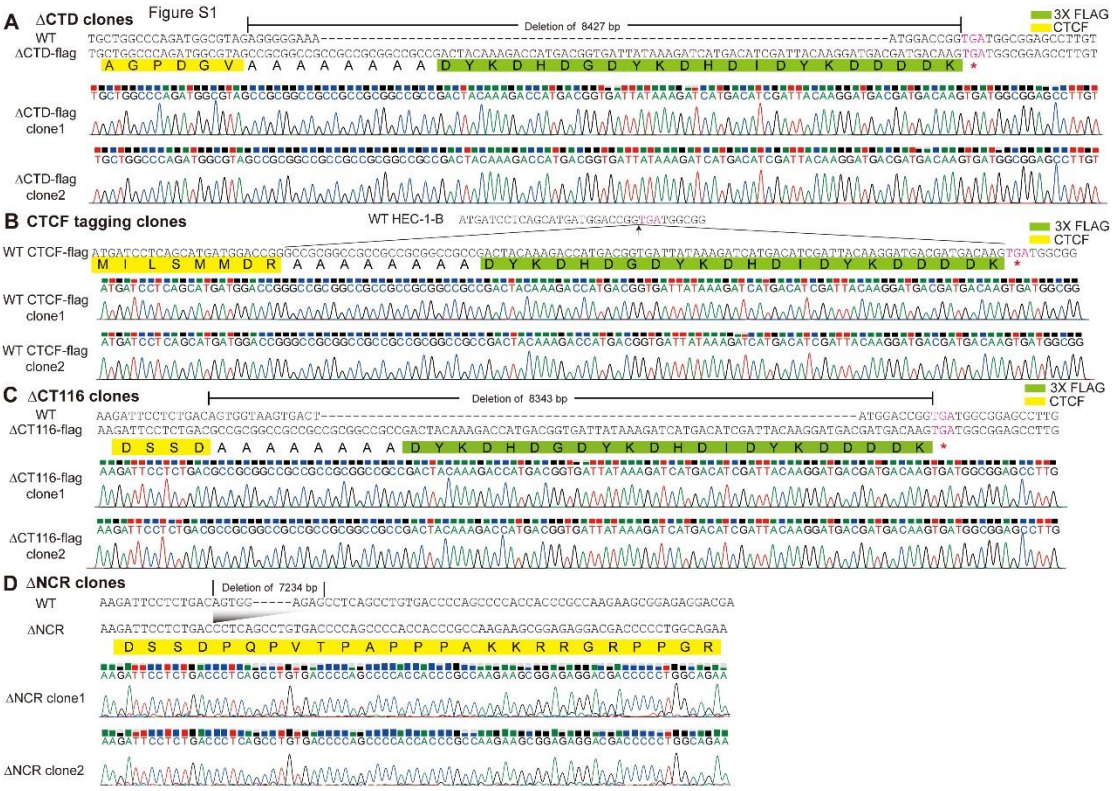


Figure S1. Genotyping of single-cell CRISPR clones of CTD, CT116, or NCR deletion as well as wild-type CTCF FLAG-tagging cell clones. Related to Figures 1, 2, and 4.

(A-D) Genotyping of the two ΔCTD-FLAG (A), WT-FLAG (B), ΔCT116-FLAG (C), or ΔNCR (D) singe-cell clones by Sanger sequencing.

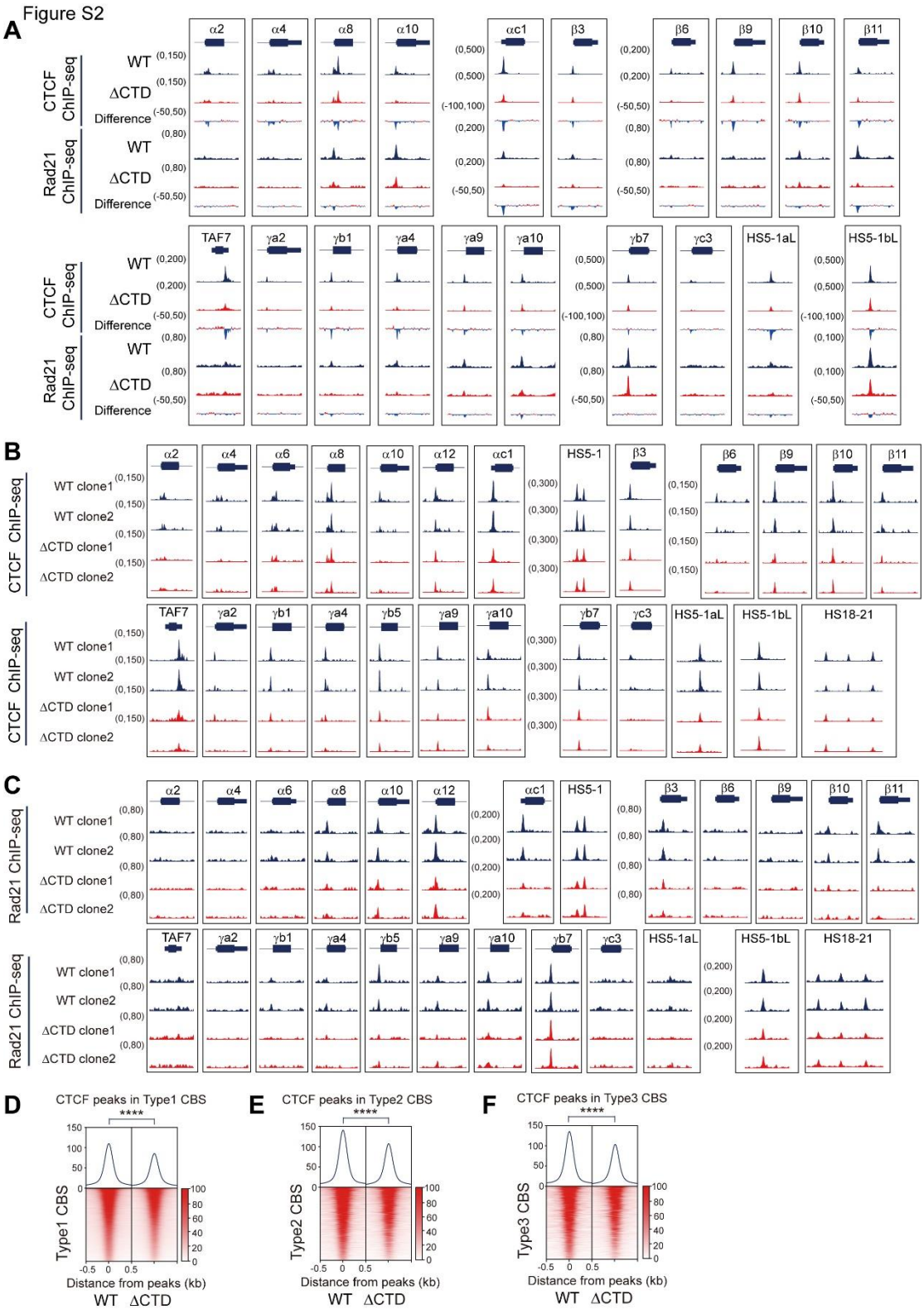


Figure S2 CTCF and Rad21 ChIP-seq profiles in ΔCTD and WT cells. Related to Figure 1. (A) CTCF and Rad21 ChIP-seq profiles for WT and ΔCTD cells at the *cPCDH* locus, indicating decreased CTCF binding upon CTD deletion. (B and C) CTCF (B) or Rad21 (C) ChIP-seq profiles for the two single-cell clones of WT or ΔCTD at the *cPCDH* locus. (D-F) Heatmaps of CTCF ChIP-seq signals at the three types of CTCF motifs upon CTD deletion. Student's *t* test, ****$p < 0.0001$.

Figure S3

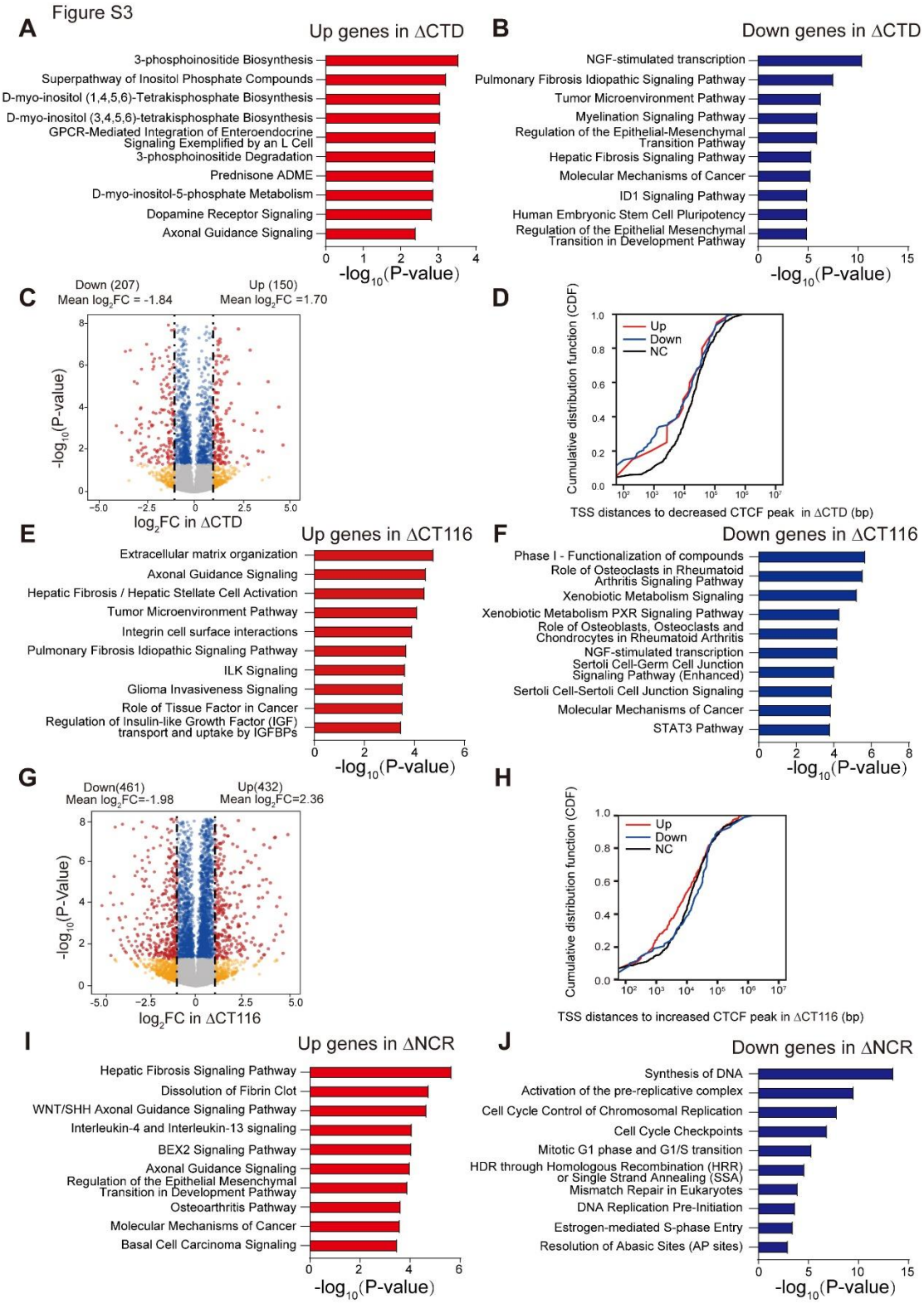

A
Up genes in ΔCTD
3-phosphoinositide Biosynthesis
Superpathway of Inositol Phosphate Compounds
D-myo-inositol (1,4,5,6)-Tetrakisphosphate Biosynthesis
D-myo-inositol (3,4,5,6)-tetrakisphosphate Biosynthesis
GPCR-Mediated Integration of Enteroendocrine Signaling Exemplified by an L Cell
3-phosphoinositide Degradation
Prednisone ADME
D-myo-inositol-5-phosphate Metabolism
Dopamine Receptor Signaling
Axonal Guidance Signaling
-log10(P-value)
B
Down genes in ΔCTD
NGF-stimulated transcription
Pulmonary Fibrosis Idiopathic Signaling Pathway
Tumor Microenvironment Pathway
Myelination Signaling Pathway
Regulation of the Epithelial-Mesenchymal Transition Pathway
Hepatic Fibrosis Signaling Pathway
Molecular Mechanisms of Cancer
ID1 Signaling Pathway
Human Embryonic Stem Cell Pluripotency
Regulation of the Epithelial Mesenchymal Transition in Development Pathway
-log10(P-value)
C
Down (207)
Mean log2FC = -1.84
Up (150)
Mean log2FC =1.70
-log10(P-value)
log2FC in ΔCTD
D
Up
Down
NC
Cumulative distribution function (CDF)
TSS distances to decreased CTCF peak in ΔCTD (bp)
E
Up genes in ΔCT116
Extracellular matrix organization
Axonal Guidance Signaling
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Tumor Microenvironment Pathway
Integrin cell surface interactions
Pulmonary Fibrosis Idiopathic Signaling Pathway
ILK Signaling
Glioma Invasiveness Signaling
Role of Tissue Factor in Cancer
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by IGFBPs
-log10(P-value)
F
Down genes in ΔCT116
Phase I - Functionalization of compounds
Role of Osteoclasts in Rheumatoid Arthritis Signaling Pathway
Xenobiotic Metabolism Signaling
Xenobiotic Metabolism PXR Signaling Pathway
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis
NGF-stimulated transcription
Sertoli Cell-Germ Cell Junction Signaling Pathway (Enhanced)
Sertoli Cell-Sertoli Cell Junction Signaling
Molecular Mechanisms of Cancer
STAT3 Pathway
-log10(P-value)
G
Down(461)
Mean log2FC=-1.98
Up(432)
Mean log2FC=2.36
-log10(P-Value)
log2FC in ΔCT116
H
Up
Down
NC
Cumulative distribution function (CDF)
TSS distances to increased CTCF peak in ΔCT116 (bp)
I
Up genes in ΔNCR
Hepatic Fibrosis Signaling Pathway
Dissolution of Fibrin Clot
WNT/SHH Axonal Guidance Signaling Pathway
Interleukin-4 and Interleukin-13 signaling
BEX2 Signaling Pathway
Axonal Guidance Signaling
Regulation of the Epithelial Mesenchymal Transition in Development Pathway
Osteoarthritis Pathway
Molecular Mechanisms of Cancer
Basal Cell Carcinoma Signaling
-log10(P-value)
J
Down genes in ΔNCR
Synthesis of DNA
Activation of the pre-replicative complex
Cell Cycle Control of Chromosomal Replication
Cell Cycle Checkpoints
Mitotic G1 phase and G1/S transition
HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA)
Mismatch Repair in Eukaryotes
DNA Replication Pre-Initiation
Estrogen-mediated S-phase Entry
Resolution of Abasic Sites (AP sites)
-log10(P-value)

Figure S3 RNA-seq analyses of ΔCTD, ΔCT116, and ΔNCR cell clones. Related to Figures 1, 2, and 4.

(A and B) Ingenuity pathway analysis (IPA) of up- (A) and down- (B) regulated genes upon CTD deletion. (C) Volcano plot of differential gene expression analyses upon ΔCTD deletion. Red dots, gene expression changed upon CTD deletion ($log_2FC > 1$ and adjusted p value < 0.05). Blue dots, genes only passed adjusted p value < 0.05. Yellow dots, gene only passed $log_2FC > 1$. (D) TSS distances of up-, down-, or none-regulated (NC) genes to the closest decreased CTCF peaks in ΔCTD cells (data were shown as a cumulative distribution function, CDF). (E and F) IPA analyses of up- (E) and down- (F) regulated genes in ΔCT116 cells. (G) Volcano plot of differential gene expression analyses of WT and ΔCT116 cells. (H) TSS distances of up-, down-, or none-regulated (NC) genes to the closest increased CTCF peaks in ΔCT116 cells (data were shown as a cumulative distribution function, CDF). (I and J) IPA analyses of up- (I) and down- (J) regulated genes upon NCR deletion.

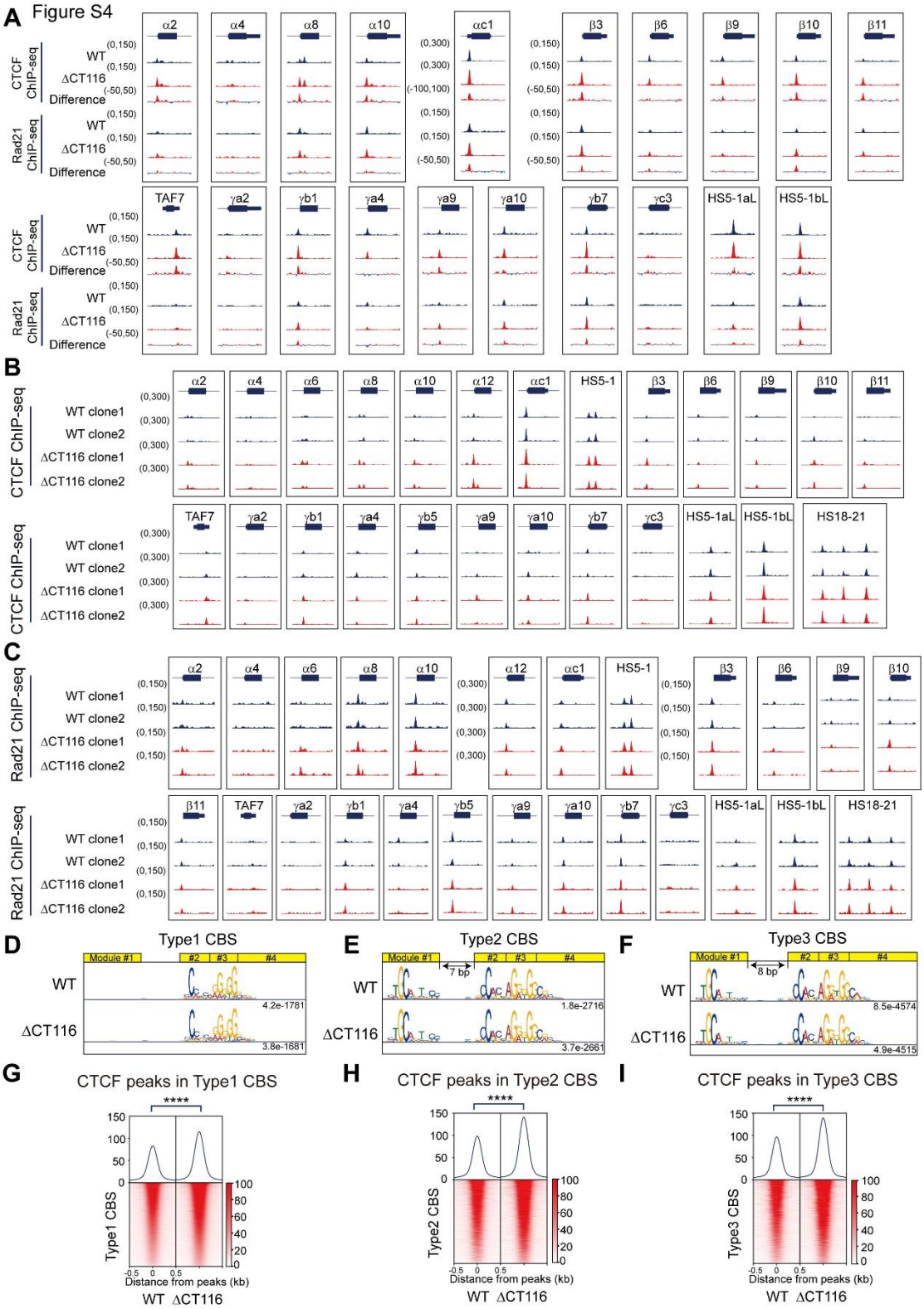


Figure S4. CTCF and Rad21 ChIP-seq profiles in ΔCT116 and WT cells. Related to Figure 2. (A) CTCF and Rad21 ChIP-seq profiles of WT and ΔCT116 cells at the *cPCDH* locus, showing increased CTCF binding in ΔCT116 cells. (B and C) CTCF (B) or Rad21 (C) ChIP-seq profiles of the two single-cell clones of WT and ΔCT116 at the *cPCDH* locus. (D-F) Three types of CTCF motifs of WT and ΔCT116 cells. (G-I) Heatmaps of CTCF ChIP-seq signals at the three types of CTCF motifs of WT and ΔCT116 cells. Student's *t* test, ****$p < 0.0001$.

Figure S5

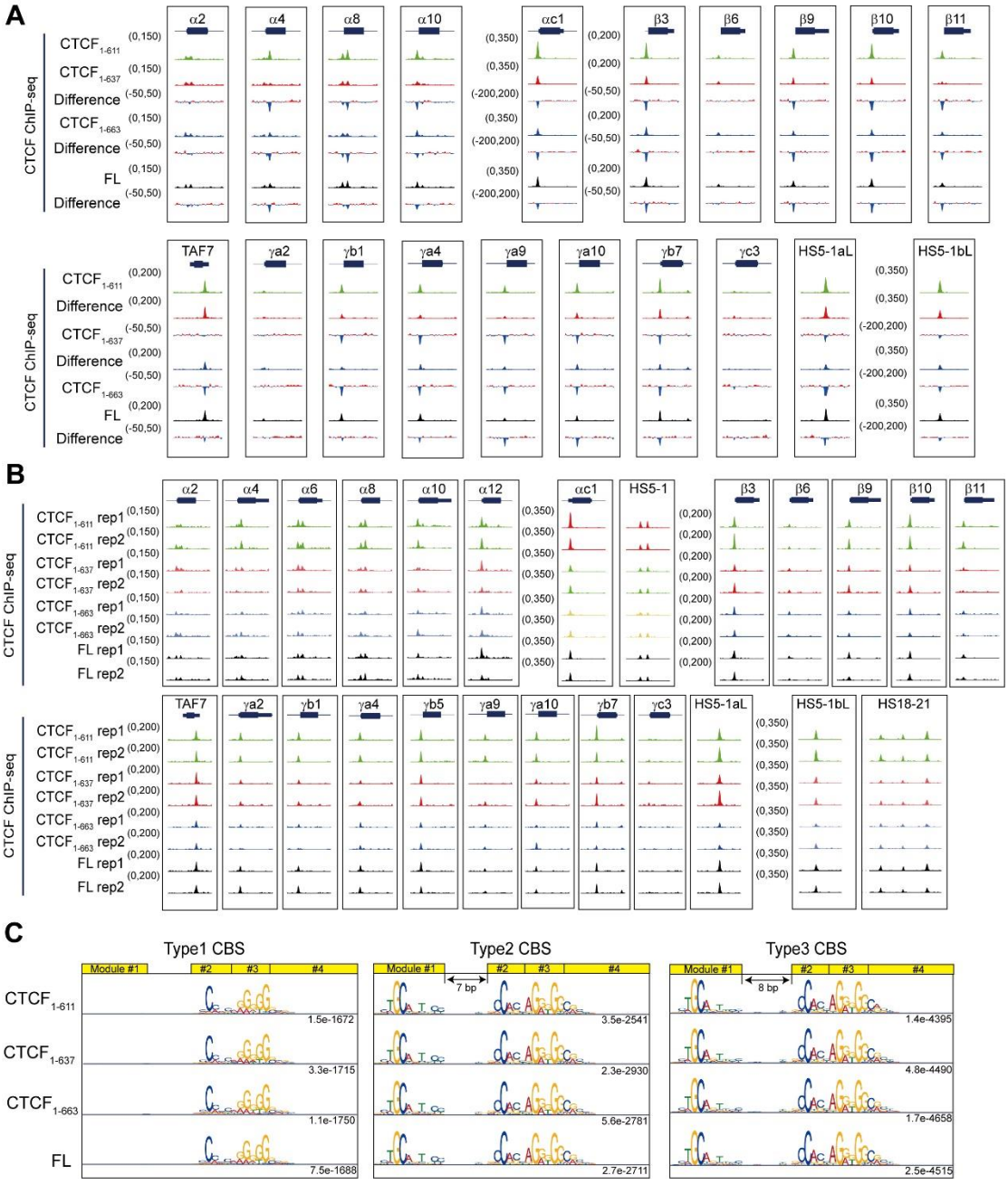


Figure S5. ChIP-seq profiles of a series of truncated CTCF proteins. Related to Figure 3. (A) ChIP-seq of different truncated CTCF proteins at the *cPCDH* locus, indicating that $CTCF_{1-611}$ has the highest binding strength. (B) Biological replicates of ChIP-seq experiments with truncated CTCF proteins. (C) Motif analyses of truncated CTCF proteins, showing no alteration of the three types of CTCF motifs.

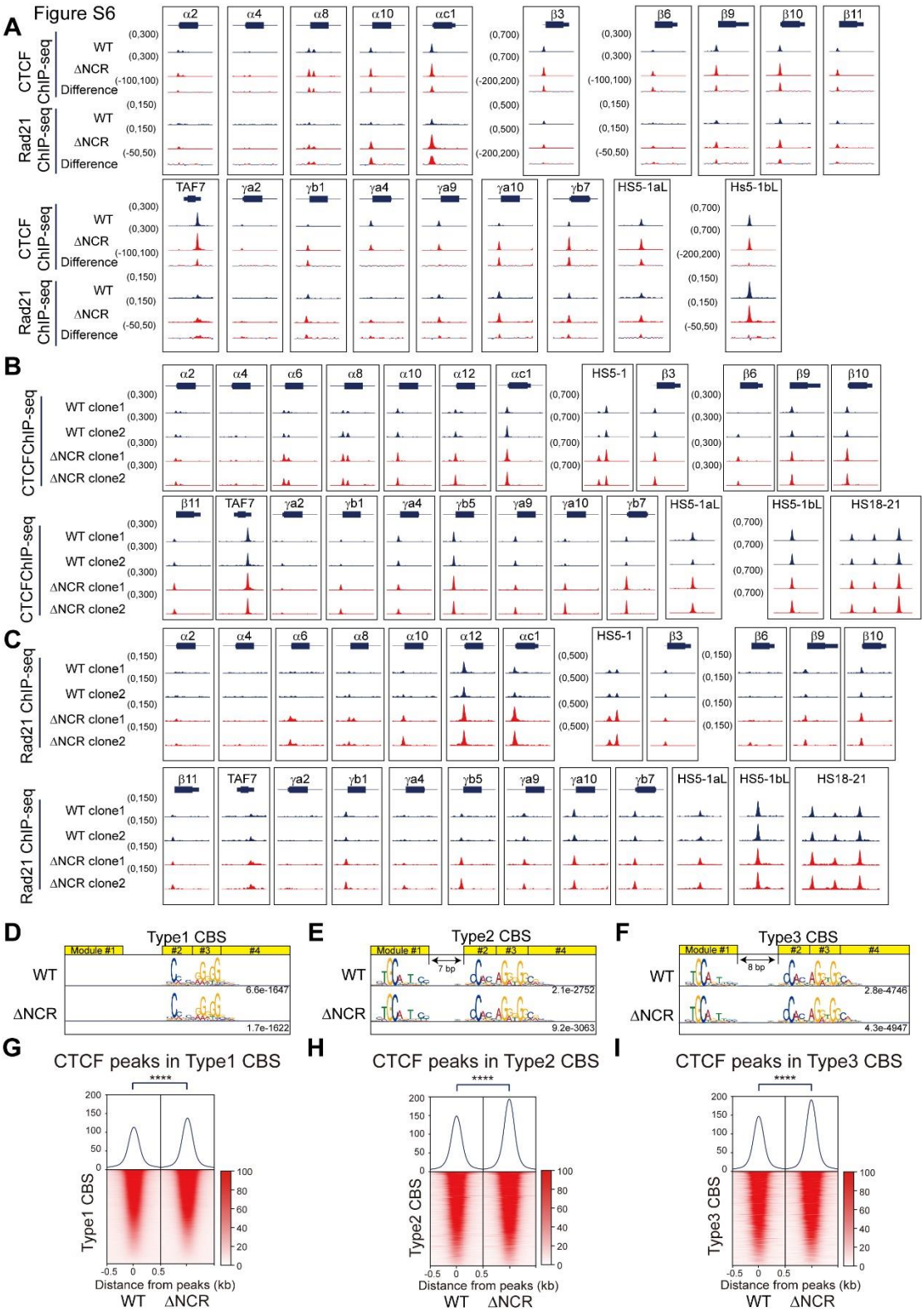


Figure S6. CTCF and Rad21 ChIP-seq profiles upon NCR deletion. Related to Figure 4. (A) CTCF and Rad21 ChIP-seq profiles at the *cPCDH* gene complex upon NCR deletion. (B and C) CTCF (B) or Rad21 (C) ChIP-seq profiles of the two single-cell clones of WT and ΔNCR at the *cPCDH* locus. (D-F) Three types of CTCF motifs upon NCR deletion. (G-I) Heatmaps of CTCF ChIP-seq signals at the three types of CTCF motifs upon NCR deletion. Student's *t* test, ****$p < 0.0001$.

Figure S7

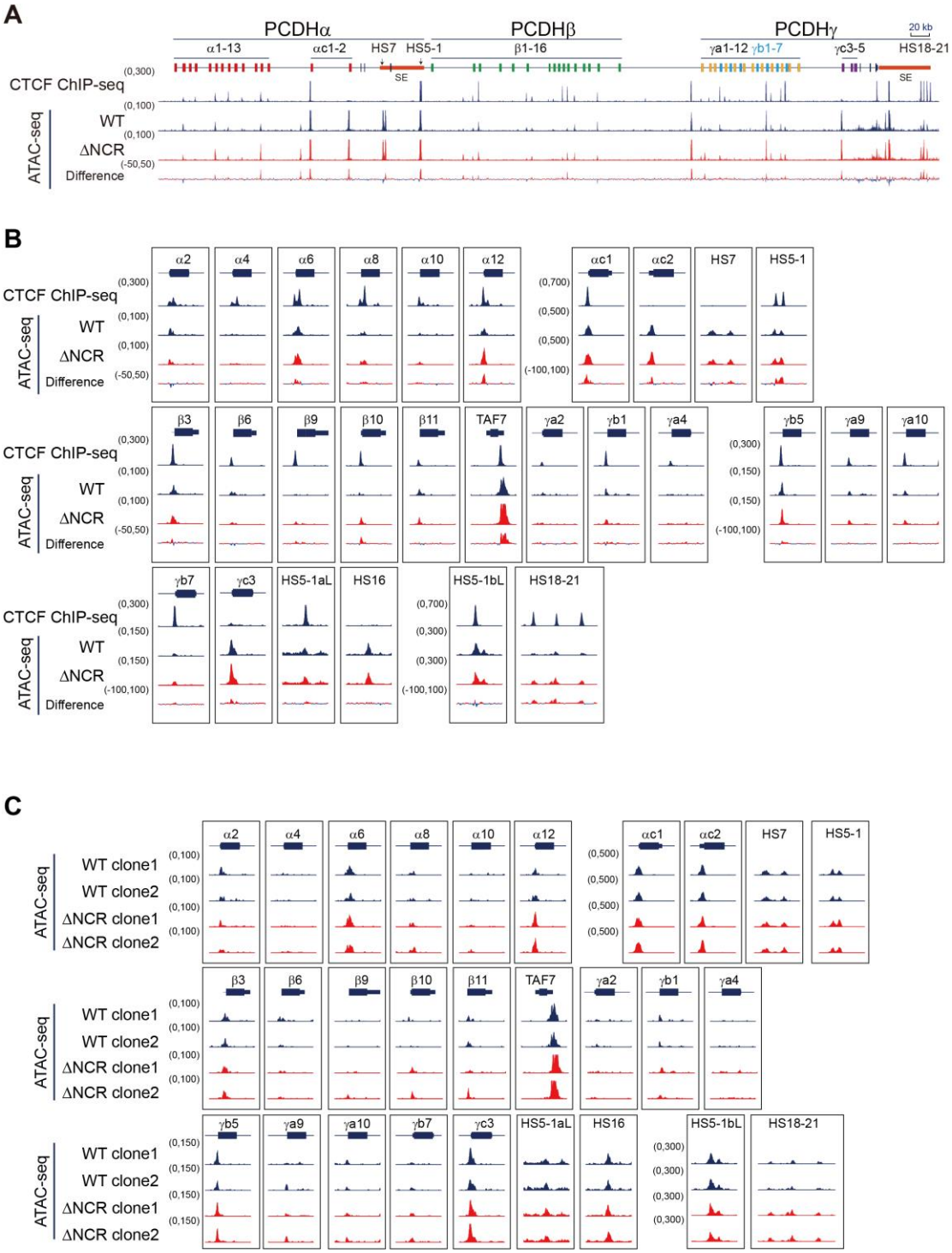


Figure S7. ATAC-seq profiles upon NCR deletion. Related to Figure 4.

(A) ATAC-seq indicates increased chromatin accessibility at the *cPCDH* locus upon NCR deletion. (B) Close-up of ATAC-seq profiles at the *cPCDH* gene complex. (C) ATAC-seq profiles of the two single-cell clones of WT and ΔNCR at the *cPCDH* gene complex.

Figure S8

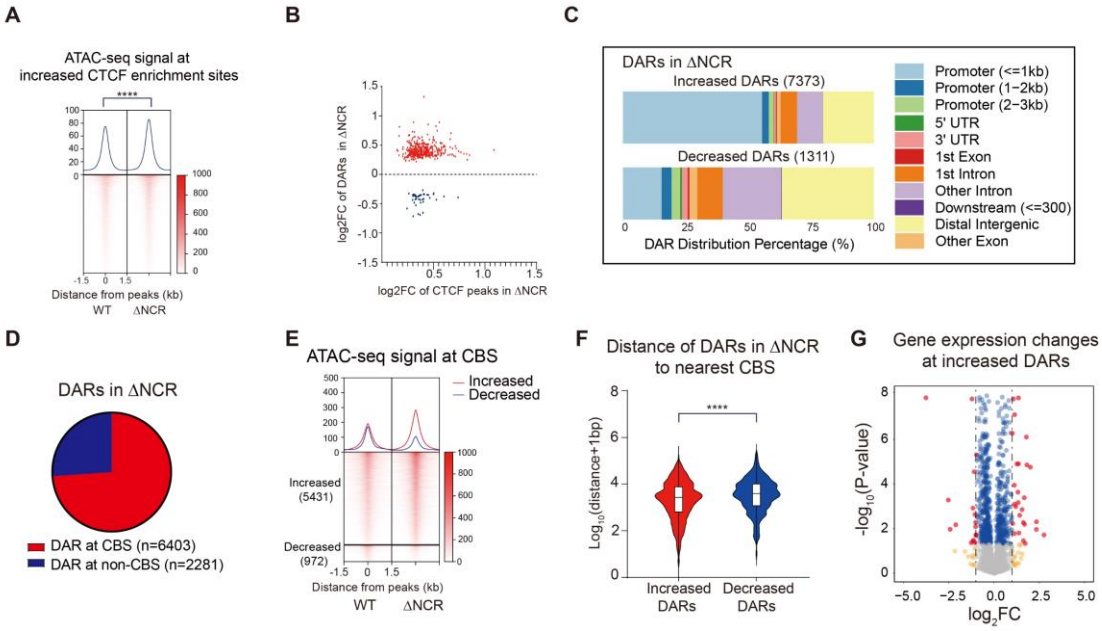


Figure S8 Deletion of NCR rewires chromatin accessibility. Related to Figure 4.

(A) Heatmaps of ATAC-seq signals at sites with increased CTCF enrichments, indicating increased chromatin accessibility upon NCR deletion. Student's *t* test, ****$p < 0.0001$. (B) Scatter plot shows the $\log_2$FC of increased CTCF enrichments (x axis) and the corresponding $\log_2$FC of DARs (y axis) upon NCR deletion. (C) Distributions of increased and decreased DARs with gene elements upon NCR deletion. (D) Distribution of DARs in term of whether overlapping with CBS elements. (E) Heatmaps of DARs overlapped with CBS elements. (F) Violin plot of $\log_{10}$(distance+1bp) distribution of increased and decreased DARs to the nearest CBS element upon NCR deletion, showing that increased chromatin accessible regions are closer to CBS elements. (G) Volcano plot showing that increased DARs are associated with more up-regulated genes upon NCR deletion.