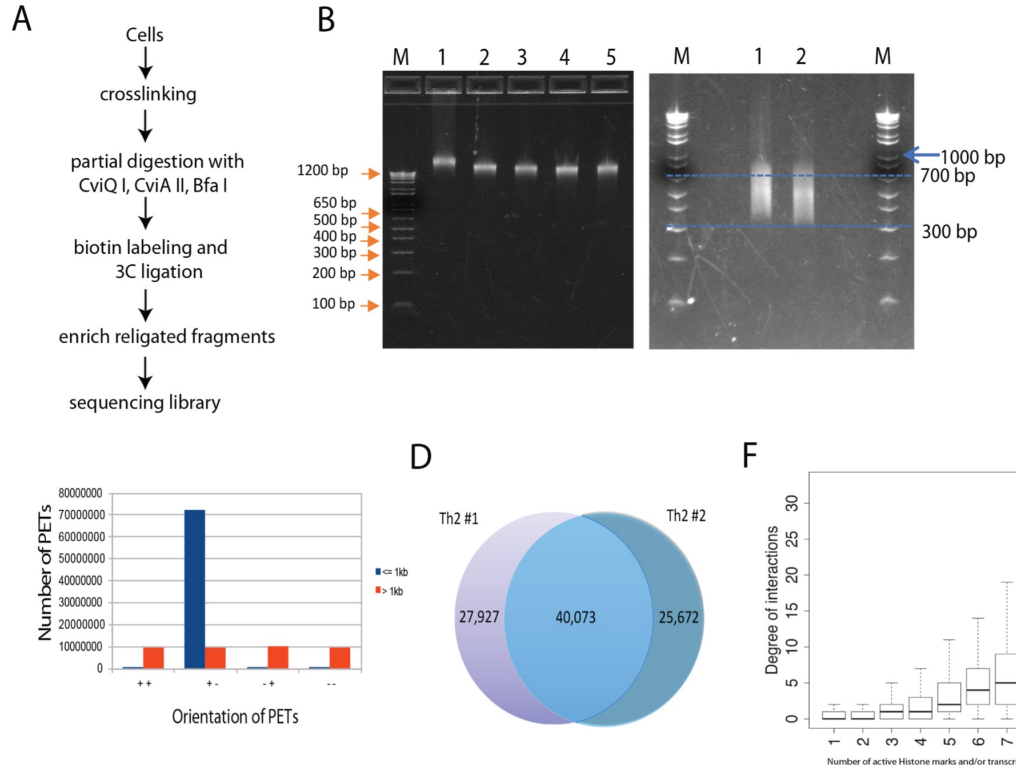
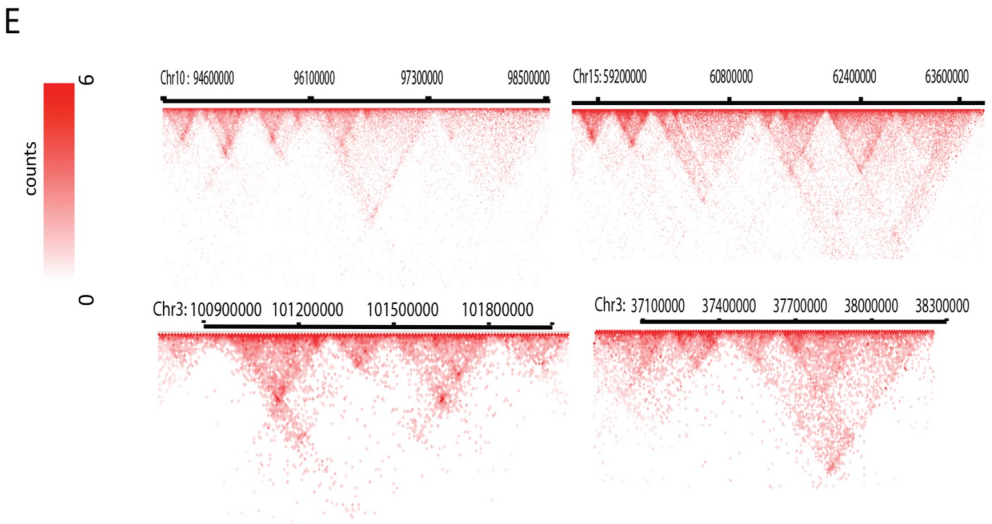
Supplemental Information

CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression

Gang Ren, Wenfei Jin, Kairong Cui Joseph Rodrigez, Gangqing Hu, Daniel R. Larson and Keji Zhao

Figure S1

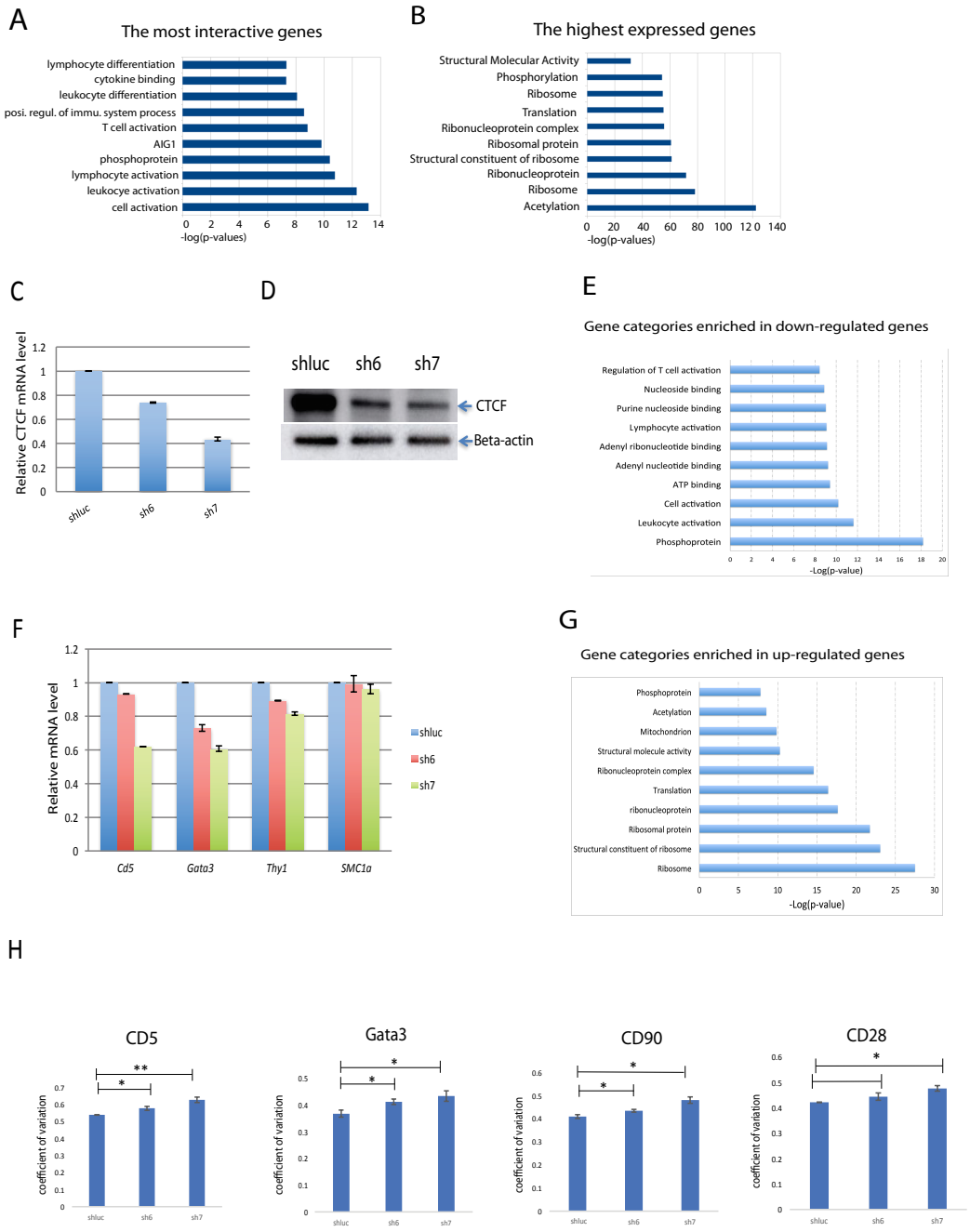

A
Cells
crosslinking
partial digestion with CviQ I, CviA II, Bfa I
biotin labeling and 3C ligation
enrich religated fragments
sequencing library
B
M 1 2 3 4 5
1200 bp
650 bp
500 bp
400 bp
300 bp
200 bp
100 bp
M 1 2 M
1000 bp
700 bp
300 bp
C
Number of PETs
80000000
70000000
60000000
50000000
40000000
30000000
20000000
10000000
0
++ +- -+ --
Orientation of PETs
<= 1kb
> 1kb
D
Th2 #1
Th2 #2
27,927
40,073
25,672
F
Degree of interactions
0 5 10 15 20 25 30
1 2 3 4 5 6 7 8 9
Number of active Histone marks and/or transcription factors


E
counts
6
0
Chr10: 94600000 96100000 97300000 98500000
Chr15: 59200000 60800000 62400000 63600000
Chr3: 100900000 101200000 101500000 101800000
Chr3: 37100000 37400000 37700000 38000000 38300000

Supplemental Figure S1. Related to Figure 1. Three enzyme Hi-C (3e Hi-C) is a reliable method to detect genome wide chromatin interaction.

A. Schema of 3e Hi-C method. Cross-linked cells were permeabilized and digested with a pool of three 4bp-cutters for 20 minutes. The DNA ends were labeled with biotin and purified for generation of sequencing libraries similar to Hi-C.
B. Agarose gel images showing the sizes of DNA fragments after three enzymes digestion (left panel) and the size range of DNA fragments isolated for sequencing by NGS after library preparation (right panel, 300bp to 700bp).
C. The orientation distribution of the pair end tags (PETs) from sequenced libraries (blue, <= 1kb; red, >1kb).
D. Venn diagram showing the Reproducibility of the promoter-enhancer interactions between two biological replicates in mouse Th2 cell.
E. 3e Hi-C interaction frequencies displayed as two-dimensional heat maps for the Chromosome10: 94600000–98500000, Chr15: 59200000–63600000, Chr3:100900000–101800000, and Chr3: 37100000–38300000 regions.
F. The degree of chromatin interaction at regulatory regions is correlated with the number of active histone modification marks and/or transcription factor binding. We examined the presence of 9 active marks and transcription factors (H3K27ac, H3K4me1, H3K4me2, H3K4me3, p300, Gata3, Stat6, CTCF and Cohesin) at each regulatory region and plotted the interaction density (Y-axis).

Figure S2

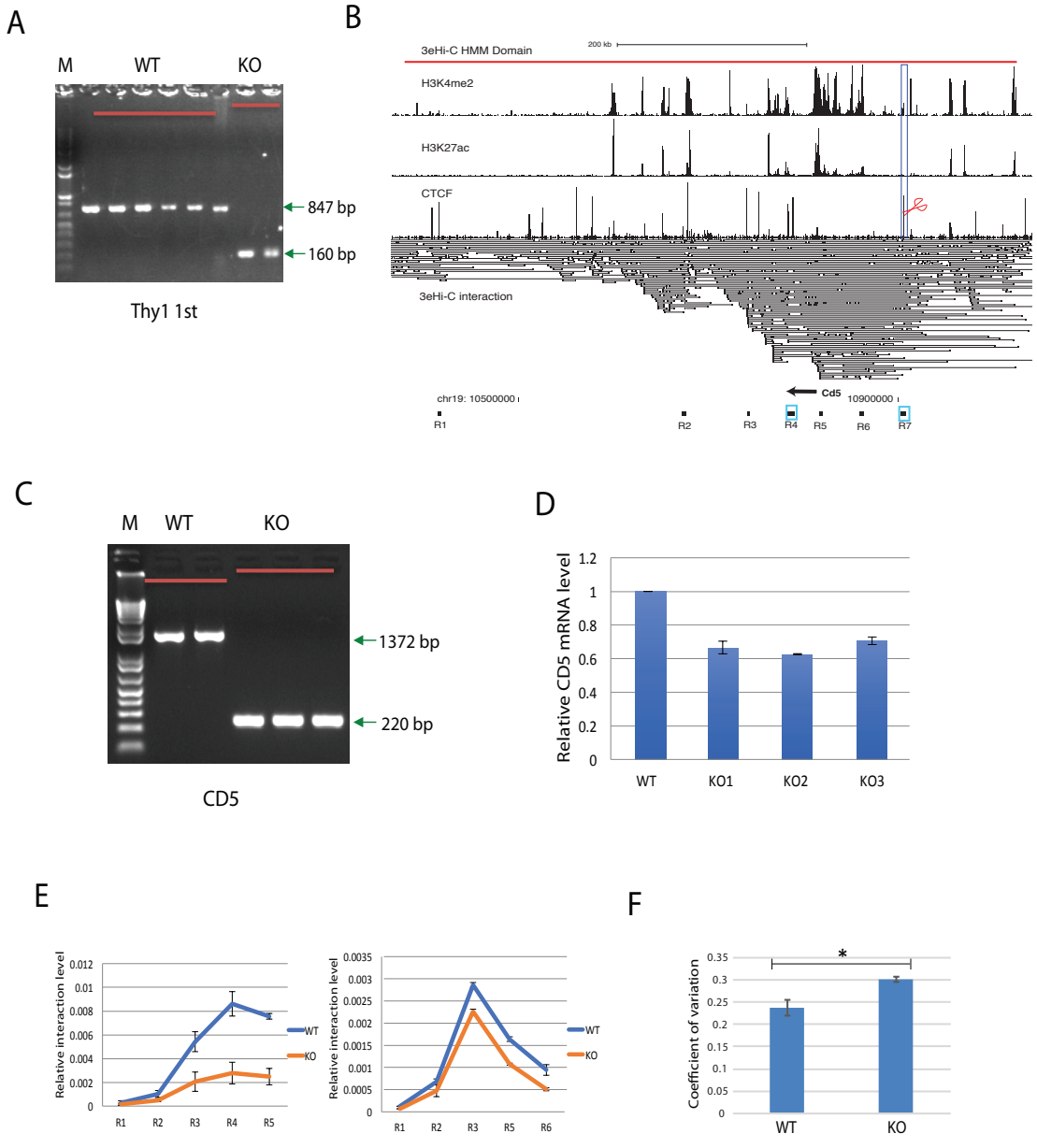

A
The most interactive genes
lymphocyte differentiation
cytokine binding
leukocyte differentiation
posi. regul. of immu. system process
T cell activation
AIG1
phosphoprotein
lymphocyte activation
leukocyte activation
cell activation
0 2 4 6 8 10 12 14
-log(p-values)
B
The highest expressed genes
Structural Molecular Activity
Phosphorylation
Ribosome
Translation
Ribonucleoprotein complex
Ribosomal protein
Structural constituent of ribosome
Ribonucleoprotein
Ribosome
Acetylation
0 20 40 60 80 100 120 140
-log(p-values)
C
Relative CTCF mRNA level
1.2
1
0.8
0.6
0.4
0.2
0
shluc
sh6
sh7
D
shluc sh6 sh7
CTCF
Beta-actin
E
Gene categories enriched in down-regulated genes
Regulation of T cell activation
Nucleoside binding
Purine nucleoside binding
Lymphocyte activation
Adenyl ribonucleotide binding
Adenyl nucleotide binding
ATP binding
Cell activation
Leukocyte activation
Phosphoprotein
0 2 4 6 8 10 12 14 16 18 20
-Log(p-value)
F
Relative mRNA level
1.2
1
0.8
0.6
0.4
0.2
0
Cd5
Gata3
Thy1
SMC1a
shluc
sh6
sh7
G
Gene categories enriched in up-regulated genes
Phosphoprotein
Acetylation
Mitochondrion
Structural molecule activity
Ribonucleoprotein complex
Translation
ribonucleoprotein
Ribosomal protein
Structural constituent of ribosome
Ribosome
0 5 10 15 20 25 30
-Log(p-value)
H
CD5
Gata3
CD90
CD28
coefficient of variation
shluc
sh6
sh7

Supplemental Figure S2. Related to Figure 2. Knockdown of CTCF results in decreased expression of some T lineage-specific genes and increased cell-to-cell variation of expression in mouse EL4 cells.

A. The genes with most long-distance chromatin interactions are highly enriched in immune functions. The gene ontology analysis was performed for 650 most interactive genes using DAVID. The top 10 enriched pathways and categories were plotted.
B. The highest expressed genes are enriched in house-keeping functions. The gene ontology analysis was performed for 650 highest expressed genes using DAVID. The top 10 enriched pathways and categories were plotted.
C. Quantitative real-time RT-PCR gene expression analysis of CTCF expression in control and knockdown (sh6, and sh7) EL4 cells. The EL4 cells were infected with retroviral particles encoding GFP and an shRNA targeting CTCF or a control sequence for 5 days. The GFP positive cells were sorted and analyzed for CTCF mRNA levels. The CTCF mRNA levels were normalized to GAPDH. Data shows average expression in two independent samples.
D. Western blotting analysis of CTCF in EL4 cells infected with shRNA retroviral particles targeting CTCF (sh6 and sh7) or the luciferase gene (shLuc).
E. Down-regulated genes are enriched in gene categories associated with T cell functions. RNA-Seq assays were performed to identify gene expression changes in the CTCF knockdown EL4 cells. Gene ontology analysis was performed using DAVID for down-regulated 819 genes.
F. The *Cd5*, *Thy1*, *Gata3*, and *Cohesin (SMC1a)* mRNA expression levels were measured in CTCF knockdown and control cells by real-time RT-PCR. mRNA levels were normalized to GAPDH. Data show average expression of two independent experiments and are represented as mean ± SEM.
G. Up-regulated genes are enriched in gene categories associated with house-keeping functions. RNA-Seq assays were performed to identify gene expression changes in the CTCF knockdown EL4 cells. Gene ontology analysis was performed using DAVID for up-regulated 879 genes.
H. Flow cytometry analysis revealed increased cell-to-cell variation of expression of CD5, Gata3, CD90 (encoded by *Thy1*) and CD28 in CTCF knockdown cells compared to the

control cells. Data show average of two independent experiments and are represented as mean ± SEM. The P-values of difference between the KO and control cells are indicated above the columns. *P<0.05, **P<0.01.

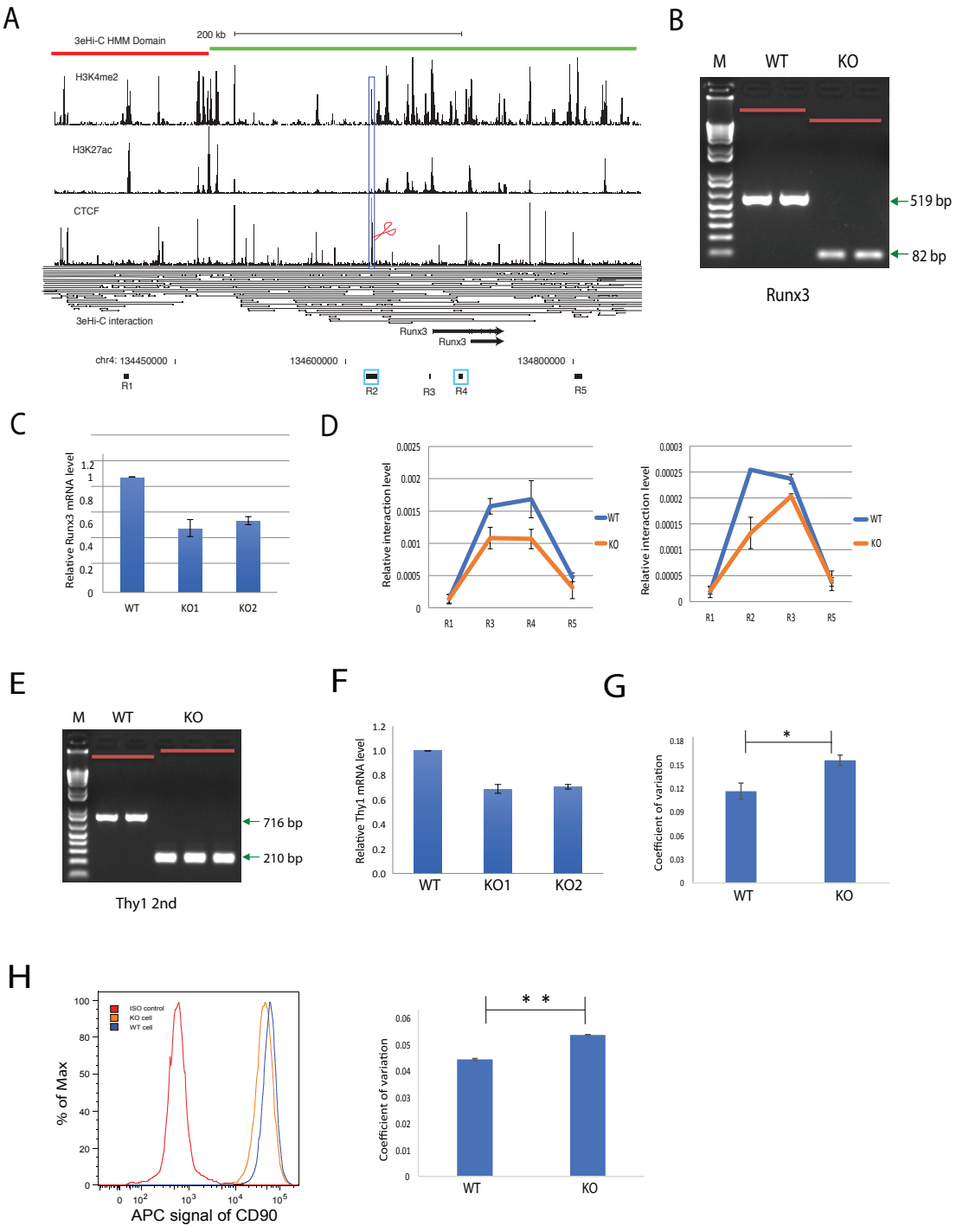
Figure S3


A
M
WT
KO
847 bp
160 bp
Thy1 1st
B
3eHi-C HMM Domain
200 kb
H3K4me2
H3K27ac
CTCF
3eHi-C interaction
chr19: 10500000
Cd5
10900000
R1
R2
R3
R4
R5
R6
R7
C
M
WT
KO
1372 bp
220 bp
CD5
D
Relative CD5 mRNA level
1.2
1
0.8
0.6
0.4
0.2
0
WT
KO1
KO2
KO3
E
Relative interaction level
0.012
0.01
0.008
0.006
0.004
0.002
0
R1
R2
R3
R4
R5
WT
KO
Relative interaction level
0.0035
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
R1
R2
R3
R5
R6
WT
KO
F
Coefficient of variation
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
*
WT
KO

Supplemental Figure S3. Related to Figure 3 and Figure 4. CTCF binding site at the *Cd5* gene locus contributes to the functional interaction between *Cd5* promoter with its enhancers and the expression noise control.

A. Genotyping of EL4 clones with deletion of the 1st CTCF binding site at the *Thy1* gene locus. PCR products of the expected sizes for both wild type and deletion clones are indicated on the right of the panel. Six wild type and two knockout clones are shown.
B. Genome browser image showing the TAD domain, ChIP-Seq signals for H3K4me2, H3K27ac and CTCF (top panels) at the *Cd5* gene locus. The 3e Hi-C-detected interactions are displayed below the ChIP-Seq tracks. The deleted CTCF site is highlighted by a blue box and scissors. The regions (R1 to R7) examined by the 3C assays are indicated below the interaction map.
C. Genotyping of EL4 clones with deletion of the CTCF binding site at the *Cd5* gene locus
D. Deletion of the CTCF binding site decreases *Cd5* mRNA levels. Total RNAs isolated from the wild type or CTCF site deletion EL4 clones were analyzed by quantitative reverse-transcription PCR and normalized to GAPDH. Three CTCF binding site deletion clones were analyzed. Data are represented as mean ± SEM.The CTCF site deletion compromised the enhancer-promoter interaction at the *Cd5* gene locus. The left panel shows the relative chromatin interaction intensity of the CTCF site with various enhancer regions indicated in panel B (R2 to R6 regions, R1 is a control region). The right panel shows the relative chromatin interaction intensity of the *Cd5* gene with various potential enhancer regions surrounding the gene. R7 is the anchor site for the left penal 3C analysis, R4 is the anchor site for the right penal 3C analysis.
E. Deletion of the CTCF site results in increased cell-to-cell variation in the number of *Cd5* mRNAs per cell by single-molecule RNA-FISH assay. Data show average of two independent experiments and are represented as mean ± SEM. The P-values of difference between the KO and control cells are indicated above the columns. **P*<0.05, ***P*<0.01.

Figure S4


A
3eHi-C HMM Domain
200 kb
H3K4me2
H3K27ac
CTCF
3eHi-C interaction
Runx3
Runx3
chr4: 134450000
134600000
134800000
R1
R2
R3
R4
R5
B
M
WT
KO
519 bp
82 bp
Runx3
C
Relative Runx3 mRNA level
WT
KO1
KO2
D
Relative interaction level
WT
KO
R1
R3
R4
R5
Relative interaction level
WT
KO
R1
R2
R3
R5
E
M
WT
KO
716 bp
210 bp
Thy1 2nd
F
Relative Thy1 mRNA level
WT
KO1
KO2
G
Coefficient of variation
*
WT
KO
H
ISO control
KO cell
WT cell
% of Max
APC signal of CD90
Coefficient of variation
* *
WT
KO

Supplemental Figure S4. Related to Figure 3 and figure 4. CTCF binding sites at the *Runx3 and Thy1* gene loci contribute to the functional interaction between promoters with their enhancers and noise control of gene expression.

A. Genome browser image showing the TAD domain, ChIP-Seq signals for H3K4me2, H3K27ac and CTCF (top panels) at the *Runx3* gene locus. The 3e Hi-C-detected interactions are displayed below the ChIP-Seq tracks. The deleted CTCF site is highlighted by a blue box and scissors. The regions (R1 to R5) examined by the 3C assays are indicated below the interaction map.
B. Genotyping of EL4 clones with deletion of the CTCF binding site at the *Runx3* gene locus.
C. Deletion of the CTCF binding site decreases *Runx3* mRNA levels. Total RNAs isolated from the wild type or CTCF site deletion EL4 clones were analyzed by quantitative reverse-transcription PCR and normalized to GAPDH. Two CTCF binding site deletion clones were analyzed. Data are represented as mean ± SEM.
D. The CTCF binding site deletion compromised the enhancer-promoter interaction at the *Runx3* gene locus. The left panel shows the relative chromatin interaction intensity of the CTCF site with various enhancer regions indicated in panel A (R2 to R5 regions, R1 is a control region). The right panel shows the relative chromatin interaction intensity of the *Runx3* gene promoter with various enhancer regions in this region. R2 is the anchor site for the left penal 3C analysis, R4 is the anchor site for the right penal 3C analysis. Data show average of two independent experiments and are represented as mean ± SEM.
E. Genotyping of EL4 clones with deletion of the 2nd CTCF binding site at the *Thy1* gene locus.
F. Deletion of the 2nd CTCF binding site of the *Thy1* locus modestly decreases *Thy1* mRNA levels. Total RNAs isolated from the wild type or CTCF site deletion EL4 clones were analyzed by quantitative reverse-transcription PCR assays and normalized to GAPDH. Two KO clones were analyzed. Data are represented as mean ± SEM.
G. Deletion of the 2nd CTCF binding site of the *Thy1* locus leads to increased cell-to-cell variation in the number of *Thy1* mRNAs per cell by single-molecule RNA-FISH assay. Data show average of two independent experiments and are represented as mean ± SEM.

The P-values of difference between the KO and control cells are indicated above the columns. *P<0.05, **P<0.01.

H. Deletion of the 2nd CTCF of the *Thy1* locus leads to increased cell-to-cell variation in the CD90 protein level. Left panels: Flow cytometry analysis shows the distribution of CD90 protein expression with the x-axis indicating the expression level and y-axis indicating the cell density. Right panel shows the increased coefficient of variation of CD90 in the 2nd CTCF binding site deletion cells as compared to the control cells. Data shows average of two independent experiments and are represented as mean ± SEM.

Supplemental Table S1. Related to Figure 1 and Figure 2.

Sheet 1: Sequencing information of the 3e Hi-C libraries for Th2 cells

Sheet 2: Public datasets used in this study

Sheet 3: The highest expressed genes in mouse Th2 cells

Sheet 4: List of oligonucleotides used in this study

Sheet 5: The most interactive genes in mouse Th2 cells

Sheet 6: The down-regulated genes in CTCF knockdown EL4 cells

Sheet 7: The up-regulated genes in CTCF knockdown EL4 cells

Supplemental Table S2. Related to Figure 3 and Figure 4.

Sheet 1 to 4: The APC fluoresce single values of CD5, CD90, Gata3 and CD28 in each of wild type and CTCF knockdown EL4 cells as measured by FACS

Sheet 5 and 6: The APC fluoresce single values of CD90 in each of *Thy1* 1st site, *Thy1* 2nd site KO EL4 cells and wild type control cells as measured by FACS

Sheet 7: The probe sequences used for detecting *Thy1*, *Cd5* and CTCF mRNA molecules by single-molecule RNA-FISH assays

Sheet 8, 9, and 10: Number of *Thy1*, *Cd5* and *CTCF* mRNA molecules in each of wild type and the 1st CTCF binding site deletion, the 2nd CTCF binding site deletion, and the *Cd5* locus CTCF binding site deletion EL4 cells measured by RNA-FISH assays