Exploring the Changing Landscape of Cell-to-Cell Variation After CTCF Knockdown via Single Cell RNA-seq

Wei Wang[1], Gang Ren[2], Ni Hong[1*], and Wenfei Jin[1*]

[1] Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China;

[2] Systems Biology Center, Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA;

*Corresponding authors

E-mail: jinwf@sustech.edu.cn or hongn@mail.sustech.edu.cn

Abstract

Background: CCCTC-Binding Factor (CTCF), also known as 11-zinc finger protein, participates in many cellular processes, including insulator activity, transcriptional regulation and organization of chromatin architecture. Based on single cell flow cytometry and single cell RNA-FISH analyses, our previous study showed that deletion of CTCF binding site led to a significantly increase of cellular variation of its target gene. However, the effect of CTCF on genome-wide landscape of cell-to-cell variation is unclear.

Results: We knocked down CTCF in EL4 cells using shRNA, and conducted single cell RNA-seq on both wild type (WT) cells and CTCF-Knockdown (CTCF-KD) cells using Fluidigm C1 system. Principal component analysis of single cell RNA-seq data showed that WT and CTCF-KD cells concentrated in two different clusters on PC1, indicating gene expression profiles of WT and CTCF-KD cells were systematically different. Interestingly, GO terms including regulation of transcription, DNA binding, Zinc finger and transcription factor binding were significantly enriched in CTCF-KD-specific highly variable genes, indicating tissue-specific genes such as transcription factors were highly sensitive to CTCF level. The dysregulation of transcription factors potentially explain why knockdown of CTCF lead to systematic change of gene expression. In contrast, housekeeping genes such as rRNA processing, DNA repair and tRNA processing were significantly enriched in WT-specific highly variable genes, potentially due to a higher cellular variation of cell activity in WT cells compared to CTCF-KD cells. We further found cellular variation-increased genes were significantly enriched in down-regulated genes, indicating CTCF knockdown simultaneously reduced the expression levels and increased the expression noise of its regulated genes.

Conclusions: To our knowledge, this is the first attempt to explore genome-wide landscape of cellular variation after CTCF knockdown. Our study not only advances our understanding of CTCF function in maintaining gene expression and reducing expression noise, but also provides a framework for examining gene function.

Keywords: Single cell RNA-seq; Cell-to-cell variation; CTCF; Change of cellular variation; CTCF knockdown

Background

CCCTC-binding factor (CTCF) is an 11-zinc finger protein that directionally binds to a well-defined DNA motif [1, 2]. Although CTCF was initially reported as a transcription factor [3, 4], subsequent studies found it served as an insulator [5-7]. Nowadays, CTCF has been reported being involved in multiple cellular processes, such as transcriptional regulation, insulator activity, epigenetic regulation, organization of chromatin architecture and X chromosome inactivation [1, 2, 8-14]. CTCF activates and silences gene expression by preventing the spread of heterochromatin and blocking of unrelated enhancer-promoter interactions [1, 15]. Interestingly, mammalian genome is organized into thousands of highly self-interacting topologically associated domains (TADs) with CTCF demarcating individual TAD boundary [16]. Analyses based on high-resolution interaction matrix further identified ~10,000 chromatin loops ranging ~185Kb in human genome, anchoring by convergent CTCF-binding motif-pair at TAD boundaries [13]. In addition, chromatin loops mediated by CTCF and cohesin can tether distal enhancers to gene promoters and regulate its target gene expression [13, 14, 17, 18]. Further studies showed CTCF-mediated DNA loop could determine the chromatin architecture, with anchor-genes almost exclusively being housekeeping genes, while loop-genes being tissue-specific genes [14, 19]. For instance, inversion of a CTCF-binding site reconfigured the topology of chromatin loops and activated gene expression by creating a new chromatin loop [20, 21].

Recent studies showed that disruption of CTCF binding in mammalian cells resulted in loss of TADs [18], genomic instability [22], developmental failure [23, 24] and other malfunctions. Our recent study demonstrated that CTCF played an important role in stabilizing enhancer-promoter interaction and reducing the gene expression noises in mammalian cells [17]. In particular, we found that CTCF-KD or deletion of CTCF binding sites led to increased variation of cellular expression of GATA3, CD28, CD90 and CD5 [17]. However, the genome-wide change of cell-to-cell variation after CTCF-KD remains unknown. In this study, we conducted single cell RNA-seq on both WT and CTCF-KD cells to investigate the changing landscape of cell-to-cell variation at a genome-wide scale. Interestingly, GO terms including regulation of transcription, DNA binding and Zinc finger were significantly

enriched in CTCF-KD specific highly variable genes. We also found that cellular variation-increased genes were significantly enriched in down-regulated genes, indicating knockdown of CTCF simultaneously reduced the expression level and increased the expression noises of its regulated genes.

Results

Efficient CTCF knockdown and single cell RNA-seq

We knocked down CTCF in EL4 cells by short hairpin RNA (shRNA). Western blotting showed a dramatic decrease of the CTCF protein level in shCTCF #1 and shCTCF #2 compared to shRNA luciferase controls (shLuc) (Fig. 1A). Quantitative polymerase chain reaction (qPCR) revealed the mRNA levels in shCTCF #1 and shCTCF #2 had been reduced to 38% and 40% of that in shLuc, respectively (Fig. 1B). These results confirmed the efficient knockdown of CTCF expression in shCTCF#1 and shCTCF#2, with a significant reduction consistently at both RNA and protein level.

In order to investigate the changing landscape of cell-to-cell variation after CTCF knockdown, we successfully conducted single cell RNA-seq for shLuc#1, shLuc #2, shCTCF #1 and shCTCF#2 using 4 integrated fluidics circuits (IFCs) (Fig. 1C). We noticed the gene expression level of pooled shLuc single cells were highly correlated with that of bulk data from our previous study [17] (r^2=0.86; Fig. S1A). The gene expression of pooled single cells from shLuc #1 was also highly correlated with that of shLuc #2 (r^2=0.87; Fig. S1B). In addition, the gene expression of pooled single cells repeats and bulk cell repeats in CTCF-KD cells were highly correlated (Fig. S1C-S1D).

Systematic differences between WT cells and CTCF-KD cells

A total of 95 cells, including 24 cells from shLuc #1, 24 cells from shCTCF #1, 23 cells from shLuc #2 and 24 cells from shCTCF #2, were kept for further analyses after quality control. We conducted principal component analysis (PCA) on 11,361 genes shared by those 95 cells (Fig. 2A). The coordination of cells from experiment1 and experiment2 on PCA projection is not significantly different (P=0.8; student's

4

t-test), indicating no obvious batch effect between experiment1 and experiment2. Further analysis showed that WT cells and CTCF-KD cells were distinguishable on PCA projection and concentrated in two different clusters on PC1 (Fig. 2B, 2C), implying the gene expression profiles of CTCF-KD and WT cells were systematically different. We also noticed a correlation between CTCF expression level and its coordination on the PCA projection (using the first 10 PCs), among which CTCF expression level and PC2 exhibited the highest correlation (Fig. S2A) (r^2=0.18, P = 0.22×10^{-6}).

We further calculated the differential gene expression between WT and CTCF-KD cells using edgeR [25]. We identified 195 up-regulated and 107 down-regulated genes in CTCF-KD cells compared to WT cells (Fig. S2B). Heatmap of the most differentially expressed genes between WT cells and CTCF-KD cells exhibited a cellular heterogeneity within the same cell population (Fig. 2D). The most enriched gene categories in down-regulated genes include glycolytic processing, Prolyl 4-hydroxylase α subunit, iron-dependent dioxygenase and carbon metabolism (Fig. S2D). Whereas the most enriched gene categories in up-regulated genes include RNA binding, ribosome biogenesis, WD40 repeat domain and RNA processing (Fig. S2C), consistent to our recent study based on bulk data in some way [17].

CTCF knockdown changed the landscape of cell-to-cell variation

In order to distinguish true signals of cellular variation from technical noise, we calculated the expression noise of each gene (σ^2/μ^2) [26, 27]. The expression noises exhibited two distinct scaling properties: negative association with expression at low expression levels and no association at high expression levels ($\log_2$TPM>1) (Fig. 3A). We filtered out low expressed genes ($\log_2$TPM≤1) to reduce the impact of technological noise, resulting in 7,843 genes for further analysis (Fig. 3A). Coefficient of variation (CV) was calculated to measure the cell-to-cell variation of each gene across the cell populations. The distribution of changes of cell-to-cell variation pre-and post-CTCF knockdown followed normal distribution (Fig. S3). We identified 602 cellular variation increased genes and 890 cellular variation decreased genes after CTCF knockdown by mean±SD (Fig. 3B).

GO analyses showed that variation-increased genes were significantly enriched in GO terms such as regulation of transcription, DNA binding, zinc finger proteins, covalent chromatin modification and transcription factor binding (Fig. 3C). In fact, almost all genes in DNA binding, zinc finger proteins and transcription factor binding are transcription factors. The significant enrichment of transcription factors in CTCF-KD-specific highly variable genes potentially indicates that transcription factors are highly sensitive to CTCF level and are tend to be cellular variation increased genes. The dysregulation of a lot transcription factor potentially explain why knockdown of CTCF lead to systematic change of gene expression. In contrast, variation-decreased genes were significantly enriched in housekeeping genes related GO terms such as rRNA processing, DNA repair, tRNA processing, and RNA modification (Fig. 3D). The enrichment of housekeeping genes in WT-specific highly variable genes potentially indicates a higher cellular variation of cell activity in WT cells compared to CTCF-KD cells.

CTCF Knockdown simultaneously altered expression levels and cellular variations of its regulated genes

We identified 302 expression-changed genes and 1,490 cellular variation-changed genes pre-and-post CTCF knockdown. It is interesting to examine whether those cellular variation-changed genes were enriched in expression-changed genes. Venn diagram showed that 47 genes out of total 107 down-regulated genes exhibited increased cellular variation (Fig. 4), which were significantly over-represented ($P=0.29\times10^{-23}$, χ^2 test), indicating CTCF knockdown simultaneously reduced the expression level and increased the gene expression noise. Among those genes with decreased expression and increased cellular variation, EGR1 and JUNB played an important role in maintaining the cell type-specific gene regulation. For instance, EGR1 belongs to the EGR family of C2H2-type zinc-finger proteins, and encodes a nuclear protein that participates in transcriptional regulation.

Meanwhile, there were 96 genes out of the total 195 up-regulated genes exhibiting decreased cellular variation (Fig. 4), which were also significantly over-represented ($P=0.48\times10^{-23}$, χ^2 test). The 96 genes with decreased cellular

variation and increased expression level were significantly enriched in poly(A) RNA binding, rRNA processing, WD40, purine nucleobase biosynthetic processing, rRNA methylation and RNA methyltransferase activity. It is obvious that those enriched GO terms were associated with basic cellular functions belonging to housekeeping genes. Taken together, our results clearly indicate that distortion of CTCF expression could simultaneously change the gene expression level and cell-to-cell variation of its regulated genes.

Furthermore, we identified CTCF binding sites using CTCF ChIP-seq data in WT EL4 cells from our previous study [17]. We identified each gene-associated CTCF by counting the CTCF binding sites within 20Kb of the transcriptional start site (TSS) for each gene. The numbers of gene-associated CTCF of variation-increased genes are significantly higher than that of variation-decreased genes (P= 0.0033; Wilcoxon test), and are significantly higher than that of variation-unchanged genes (P= 0.16×10^{-6}; Wilcoxon test). The numbers of gene-associated CTCF of variation-decreased genes are not significantly different from that of variation-unchanged genes (P= 0.5; Wilcoxon test). These results suggest that genes regulated by multiple CTCF binding sites tend to possess higher cellular variation after CTCF knockdown.

Discussion

CTCF plays an important role in chromatin structure organization and regulation of gene expression [13-16]. In this study, we used single cell RNA-seq to analyze genome-wide gene expression profiles of WT and CTCF-KD cells at single cell resolution. Indeed, WT cell population and CTCF-KD cell population showed distinct concentration on PC1, indicating that knockdown of CTCF resulted in a systematic impact on the genome-wide gene expression profile. These results further implied that CTCF contributed to key functions in controlling the genome-wide gene regulation. We generated the genome-wide landscape of cell-to-cell variation in both WT and CTCF-KD cells. After comparing cell-to-cell variations between WT and CTCF-KD cells, we identified those genes showing significant change of cellular variation after CTCF knockdown. Interestingly, the cellular variation-increased genes are

significantly enriched in expression-decreased genes, suggesting CTCF-medicated promoter-enhancer interaction did not only play an important role in maintaining the expression of its regulated genes, but also reduced their expression noise.

In this study, we identified a lot gene with an obvious change of cellular variation after CTCF knockdown. Interestingly, the variation-increased genes were significantly enriched in GO terms such as chromatin DNA binding, zinc finger proteins and zinc ion binding, indicating the expression noise of those zinc finger proteins were strongly increased after CTCF knockdown. The increased cellular variation of zinc finger proteins potentially indicates a high sensitivity of zinc finger proteins to CTCF expression level or cellular environmental change within the cell. In fact, the majority of those zinc finger proteins were transcription factors that played an important role in the regulation of cell type-specific gene expression. Our observation that CTCF knockdown fluctuated expression of a lot transcriptional factor further explain why disruption of CTCF expression led to pronounced biological effects such as development failure [23, 24]. Taken together, our findings provide convincing evidence that CTCF serves as a key player in stabilizing the gene expression noise of zinc finger related genes.

Conclusion

We conducted single cell RNA-seq on both wild type (WT) cells and CTCF-Knockdown (CTCF-KD) cells using Fluidigm C1 system. Principal component analysis of single cell RNA-seq data showed that WT and CTCF-KD cells concentrated in two different clusters on PC1, indicating gene expression profiles of WT and CTCF-KD cells were systematically different. Interestingly, GO terms including regulation of transcription, DNA binding, Zinc finger and transcription factor binding were significantly enriched in CTCF-KD-specific highly variable genes, indicating tissue-specific genes such as transcription factors were highly sensitive to CTCF level. The dysregulation of transcription factors potentially explain why knockdown of CTCF lead to systematic change of gene expression. In contrast, housekeeping genes such as rRNA processing, DNA repair and tRNA processing were significantly enriched in WT-specific highly variable genes, potentially due to a

higher cellular variation of cell activity in WT cells compared to CTCF-KD cells. We further found cellular variation-increased genes were significantly enriched in down-regulated genes, indicating CTCF knockdown simultaneously reduced the expression levels and increased the expression noise of its regulated genes. To our knowledge, this is the first attempt to explore genome-wide landscape of cellular variation after CTCF knockdown. Our study not only advances our understanding of CTCF function in maintaining gene expression and reducing expression noise, but also provides a framework for examining gene function.

Methods

Cell culture

EL4 cells were cultured in DMEM (GIBCO Invitrogen) supplemented with 50 IU/mL penicillin, 50 mg/mL streptomycin (GIBCO Invitrogen) and 10% heat-inactivated calf serum (Sigma, USA). Cultures were maintained by replacement of fresh medium every 3 days, and cell density was kept between 1×10^5 and 1×10^6 cells/mL.

Knockdown of CTCF by shRNA

Knockdown of CTCF was performed using Lentiviral-mediated short hairpin RNA (shRNA) in EL4 cells as described previously [17]. Briefly, 293T cells were co-transfected with an envelope plasmid (pLP/VSVG) to generate lentiviral particles. The medium containing lentiviral particles was harvested after 48 hours transfection. EL4 cells were infected with the harvested shLuc and shCTCF retroviral particles packaged in GP2-293. Some GFP+ cells were sorted out to check the knockdown efficiency using RT-qPCR and Western blotting after 5 days of infection. The cell populations displaying efficient knockdown of CTCF were used for single cell RNA-seq.

The following shRNA sequences were used for CTCF knockdown: mouse CTCF-shRNA 1: 5'-GGTGCAATTGAGAACATTATA; mouse CTCF-shRNA 2: 5'-TGGACGATACCCAGATCATAA.

Western blot analyses

After thorough washing, the knockdown (shCTCF) cells and control (shLuc) were harvested for Western blotting analyses. Protein concentration of the cell lysates was measured using BCA kit (Pierce, Rockford, IL). Protein samples (40μg/lane) were applied to SDS-PAGE followed by Western blotting against anti-CTCF antibody (07-729, Millipore), and anti-GAPDH antibody (sc-1616, Santa Cruz Biotechnology).

Quantitative Real-time PCR

Total RNAs from the knockdown (shCTCF#1 and shCTCF#2) and control (shLuc) cells were extracted using miRNeasy Micro Kit (QIAGEN). cDNA was synthesized by using oligo (dT)20 and SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturers' instructions. RT-qPCR samples were mixed with the following Taqman probes mixture (Applied Biosystems) and run on a LightCycler 96 (Roche): Gapdh: Mm03302249_g1; CTCF: Mm00484027_m1. Results were normalized to the mRNA level of Gapdh.

Single cell RNA sequencing

Fluidigm C1^{TM} Single-Cell Autoprep System (Fluidigm, South San Francisco, CA, USA) was used for single cell RNA-seq. In the initial experiment, shLuc#1 or shCTCF#1 were uploaded to a C1 integrated fluidics circuit (IFC) for cell capture, respectively. We checked the IFC to count the number of captured cells, and to distinguish between live and dead cells for later data processing. After successful completion of the second knockdown, cells from shLuc#2 and shCTCF#2 were treated similarly to the initial experiment. Single-cell RNA-seq with SMARTer protocol (Clontech, Mountain View, CA, USA) was prepared following Fluidigm manual 'Using the C1 Single-Cell Auto Prep System to generate mRNA from Single Cells and Libraries for Sequencing'. The wells containing either zero or double cells were filtered out. We selected 24 cells with the highest quality from each IFC. The DNA materials obtained from the 96 single cells were sequenced on Illumina HiSeq 3000, as illustrated in Fig. 1C.

Reads mapping and quality control

Quality of the reads was assessed using FASTQC. All reads were aligned to the mouse genome (Ensemble version GRCm38.89) utilizing STAR v.2.5.2 [28, 29]. Unique mapping reads were allowed (using parameter *--outFilterMultimapNmax*). The alignments were used as input in HTSEQ v.0.9.1 [30] to count the number of reads mapping to each of the 24,057 ref-seq genes in each cell. We filtered out those low-quality cells from our dataset based on a threshold for a minimum of 3000 unique genes per cell. Transcripts per million (TPM) was used to normalize the gene expression level and log2 transformed. Furthermore, genes with log2(TPM+1)>1 in less than two individual cells were filtered out, leaving a total of 95 samples and 11,361 genes for further analyses.

Statistical analyses and gene enrichment analyses

For identification of genes with biologically significant cell-to-cell variation, we used $\eta^2 = \sigma^2/\mu^2$ (σ denotes standard deviation; μ denotes mean) to measure the noise of gene expression [26, 27, 31]. We filtered out those genes exhibiting a low expression ($\log_2 TPM \leq 1$), since the expression noise is inversely proportional to the expression when gene expression level is low ($\log_2 TPM \leq 1$) (Fig. 3A), leading to 7,843 genes remaining for further analyses. To examine any possible enrichment of particular gene categories and pathways in certain gene lists, GO enrichment analysis was performed using DVAID [32, 33]. GO categories with Benjamini < 0.05 were considered as significant.

In this study, we used coefficient of variation (CV) to calculate the cell-to-cell variation. The variation increased genes is calculated by

$$CV_{KD} - CV_{WT} > mean_{CV_{KD}-CV_{WT}} + sd_{CV_{KD}-CV_{WT}}$$

while the variation decreased genes is calculated by

$$CV_{KD} - CV_{WT} < mean_{CV_{KD}-CV_{WT}} - sd_{CV_{KD}-CV_{WT}}$$

Acknowledgements

computation was supported by Center for Computational Science and Engineering of Southern University of Science and Technology.

Author's contributions

W.J. conceived the project. R.G. performed the experiments and W.W. analyzed the data. N.H. and W.J. supervised the study. W.W., W.J and N.H. wrote the manuscript.

Data Availability

The single cell RNA-seq data of EL4 cells have been deposited in the Gene Expression Omnibus database with accession number GSE135769.

Conflict of Interests

The authors declare no conflict of interests.

References

1.   Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007, 128(6):1231-1245.

2.   Ong CT, Corces VG: CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 2014, 15(4):234-246.

3.   Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH: A novel sequence-specific DNA binding protein which

interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene 1990, 5(12):1743-1753.

4.   Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanenkov VV: 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 1993, 13(12):7612-7624.

5.   Chung JH, Bell AC, Felsenfeld G: Characterization of the chicken beta-globin insulator. Proc Natl Acad Sci U S A 1997, 94(2):575-580.

6.   Bell AC, Felsenfeld G: Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 2000, 405(6785):482-485.

7.   Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G: CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell 2004, 13(2):291-298.

8.   Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K: Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res 2009, 19(1):24-32.

9.   Van Bortle K, Ramos E, Takenaka N, Yang J, Wahi JE, Corces VG: Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res 2012, 22(11):2176-2187.

10.   Taslim C, Chen Z, Huang K, Huang TH, Wang Q, Lin S: Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. Nucleic Acids Res 2012, 40(11):4754-4764.

11.   Bonora G, Plath K, Denholtz M: A mechanistic link between gene regulation and genome architecture in mammalian development. Curr Opin Genet Dev 2014, 27:92-101.

12. Rao SSP, Huang SC, Glenn SHB, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID et al: Cohesin Loss Eliminates All Loop Domains. Cell 2017, 171(2):305-320.e324.

13. Rao SP, Huntley MH, Durand NC, Stavenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES et al: A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 2014, in-press(159):1-16.

14. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B et al: CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 2015, 163(7):1611-1627.

15. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F et al: CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 2011, 43(7):630-638.

16. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012, 485(7398):376-380.

17. Ren G, Jin W, Cui K, Rodrigez J, Hu G, Zhang Z, Larson DR, Zhao K: CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. Mol Cell 2017, 67(6):1049-1058.e1046.

18. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG: Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 2017, 169(5):930-944.e922.

19. Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IWF et al: Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A 2014, 111(3):996-1001.

20. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y et al: CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell 2015, 162(4):900-910.

21. Kang JY, Song SH, Yun J, Jeon MS, Kim HP, Han SW, Kim TY: Disruption of CTCF/cohesin-mediated high-order chromatin structures by DNA methylation downregulates PTGS2 expression. Oncogene 2015, 34(45):5677-5684.

22. Lang F, Li X, Zheng W, Li Z, Lu D, Chen G, Gong D, Yang L, Fu J, Shi P et al: CTCF prevents genomic instability by promoting homologous recombination-directed DNA double-strand break repair. Proc Natl Acad Sci U S A 2017, 114(41):10912-10917.

23. Shin JO, Lee JJ, Kim M, Chung YW, Min H, Kim JY, Kim HP, Bok J: CTCF Regulates Otic Neurogenesis via Histone Modification in the Neurog1 Locus. Mol Cells 2018, 41(7):695-702.

24. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T: CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep 2012, 2(2):345-357.

25. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010, 26(1):139-140.

26. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS: Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 2006, 441(7095):840-846.

27. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS: Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 2010, 329(5991):533-538.

28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013, 29(1):15-21.

29.   Dobin A, Gingeras TR: Mapping RNA-seq Reads with STAR. Curr Protoc Bioinformatics 2015, 51:11 14 11-19.

30.   Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015, 31(2):166-169.

31.   Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC et al: Accounting for technical noise in single-cell RNA-seq experiments. Nature methods 2013, 10(11):1093-1095.

32.   Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA: DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 2012, 28(13):1805-1806.

33.   Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 2009, 4(1):44-57.

Figure Legend

Fig 1. Knockdown of CTCF and schema of single cell sequencing.
A. Western blot analysis of CTCF in luciferase control (shLuc) and CTCF-KD cells (shCTCF#1 and shCTCF#2). EL4 cells were infected with retroviral particles encoding GFP and an shRNA targeting CTCF or a control sequence for 5 days.
B. Real-time quantitative PCR (RT-qPCR) analysis of CTCF expression in luciferase control (shLuc) and knockdown (shCTCF#1 and shCTCF#2) cells. The expression level of CTCF was normalized to GAPDH.
C. Schema of single cell RNA sequencing using Fluidigm C1 system.

Fig 2. There is systematic difference between CTCF-KD and WT cells.
A. No significant batch effect among the experimental repeats based on PCA analysis.
B. CTCF-KD cells were largely distinguishable from WT cells on PCA projection.
C. Distribution of individual WT cells and individual CTCF-KD cells on PC1.
D. Heatmap of differentially expressed genes (TOP 20) between WT cells and CTCF-KD cells.

Fig 3. Identification and analyses of genes showing cellular variation-changed after CTCF KD.
A. The relationship between expression level and noise level of reference genes. Genes with low cellular variation was used for further analyses.
B. Scatter plot of the cellular variation-changed genes after CTCF KD. Blue and red indicate the variation decrease and variation increase, respectively.
C. The top 15 gene categories enriched in variation increased gene.
D. The top 15 gene categories enriched in variation decreased gene.

Fig 4. Genes showing cellular variation change are more likely to be differentially expressed genes. Genes with expression decreased and cellular variation increased were significantly over-represented ($P=0.29\times10^{-23}$, χ^2 test). Genes with decreased cellular variation and increased expression level were significantly over-represented ($P=0.48\times10^{-23}$, χ^2 test).

17

Supplementary figures

Fig S1. High reproducibility of single cell RNA-seq data.
(A) Scatter plot showing the correlation of gene expressions between pooled single cells and bulk data in shLuc.
(B) Scatter plot showing the correlation of gene expression between shLuc #1 and shLuc#2.
(C) Scatter plot showing the correlation of gene expressions between pooled single cells and bulk data in CTCF-KD cells.
(D) Scatter plot showing the correlation of gene expressions between shCTCF#1 and shCTCF#2.

Fig S2. Analysis of differentially expressed genes between WT cells and CTCF KD cells.
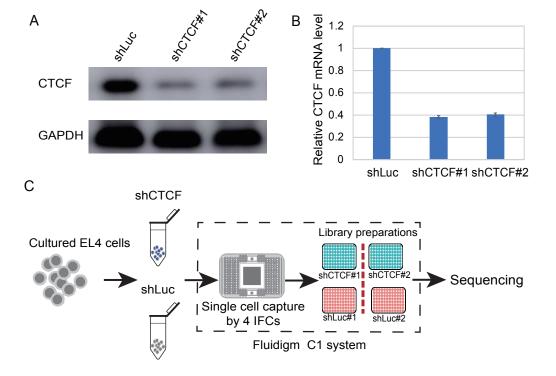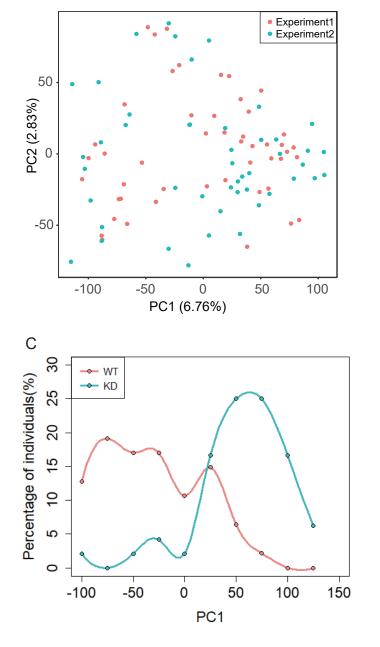(A) The expression of CTCF is correlated with the cell coordination on PC2.
(B) Differentially expressed genes were plotted in MAplot. Significantly up-regulated genes and down-regulated genes were indicated by red and blue, respectively.
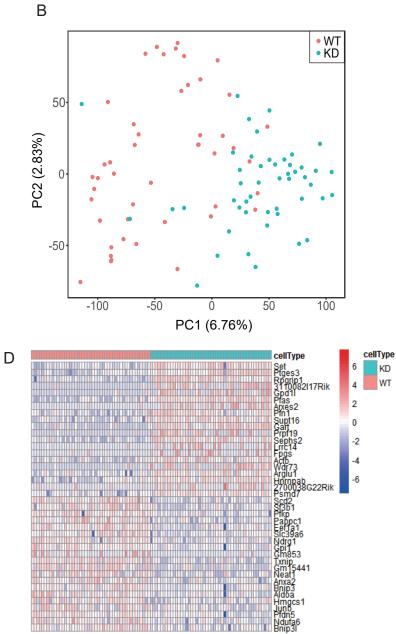(C) The top 15 enriched GO terms in the 195 up-regulated genes.
(D) The top 15 enriched GO terms in the 107 down-regulated genes.

Fig. S3. Histogram showing that CV difference of gene expression between CTCF-KD cells and WT cells followed normal distribution.