8-24-2017

In Situ Capture of Chromatin Interactions by Biotinylated dCas9.

Xin Liu

Yuannyu Zhang

Yong Chen
Rowan University

Mushan Li

Feng Zhou

See next page for additional authors

Recommended Citation

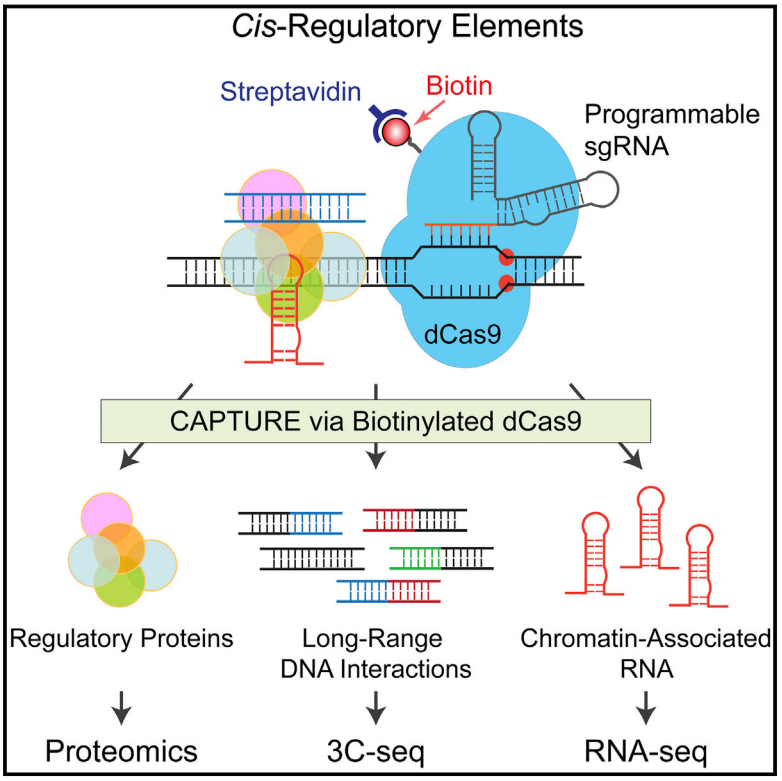
Xin Liu, Yuanyu Zhang, Yong Chen, Mushan Li, Feng Zhou, Kailong Li, Hui Cao, Min Ni, Yuxuan Liu, Zhimin Gu, Kathryn E. Dickerson, Shiqi Xie, Gary C. Hon, Zhenyu Xuan, Michael Q. Zhang, Zhen Shao, Jian Xu. (2017) In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell 170 (5), 1028-1043.e19.

Resource

Cell

In Situ Capture of Chromatin Interactions by Biotinylated dCas9

Graphical Abstract



Highlights

- dCas9 capture enables analysis of locus-specific chromatin-regulating proteome
- dCas9 capture identifies high-resolution locus-specific long-range DNA interactions
- Capture of disease-associated CREs uncovers mechanisms for β-globin disorders
- Capture of super-enhancers reveals composition-based hierarchical structure

Authors

Xin Liu, Yuannyu Zhang, Yong Chen, ..., Michael Q. Zhang, Zhen Shao, Jian Xu

Correspondence

zhou_feng@fudan.edu.cn (F.Z.), jian.xu@utsouthwestern.edu (J.X.)

In Brief

A biotinylated dCas9-based method for simultaneously studying long-range DNA interactions and chromatin-associated proteins at any genomic location.

Liu et al., 2017, Cell *170*, 1028–1043
August 24, 2017
http://dx.doi.org/10.1016/j.cell.2017.08.003

CellPress

Cell

Resource

In Situ Capture of Chromatin Interactions by Biotinylated dCas9

Xin Liu,[1,7] Yuannyu Zhang,[1,7] Yong Chen,[2,7] Mushan Li,[3,7] Feng Zhou,[4,7,*] Kailong Li,[1] Hui Cao,[1] Min Ni,[1] Yuxuan Liu,[1] Zhimin Gu,[1] Kathryn E. Dickerson,[1] Shiqi Xie,[5] Gary C. Hon,[5] Zhenyu Xuan,[2] Michael Q. Zhang,[2,6] Zhen Shao,[3] and Jian Xu[1,8,*]

[1]Children's Medical Center Research Institute, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[2]Department of Biological Sciences, Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080, USA

[3]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[4]Liver Cancer Institute, Zhongshan Hospital, Key Laboratory of Carcinogenesis and Cancer Invasion, Minister of Education, and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

[5]Cecil H. and Ida Green Center for Reproductive Biology Sciences, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[6]MOE Key Laboratory of Bioinformatics; Bioinformatics Division and Center for Synthetic and Systems Biology, TNLIST; Department of Automation, Tsinghua University, Beijing 100084, China

[7]These authors contributed equally

[8]Lead Contact

*Correspondence: zhou_feng@fudan.edu.cn (F.Z.), jian.xu@utsouthwestern.edu (J.X.)

http://dx.doi.org/10.1016/j.cell.2017.08.003

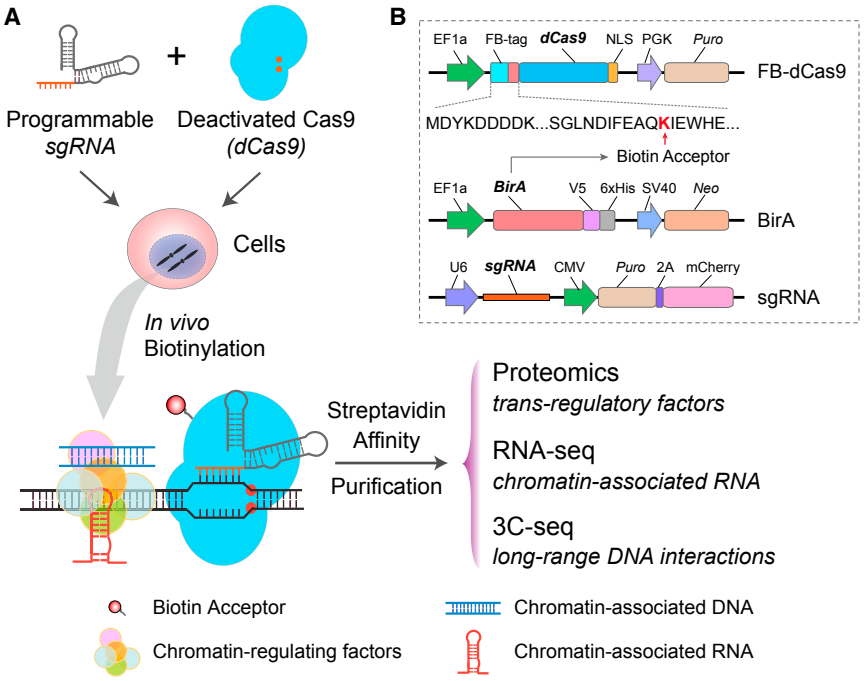
SUMMARY

Cis-regulatory elements (CREs) are commonly recognized by correlative chromatin features, yet the molecular composition of the vast majority of CREs in chromatin remains unknown. Here, we describe a CRISPR affinity purification *in situ* of regulatory elements (CAPTURE) approach to unbiasedly identify locus-specific chromatin-regulating protein complexes and long-range DNA interactions. Using an *in vivo* biotinylated nuclease-deficient Cas9 protein and sequence-specific guide RNAs, we show high-resolution and selective isolation of chromatin interactions at a single-copy genomic locus. Purification of human telomeres using CAPTURE identifies known and new telomeric factors. *In situ* capture of individual constituents of the enhancer cluster controlling human β-globin genes establishes evidence for composition-based hierarchical organization. Furthermore, unbiased analysis of chromatin interactions at disease-associated *cis*-elements and developmentally regulated super-enhancers reveals spatial features that causally control gene transcription. Thus, comprehensive and unbiased analysis of locus-specific regulatory composition provides mechanistic insight into genome structure and function in development and disease.

INTRODUCTION

Temporal and tissue-specific gene expression depends on *cis*-regulatory elements (CREs) and associated *trans*-acting factors. In contrast to protein-coding genes, our understanding of *cis*-regulatory DNA is very limited. Analyses of the human epigenome have revealed more than one million DNase I hypersensitive sites (DHS), many of which act as transcriptional enhancers (Thurman et al., 2012); however, the regulatory composition of the vast majority of these elements remain unknown, largely due to the limitations of the technologies previously employed to study CREs.

Cis-regulatory DNA is bound and interpreted by protein and RNA complexes and is organized as a 3D structure through long-range chromatin interactions. Identifying the complete composition of a specific CRE *in situ* can provide unprecedented insight into the mechanisms regulating its activity. However, purifying a small chromatin segment from the cellular milieu represents a major challenge—the protein complexes isolated with the targeted chromatin constitute only a small fraction of the co-purified proteins, most of which are non-specific associations. As such, major challenges have limited the application of existing approaches in purifying a specific genomic locus. Chromatin immunoprecipitation (ChIP) assays have provided crucial insights into the genome-wide distribution of TFs and histone marks, but it relies on *a priori* identification of molecular targets and is confined to examining single TFs. Targeted purification of genomic loci with engineered binding sites has been employed to identify single locus-associated proteins, yet it requires knockin gene targeting, which remains inefficient. DNA sequence-specific molecules such as locked nucleic acids (LNAs) (Déjardin and Kingston, 2009) and transcription activator-like (TAL) proteins (Fujita et al., 2013) have been used to enrich large chromatin structures, but these approaches do not enrich for a single genomic locus and cannot be adapted for multiplexed applications. The development of the CRISPR system containing an inactive Cas9 nuclease facilitated sequence-specific enrichment of native genomic regions (Fujita and Fujii, 2013; Waldrip et al., 2014);



G

Protein	Ratio	Function
APEX1	5.33	
POT1	1.99	Telomere Maintenance
TERF2	4.37	
TERF2IP	2.09	
MSH2	9.62	DNA Damage Repair
RAD50	2.39	
TP53BP1	1.57	
CDC73	1.89	Cell Cycle
MCM5	2.22	
PRIM1	10.28	
ATRX	10.74	Chromatin
CBX5	2.31	

Figure 1. *In Situ* Capture of Locus-Specific Chromatin Interactions by Biotinylated dCas9

(A) Schematic of dCas9-mediated capture of chromatin interactions.

(B) The three components of the CAPTURE system: a FB-dCas9, a biotin ligase BirA, and target-specific sgRNAs.

(C) Schematic of dCas9-mediated capture of human telomeres.

(D) Labeling of human telomeres in MCF7 cells. Scale bar, 5 μm.

(E) qPCR analysis shows significant enrichment of telomere DNA. Results are mean ± SEM of three experiments and analyzed by two-tailed t test. $^{**}p < 0.01$.

(F) Western blot shows enrichment of TERF2 in sgTelomere-expressing but not control K562 cells with dCas9 alone (no sgRNA) or the non-targeting sgGal4.

(G) iTRAQ-based proteomic analysis of telomere-associated proteins. Representative proteins and the mean iTRAQ ratios are shown.

See also Table S3.

however, these studies were limited to antibody-based purification, whereas the genome-scale specificity and the utility in identifying the *cis*- and *trans*-regulatory components were not evaluated.

Here, we developed an approach to isolate CRE-regulating proteins and long-range DNA interactions by repurposing the CRISPR/Cas9 system. Using human telomeres, β-globin cluster, and embryonic stem cell (ESC) super-enhancers (SEs), we identified *trans*-acting proteins at a single genomic locus. By combining dCas9 capture with chromatin interaction assays, we revealed locus-specific DNA interactions critical for regulatory function. *In situ* capture of SE constituents and disease-associated *cis*-elements provides insight into composition-based hierarchical regulation. Hence, the unbiased analysis of CRE-regulating proteome and 3D interactome by *in situ* capture has the potential to uncover the causal relationship between organizational structure and transcriptional function in a mammalian genome.

RESULTS

In Situ Capture of Chromatin Interactions by dCas9-Mediated Affinity Purification

To facilitate the analysis of native CREs, we developed a method to isolate chromatin interactions *in situ* (Figure 1A). The core components of CRISPR include Cas9 and a single guide RNA (sgRNA), which serves to direct Cas9 to a target genomic sequence (Cong et al., 2013; Mali et al., 2013). We engineered an N-terminal FLAG and biotin-acceptor-site (FB)-tagged deactivated Cas9 (dCas9) (Figure 1B). Upon *in vivo* biotinylation of dCas9 by the biotin ligase BirA together with sequence-specific sgRNAs in mammalian cells, the genomic locus-associated macromolecules are isolated by high-affinity streptavidin purification. The purified protein, RNA, and DNA complexes are identified and analyzed by mass spectrometry (MS)-based proteomics and high-throughput sequencing for study of native CRE-regulating proteins, RNA, and long-range DNA interactions, respectively (Figure 1A).

This approach has several advantages:

(1) High sensitivity—the affinity between biotin and streptavidin with $K_d = 10^{-14}$ mol/L is >1,000-fold higher than antibody-mediated interactions (Kim et al., 2009a; Schatz,

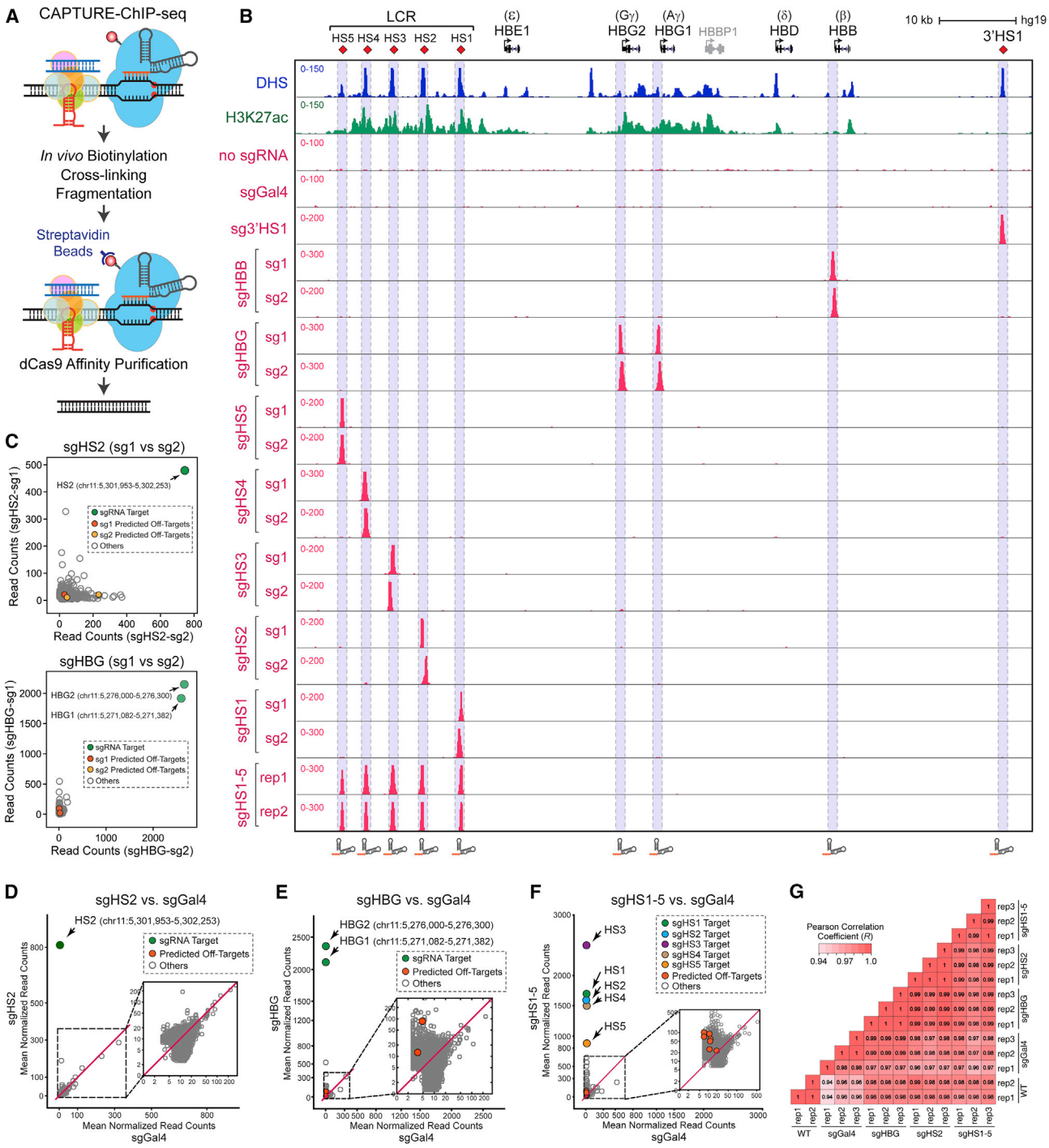


Figure 2. Biotinylated dCas9-Mediated Capture of the β-Globin Cluster

(A) Schematic of CAPTURE-ChIP-seq.

(B) Density maps are shown for CAPTURE-ChIP-seq at the β-globin cluster (chr11:5,222,500–5,323,700; hg19) in K562 cells, together with DHS and H3K27ac ChIP-seq profiles. Two independent sgRNAs (sg1 and sg2) or replicate experiments (rep1 and rep2) are shown. Cells expressing dCas9 only (no sgRNA) or dCas9 with sgGal4 were analyzed as controls.

(C) Genome-wide analysis of dCas9 binding in cells expressing two sgRNAs (sg1 and sg2) for HS2 or *HBG*. Data points for the sgRNA target regions and the predicted off targets are shown as green, red, and orange, respectively. The x and y axis denote the mean normalized read counts from N = 2 to 5 CAPTURE-ChIP-seq experiments.

(legend continued on next page)

1993), thus allowing for more efficient and stable capture of protein-DNA complexes.

(2) High specificity—this approach avoids using antibodies, which significantly reduces non-specific binding. In addition, the extraordinary stability of biotin-streptavidin allows for stringent purification to eliminate protein contamination.

(3) Adaptability for multiplexed approaches—the dCas9/sgRNA system can be manipulated by altering sgRNA sequences or combinations, thus allowing for medium- to high-throughput analysis of chromatin interactions.

Taken together, this new approach, which we named CAPTURE (CRISPR affinity purification *in situ* of regulatory elements), has the potential to expedite the analysis of chromatin-templated events by characterizing the entire set of interacting macromolecules and how composition changes during cellular differentiation.

In Situ CAPTURE of Human Telomeres

As a proof-of-principle, we used CAPTURE to isolate human telomeres in K562 cells (Figure 1C). We employed a validated telomere-targeting sgRNA (sgTelomere; Figure 1C) (Chen et al., 2013), which displayed specific labeling of telomeres by the dCas9-EGFP fusion protein, in contrast to the diffuse nucleolar localization of the non-targeting dCas9-EGFP (Figure 1D). Upon stable co-expression of sgTelomere and biotinylated dCas9, we observed significant enrichment of telomeric DNA (Figure 1E). The known telomere-associated protein TERF2 was highly enriched in sgTelomere-expressing samples but not in control samples expressing dCas9 alone (no sgRNA) or the non-targeting sgGal4 (Figure 1F). Most importantly, by iTRAQ-based proteomics, we identified many known telomere maintenance proteins (Déjardin and Kingston, 2009; Lewis and Wuttke, 2012) and new telomere-associated proteins (Figure 1G; Table S3).

In Situ CAPTURE of β-Globin Cluster

To validate the CAPTURE approach for identifying single copy CREs, we focused on the human β-globin cluster containing five β-like globin genes controlled by a shared enhancer cluster (locus control region [LCR]) with five discrete DHS (HS1–HS5). We designed two or three independent sgRNAs for each promoter (*HBG1*, *HBG2*, and *HBB*), enhancer (HS1–HS4), or insulator (HS5) (Tables S1 and S2). Upon co-expression of sgRNAs and dCas9, K562 chromatin was cross-linked and purified, followed by sequencing of the captured DNA (CAPTURE-ChIP-seq; Figure 2A). We observed specific and significant enrichment of discrete sgRNA-targeted regions (Figure 2B). For example, expression of two sgRNAs for HS1 (sgHS1-sg1 and sg2) led to significant enrichment of HS1 but no other enhancers. Because of the sequence similarity between *HBG1* and *HBG2*, the sgRNAs targeting *HBG* promoters (sgHBG-sg1 and sg2) do not distinguish the two genes. Consistently, co-expression of sgHBG and dCas9 resulted in significant enrichment of both *HBG* genes. In contrast, binding of dCas9 to β-globin cluster was undetectable when expressed alone (no sgRNA) or with the non-targeting sgGal4. Importantly, co-expression of five sgRNAs (sgHS1–5) led to simultaneous capture of all five LCR enhancers, demonstrating that the CAPTURE system can be adapted for multiplexed analysis of independent CREs. Furthermore, by comparing ChIP-seq intensity using two or three independent sgRNAs, we observed highly specific enrichment of each captured region with minimal off targets (Figures 2C and S1D). Given the consistent performance, hereafter we focus on one sgRNA (sg1; Table S2) for each region unless otherwise specified.

Genome-wide Enrichment and Specificity of CAPTURE

To identify locus-specific interactions, it is critical to evaluate the on-target enrichment and off-target effects. We first compared CAPTURE-ChIP-seq with dCas9 or FLAG antibody-based ChIP-seq using sgHS2 and sgHBG, and we observed significantly higher binding intensity by CAPTURE-ChIP-seq (Figure S1A; Table S1). Among the top 100 peaks by sgHS2, CAPTURE-ChIP-seq led to 18- or 284-fold on-target enrichment compared to dCas9 or FLAG-based ChIP-seq, respectively (Figure S1B). At the global scale, CAPTURE-ChIP-seq resulted in highly specific enrichment of HS2 or *HBG* with many fewer off targets than antibody-based ChIP-seq (Figure S1C). These results provide evidence that the CAPTURE approach allows for more efficient purification of targeted chromatin through improved on-target enrichment and elimination of potential off targets.

We next assessed the genome-wide specificity by comparing dCas9 binding in cells expressing target-specific sgRNAs or sgGal4. Specifically, recruitment of dCas9 by sgHS2 resulted in highly specific enrichment of HS2 with no additional significant dCas9 binding (Figure 2D). Similarly, recruitment of dCas9 by sgHBG led to specific enrichment of *HBG1* and *HBG2*, whereas none of the predicted off targets were significantly enriched (Figure 2E). Moreover, multiplexed capture by sgHS1–5 resulted in identification of LCR enhancers as the top enriched binding sites (Figure 2F). Similar results were obtained with 12 other sgRNAs (Figures S1D and S1E; Table S1). RNA-seq in target-specific sgRNAs, sgGal4, and wild-type (WT) K562 cells revealed minimal transcriptomic changes (Figures 2G and S1F). The expression of β-globin mRNAs remained unchanged (Figure S1G), suggesting that the dCas9 capture did not interfere with the expression of endogenous genes. Together, these analyses establish that the CAPTURE system is highly specific to target loci and can be used to isolate locus-specific regulatory components.

(D–F) Genome-wide differential analysis of dCas9 binding in cells expressing sgHS2, sgHBG, or sgHS1–5 versus sgGal4. Data points for the sgRNA target regions and the predicted off targets are shown as green and red, respectively. N = 5, 4, 6, and 4 CAPTURE-ChIP-seq experiments for sgHS2, sgHBG, sgHS1–5 and sgGal4, respectively.

(G) RNA-seq analysis was performed in cells expressing dCas9 with sgHS2, sgHBG, sgHS1–5, sgGal4, or WT K562 cells. The Pearson correlation coefficient (*R*) value is shown.

See also Figure S1 and Tables S1 and S2.

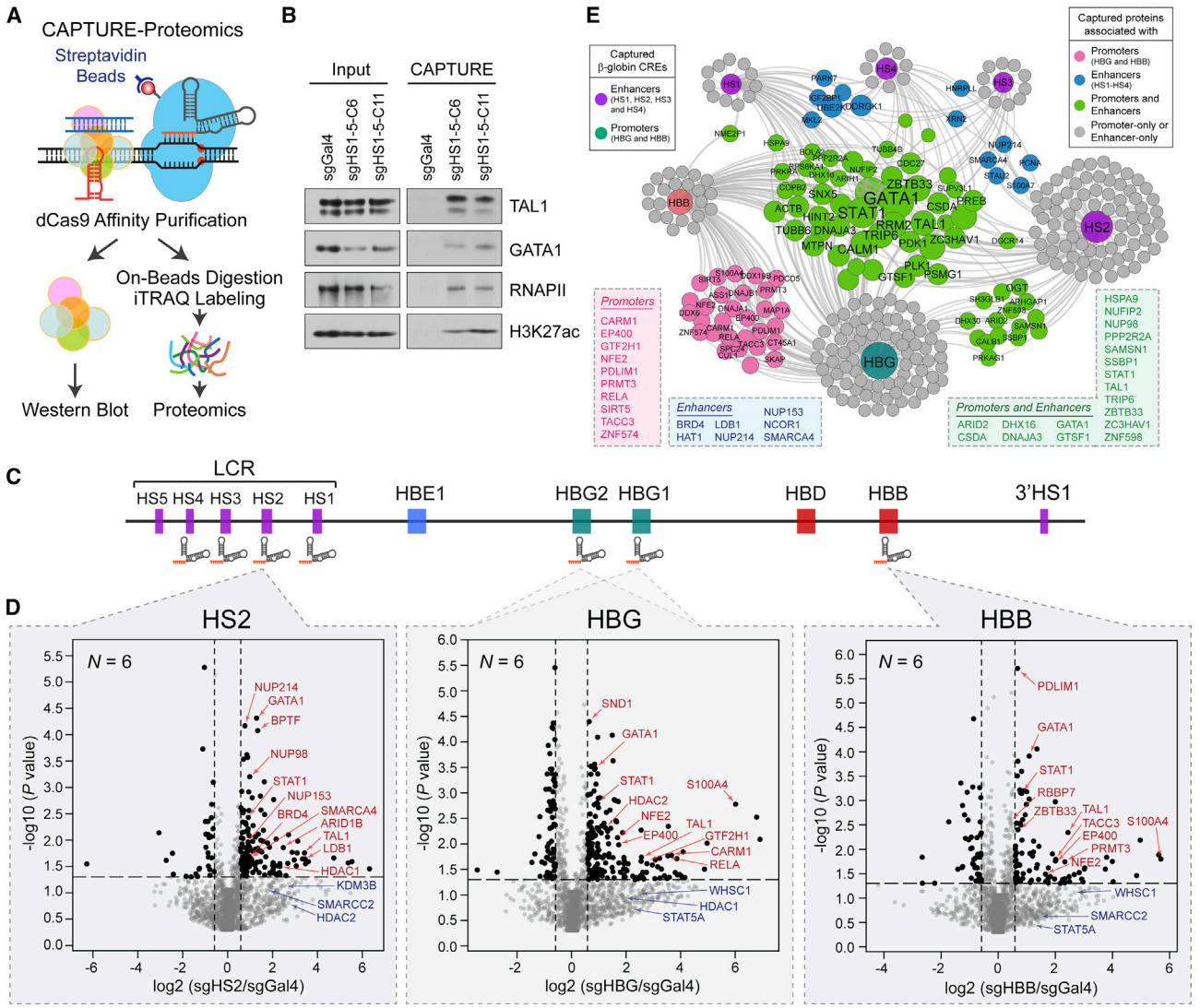


Figure 3. CAPTURE-Proteomics Identify β-Globin CRE-Associated Protein Complexes

(A) Schematic of CAPTURE-Proteomics.

(B) Western blot analysis of captured proteins in sgHS1–5 or sgGal4-expressing K562 cells.

(C) Schematic of the β-globin cluster and sgRNAs used for CAPTURE-Proteomics.

(D) CAPTURE-Proteomics identified β-globin CRE-associated proteins. Volcano plots are shown for the iTRAQ proteomics of purifications in sgHS2, sgHBG, or sgHBB versus sgGal4-expressing cells. Relative protein levels in target-specific sgRNAs versus sgGal4 are plotted on the x axis as mean log2 iTRAQ ratios across *N* replicate experiments. Negative log10 transformed *P* values are plotted on the y axis. Significantly enriched proteins ($p \leq 0.05$, iTRAQ ratio ≥ 1.5) are denoted by black dots, all others by gray dots. Dotted lines indicate 1.5-fold ratio (x axis) and *P* value of 0.05 (y axis). Representative chromatin-regulating proteins are denoted by red arrowheads. Representative proteins with iTRAQ ratio ≥ 1.5 and $p > 0.05$ are denoted by blue arrowheads.

(E) Connectivity network of CAPTURE-Proteomics-identified proteins converged by β-globin CREs. The connectivity was built using interactions (gray lines) between proteins and CREs. Colored nodes denote proteins enriched at single or multiple CREs. Size of the circles denotes the frequency of interactions. Inset tables show the lists of representative proteins associated with the β-globin promoters (red), enhancers (blue), or both (green).

See also Figures S2 and S3 and Table S4.

CAPTURE-Proteomics Identify *Trans*-acting Regulators of β-Globin Genes

A major challenge for proteomic analysis of a single genomic locus is the need for a sufficient amount of purified proteins. Hence, we optimized several components of the procedures including protein purification, peptide isolation, and quantitative proteomic profiling, and we developed the "CAPTURE-Proteomics" approach to identify locus-specific protein complexes (Figures 3A and S2A). We first performed purification in control cell lines to categorize the endogenous biotinylated proteins and/or dCas9-associated non-specific proteins (Figure S2B). Specifically, we identified proteins purified from K562 cells expressing BirA only, BirA with dCas9, BirA with dCas9 and sgGal4, and BirA with dCas9 and β-globin CRE-specific sgRNAs

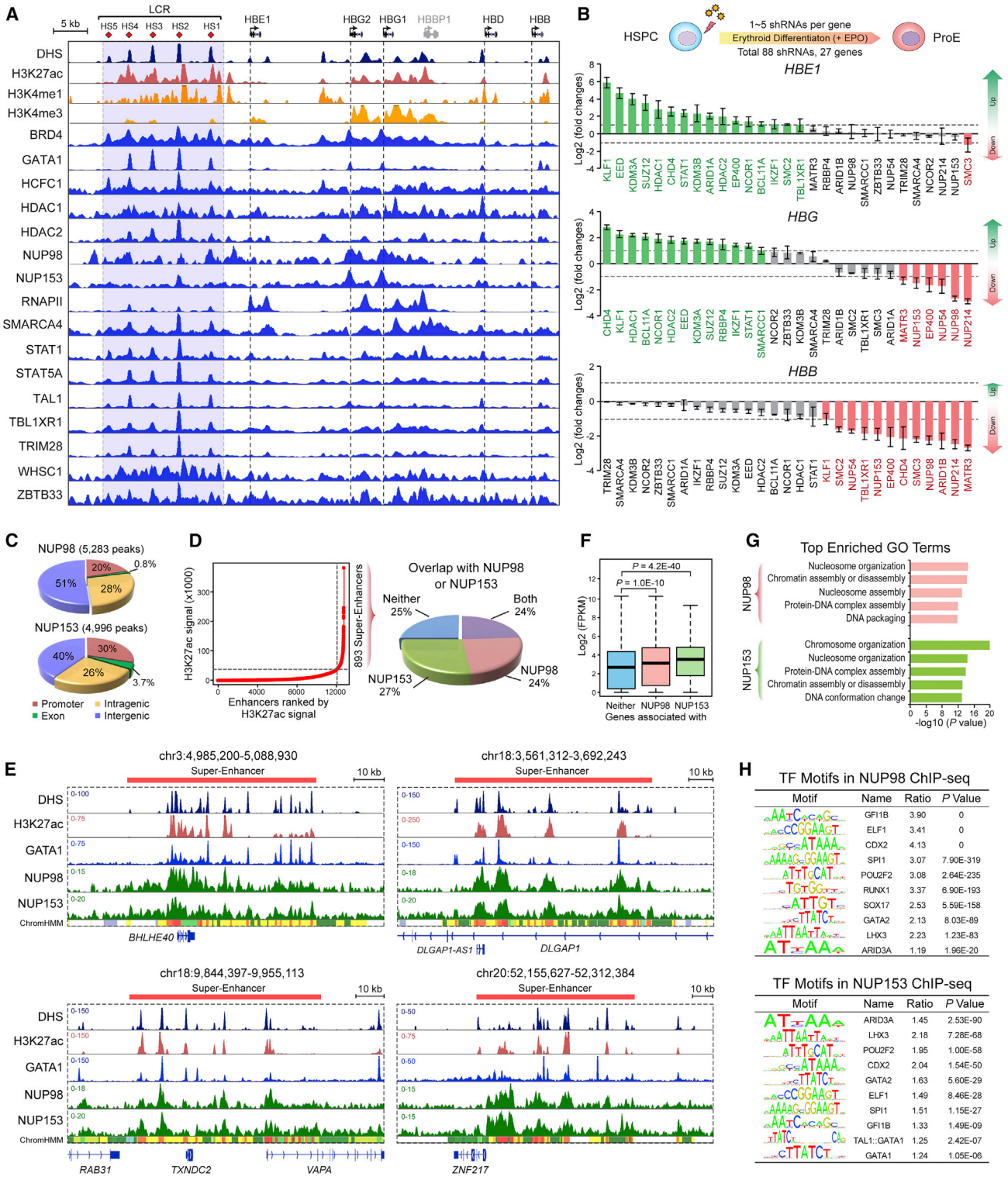


Figure 4. CAPTURE-Proteomics Identify Known and New Regulators of β-Globin Genes and Erythroid Enhancers

(A) ChIP-seq analysis of the identified regulators in K562 cells.

(B) RNAi screen of the identified regulators in human primary erythroid cells. Data are plotted as log2(fold change) of the β-globin mRNA in each shRNA experiment relative to the non-targeting shNT control. Genes are ranked based on the changes in *HBE1*, *HBG*, or *HBB* expression. shRNAs against *BCL11A* and *KLF1* were analyzed as controls. Results are mean ± SEM of all shRNAs for each gene from four experiments.

(legend continued on next page)

in which the endogenous β-globin cluster was deleted (BirA-dCas9-sgAll-Globin-KO; Method Details). Compiled from three experiments, we identified 304–468 proteins from individual controls, including 277 ''high-confidence non-specific proteins'' present in all controls (Figure S2B; Table S4).

We next determined whether known β-globin regulators can be isolated. Co-expression of dCas9 with sgHS1–5 led to significant enrichment of the erythroid TFs (GATA1 and TAL1) required for globin enhancers, together with RNA polymerase II (RNAPII) and acetylated H3K27 (H3K27ac) (Figure 3B). We then performed iTRAQ-based quantitative proteomics of captured β-globin CREs (Figure 3C). Relative protein abundance associated with the captured CRE versus sgGal4 was determined by the ratio of the iTRAQ reporter ion intensity. The significance of enrichment (*P* value) for each protein was calculated by t test of the $\log_2$ iTRAQ ratios in replicate experiments. We surveyed the distribution of high-confidence non-specific proteins in all experiments and observed that 78.3% and 79.8% of them had iTRAQ ratio < 1.5 and *P* value > 0.05 (Figure S2C). Therefore, we employed the iTRAQ ratio ≥ 1.5 and *P* value ≤ 0.05 as the cutoffs and identified 25–164 candidate locus-specific proteins (Figures 3D, S2D, and S2E; Table S4).

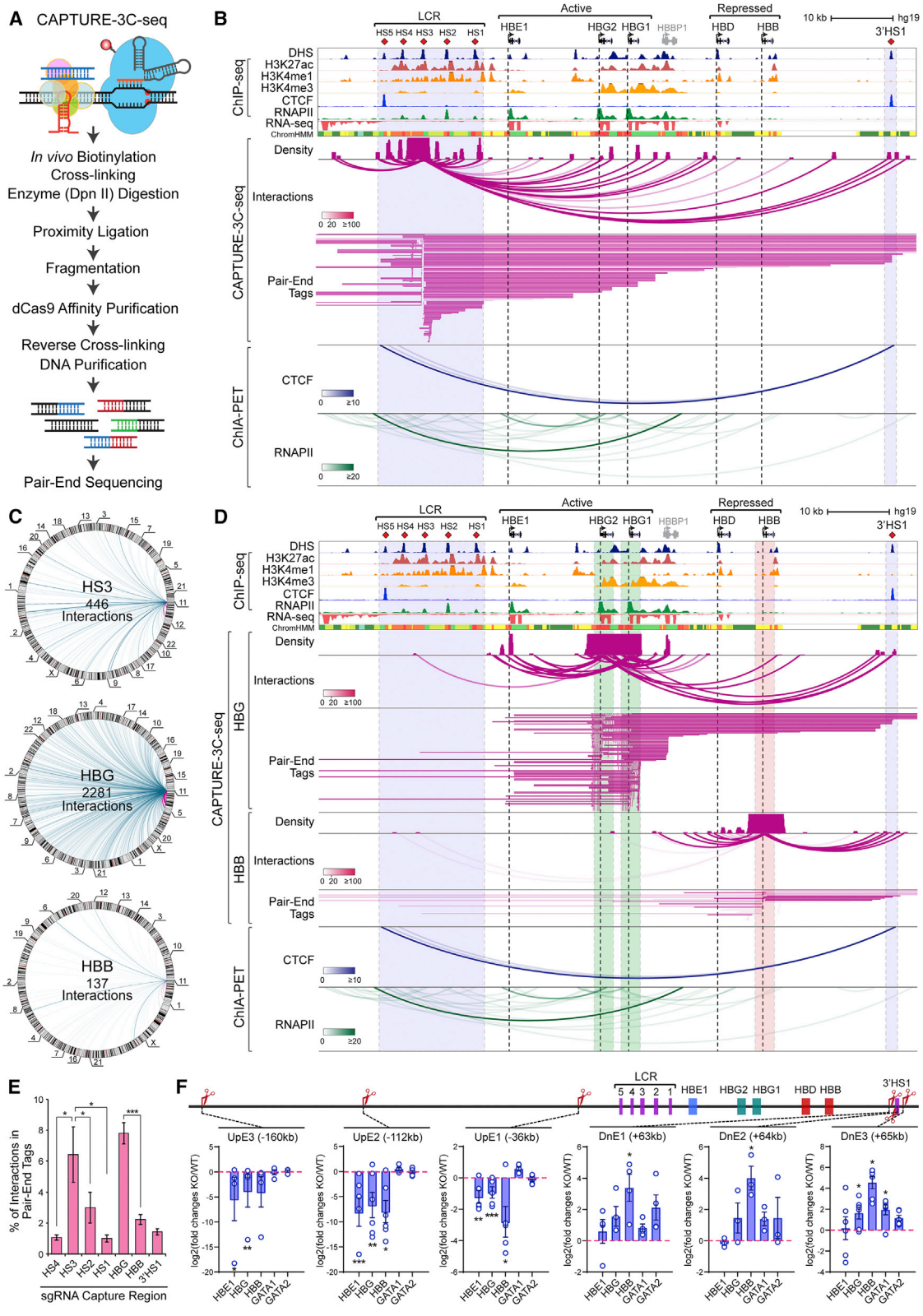
Using CAPTURE-Proteomics, we identified many known factors including GATA1, TAL1, NFE2, components of the SWI/SNF (ARID1A, ARID1B, SMARCA4, and SMARCC1), and NuRD (CHD4, RBBP4, RBBP7, HDAC1, and HDAC2) complexes (Kim et al., 2009b; Miccio and Blobel, 2010; Xu et al., 2013) at β-globin CREs. More importantly, by locus-specific proteomics, we identified new β-globin CRE-associated complexes including the nucleoporins (NUP98, NUP153, and NUP214), components of the large multiprotein nuclear pore complexes (NPCs), at LCR enhancers (Figures 3D and 3E). In addition, BRD4 and LDB1 were identified at LCR enhancers, whereas the NuA4 acetyltransferase (EP400) and transcriptional initiation complex (GTF2H1) were found at β-globin promoters. Furthermore, we observed that the *HBG* and *HBB* promoters shared many interacting proteins and clustered closely in protein-DNA connectivity networks (Figure 3E, S3A, and S3B). By contrast, the distal enhancers (HS1, HS3, and HS4) clustered together to form a distinct subdomain through enhancer-associated proteins, whereas HS2 shared interacting proteins with both subdomains. These analyses provide initial evidence for the composition-based hierarchical organization of the β-globin CREs.

Identification of New Regulators of β-Globin Genes and Erythroid Enhancers

We validated the binding of a subset of the identified proteins in K562 cells by ChIP-seq (Figure 4A; Table S1). Importantly, among the factors not previously implicated in β-globin regulation, we confirmed the nucleoporins (NUP98 and NUP153), STAT proteins (STAT1 and STAT5A), TBL1XR1, HCFC1, TRIM28/KAP1, WHSC1/NSD2, and ZBTB33/KAISO to be significantly enriched at one or multiple LCR enhancers by CAPTURE-Proteomics and ChIP-seq. To establish the functional roles, we performed RNAi-mediated loss-of-function analysis in human primary erythroid cells (Figures 4B, S3G, and S3H; Table S2). Specifically, depletion of 17 of 27 factors led to significant upregulation or downregulation of *HBG* (≥2-fold; Figure 4B). Similarly, depletion of 15 or 11 of 27 factors led to significant changes in *HBB* or *HBE1* (≥2-fold), respectively. Notably, depletion of NUP98, NUP153, and NUP214 led to marked downregulation of *HBG* (2.8- to 7.3-fold) and *HBB* (3.3- to 5.6-fold), suggesting that the NUP proteins are directly or indirectly required for the activation of β-globin genes.

The peripheral NUPs, including NUP98, NUP153, and NUP214, extend from the membrane-embedded NPC scaffold to regulate nuclear trafficking. While a few NUPs were found to be associated with transcriptionally active genes or regulatory elements (Capelson et al., 2010; Ibarra et al., 2016; Kalverda et al., 2010), their roles in erythroid enhancers remained unknown. Hence, we performed NUP98 and NUP153 ChIP-seq in K562 cells and identified 5,283 and 4,996 binding sites in gene-proximal promoters and distal elements (Figure 4C). Notably, NUP98 and NUP153 binding sites are highly enriched at erythroid SEs (Figures 4D and 4E), associated with gene activation (Figure 4F), nucleosome organization, and DNA packaging (Figure 4G), highlighting their potential roles in regulating chromatin organization and/or enhancer activities. Moreover, NUP98/NUP153 binding sites are enriched for motifs associated with hematopoietic TFs, chromatin factors, and homeobox proteins (Figure 4H), suggesting that NUPs may cooperate with lineage TFs and chromatin regulators in gene transcription. Another identified protein BRD4 binds acetylated histones and plays a critical role in chromatin regulation. Inhibition of BRD4 by a small molecule JQ1 abrogates its function (Filippakopoulos et al., 2010). BRD4 and related BET proteins (BRD2 and BRD3) are required for globin gene transcription in mouse erythroid cells (Stonestrom et al., 2015). Consistently, inhibition of BET proteins by JQ1 in human erythroid cells significantly decreased β-globin mRNAs and BRD4 occupancy without apparent effects on erythroid differentiation (Figures S3C–S3F). Together, these results not only establish new regulators of β-globin enhancers but demonstrate the potential of the CAPTURE approach for unambiguous identification of protein complexes specifically associated with a single genomic locus, such as an enhancer, *in situ*.

(C) Genome-wide distribution of NUP98 and NUP153 ChIP-seq peaks in promoters (−2 kb–1 kb of TSS), exons, intragenic and intergenic regions.
(D) NUP98 and NUP153 associate with erythroid SEs. SEs were identified by ROSE (Whyte et al., 2013) using the H3K27ac ChIP-seq signal.
(E) Representative SE loci co-occupied by NUP98 and NUP153. DHS, ChIP-seq, and chromatin state (ChromHMM) data are shown. Red bars denote the annotated SEs.
(F) NUP98- and NUP153-associated genes show significantly higher mRNA expression. Boxes show median of the data and quartiles, and whiskers extend to 1.5× of the interquartile range. *P* values were calculated by a two-side t test.
(G) Enriched gene ontology (GO) terms associated with NUP98- or NUP153-occupied regions.
(H) Motif analysis of NUP98 or NUP153 binding sites.



(legend on next page)

Capture of Long-Range DNA Interactions by Biotinylated dCas9

Enhancers regulate designated promoters over distances by long-range DNA interactions, or chromatin loops. Long-range chromatin interactions have been observed by chromosome conformation capture (3C) (Dekker et al., 2002) and derivative methods including 4C (Simonis et al., 2006; Zhao et al., 2006), 5C (Dostie et al., 2006), and Hi-C (Lieberman-Aiden et al., 2009), as well as fluorescence *in situ* hybridization (FISH) (Osborne et al., 2004). However, these methods are either limited to pre-defined chromatin domains or of low resolution and lacking functional details. For large-scale, *de novo* analysis of chromatin interactions, the ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) approach has been developed (Fullwood et al., 2009; Li et al., 2012). While this method provides unprecedented insight into the principles of 3D genomic architectures, the reliance on specific target proteins and antibodies limits its application in studying a single genomic locus.

To overcome these limitations, we sought to combine chromatin interaction assays with the high-affinity dCas9 capture to unbiasedly identify single genomic locus-associated long-range interactions (CAPTURE-3C-seq; Figure 5A). Specifically, upon co-expression of dCas9 and sgRNAs, long-range chromatin interactions were cross-linked, followed by DpnII digestion and proximity ligation of distant DNA fragments. After fragmentation, locus-specific interactions were captured by dCas9 and analyzed by pair-end sequencing to identify the tethered long-range interactions. Of note, this approach does not involve any pre-selection steps such as PCR-based amplification (Simonis et al., 2006; Zhao et al., 2006) or oligonucleotide-based capture (Hughes et al., 2014), and all interactions brought together by dCas9-tethered DNA were captured in a single experiment.

CAPTURE-3C-Seq of Locus-Specific DNA Interactions at β-Globin Cluster

Using this approach, we first identified long-range interactions at β-globin LCR by targeting dCas9 to HS3 (Figures 5B and 5C; Table S1). From 6,074 pair-end tags (PETs), we identified 446 long-range interactions, including 232 (52.0%) intra-chromosomal interactions, 208 (46.6%) interactions within 1 Mb from HS3, and 126 (28.3%) within the β-globin cluster (Table S5). To quantitatively analyze interactions, we employed the FDR-controlled Bayes factor (BF) to identify "high-confidence interactions" (Figures S4A and S4B; Method Details). Notably, the interaction frequencies were significantly higher between HS3 and the active genes (*HBG1* and *HBG2*) than the repressed gene (*HBB*), suggesting that the enhancer-promoter loop formation correlates with transcriptional activities. By comparing with CTCF and RNAPII ChIA-PET data (Consortium, 2012; Li et al., 2012), we identified CTCF- or RNAPII-mediated interactions and many new interactions (Figure 5B). By comparing the normalized number and frequency of interactions captured by CAPTURE-3C-seq, ChIA-PET, and Hi-C, we observed that CAPTURE-3C-seq displayed the highest percentage of unique PETs and on-target enrichment (Figure S4C). Compared to 4C-based approach (Schwartzman et al., 2016), CAPTURE-3C-seq displayed a higher percentage of unique PETs but comparable or slightly lower on-target enrichment (Figure S4C).

We then compared the long-range interactions at the active (*HBG*) and repressed (*HBB*) genes (Figure 5D). CAPTURE-3C-seq of *HBG* revealed 215 long-range interactions connecting with most of the β-globin CREs including HS3, *HBE1*, and 3′HS1. Notably, 164 of 215 (76.3%) interactions were between the active *HBG* and *HBE1* genes, whereas no interactions were detected between *HBG* and the repressed *HBB* or *HBD* gene, suggesting that the active genes are interconnected and coregulated through long-range DNA interactions. By contrast, the interactions at *HBB* were predominantly with the proximal *HBD* and 3′HS1.

In CAPTURE-3C-seq, it is critical to rule out that the difference in the position of sgRNA target sites may cause variations in capture efficiency. Therefore, we designed sgRNAs with varying distance to the DpnII site at HS2 or HS3 enhancer (Figure S5A). Importantly, sgRNAs at various positions consistently showed higher frequency of DNA interactions at HS3 than the neighboring HS2 enhancer (Figure S5B). Finally, we compared the interactions captured at discrete β-globin CREs and identified a high-resolution, locus-specific interaction map (Table S5; Figures 5E and S6). While some interactions were shared, most were specific to individual elements. Of note, while HS2, HS3, and HS4 are all required for β-globin gene activation (Fraser et al., 1993; Morley et al., 1992; Navas et al., 1998), HS2 and HS4 contained many fewer interactions than HS3 (Figure 5E, S5, and S6), suggesting that they may cooperate through distinct regulatory composition.

Identification of *De Novo* CREs for β-Globin Genes

Through unbiased capture of HS3, we identified several *de novo* CREs with unknown roles in globin gene regulation (Figures 5F and S7A). By CRISPR-mediated knockout (KO) using paired

Figure 5. CAPTURE-3C-Seq Identifies Locus-Specific Long-Range DNA Interactions

(A) Schematic of CAPTURE-3C-seq.

(B) Browser view of the long-range interactions at HS3 (chr11:5,222,500–5,323,700; hg19) is shown. Contact profiles including the density map, interactions (or loops), and PETs are shown. The statistical significance of interactions was determined by the Bayes factor (BF) and indicated by the color scale bars. ChIA-PET, DHS, ChIP-seq, RNA-seq, and ChromHMM data are shown.

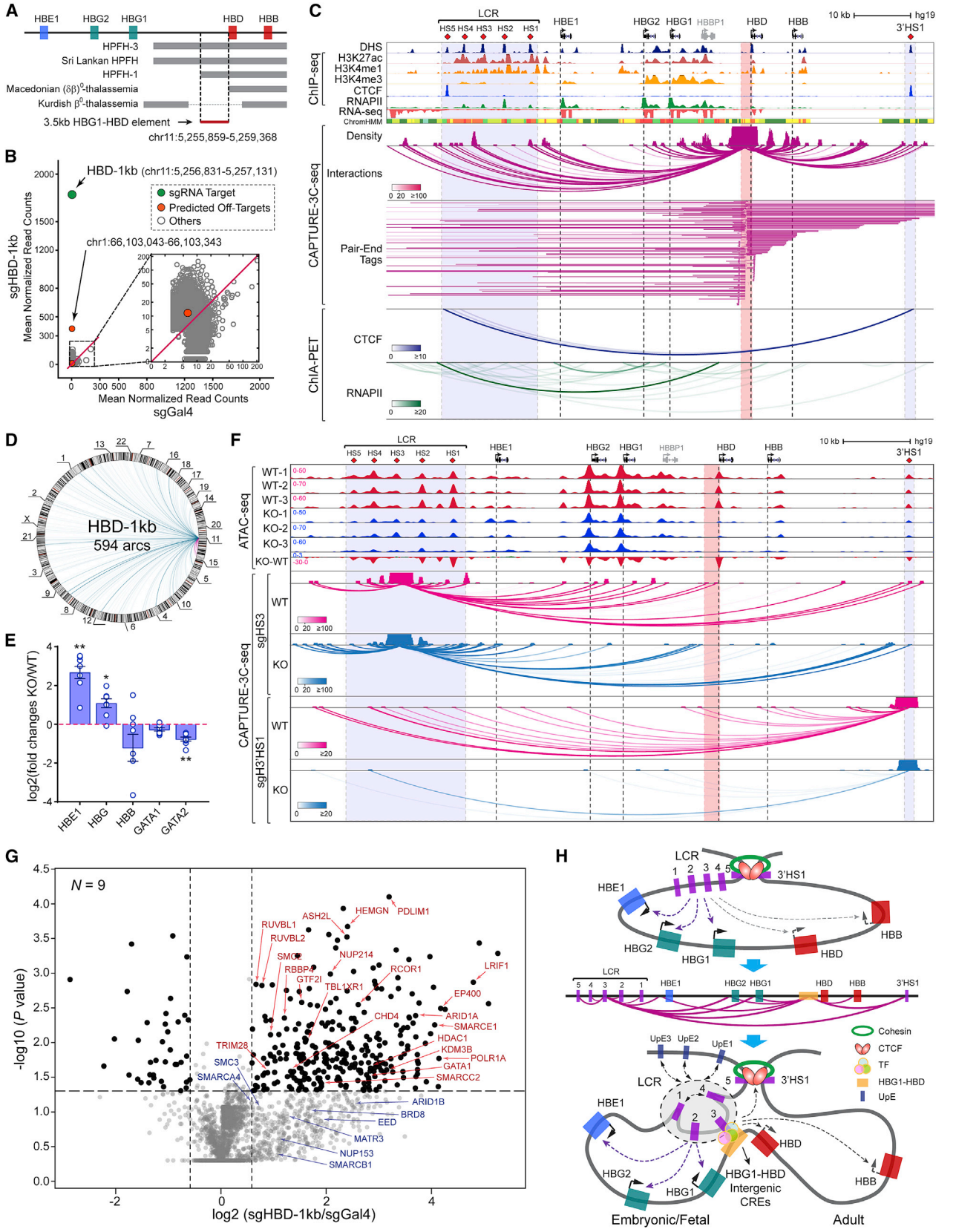
(C) Circlet plots of the long-range interactions are shown. The numbers of identified inter- (blue lines) and intrachromosomal (purple lines) interactions are shown.

(D) Browser view of the long-range interactions at the active *HBG* (green shaded lines) and the repressed *HBB* promoters (red shaded lines) is shown.

(E) The fraction of identified interactions relative to the total PETs at each captured region is shown. Results are mean ± SEM of two or three experiments and analyzed by a two-sided t test. *$p < 0.05$, ***$p < 0.001$.

(F) KO of *de novo* CREs impaired the expression of β-globin genes. The log2(fold change) of the mRNA expression in KO versus WT cells are shown. Each circle denotes an independent single-cell-derived KO clone. A diagram depicting the upstream (UpE1, UpE2, and UpE3) and downstream (DnE1, DnE2, and DnE3) CREs is shown on the top. Results are mean ± SEM of independent clones and analyzed by a two-sided t test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

See also Figures S4, S5, and S6, and Table S5.



(legend on next page)

sgRNAs, we observed that deletion of the UpE3 element located 160 kb upstream of *HBE1* led to significant downregulation of β-globin mRNAs (Figure 5F). Similarly, KO of UpE2 (−112 kb) and UpE1 (−36 kb) resulted in significant downregulation of β-globin genes. By contrast, KO of three downstream elements (DnE1, DnE2, and DnE3) overlapping with the CTCF-associated insulator resulted in significant upregulation of the repressed *HBB* gene, whereas the expression of *HBE1*, *HBG*, *GATA1*, and *GATA2* remained largely unaffected. The identification of new β-globin CREs illustrates the presence of additional distal *cis*-elements not recapitulated in studies using mouse models (Hardison et al., 1997; Navas et al., 1998; Peterson et al., 1998).

In Situ CAPTURE of a Disease-Associated CRE

Disease-associated CREs are commonly recognized by correlative chromatin features, yet limited insight has been gained into their regulatory composition. One example is the 3.5 kb *HBG1-HBD* intergenic region required for the silencing of fetal β-globin genes (Figure 6A). Genetic mapping studies showed that deletion of this region in humans, including in hereditary persistence of fetal hemoglobin 1 (HPFH-1), HFPH-3, and Sri Lankan HPFH patients, led to reactivation of *HBG*. By contrast, in patients that retained the intergenic region, including Macedonian $(\delta\beta)^0$-thalassemia and Kurdish β^0-thalassemia, *HBG* silencing was maintained (Sankaran et al., 2011). While these studies established the *HBG1-HBD* intergenic region as a critical disease-associated CRE, the underlying regulatory components remained unclear.

Therefore, we designed three sgRNAs targeting the 3.5 kb HBG1-HBD intergenic element (HBD-1kb, HBD-1.5kb and HBD-2kb; Figure S7B). The specificity of the sgRNAs was confirmed by CAPTURE-ChIP-seq (Figure 6B). By CAPTURE-3C-seq, we observed that the HBD-1kb region contained significantly higher frequency of long-range interactions than the neighboring HBD-1.5kb and HBD-2kb regions (Figure S7B). These interactions connected HBD-1kb with most β-globin CREs, including the HS1–HS4 enhancers, β-globin genes, and insulators (Figures 6C and 6D). Notably, KO of HBD-1kb in K562 cells resulted in upregulation of *HBE1* and *HBG*, whereas *HBB* was largely unaffected (Figure 6E). HBD-1kb KO also led to marked decreases in chromatin accessibility at the *HBG* and *HBD* promoters; HS1, HS2, and HS4 enhancers; and 3′HS1 (Figure 6F). Furthermore, by CAPTURE-3C-seq, we observed significant changes in the frequency of long-range interactions at several CREs (Figure 6F), suggesting that the *HBG1-HBD* intergenic region is required for the proper chromatin configuration and the expression of β-globin genes.

By CAPTURE-Proteomics of the *HBG1-HBD* intergenic region, we identified components of the SWI/SNF and NuRD complexes, transcriptional co-activators (EP400, KDM3B, and ASH2L), co-repressors (RCOR1, TBL1XR1, LRIF1, and TRIM28/KAP1), cohesin (SMC3), nucleoporins (NUP153 and NUP214) and TFs (GATA1 and STAT1) (Figure 6G; Table S4). The identification of the SWI/SNF and cohesin proteins is consistent with their function in regulating chromatin looping (Kagey et al., 2010; Kim et al., 2009b). The presence of co-activators and co-repressors may be related to the interactions with both active and repressed β-globin genes (Figure 6C). Notably, most of the HBD-1kb-associated proteins were not identified at the neighboring HBD-1.5kb or HBD-2kb region (Figure S7C).

Together, our studies support a refined model for the spatial organization of the β-globin CREs (Figure 6H). The β-globin genes are coordinately regulated in an insulated neighborhood between HS5 and 3′HS1. The *HBG1-HBD* intergenic region functions as a major interaction hub linking enhancers and insulators to establish two subdomains: an embryonic/fetal subdomain containing *HBE1*, *HBG1*, and *HBG2* genes and an adult subdomain containing *HBD* and *HBB*. HS2 and other LCR enhancers cooperate with associated regulators to activate the embryonic/fetal or adult genes in a developmental-stage-specific manner. Thus, our in-depth analyses of locus-specific interactions at the β-globin cluster by *in situ* CAPTURE not only reveal new spatial features for the composition-based hierarchical control of a lineage-specific enhancer cluster but also establish new approaches for molecular dissection of disease-associated CREs.

In Situ CAPTURE of Developmentally Regulated SEs

To demonstrate the utility of CAPTURE across cell models, we analyzed lineage-specific SEs during mouse ESC differentiation. We generated a site-specific knockin allele containing FB-dCas9-EGFP and BirA through FLPe-mediated recombination (Beard et al., 2006) (Figure 7A). After confirming the

Figure 6. Biotinylated dCas9-Mediated *In Situ* Capture of a Disease-Associated CRE

(A) Schematic of the 3.5 kb intergenic element (chr11:5,255,859–5,259,368; hg19) along with the deletions mapped in prior studies.

(B) Genome-wide specificity of sgHBD-1kb was measured by CAPTURE-ChIP-seq. N = 2 and 4 experiments for sgHBD-1kb and sgGal4.

(C) Browser view of the long-range interactions at HBD-1kb (red shaded lines) is shown.

(D) Circlet plot of the long-range interactions at HBD-1kb is shown.

(E) HBD-1kb KO impaired the expression of β-globin genes. Results are mean ± SEM of independent KO clones and analyzed by a two-sided t test. *$p < 0.05$, **$p < 0.01$.

(F) HBD-1kb KO led to altered chromatin accessibility and long-range interactions. Results from three ATAC-seq experiments in WT or KO cells are shown. Regions showing increased or decreased ATAC-seq signals in KO relative to WT cells (KO-WT) are depicted in green and red, respectively. HS3- or 3′HS1-mediated long-range interactions were determined by CAPTURE-3C-seq.

(G) CAPTURE-Proteomics identified HBD-1kb-associated proteins. Volcano plot is shown for the iTRAQ proteomics of purifications in sgHBD-1kb versus sgGal4-expressing cells.

(H) The model of composition-based organization of the β-globin cluster. Top: A previously described model depicting an active chromatin hub (ACH) formed through spatial organization of β-globin CREs (Palstra et al., 2003; Tolhuis et al., 2002). Middle: Two-dimensional representation of the long-range DNA interactions (purple lines) identified at HS3 and the *HBG1-HBD* intergenic CREs (yellow square) by CAPTURE. Bottom: A refined model depicting the composition-based spatial and hierarchical organization of the β-globin CREs.

See also Figure S7 and Tables S4 and S5.

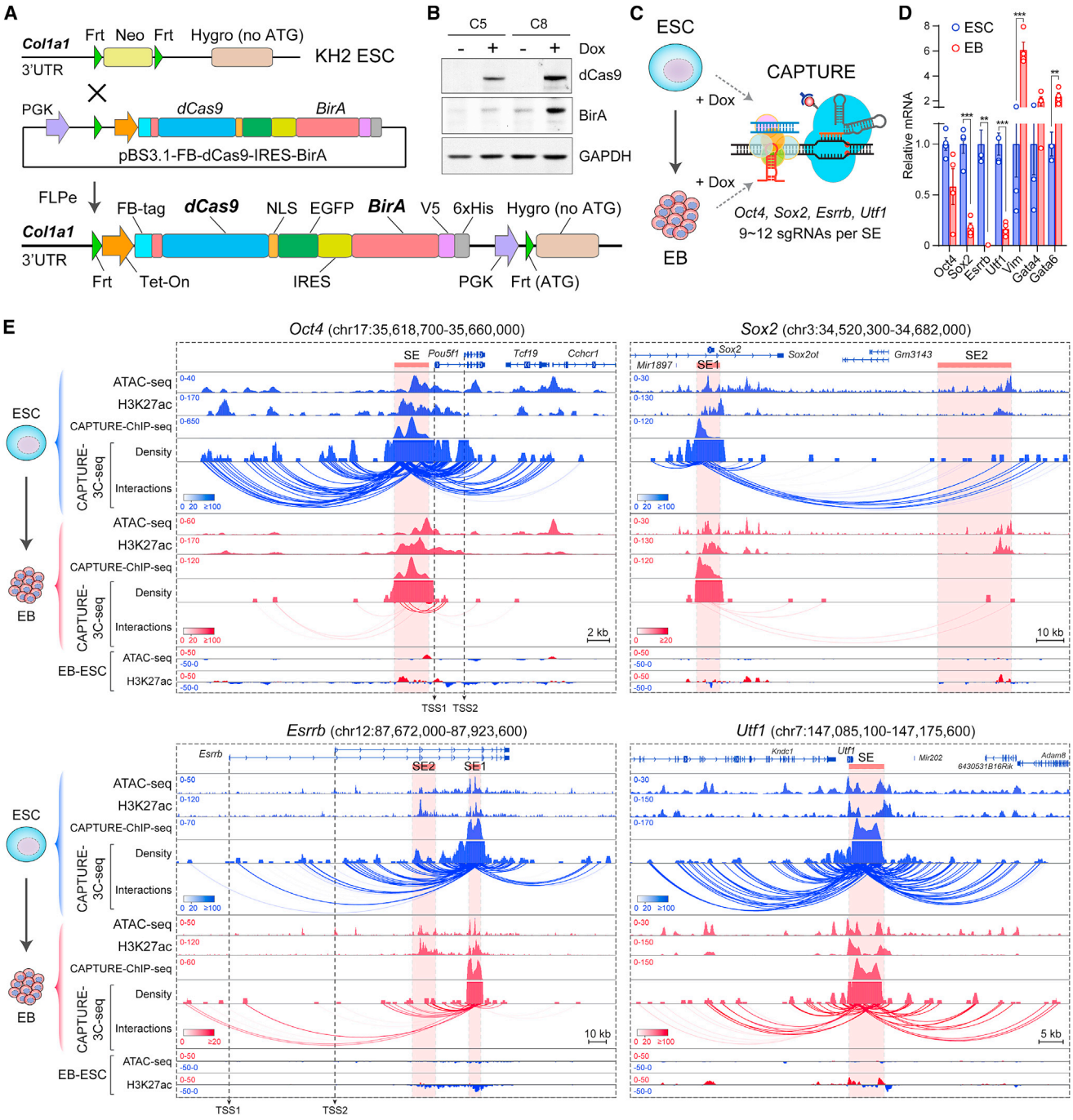


Figure 7. Multiplexed CAPTURE of Developmentally Regulated SEs during Differentiation

(A) Schematic of site-specific knockin of tetracycline-inducible FB-dCas9-EGFP and BirA.

(B) Dox-inducible expression of dCas9 and BirA proteins was confirmed by western blot in two independent knockin ESC lines.

(C) Schematic of multiplexed CAPTURE of ESC-specific SEs in ESCs and EBs.

(D) Differentiated EBs were characterized by downregulation of ESC-associated genes (*Oct4*, *Sox2*, *Esrrb*, and *Utf1*) and upregulation of differentiation-associated genes (*Vim*, *Gata4*, and *Gata6*). Results are mean ± SEM of 3 or 4 experiments and analyzed by a two-sided t test. $^{**}p < 0.01$, $^{***}p < 0.001$.

(E) Browser view of SE-associated long-range interactions captured by CAPTURE-3C-seq in ESCs and EBs. Regions showing increased or decreased ATAC-seq or H3K27ac ChIP-seq signals in EBs relative to ESCs (EB-ESC) are depicted in red and blue, respectively. Red bars denote the annotated SEs. Dashed lines denote the alternative TSS of transcript variants for *Oct4* (*Pou5f1*) and *Esrrb*.

doxycycline (Dox)-inducible expression of dCas9 and BirA proteins (Figure 7B), ESCs were differentiated to embryoid bodies (EBs). We designed multiplexed sgRNAs targeting four ESC-specific SEs (*Oct4*, *Sox2*, *Esrrb*, and *Utf1*; Figure 7C). Upon differentiation, the expression of the SE-linked genes was significantly downregulated (Figure 7D). We then analyzed SE-associated long-range interactions and chromatin features (Figure 7E). Strikingly, *in situ* CAPTURE of distinct SEs revealed frequent long-range interactions between SEs and their gene targets in ESCs, whereas the interactions were significantly less or absent in EBs. More importantly, the significant changes in SE-mediated long-range interactions, together with minimal or no changes in chromatin accessibility or H3K27ac, demonstrate that the loss of enhancer-promoter contacts precedes changes in chromatin landscape during differentiation. These findings support a model in which enhancer-promoter loop formation causally underlies gene activation (Deng et al., 2012; Deng et al., 2014). Many long-range interactions were between different SEs (*Sox2* and *Esrrb*; Figure 7E) or between SEs and promoters of transcript variants (*Oct4* and *Esrrb*). Furthermore, while most long-range interactions were absent or weakened in EBs, some were maintained, indicating a dynamic and hierarchal regulation of SE interactions in response to cellular differentiation. Taken together, our studies demonstrate that the CAPTURE approaches work effectively in human cells and transgenic mouse ESCs, raising the prospect of using biotinylated dCas9 in purification of CRE-associated chromatin interactions across cellular conditions *in situ* and in developing tissues *in vivo*.

DISCUSSION

In Situ CAPTURE of Locus-Specific Interactions

Current technologies in studying chromatin structure rely on 3D genome-mapping approaches. The basic principle is nuclear proximity ligation that allows detection of distant interacting DNA tethered together by higher-order architectures. ChIA-PET was designed to detect genome-wide chromatin interactions mediated by specific protein factors. Hi-C was developed to capture all chromatin contacts, particularly large-scale structures including the topologically associated domains (TADs) (Dixon et al., 2012); however, it lacked the level of resolution required for locus-specific interactions as well as the information of the *trans*-acting factors mediating such interactions. Hence, the CAPTURE method provides a complementary approach for high-resolution, unbiased analysis of locus-specific proteome and 3D interactome that is not dependent on predefined proteins, available reagents, or *a priori* knowledge of the target loci. The CAPTURE approach has several unique features, including the ability to specifically detect macromolecules at an endogenous locus with minimal off targets, to identify combinatorial protein-DNA interactions, and to dissect the disease-associated or developmentally regulated *cis*-elements.

Important Considerations for *In Situ* CAPTURE

For selective capture of locus-specific chromatin interactions, the following parameters need to be carefully evaluated. First, the sgRNA target sequences should locate in close proximity to the captured element to maximize the capture efficiency, but not overlap with TF binding sites to avoid interference with protein-DNA interactions. Second, the on-target enrichment and genome-wide specificity by independent sgRNAs should be evaluated to minimize off targets. Third, the study of locus-specific proteome requires the identification of non-specific proteins in control cells for quantitative and statistical analysis. Fourth, the analysis of CRE-mediated long-range DNA interactions requires the design of sgRNAs in close proximity to DpnII sites. Finally, the use of multiplexed sgRNAs targeting multiple CREs at the same enhancer or multiple enhancers helps distinguish consistent interactions from rare interactions of individual sgRNAs; however, the selection of multiplexed sgRNAs requires comparable on-target enrichment for each sgRNA to minimize variation in capture efficiency.

Multiplexed CAPTURE of SE Composition

Intensively marked clusters of enhancers or SEs have been described, yet the underlying principles of enhancer clustering remained unclear. Here, we focus on an erythroid-specific SE, or LCR, controlling the expression of β-globin genes. The β-globin LCR consists of five DHS, three of which display enhancer activities. Specifically, HS2 behaves as a classical enhancer in reporter assays (Fraser et al., 1993; Morley et al., 1992), whereas the enhancer activities of HS3 and HS4 can only be detected in the context of chromatin (Hardison et al., 1997; Navas et al., 1998). By *in situ* capture of β-globin CREs, our studies uncover distinguishing features in the regulatory composition of SE constituents. Importantly, the *HBG* and *HBB* promoters shared many interacting proteins and clustered closely, whereas the HS1, HS3, and HS4 enhancers clustered to form a distinct subdomain. HS2 shared interacting proteins with both subdomains. Furthermore, HS3 contains significantly more long-range interactions than the nearby enhancers. Hence, our results support a model for the hierarchical organization of the β-globin LCR, in which HS2 functions as a conventional enhancer by providing binding sites for *trans*-acting factors, whereas HS3 mediates long-range chromatin looping. Hence, the SE constituents cooperate through distinct regulatory composition to function within the same SE cluster. These findings also help explain the distinct requirement of HS2 and HS3 for the transgenic versus endogenous β-globin gene expression. Thus, the CAPTURE approach provides a platform for the systematic dissection of SE constituents and the underlying formative composition controlling enhancer structure-function.

Finally, the CAPTURE system can be adapted for multiplexed analysis of multiple CREs at the same enhancer or multiple enhancers, thus allowing for high-throughput capture of locus-specific interactions. High-resolution, multiplexed analysis of chromatin interactions at developmentally regulated enhancers provides evidence for the causality of chromatin looping and enhancer activities. Conversely, unbiased analysis of promoter-associated interactions will help identify the complete set of constitutive or tissue-specific distal CREs, thus allowing for comprehensive analysis of regulatory CREs of any gene. The vast majority of disease-associated variants reside within

non-coding elements and exert effects through long-range regulation of gene expression. The unbiased analysis of chromatin-templated hierarchical events will help define the underlying regulatory principles, thus advancing our mechanistic understanding of the non-coding genome in human disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cells and Cell Culture
- METHOD DETAILS
  - sgRNA Cloning and Transduction
  - CAPTURE-ChIP-seq
  - CAPTURE-Proteomics
  - CAPTURE-3C-seq
  - CRISPR Imaging of Human Telomeres
  - RNA-seq and qRT-PCR Analysis
  - ChIP-seq Analysis
  - ATAC-seq Analysis
  - Flow Cytometry
  - Cytospin
  - CRISPR/Cas9-Mediated Knockout of *Cis*-Regulatory Elements
  - Generation of Tetracycline-Inducible dCas9 Knockin ESCs
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.003.

AUTHOR CONTRIBUTIONS

Conceptualization: X.L. and J.X.; Methodology: X.L., Y.Z., Y.C., M.L., F.Z., K.L., H.C., M.N., Y.L., Z.G., K.E.D., S.X., G.C.H., Z.X., M.Q.Z., Z.S., and J.X.; Investigation: X.L., F.Z., K.L., H.C., and J.X.; Writing – Original Draft: X.L., Y.Z., Y.C., M.L., and J.X.; Writing – Review & Editing: J.X.; Funding Acquisition: F.Z., M.Q.Z., Z.S., and J.X.; Supervision, Z.X., M.Q.Z., Z.S., and J.X.

ACKNOWLEDGMENTS

We thank B. Chen and B. Huang at UCSF for the CRISPR imaging reagents and discussion; L. Wang, Y. Du, and D. Trudgian at UTSW BioHPC for assistance; and S.T. Smale at UCLA for critical reading of the manuscript and discussion. This work was supported by the NIH grant R01MH102616, the Cecil H. and Ida Green Endowment, the SKR&DPC grant (2017YFA0505503), and the NSFC grants (91519326, 31671384) to M.Q.Z.; the NSFC grant (31670836), the National Key R&D Program of China (2017YFA0505100), and Shanghai Institute of Higher Learning (TP2015003) to F.Z.; the "100-Talent" Program of Chinese Academy of Sciences (Y516C11851) to Z.S.; the NIH grants K01DK093543, R03DK101665, and R01DK111430, a Cancer Prevention and Research Institute of Texas (CPRIT) New Investigator award (RR140025), the American Cancer Society award and the Harold C. Simmons Comprehensive Cancer Center at UT Southwestern (IRG-02-196), the Welch Foundation grant I-1942-20170325, and an American Society of Hematology Scholar award to J.X.

Received: October 27, 2016
Revised: May 23, 2017
Accepted: August 1, 2017
Published: August 24, 2017

REFERENCES

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23–28.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell *140*, 372–383.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Consortium, T.E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Déjardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic loci. Cell *136*, 175–186.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., and Blobel, G.A. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849–860.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods *4*, 207–214.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067–1073.

Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993). Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev. *7*, 106–113.

Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., and Fujii, H. (2013). Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci. Rep. *3*, 3171.

Fujita, T., and Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem. Biophys. Res. Commun. *439*, 132–136.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N., and Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene *205*, 73–94.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.C., et al. (2016). Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. Dev. Cell *36*, 9–23.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46*, 205–212.

Ibarra, A., Benner, C., Tyagi, S., Cool, J., and Hetzer, M.W. (2016). Nucleoporin-mediated regulation of cell identity genes. Genes Dev. *30*, 2253–2258.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kalverda, B., Pickersgill, H., Shloma, V.V., and Fornerod, M. (2010). Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. Cell *140*, 360–371.

Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. J. Am. Stat. Assoc. *90*, 773–795.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009a). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nat. Protoc. *4*, 506–517.

Kim, S.I., Bultman, S.J., Kiefer, C.M., Dean, A., and Bresnick, E.H. (2009b). BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. Proc. Natl. Acad. Sci. USA *106*, 2259–2264.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lewis, K.A., and Wuttke, D.S. (2012). Telomerase and telomere-associated proteins: structural insights into mechanism and evolution. Structure *20*, 28–39.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., et al. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat. Methods *12*, 71–78.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Miccio, A., and Blobel, G.A. (2010). Role of the GATA-1/FOG-1/NuRD pathway in the expression of human beta-like globin genes. Mol. Cell. Biol. *30*, 3460–3470.

Morley, B.J., Abbott, C.A., Sharpe, J.A., Lida, J., Chan-Thomas, P.S., and Wood, W.G. (1992). A single beta-globin locus control region element (5′ hypersensitive site 2) is sufficient for developmental regulation of human globin genes in transgenic mice. Mol. Cell. Biol. *12*, 2057–2066.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953.

Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H., and Stamatoyannopoulos, G. (1998). Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion. Mol. Cell. Biol. *18*, 4188–4196.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. *36*, 1065–1071.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. *35*, 190–194.

Peterson, K.R., Navas, P.A., Li, Q., and Stamatoyannopoulos, G. (1998). LCR-dependent gene expression in beta-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. Hum. Mol. Genet. *7*, 2079–2088.

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.

Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A., and Bender, M.A. (2011). A functional element necessary for fetal hemoglobin silencing. N. Engl. J. Med. *365*, 807–814.

Schatz, P.J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. Nat Biotechnol. *11*, 1138–1143.

Schwartzman, O., Mukamel, Z., Oded-Elkayam, N., Olivares-Chauvet, P., Lubling, Y., Landan, G., Izraeli, S., and Tanay, A. (2016). UMI-4C for quantitative and targeted chromosomal contact profiling. Nat. Methods *13*, 685–691.

Shao, Z., Zhang, Y., Yuan, G.C., Orkin, S.H., and Waxman, D.J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. *13*, R16.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Stonestrom, A.J., Hsu, S.C., Jahn, K.S., Huang, P., Keller, C.A., Giardine, B.M., Kadauke, S., Campbell, A.E., Evans, P., Hardison, R.C., and Blobel, G.A. (2015). Functions of BET proteins in erythroid gene expression. Blood *125*, 2825–2834.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell *10*, 1453–1465.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods *9*, 969–972.

Waldrip, Z.J., Byrum, S.D., Storey, A.J., Gao, J., Byrd, A.K., Mackintosh, S.G., Wahls, W.P., Taverna, S.D., Raney, K.D., and Tackett, A.J. (2014). A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics *9*, 1207–1211.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Xu, J., Bauer, D.E., Kerenyi, M.A., Vo, T.D., Hou, S., Hsu, Y.J., Yao, H., Trowbridge, J.J., Mandel, G., and Orkin, S.H. (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. USA *110*, 6518–6523.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., and Orkin, S.H. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993–996.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

Zhou, F., Lu, Y., Ficarro, S.B., Adelmant, G., Jiang, W., Luckey, C.J., and Marto, J.A. (2013). Genome-scale proteome quantification by DEEP SEQ mass spectrometry. Nat. Commun. *4*, 2171.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Cas9	Abcam	Cat#ab191468
M2-FLAG	Sigma-Aldrich	Cat#F1804; RRID:AB_262044
TERF2	Santa Cruz	Cat#sc-9143; RRID: AB_2201333
GATA1	Abcam	Cat#ab28839; RRID:AB_2108281
TAL1	Santa Cruz	Cat#sc-12984; RRID:AB_2199699
V5-HRP	Life Technologies	Cat#R961-25; RRID:AB_2556565
RNAPII	Covance	Cat#MMS-126R; RRID:AB_10013665
ARID1B	Abcam	Cat#ab57461; RRID:AB_2243092
BCL11A	Abcam	Cat#ab19487; RRID:AB_444947
BRD4	Bethyl Labs	Cat#A301-985A; RRID:AB_1576498
HDAC1	Millipore	Cat#06-720; RRID:AB_2295297
HDAC2	Santa Cruz	Cat#sc-7899; RRID:AB_2118563
NUP98	Cell Signaling	Cat#2598S; RRID:AB_2267700
NUP153	BioLegend	Cat#906201; RRID:AB_2565060
SUZ12	Active Motif	Cat#39357; RRID:AB_2614929
H3K27ac	Abcam	Cat#ab4729; RRID:AB_2118291
Bacterial and Virus Strains		
Home-made *E. coli* DH5-alpha	This paper	N/A
Stellar competent cells	Clontech Laboratories	Cat#636766
Chemicals, Peptides, and Recombinant Proteins		
G418	Sigma-Aldrich	Cat#A1720
Puromycin	Thermo-Fisher	Cat#A1113802
(+)-JQ1	Sigma-Aldrich	Cat#SML1524
DpnII	New England Biolabs	Cat#R0543L
T4 DNA ligase	New England Biolabs	Cat#M0202L
Home-made Tn5 transposase	This paper	N/A
Streptavidin agarose beads	Thermo-Fisher	Cat#SA10004
Dynabeads MyOne streptavidin T1	Thermo-Fisher	Cat#65601
Dynabeads protein A	Thermo-Fisher	Cat#10002D
Dynabeads protein G	Thermo-Fisher	Cat#10004D
Critical Commercial Assays		
iScript reverse transcription supermix	Bio-Rad	Cat#1708840
iTaq SYBR® Green supermix	Bio-Rad	Cat#1725124
TruSeq RNA sample preparation kit	Illumina	Cat#RS-122-2001
Ovation RNA-seq system	NuGEN	Cat#0340-32
NEBNext ChIP-seq library prep kit	New England Biolabs	Cat#E6240L
QIAGEN PCR purification kit	QIAGEN	Cat#28104
KAPA HiFi HotStart DNA polymerase	KAPA Biosystems	Cat#KK2502
Deposited Data		
Raw and processed sequencing data	This paper	GEO: GSE88817; Table S1
iTRAQ-based mass spectrometry	This paper	Tables S3 and S4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: K562 cells	ATCC	Cat#CCL-243; RRID:CVCL_0004
Human: CD34+ HSPCs	CCEH, Fred Hutch Cancer Research Center	N/A
Mouse: KH2 ESC line	Dr. Stuart Orkin at Dana-Farber Cancer Institute	N/A
Oligonucleotides		
sgRNAs for CAPTURE	This paper	Table S2
Primers for qRT-PCR	This paper	Table S2
sgRNAs for knockout	This paper	Table S2
shRNAs for knockdown	This paper	Table S2
Genotyping primers	This paper	Table S2
Recombinant DNA		
pEF1a-FB-dCas9-puro	This paper	Addgene #100547
pEF1a-BirA-V5-neo	Kim et al., 2009a	Addgene #100548
pBS3.1-FB-dCas9-GFP-IRES-BirA-V5	This paper	N/A
pCAGGS-FLPe-puro	Dr. Stuart Orkin	N/A
pSLQ1651-sgRNA(F+E)-sgGal4	This paper	Addgene #100549
pSLQ1651-sgTelomere(F+E)	Chen et al., 2013	Addgene #51024
Software and Algorithms		
Bowtie	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MACS	Zhang et al., 2008	http://liulab.dfci.harvard.edu/MACS/
Tophat	Trapnell et al., 2009	https://ccb.jhu.edu/software/tophat/index.shtml
Other		
Illumina NextSeq500 instrument	Illumina	Cat#SY-415-1001
Agilent 2200 TapeStation instrument	Agilent Technologies	Cat#G2965AA
Qubit quantitation instrument	Thermo-Fisher	Cat#Q33216

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact, Jan Xu (jian.xu@utsouthwestern.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cells and Cell Culture

Human female K562 cells were obtained from ATCC and cultured in IMDM medium containing 10% FBS and 1% penicillin/streptomycin. pEF1α-FB-dCas9 and pEF1α-BirA-V5 vectors were co-transfected into K562 cells by nucleofection using the ECM 830 Square Wave Electroporation System (Harvard Apparatus, Holliston, MA). Cells were plated in 96-well plates and treated with 1 μg/ml of puromycin (Sigma) and 600 μg/ml of G418 (Sigma) 48-72 h post-transfection. Single-cell-derived clones were isolated and examined by western blot analysis to screen for FB-dCas9 and BirA-expressing stable clones. Human primary adult erythroid progenitor cells were generated *ex vivo* from CD34+ HSPCs as previously described (Huang et al., 2016). Primary HSPCs from both sexes were used in this study. For inhibition of BRD4, K562 or primary human erythroid progenitor cells were treated with the vehicle control (DMSO), JQ1 (0.25 μM or 1 μM) for 2 or 6 h before harvesting for ChIP-seq or qRT-PCR analyses. Mouse male embryonic stem cells (ESCs) were cultured on primary embryonic fibroblasts and differentiated to embryoid bodies (EBs) by LIF withdrawal for 8 days. All cultures were incubated at 37°C in 5% CO2. All cell lines were tested for mycoplasma contamination. No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI BioSample.

METHOD DETAILS

sgRNA Cloning and Transduction

Single guide RNAs (sgRNAs) for site-specific targeting of genomic regions were designed to minimize off-target cleavage based on publicly available filtering tools (http://crispr.genome-engineering.org/crispr/). To minimize potential interference between dCas9 and *trans*-acting factors, sgRNAs were designed to target the proximity of *cis*-elements. We also adapted an optimized sgRNA design by including the A-U pair flip and a 5bp extension of the hairpin as previously described (Chen et al., 2013). The sgRNAs were cloned into the lentiviral U6-driven expression vector by amplifying the insertions using a common reverse primer and unique forward primers containing the protospacer sequence, as previously described (Chen et al., 2013). Briefly, the forward primers were mixed with equal amount of reverse primer to PCR amplify sgRNA fragments using pSLQ1651 vector as the template. The PCR amplicon and the sgRNA vector containing a mCherry reporter gene were digested by restriction enzymes BstXI and XhoI for 3 h. The digested DNA were then purified, and ligated to the digested sgRNA vector using T4 DNA ligase. Insertion of sgRNA was validated by Sanger sequencing. Lentiviruses containing sgRNAs were packaged in HEK293T cells as previously described (Huang et al., 2016). Briefly, 2 μg of pΔ8.9, 1 μg of VSV-G and 3 μg sgRNA vectors were co-transfected into HEK293T cells seeded in 10 cm Petri dish. Lentiviruses were harvested from the supernatant 48-72 h post-transfection. FB-dCas9 and BirA-expressing K562 stable cells were then transduced with sgRNA-expressing lentiviruses in 6-well plates. To maximize sgRNA expression, the top 1% of mCherry-positive cells were FACS sorted 48 h post-transfection. The sequences for all sgRNAs used in this study are listed in Table S2.

CAPTURE-ChIP-seq

Streptavidin Affinity Purification of dCas9-Captured DNA and Sequencing

1×10^7 FB-dCas9/BirA-expressing K562 stable cells transduced with sequence-specific or non-targeting sgRNAs were harvested, cross-linked with 1% formaldehyde for 10 min, and quenched with 0.125 M glycine for 5 min. Cells were lysed in 1 mL RIPA buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, pH 8.0), and rotated for 15 min at 4°C. Cell lysates were centrifuged at 2,300 x g for 5 min at 4°C to isolate the nuclei. Nuclei were suspended in 500 μL of 0.5% SDS lysis buffer (0.5% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) and subjected for sonication to shear chromatin fragments to an average size between 200bp and 500bp on the Branson Sonifier 450 ultrasonic processor (20% amplitude, 0.5 s on 1 s off for 30 s). Fragmented chromatin was centrifuged at 16,100 x g for 10 min at 4°C. 450 μL of supernatant was transferred to a new Eppendorf tube and added final concentration 300 mM NaCl. Supernatant was then incubated with 10 μL of MyOne Streptavidin T1 Dynabeads (Thermo-Fisher Scientific) at 4°C overnight. After overnight incubation, Dynabeads were washed twice with 1 mL of 2% SDS, twice with 1 mL of RIPA buffer with 0.5 M NaCl, twice with 1 mL of LiCl buffer (250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl, pH 8.0), and twice with 1 mL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The chromatin was eluted in SDS elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) followed by reverse cross-linking at 65°C overnight. The ChIP DNA was treated with RNase A (5 μg/ml) and protease K (0.2 mg/ml) at 37°C for 30 min, and purified using QIAquick Spin columns (QIAGEN). 1 ng of ChIP DNA was processed for library generation using the NEBNext ChIP-seq Library Prep Master Mix (New England Biolabs or NEB) following the manufacturer's protocol. Libraries were pooled and sequenced on an Illumina Nextseq500 system using the 75bp high output sequencing kit.

CAPTURE-ChIP-seq Data Analysis

ChIP-seq raw reads were aligned to human (hg19) or mouse (mm9) genome assembly using Bowtie1 (Langmead et al., 2009) with default parameters. The first 10 nucleotides and the last 3 nucleotides from each read were excluded from alignment. For all ChIP-seq samples except sgHBG, only reads that can be uniquely mapped to the genome were used for further analysis. For sgHBG samples, since the sequences of *HBG1* and *HBG2* genes are highly similar, we kept reads with two alignments. MACS was applied to each sample to perform peak calling using the ''–nomodel'' parameter (Zhang et al., 2008). Peaks that overlap with the blacklist regions annotated by the ENCODE project (Consortium, 2012), the repeat masked region (chr2:33,141,250-33,142,690; hg19), or the validated non-targeting control sgRNA (sgGal4) enriched regions (chr6:119,558,373-119,558,873, chr17:42,074,844-42,075,323, chr21:15,457,141-15,457,641, chr20:26,188,800-26,190,400, chr17:42,074,844-42,075,323 and chr11:192,110-192,410; hg19) were removed. To compare ChIP-seq signal intensities in samples prepared from cells expressing the target-specific sgRNAs versus the non-targeting sgGal4, MAnorm (Shao et al., 2012) was applied to remove systematic bias between samples and then calculate the normalized ChIP-seq read densities of each peak for all samples. The window size was 300bp which matched the average width of the identified ChIP-seq peaks.

CAPTURE-ChIP-qPCR

For CAPTURE-ChIP-qPCR analysis, 0.5 to 1×10^7 FB-dCas9/BirA K562 stable cells transduced with sgTelomere were used. The captured DNA was isolated using the protocol described for CAPTURE-ChIP-seq except was analyzed by quantitative PCR (qPCR). For input samples, 80 μl of SDS elution buffer was added into 20 μl of the sheared chromatin. The samples were incubated at 65°C overnight to reverse cross-linking. DNA fragments were purified with the QIAquick PCR Purification Kit and eluted with 100 μL of EB buffer. Primers targeting human telomere sequences or a single copy gene 34B11 as a control were used for qPCR analysis. Primer sequences are listed in Table S2.

CAPTURE-Proteomics

We performed multiplexed isobaric tag for relative and absolute quantitation (iTRAQ)-based quantitative proteomic analysis of the isolated protein complexes. Briefly, the trypsin-digested peptides were labeled with 4-plex iTRAQ reagents (AB Sciex). After labeling, all peptides were mixed and loaded into an online three dimensional chromatography platform for in-depth proteome quantification as previously described (Zhou et al., 2013) with the following modifications. First, we performed in-solution, on-bead digestion of the purified samples to minimize sample loss associated with gel-based protocols. Second, we used the high-pH reversed phase (RP) and strong anion exchange separation stages coupled with a narrow-bore low-pH RP analytical column to achieve extreme separation of peptides in a nanoflow regime. Third, we chose the final dimension column geometry to maintain the integrity of chromatographic separation at ultra-low effluent flow rates to maximize electrospray ionization efficiency. Finally, we implemented all separation stages in microcapillary format coupled to the spectrometer, thus providing automated, efficient capture and transfer of peptides.

dCas9 Affinity Purification

0.25 to 1 × 10^9 FB-dCas9/BirA K562 stable cells transduced with sequence-specific sgRNAs or non-targeting sgRNA (sgGal4) were harvested, cross-linked with 2% formaldehyde for 10 min, and quenched with 0.25 M glycine for 5 min. Cells were washed twice with PBS, lysed with 10 mL of cell lysis buffer (25 mM Tris-HCl, 85 mM KCl, 0.1% Triton X-100, pH 7.4, freshly added 1 mM DTT and 1:200 proteinase inhibitor cocktail (Sigma)), and rotated for 15 min at 4°C. Cell lysates were centrifuged at 2,300 x g for 5 min at 4°C to isolate the nuclei. The nuclei were resuspended in 5 mL nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 4% SDS, pH 7.4, freshly added 1 mM DTT and 1:200 proteinase inhibitor cocktail) and incubated for 10 min at room temperature. Nuclei suspension was then mixed with 15 mL of 8 M urea buffer and centrifuged at 16,100 x g for 25 min at room temperature. Nuclei pellets were then resuspended in 5 mL nuclear lysis buffer and mixed with 15 mL of 8 M urea buffer, and centrifuged at 16,100 x g for 25 min at room temperature. The samples were washed twice more in 5 mL nuclear lysis buffer and mixed with 15 mL of 8 M urea buffer, followed by centrifugation at 16,100 x g for 25 min at room temperature. Pelleted chromatin was then washed twice with 5 mL cell lysis buffer. Chromatin pellet was resuspended in 5 mL of IP binding buffer without NaCl (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, 10% glycerol, pH 7.5, freshly added proteinase inhibitor) and aliquoted into Eppendorf tubes. Chromatin suspension was then subjected to sonication to an average size ~500bp on the Branson Sonifier 450 ultrasonic processor (10% amplitude, 0.5 s on 1 s off for 1 min). Fragmented chromatin was centrifuged at 16,100 x g for 25 min at 4°C. Supernatant was combined and final concentration 150 mM NaCl was added to the sheared chromatin. To prepare the streptavidin beads for affinity purification, 250 μL to 1 mL of streptavidin agarose slurry (Life Technologies) was washed 3 times in 1 mL of IP binding buffer and added to soluble chromatin. After overnight incubation at 4°C, streptavidin beads were collected by centrifugation at 800 x g for 3 min at 4°C. The beads were then washed 5 times with 1 mL of IP binding buffer (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, 10% glycerol, 150~300 mM NaCl, pH 7.5, freshly added proteinase inhibitor) and resuspended in 100 μL of 1 x XT sample loading buffer (Bio-Rad) containing 1.25% 2-mercaptoethanol followed by incubation at 100°C for 20 min. The proteins were separated by SDS-PAGE and analyzed by western blot.

In-Solution Digestion and Peptide Isolation

To improve the sensitivity and minimize sample loss associated with in-gel digestion, we performed in-solution on-beads trypsin digestion. Briefly, after overnight incubation of streptavidin beads with chromatin, the beads were washed 5 times with detergent-free IP binding buffer (20 mM Tris-HCl, 1 mM EDTA, 150 mM NaCl, 10% glycerol, pH 7.5). The beads were resuspended in 500 μL of 0.5 M Tris (pH 8.5) and incubated with final concentration 20 mM TCEP (tris(2-carboxyethyl)phosphine, Sigma, made freshly as 0.5M stock in 2M NaOH) at room temperature for 1 hour. The beads were then mixed with 4 μL of MMTS (S-Methyl methanethiosulfonate, Sigma) and incubated for 20 min at room temperature. The beads suspension was then digested with 20 μg of Trypsin (Promega) at 37°C overnight. After trypsin digestion, the beads were loaded to the cellulose acetate filter spin cup (0.45 μm pore size, Pierce) and centrifuged at 12,000 x g for 2 min at room temperature to collect flow-through containing peptides. The peptide solution was mixed with final concentration 3 M NaCl and boiled at 95°C for 1 hour to reverse formaldehyde cross-linking. Digested peptides were dried using a SpeedVac (Thermo-Fisher Scientific), reconstituted in 200 μL of 0.1% trifluoroacetic acid (TFA) and loaded onto a pre-equilibrated Oasis HLB elute plate (Waters Corporation). After discarding the flow-through, the columns were washed with 800 μL of 0.1% TFA, followed by another wash with 200 μL of ddH_2O. The desalted peptides were then eluted with 50 μL of 70% acetonitrile and labeled with multiplexed isobaric tags using the iTRAQ Reagents-4Plex Multiplex Kit (SCIEX) according to the manufacturer's protocol.

Multi-Dimension Separation and Data Acquisition

Nanoscale three dimensional online chromatography platform consists of first dimension reversed phase (RP) column (100 μm I.D. capillary packed with 10 cm of 5 μm dia. XBridge (Waters Corp., Milford, MA) C18 resin), second dimension strong anion exchange (SAX) column (100 μm I.D. 10 cm of 10μm dia. POROS10HQ (AB Sciex, Foster City, CA) resin) and third dimension reversed phase column (15 μm I.D. 50 cm of 3 μm dia. Monitor C18 (Column Engineering, Ontario, CA), integrated 1 μm dia. emitter tip). The final dimension ran at 1-2 nL/min with a ~280 min gradient from 2% B to 50% B (A = 0.1% formic acid, B = acetonitrile with 0.1% formic acid). The downstream TripleTOF 5600+ (AB Sciex, Foster City, CA) was set in data-dependent acquisition (DDA) mode for data acquisition. Top 50 precursors (charge state +2 to +4, > 70 counts) in each MS scan (800 ms, scan range 550-1500 m/z) were subjected to MS/MS (maximum time 250 ms, scan range 100-1400 m/z). Electrospray voltage was 2.4 kV.

Data Processing and Protein Quantification

The mass spectrometry data was subjected to search against SwissProt database (downloaded on 10/02/2016) with ProteinPilot V4.5 (AB Sciex, Framingham, MA). Official HGNC Gene Symbols were included in the database. The search parameter was set to "iTRAQ 4-plex (peptides labeling) with 5600 TripleTOF." In this study, we also removed peptides that can be assigned to more than one gene. The peptide spectra match (PSM) false discovery rate (FDR) was used to filter the peptides identified for further analysis. Specifically, FDR is the statistical model used to evaluate the confidence level of peptide identification based on the well-established target-decoy search strategy (Elias and Gygi, 2007). The target-decoy search strategy requires repeated search using identical parameters against a 'decoy' database in which the target sequences have been reversed or randomized. The number of matches found in 'decoy' database is used as an estimate of the number of false positives (FP) that are present in the 'target' database. The number of true positive (TP) matches in the 'target' database and the number of FP matches in the 'decoy' database are then used to calculate the False Discovery Rate (FDR) = FP / (FP + TP). Only those peptides with scores at or below a PSM FDR threshold of 1% were kept for data analysis. After that, we summed the intensity of each iTRAQ reporter ion for the peptides that can only be assigned to single gene to generate the iTRAQ intensity value for each gene. We then removed genes with weak quantification signal (total signal intensity of iTRAQ reporter ions $\leq$ 50). To compare between independent experiments and individual samples, the ion intensity of iTRAQ mass spectrometry signal was normalized based on the cumulative intensity of the high-confidence non-specific proteins (Figure S2B) identified from four control cell lines expressing the non-targeting sgRNAs (sgGal4) and/or dCas9 and the bait protein (dCas9). Specifically, for each individual target-specific sgRNA and the corresponding control samples, the log2 ratios of iTRAQ reporter ion intensities of all detected non-specific proteins were plotted against the average intensities between two profiles. Principal component analysis (PCA) was applied to the plot to not only rescale the average log2 ratios of these proteins to zero, but also minimize the total variation of observed log2 ratios. Then the principal components were applied to the log2 ratios and the average intensities of all detected proteins, and the projection of their log2 ratios to the second principal component was taken as the normalized log2 ratios of iTRAQ intensities between two profiles. After the global normalization of each sample, the ratios of the iTRAQ reporter ion intensity for each protein in target-specific sgRNA samples relative to the non-targeting sgGal4 sample were collected across replicate experiments. Only proteins detected in at least 3 replicates (at least 2 replicates for sgHBD-1.5kb and sgHBD-2kb) were subjected to statistical analysis, in which a *P* value was calculated to measure the statistical significance of the log2 iTRAQ ratios of each identified protein in the replicate experiments by t test. After removing the non-specific proteins identified from control experiments, the iTRAQ ratio and *P* value for the remaining proteins were calculated in each replicate experiment. To determine the ratio and *P* value cutoffs used to identify significantly enriched locus-specific proteins, we surveyed the distribution of the "high-confidence non-specific proteins" in all proteomic experiments, and observed that 78.3% and 79.8% of the 'high-confidence non-specific proteins' displayed iTRAQ ratio less than 1.5-fold and *P* value more than 0.05 (Figure S2C). Based on these analyses, a protein was considered to be significantly enriched if the iTRAQ ratio $\geq$ 1.5 and *P* value $\leq$ 0.05 in samples prepared from cells expressing sequence-specific sgRNAs versus the non-targeting sgGal4 control.

Connectivity Network Analysis

The connectivity network was built by Gephi (version 0.9.1) using all interactions between the dCas9-captured locus-specific proteins and the β-globin CREs (*HBG* and *HBB* promoters, and HS1-HS4 enhancers). Colored nodes represent proteins significantly enriched at single or multiple promoter and/or enhancer regions. Size of the circles represents the frequency of interactions.

CAPTURE-3C-seq

3C Library Preparation and Sequencing

1 to 5 × 10^7 cells were cross-linked with 2 mM EGS (ethylene glycol bis(succinimidyl succinate)) (Thermo-Fisher Scientific) for 45 min and 1% formaldehyde for 15 min at room temperature. Cross-linking was quenched with 0.25 mM of glycine for 10 min at room temperature, followed by two washes with PBS. Cells were resuspended in ice-cold 1 mL of RIPA buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, pH 8.0, freshly added 1 mM DTT, and 1:200 proteinase inhibitor cocktail) and rotated for 15 min at 4°C. Cell lysates were centrifuged at 2,300 x g for 5 min at 4°C to isolate the nuclei. Nuclei were then resuspended in 500 μL of 1.2 x NEBuffer DpnII buffer containing 0.25% SDS and incubated for 10 min at 65°C, followed by 1 h incubation after adding 100 μL of 10% Triton X-100 (final concentration 1.67%). Nuclei were digested using 300 U of DpnII (NEB) on a Thermomixer (Eppendorf) overnight at 37°C. DpnII digestion was quenched by adding 44 μL of 20% SDS (final concentration 1.6%) and vortexed for 20 min at 65°C. The digested nuclei were diluted with 2.041 mL of 1.5 x T4 ligation buffer (300 μL of 10 x NEB T4 ligase buffer, 1.741 mL of ddH_2O, freshly added 1:200 proteinase inhibitor cocktail). SDS was sequestered by adding 700 μL of 10% Triton X-100 and incubating at 37°C for 1 h at 400 RPM. Nuclei were then ligated overnight by adding 15 μL of NEB T4 DNA ligase (final concentration 30 weiss U/ml) with rotation overnight at 16°C. The nuclei were collected by centrifuge at 2,300 g for 5 min at 4°C, and resuspended in 500 μL 0.5% SDS lysis buffer (0.5% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) followed by sonication to shear chromatin fragments to an average size ~500bp on the Branson Sonifier 450 ultrasonic processor (10% amplitude, 0.5 s on 1 s off for 30 s). Chromatin fragments were centrifuged at 16,100 x g for 10 min at 4°C. Final concentration 300 mM NaCl was added to the supernatant followed by incubation with 50 μL of MyOne Streptavidin T1 Dynabeads (Thermo-Fisher Scientific) overnight at 4°C. After overnight incubation, the Dynabeads were washed twice with 1 mL of 2% SDS, twice with 1 mL of RIPA buffer with 0.5 M NaCl, twice with 1 mL of LiCl buffer, and twice with 1 mL of TE buffer. The chromatin was resuspended in SDS elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, 0.2 mg/ml proteinase K) followed by reverse cross-linking and proteinase K digestion at

65°C overnight. The DNA was purified using QIAquick Spin columns (QIAGEN). 5 ng of CAPTURE-3C DNA was processed for library generation using the NEBNext ChIP-seq Library Prep Kit (New England Biolabs). Libraries were pooled and 38bp pair-end sequencing was performed on an Illumina Nextseq500 platform using the 75bp high output sequencing kit. To determine the specificity of CAPTURE-3C-seq, we performed two control experiments: 1) CAPTURE-3C-seq using the non-targeting sgGal4 control, and 2) CAPTURE-3C-seq using the purified, DpnII-digested genomic DNA (naked gDNA) control. The sgGal4 control was performed in parallel with other target-specific sgRNAs following the same CAPTURE-3C-seq protocol, whereas the gDNA control was performed in the absence of dCas9 affinity purification step to determine the probabilities of ligation of any DpnII-digested DNA fragments due to random collision in the ligation reaction.

CAPTURE-3C-seq Data Analysis

To identify significant interactions from sequenced read pairs, we developed a customized data processing pipeline for the mapping of raw reads and statistical analysis. All sequencing reads were mapped to human (hg19) or mouse (mm9) genome assembly. Raw reads from all replicate experiments for each sgRNA sample were merged. Pair-end reads were mapped as single-end reads by using Bowtie2 (Langmead and Salzberg, 2012) with the default parameters to avoid the build-in assumption of the relative positioning of pair-end sequences in the alignment program. Unmapped reads were tested if they contained a DpnII restriction site. The reads with digestion position were trimmed and the longer fragment with length $\geq$ 20 bp was collected and remapped. The mapped reads from both procedures were combined and the reads with low mapping quality were removed by using the cutoff of MAPQ $\geq$ 30. The mapped reads from pair-end sequencing were then paired. PCR duplicates were removed by discarding the reads with the same positions at both paired ends.

The preprocessed read pairs were used to define the interactions at each sgRNA-targeted (or bait) region to other chromosomal regions. Previous studies of 4C and Capture-C used fixed sizes of sliding window (typically ± 1 kb of targeted sites) to define the interacting regions (Hughes et al., 2014; van de Werken et al., 2012). However, the peaks of local read pairs (or self-ligations) are different from each experiment and skewness of peaks can be observed from the sgRNA-targeted regions. Hence, fixed window sizes with 2kb would have hard cutoff of bait regions and may lead to inaccurate positioning of bait regions. Therefore, we defined the bait region as the local peaks surrounding the sgRNA target site by using MACS2 with default parameters (Zhang et al., 2008). The read pairs located within the bait region were considered as self-ligated reads and filtered. After preprocessing and filtering, the resulting data is a list of count numbers of read pairs from the bait region to any chromosomal regions. A pair of reads that located within two different regions is considered an interaction. We then applied separate background models to calculate the significance for intra- and inter-chromosomal interactions.

Intra-chromosomal Model

To understand the statistical significance of enrichment for $x_d(i)$ that denotes the interaction numbers from the bait region to the chromosomal region *i* with distance *d*l*, we need to know the bias/noise background of $x_d(i)$. Here *d* is the indicator of the region that is with distance of *d*l* to the bait region, where *l* is the size of bait region. We used interaction values X_d of any two regions in the same chromosome as the background (excluding the bait region). We found (1) the means/medians of X_d were decreased when distances increased; (2) the mean and variance showed proportional relationship revealed by linear regression analysis. To better fit the underlying observations, we used the Bayesian mixture model to describe the interaction background and presented multiple models for different distance *d*. The count of interactions X_d is assumed to have been drawn from a Poisson distribution with mean λ_d, which follows a Gamma distribution with parameters α_d and β_d. e.g $X_d \sim Poisson(\lambda_d)$, $\lambda_d \sim Gamma(\alpha_d, \beta_d)$, we have

$$\Pr(X_d \mid \alpha_d, \beta_d) = \int_0^\infty \Pr(X_d \sim Poisson(\lambda_d))\Pr(\lambda_d \sim Gamma(\alpha_d, \beta_d))d\lambda_d$$

$$= \frac{\beta_d^{\alpha_d}\Gamma(\alpha_d + X_d)}{(\beta_d + 1)^{\alpha_d + X_d}\Gamma(\alpha_d)X_d!}$$

We can get X_d follows a negative binomial distribution with parameters α_d and $(\beta_d/\beta_d + 1)$. We used Maximum Likelihood Estimator (MLE) to estimate the parameters α_d and β_d. Since negative binomial distribution has a closed form of expected value, we can achieve a great practical advantage to estimate parameters by using simple mean and variance. Thus, X_d models the random collision frequency between any two chromosomal regions (with distance of *d*). We can therefore calculate *P* values by using negative binomial distribution to reflect the significance of $x_d(i)$ as $p_d(i) = P(X_d < x_d(i))$. Specifically, the bigger $p_d(i)$ indicates lower possibility of random collisions that are bigger than $x_d(i)$, suggesting higher confidence of interactions between the bait region and the chromosomal region *i*. Instead of calculating *P* values, we used the Bayes factor (BF) to compare the hypothesis H_0 that specific interactions have occurred between the bait region and a given chromosomal region ($\Pr(H_0 \mid x_d(i)) = P(X_d < x_d(i))$, e.g., the probability that random collisions are less than observed interaction $x_d(i)$), against the alternative hypothesis H_1, representing no interactions between them. The BF is defined as $BF = \Pr(x_d(i)|H_0)/\Pr(x_d(i)|H_1) = \Pr(H_0 \mid x_d(i))/\Pr(H_1 \mid x_d(i))\Pr(H_1)/\Pr(H_0)$, a strength measure for comparing two hypotheses, which provides a natural way to consider the uncertainty in hypothesis testing and controlling false discovery rate (FDR). Here, the prior odds $(\Pr(H_1)/\Pr(H_0))$ were assigned as 0.001, indicating that random collision bigger than true interactions is a rare event. According to the scale for BF, $3 \leq BF < 20$ is considered 'positive' and $20 \leq BF$ is considered 'strong'

evidence of supporting H_0 (Kass and Raftery, 1995). Here, we considered paired regions with BF of interactions more than 20 as the 'high-confidence interactions'. We set up 11 different models for different distance *d,* including 10 models for paired regions with distances ranged from 1*l* to 10*l* and one for paired regions with distances bigger than 10*l*, where *l* is the size of the bait region.

Inter-chromosomal Model

To test the significance of interactions between the bait region to the interacting regions on a different chromosome, we developed the background model by using the random collisions among inter-chromosomal region pairs (regions located on different chromosomes). Specifically, we first extended the bait region to 1 Mb and split all chromosomes into 1 Mb regions. For a region *j* of other chromosomes (excluding chr11), we counted the numbers from the bait region to region *j*. We randomly selected 1000 regions from chr11 and counted interactions from them to region *j* as the background (negative binomial distribution). Similar to the intra-chromosomal model, we also used the Bayes factor (BF) to test if interactions from the bait region and other regions were significant. All scripts are tested on Linux operating system and available on request.

Comparison of Chromatin Interactions Defined by CAPTURE-3C-seq, 4C, 5C, ChIA-PET and Hi-C

RNAPII and CTCF ChIA-PET (GSM970213 and GSM970216), UMI-4C (GSM2037371), 5C (GSM970500), DNase Hi-C (GSM1370434 and GSM1370436), and *in situ* Hi-C data (GSM1551618) were downloaded from GEO (Table S1). The raw reads from all samples were mapped by Bowtie2 using the same parameters as in CAPTURE-3C-seq. The unique read pairs with one end in bait region (PETs) were collected. We then calculated the normalized PETs of a bait region as $(PETs \cdot 10^9 / Bait\_Length \cdot Total\_reads.)$, which represents the on-target enrichment as the number of PETs per kilobases of bait region per million mapped reads. The unique PETs were defined as pair-end sequence tags with distinct genomic locations at one or both sides of the pair-end reads.

CRISPR Imaging of Human Telomeres

CRISPR imaging of human telomeres was performed as described (Chen et al., 2013). Briefly, human MCF7 cells were transduced with lentiviruses expressing a dCas9-EGFP fusion protein driven by a TRE3G promoter and the Tet-on-3G trans-activator protein. After confirming the expression of the dCas9-EGFP fusion protein by induction with doxycycline (100 ng/ml), the cells were transduced with lentiviruses expressing the telomere-specific sgRNA (sgTelomere) in an 8-well chambered coverglass. The nuclear location of dCas9-EGFP was determined on a 2-photon fluorescence microscope (Zeiss LSM780 Inverted) with 40x and 60x objective lens. The images were acquired and analyzed on the ZEN software (Zeiss).

RNA-seq and qRT-PCR Analysis

Total RNA was isolated using RNeasy Plus Mini Kit (QIAGEN) following manufacturer's protocol. RNA-seq library was prepared using the Truseq v2 LT Sample Prep Kit (Illumina) or the Ovation RNA-seq system (NuGEN). Sequencing reads from all RNA-seq experiments were aligned to human (hg19) reference genome by TopHat v2.0.13 (Trapnell et al., 2009) with the parameters:–solexaquals–no-novel-juncs. Quantitative RT-PCR (qRT-PCR) was performed using the iQ SYBR Green Supermix (Bio-Rad). Primer sequences are listed in Table S2.

ChIP-seq Analysis

ChIP-seq was performed as described (Huang et al., 2016) using the antibodies for BRD4 (A301-985A, Bethyl, lot: A301-985A-1), RNAPII (MMS-126R, Covance, lot: D12LF03144) and H3K27ac (ab4729, Abcam) in K562 erythroid cells treated with DMSO (control), or 1 μM of JQ1 for 6 hours. Antibodies for NUP98 (2598, Cell Signaling Technology, lot:4) or NUP153 (906201, BioLegend, lot:B215613) were used. Cross-linked K562 chromatin was sonicated in RIPA 0 buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 0.25% Sarkosyl, pH 8.0) to 200∼500 bp. Final concentration 150 mM NaCl was added to the chromatin and antibody mixture before incubation overnight at 4°C. ChIP-seq libraries were generated using NEBNext ChIP-seq Library Prep Master Mix following the manufacturer's protocol (New England Biolabs), and sequenced on an Illumina NextSeq500 system using the 75bp high output sequencing kit. ChIP-seq raw reads were aligned to the hg19 or mm9 genome assembly using Bowtie (Langmead et al., 2009) with the default parameters. Only tags that uniquely mapped to the genome were used for further analysis. ChIP-seq peaks were identified using MACS (Zhang et al., 2008). Gene ontology (GO) analysis was performed using GREAT (McLean et al., 2010).

ATAC-seq Analysis

5×10^4 cells were washed twice in PBS and resuspended in 500 μL lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl, 0.1% NP-40, pH 7.4). Nuclei were harvested by centrifuge at 500 x g for 10 min at 4°C. Nuclei were suspended in 50 μL of tagmentation mix (10 mM TAPS (Sigma), 5 mM MgCl, pH 8.0 and 2.5 μL Tn5) and incubated at 37°C for 30 min. Tagmentation reaction was terminated by incubating nuclei at room temperature for 2 min followed by incubation at 55°C for 7 min after adding 10 μL of 0.2% SDS. Tn5 tranposase-tagged DNA was purified using QIAquick MinElute PCR Purification kit (QIAGEN), amplified using KAPA HiFi Hotstart PCR Kit (KAPA), and sequenced on an Illumina Nextseq500 system using the 75 bp high output sequencing kit. ATAC-seq raw reads were trimmed to remove adaptor sequence and aligned to hg19 or mm9 genome assembly using Bowtie2 (Langmead et al., 2009) with k = 1 and m = 1. Only tags that uniquely mapped to the genome were used for further analysis.

Flow Cytometry

Human erythroid cell differentiation was analyzed by flow cytometry using FACSCanto. Live cells were identified and gated by exclusion of 7-amino-actinomycin D (7-AAD; BD PharMingen). The cells were analyzed for expression of cell surface receptors with antibodies specific for CD71 and CD235a conjugated to phycoerythrin (PE) and fluorescein isothiocyanate (FITC), respectively. Data were analyzed using FlowJo software (Ashland, OR).

Cytospin

Cytospin preparations from cells at various stages of erythroid differentiation were stained with May-Grunwald-Giemsa as described previously (Xu et al., 2011).

CRISPR/Cas9-Mediated Knockout of *Cis*-Regulatory Elements

The CRISPR/Cas9 system was used to introduce deletion mutations of the *cis*-regulatory elements in K562 cells following published protocols (Cong et al., 2013; Mali et al., 2013). Briefly, sequence-specific sgRNAs for site-specific cleavage of genomic targets were designed following described guidelines, and sequences were selected to minimize off-target cleavage based on publicly available filtering tools (http://crispr.mit.edu/). Oligonucleotides were annealed in the following reaction: 10 μM guide sequence oligo, 10 μM reverse complement oligo, T4 ligation buffer (1X), and 5U of T4 polynucleotide kinase with the cycling parameters of 37°C for 30 min; 95°C for 5 min and then ramp down to 25°C at 5°C/min. The annealed oligos were cloned into the pSpCas9(BB) (pX458) vector (Addgene #48138) using a Golden Gate Assembly strategy including: 100 ng of circular pX458 plasmid, 0.2 μM annealed oligos, 2.1 buffer (1X) (New England Biolabs), 20 U of BbsI restriction enzyme, 0.2 mM ATP, 0.1 mg/ml BSA, and 750 U of T4 DNA ligase (New England Biolabs) with the cycling parameters of 20 cycles of 37°C for 5 min, 20°C for 5 min; followed by 80°C incubation for 20 min. To induce deletions of candidate regulatory DNA regions, two CRISPR/Cas9 constructs were co-transfected into K562 cells by nucleofection using the ECM 830 Square Wave Electroporation System (Harvard Apparatus). Each construct was directed to flanking the target genomic regions. To enrich for deletion, the top 1%–5% of GFP-positive cells were FACS sorted 48-72 h post-transfection and plated in 96-well plates. Single-cell-derived clones were isolated and screened for CRISPR-mediated deletion of target genomic sequences. PCR amplicons were subcloned and analyzed by Sanger DNA sequencing to confirm non-homologous end-joining (NHEJ)-mediated repair upon double-strand break formation. The positive single-cell-derived clones containing deletion of the targeted sequences were expanded and processed for analysis.

Generation of Tetracycline-Inducible dCas9 Knockin ESCs

Site-specific knockin of tetracycline-inducible FLAG-biotin-acceptor-site (FB)-tagged dCas9-EGFP and BirA transgenes was generated through flippase (FLPe)-mediated recombination (Beard et al., 2006). Briefly, KH2 mouse embryonic stem cells (ESCs) harboring a targeted M2rtTA tetracycline-responsive trans-activator in the *Rosa26* locus and a modified *Col1a1* locus with an frt site and ATG-less hygromycin resistance gene were used. A targeting construct pBS3.1-FB-dCas9-IRES-BirA containing the PGK promoter, an frt site, a tetracycline-inducible minimal CMV promoter, the FB-dCas9-EGFP-IRES-BirA transgenes, and an ATG initiation codon was co-electroporated with the pCAGGS-FLPe-puro into KH2 ESCs at 500V and 25 μF using a Gene Pulser II (Bio-Rad). The cells were selected with hygromycin (140 μg/ml) after 24 hours. The positive clones were expanded and analyzed by genotyping PCR. The correctly targeted ESCs were cultured in the absence or presence of doxycycline (0.1-1 μg/ml) for 48 h and harvested for CAPTURE experiments.

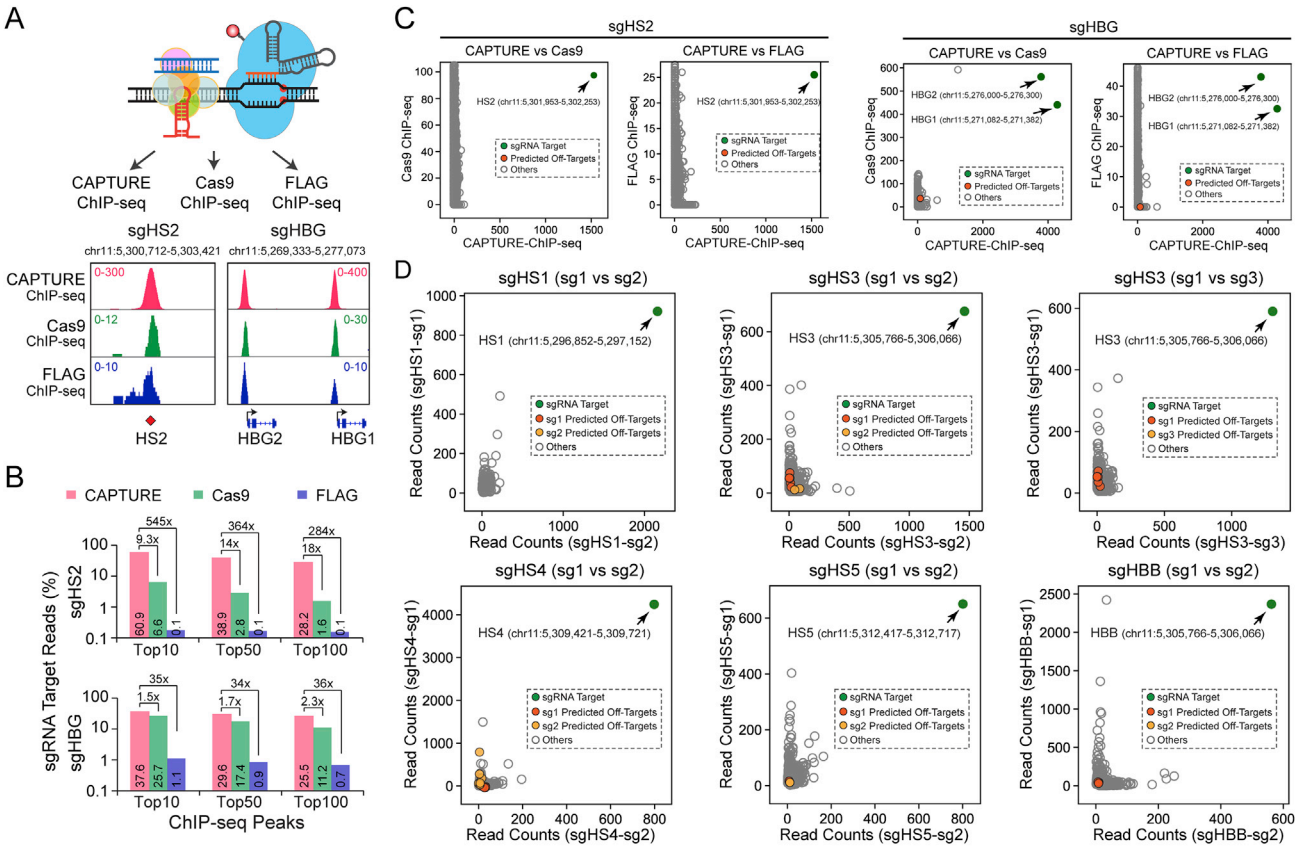
QUANTIFICATION AND STATISTICAL ANALYSIS

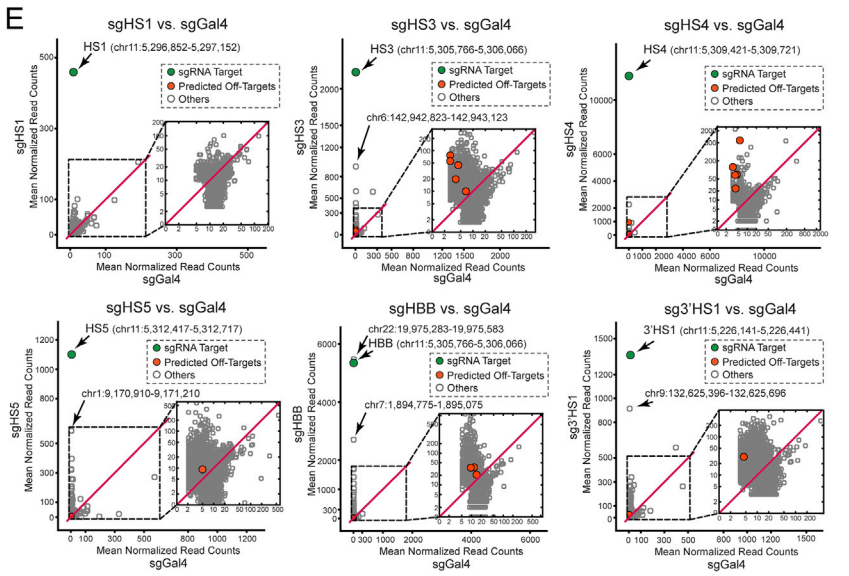
Statistical details including *N*, mean and statistical significance values are indicated in the text, figure legends, or Method Details. Error bars in the experiments represent standard error of the mean (SEM) from either independent experiments or independent samples. All statistical analyses were performed using GraphPad Prism, and the detailed information about statistical methods is specified in figure legends or Methods Details.

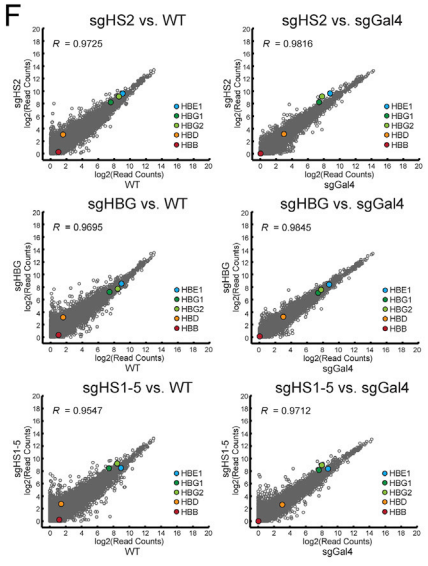
DATA AND SOFTWARE AVAILABILITY

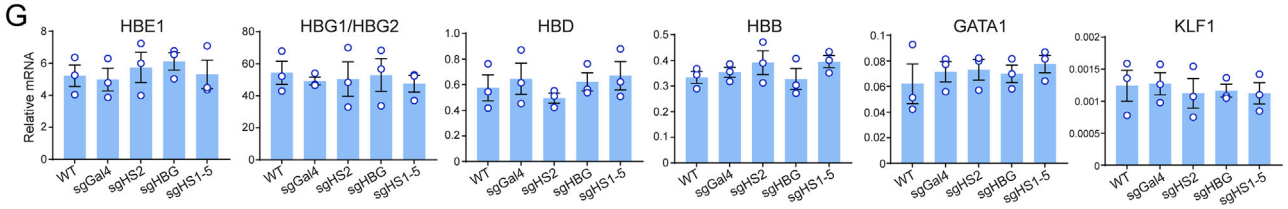
All raw and processed RNA-seq, ChIP-seq, CAPTURE-ChIP-seq, CAPTURE-3C-seq and ATAC-seq data are available in the Gene Expression Omnibus, GEO: GSE88817.

Cell









(legend on next page)

Figure S1. Genome-wide Enrichment and Specificity of dCas9-Mediated CAPTURE, Related to Figure 2

(A) CAPTURE-ChIP-seq markedly improved the on-target enrichment compared to antibody-based ChIP-seq. A schematic of the comparison at the captured HS2 enhancer and *HBG* promoters is shown on the *top*. The density maps are shown for CAPTURE-ChIP-seq, Cas9 or FLAG antibody-based ChIP-seq, respectively. The y axis denotes the normalized ChIP-seq intensity as reads per kilobases per million reads (RPKM).

(B) The fractions (%) of sgRNA on-target reads were significantly higher in CAPTURE-ChIP-seq than in Cas9 or FLAG antibody-based ChIP-seq. The fold increases in the % of on-target reads at sgHS2 or sgHBG targeted regions in the top 10, 50 or 100 ChIP-seq peaks in CAPTURE-ChIP-seq versus antibody-based ChIP-seq are shown.

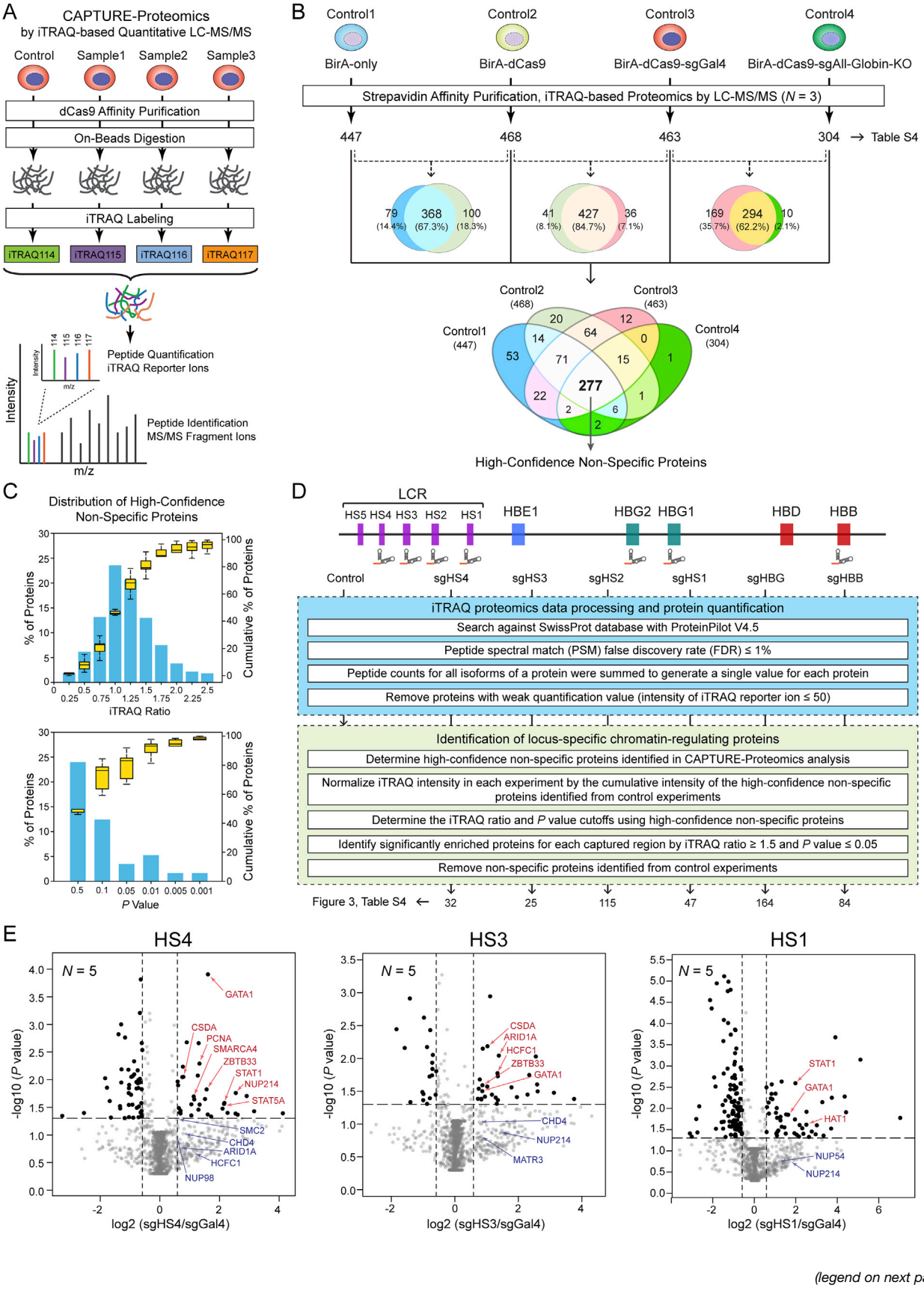
(C) CAPTURE-ChIP-seq displayed significantly less off-targets compared to antibody-based ChIP-seq. Scatterplots show the genome-wide differential analysis of dCas9 binding at sgHS2 or sgHBG targeted regions by CAPTURE-ChIP-seq, Cas9 or FLAG antibody-based ChIP-seq. Data points for the sgRNA target regions and predicted off-targets are shown as *green* and *red*, respectively. Other enriched ChIP-seq peaks are shown as *gray*. The *x*- and y axis denote the mean normalized read counts from N = 2 independent CAPTURE-ChIP-seq.

(D) Genome-wide differential analysis of dCas9 binding in cells expressing two or three independent sgRNAs (sg1, sg2 and sg3) for sgHS1, sgHS3, sgHS4, sgHS5 or sgHBB targeted regions. Data points for the sgRNA target regions and the predicted off-targets for each sgRNA are shown as *green, red* and *orange*, respectively. The *x*- and y axis denote the mean normalized read counts from N = 2 or 3 independent CAPTURE-ChIP-seq.

(E) Genome-wide differential analysis of dCas9 binding in cells expressing sgHS1, sgHS3, sgHS4, sgHS5, sgHBB, or sg3′HS1 versus the non-targeting sgGal4. Data points for the sgRNA target regions and the predicted off-targets are shown as *green* and *red*, respectively. N = 2 to 4 independent ChIP-seq experiments.

(F) Genome-wide differential gene expression analysis was performed using RNA-seq in K562 cells expressing dCas9 with sgHS2, sgHBG, sgHS1-5, the non-targeting sgGal4 or the wild-type (WT) cells. The β-like globin genes are indicated by colored data points. The Pearson correlation coefficient (*R*) value is calculated for each comparison (N = 2 or 3 independent RNA-seq experiments).

(G) Expression of β-globin mRNAs remained unchanged in K562 cells expressing biotinylated dCas9 and target-specific or non-targeting sgRNAs. The mRNA expression of β-globin genes and erythroid regulators (*GATA1* and *KLF1*) was analyzed by qRT-PCR. Results are mean ± SEM of N = 3 independent experiments.



(legend on next page)

Figure S2. CAPTURE-Proteomics Identify CRE-Associated Protein Complexes at the β-Globin Cluster, Related to Figure 3

(A) Schematic of iTRAQ-based CAPTURE-Proteomics. Samples prepared from cells expressing target-specific sgRNAs or sgGal4 were isolated by dCas9 affinity purification, followed by in-solution trypsin digestion. The resulting peptides were purified and labeled by multiplexed isobaric tags. The iTRAQ-labeled peptides were mixed, and subjected to multi-dimensional separation and high-resolution MS analysis for peptide identification and quantification.

(B) Identification of the high-confidence non-specific proteins in CAPTURE-Proteomics. Non-specific proteins were identified by streptavidin purification followed by iTRAQ-based proteomic analyses from K562 cells expressing BirA-only (Control1), BirA with dCas9 alone (Control2), BirA with dCas9 and sgGal4 (Control3), and BirA with dCas9 and 8 individual β-globin CRE-targeting sgRNAs in which the β-globin cluster was deleted (Control4, BirA-dCas9-sgAll-Globin-KO). The non-specific proteins from each experiment were defined as the proteins with iTRAQ ion intensity $\geq$ 100 in at least 2 of 3 replicate experiments. Venn diagrams show the overlap of the non-specific proteins identified from two or four samples. The 'high-confidence non-specific proteins' were defined as the proteins identified from all four control samples.

(C) The distribution of the high-confidence non-specific proteins in all CAPTURE-Proteomics experiments across iTRAQ ratios (x axis, *top*) or *P* values (x axis, *bottom*) is shown. *Blue* bars represent the percentage (%) of non-specific proteins (left y axis) in each category. Boxplots represent of the cumulative % of non-specific proteins (right y axis). Boxes show median of the data and quartiles. Whiskers show the minimum and maximum of the data.

(D) Schematic of data processing, quantification, and identification of locus-specific proteome. The numbers of the significantly enriched locus-specific proteins for each captured region are shown. A diagram of the β-globin cluster showing the positions of sgRNAs used for CAPTURE-Proteomics is shown on the *top*.

(E) CAPTURE-Proteomics identified β-globin CRE-associated proteins. Volcano plots are shown for the CAPTURE-Proteomics in sgHS1, sgHS3 or sgHS4 versus sgGal4-expressing cells. Relative protein levels in the target-specific sgRNA versus sgGal4 samples are plotted on the x axis as mean log2 iTRAQ ratios across *N* replicate experiments. Negative log10 transformed *P* values are plotted on the y axis. Significantly enriched proteins ($p \leq 0.05$; iTRAQ ratio ≥ 1.5) are denoted by *black* dots, all others by *gray* dots. Dotted lines indicate 1.5-fold ratio (x axis) and *P* value of 0.05 (y axis). Representative locus-specific chromatin-regulating proteins are denoted by *red* arrowheads. Representative proteins with iTRAQ ratio $\geq$ 1.5 and $p > 0.05$ are denoted by *blue* arrowheads. The complete lists of proteins are shown in Table S4.

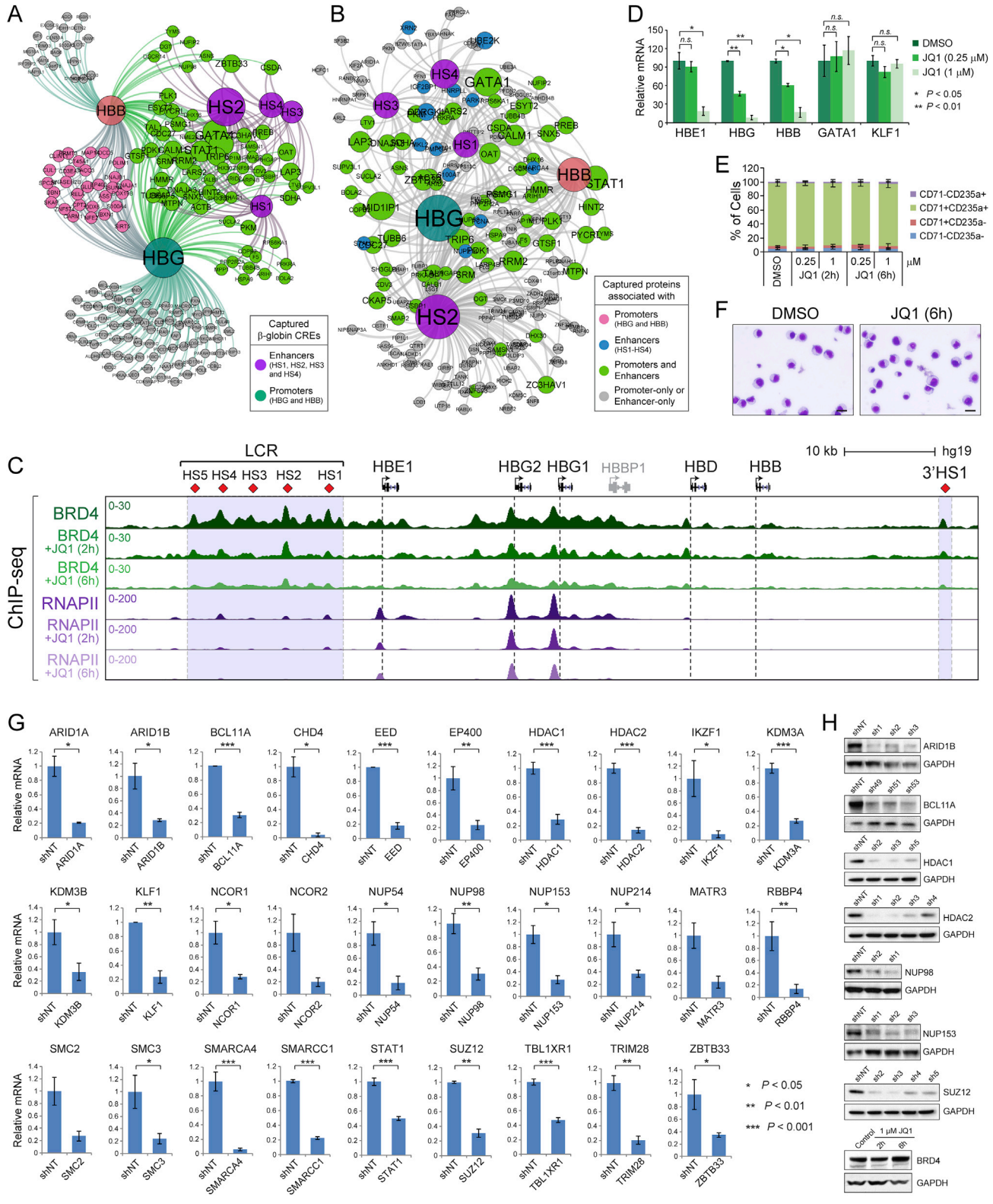


Figure S3. CAPTURE-Proteomics Identify Candidate Regulators for β-Globin CREs, Related to Figure 3

(A and B) Connectivity network of promoter- or enhancer-associated proteins converged by β-globin CREs. The connectivity was built using interactions (*gray* lines) between the identified promoter- or enhancer-associated proteins and β-globin CREs. The promoter- or enhancer-associated proteins were defined as the

(legend continued on next page)

proteins identified to be significantly enriched at any of the captured β-globin promoters (*HBG* and *HBB*) or LCR enhancers (HS1-HS4), respectively. Colored nodes denote proteins significantly enriched at single or multiple CREs. Size of the circles denotes the frequency of interactions. Inset tables show the lists of representative proteins associated with β-globin promoters (*red*), enhancers (*blue*) or both (*green*).

(C) The chromatin occupancy of BRD4 was validated by ChIP-seq. BRD4 and RNAPII ChIP-seq was performed in K562 cells treated with DMSO or 1 μM of JQ1 for 2 or 6 hours, respectively.

(D) JQ1 treatment led to significant downregulation of β-globin genes but not *GATA1* or *KLF1* in human primary erythroid cells. Results are mean ± SEM of three experiments and analyzed by a two-tailed t test. *$p < 0.05$, **$p < 0.01$, *n.s.* not significant.

(E) Erythroid maturation was assessed using the cell surface markers CD71 and CD235a.

(F) Example cytospin of DMSO or JQ1-treated erythroid cells. Scale bars, 20 μm.

(G) Validation of RNAi knockdown by qRT-PCR. Results are mean ± SEM of 1 to 5 shRNAs for each gene in 2 or 3 experiments, and analyzed by a two-sided t test.

(H) Validation of RNAi knockdown of the indicated proteins by western blot analysis in K562 cells.

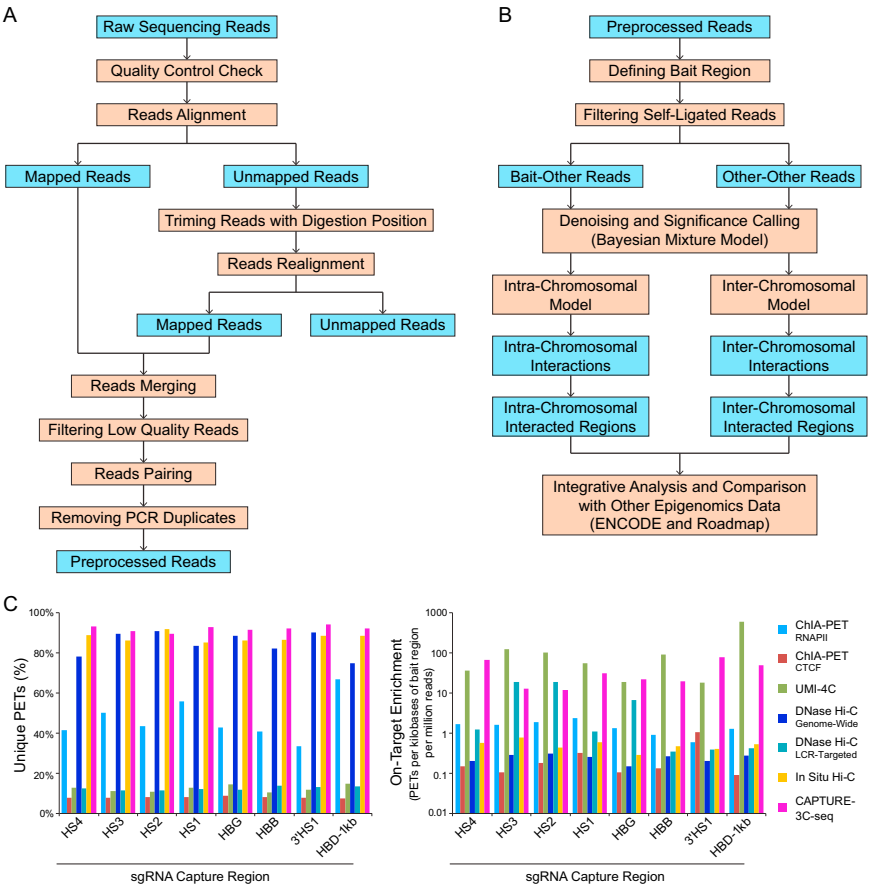


Figure S4. Data Analysis Pipelines for CAPTURE-3C-Seq, Related to Figure 5

(A) Data preprocessing pipeline for CAPTURE-3C-seq is shown. The output data files and the processing steps are shown as *blue* and *red* boxes, respectively.

(B) Statistical analysis pipeline for CAPTURE-3C-seq is shown.

(C) The comaprison between CAPTURE-ChIP-seq, ChIA-PET (RNAPII and CTCF), UMI-4C, DNase Hi-C (genome-wide or LCR-targeted) and *in situ* Hi-C is shown. Compared with RNAPII and CTCF ChIA-PET data in K562 cells (Consortium, 2012; Li et al., 2012), CAPTURE-3C-seq shows significantly higher % of unique PETs and on-target enrichment as measured by the number of PET interactions per kilobases of bait region per million mapped reads. Compared with Hi-C data in K562 cells (Ma et al., 2015; Rao et al., 2014), CAPTURE-3C-seq shows comparable or slightly higher % of unique PETs but significantly higher on-target enrichment. Compared to UMI-4C (Schwartzman et al., 2016), CAPTURE-3C-seq displayed higher % of unique PETs but comparable or slightly lower on-target enrichment. The unique PETs were defined as pair-end sequence tags with distinct genomic locations at one or both sides of the pair-end reads.

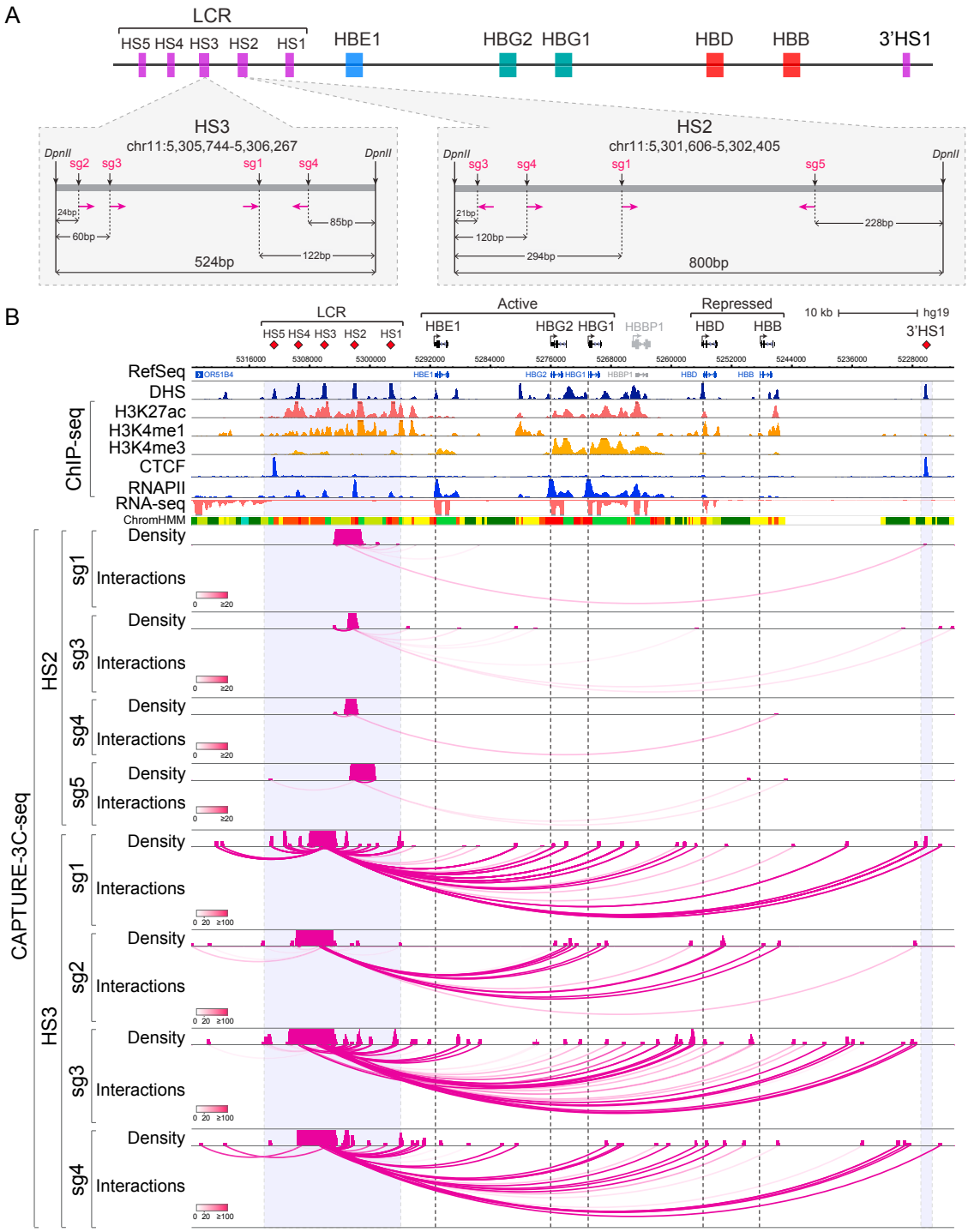


Figure S5. CAPTURE-3C-Seq of Locus-Specific DNA Interactions by Multiple sgRNAs, Related to Figure 5

(A) Schematic of CAPTURE-3C-seq analysis of HS2 or HS3-mediated long-range DNA interations by four independent sgRNAs at various positions of the captured region. The distance between sgRNAs and the DpnII sites is shown.

(B) Browser view of the long-range DNA interactions at HS2 or HS3 captured by four independent sgRNAs. Contact profiles compiled from two or three CAPTURE-3C-seq experiments for each sgRNA including the density map and interactions (or loops) are shown. The statistical significance of interactions was determined by the Bayes factor (BF), and is indicated by the darkness of each interaction loop according to the color scale bars. Interactions with BF $\geq$ 20 were considered high-confidence long-range DNA interactions. The DHS, ChIP-seq (H3K27ac, H3K4me1, H3K4me3, CTCF, and RNAPII), RNA-seq, and ChromHMM data are shown for comparison. The locations of the LCR (HS1 to HS5) and the 3′HS1 insulator are shown as *shaded* lines. The TSS for β-globin genes are shown as *dashed* line.

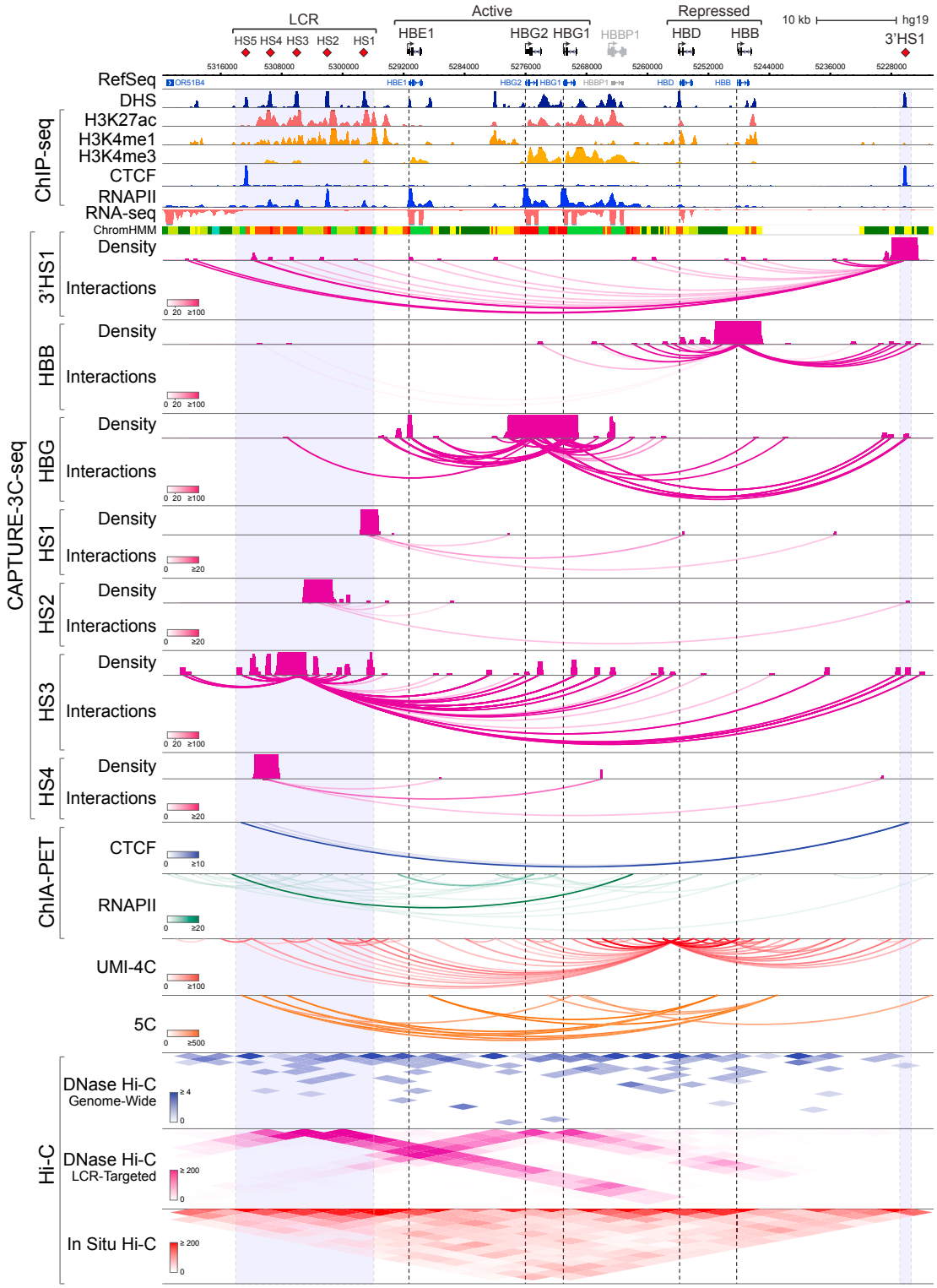


Figure S6. CAPTURE-3C-Seq of Locus-Specific DNA Interactions at Multiple β-Globin CREs, Related to Figure 5

Browser view of the long-range DNA interaction profiles at dCas9-captured β-globin CREs is shown (chr11:5,222,500-5,323,700; hg19). Contact profiles compiled from two or three CAPTURE-3C-seq experiments including the density map and interactions (or loops) are shown. ChIA-PET (Consortium, 2012; Li et al., 2012), UMI-4C (Schwartzman et al., 2016), 5C (Naumova et al., 2013), DNase Hi-C (Ma et al., 2015), *in situ* Hi-C (Rao et al., 2014), DHS, ChIP-seq, RNA-seq, and ChromHMM data are shown for comparison.

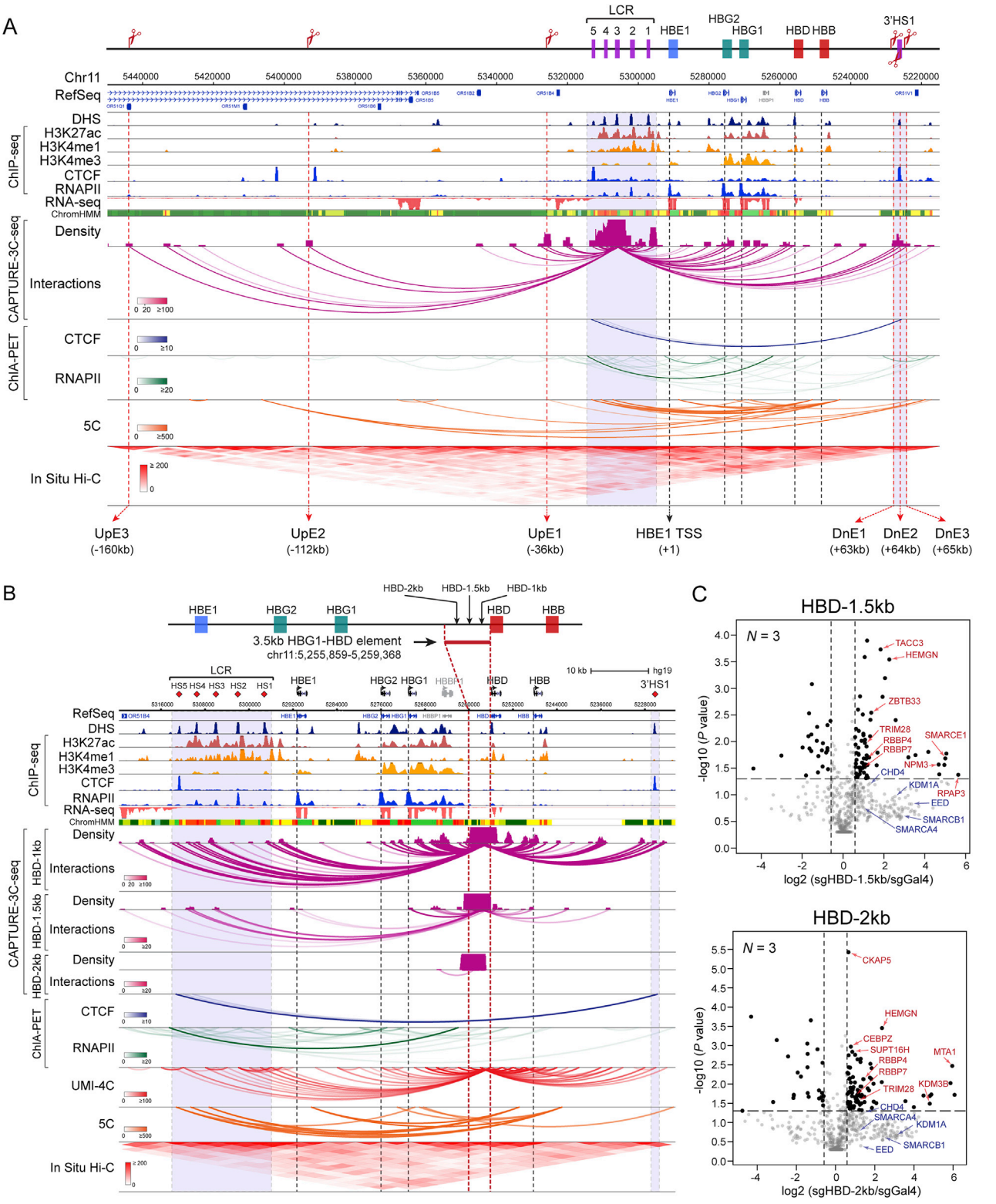


Figure S7. CAPTURE-3C-Seq of Locus-Specific DNA Interactions at HS3 and HBD-1kb, Related to Figures 5 and 6

(A) A zoom-out browser view of the long-range DNA interactions at HS3 (chr11:5,214,997-5,449,997; hg19) is shown. Contact profiles compiled from 3 experiments including the density map, interactions (or loops) and pair-end tags (PETs), along with the ChIA-PET, 5C, Hi-C, DHS, ChIP-seq, RNA-seq, and ChromHMM data are shown for comparison.

(legend continued on next page)

(B) Browser view of the long-range DNA interactions at the HBD-1kb, HBD-1.5kb and HBD-2kb regions (chr11:5,222,500-5,323,700; hg19) is shown. Schematic of the 3.5kb *cis*-element along with the deletions mapped in prior studies are shown on the top. A 3.5kb putative *cis*-element (chr11:5,255,859-5,259,368; hg19) was defined by the upstream breakpoint of the HPFH-1 deletion and the TSS of *HBD*. The sgRNAs (HBD-1kb, HBD-1.5kb and HBD-2kb) used for CAPTURE-3C-seq and CAPTURE-Proteomics are indicated by arrowheads.

(C) CAPTURE-Proteomics identified HBD-1.5kb and HBD-2kb-associated proteins. Volcano plots are shown for the iTRAQ-based proteomics of affinity purification in sgHBD-1.5kb or sgHBD-2kb versus sgGal4-expressing cells (N = 3 replicate experiments). The complete lists of identified proteins are shown in Table S4.