Supplemental Methods

Construction of Plasmids and the generation of stably transfected NE cell lines

pcDNA harboring the full-length human CHD7 transcript (Flag-His-CHD7 fusion transcript) was a kind gift from Dr. Joanna Wysocka. The CHD7 transcript was first cloned into the pENTR-D-TOPO vector (pENTR-CHD7-wt) before inserted into the PiggyBac vector PB513B-1 (System Biosciences) using LR reaction (PB513-CHD7-wt). Mutant CHD7 (CHD7G1391fs and CHD7R1494X) were generated using a modified PCR-based site-directed mutagenesis strategy (Higuchi et al. 1988) using full-length wild-type CHD7 as a template. Two primer sets (primers are listed in Supplemental Table S7) carrying the desired point mutations and one outer primer set were used to PCR amplify fragments of CHD7 from the pENTR-CHD7-wt vector. PCR products were cloned into the PB513-CHD7-wt vector to replace the corresponding regions of wild-type CHD7.

Stable transfection of AF22 cells and CHARGE-NE cells with either mutant CHD7 proteins or wild-type CHD7 was performed using a Nucleofector device (Amaxa) following the manufacturer’s guidelines. The PB513-CHD7G1391fs, PB513-CHD7R1494X and PB513-CHD7-wt vectors were co-electroporated with pCMV-hyPBase vector (Yusa et al. 2011) at a ratio of 2.5:1 into the respective NE cell lines. Control cell lines were generated using an empty PB513B-1 vector. Stable clones were selected with puromycin (1 µg/mL) 48 hours after electroporation. Resistant cells were selected for at least 3 passages before further analysis.

CRISPR-Cas9-mediated protein tagging

In general, the CRISPR-Cas9 system was exploited as previously described (Cong et al. 2013; Ran et al. 2013). Plasmids expressing wild-type Cas9 under the control of the CBh promoter and a human U6 promoter-driven sgRNA cloning vector were purchased from Addgene (http://www.addgene.org/). sgRNA was designed to target sequences
immediately downstream of the stop codon of CHD7 using the CRISPR tool (http://crispr.mit.edu). sgRNA was constructed using the primers listed in Supplemental Table S7. The SURVEYOR assay was performed to examine the targeting efficiency and specificity of gRNA. The cleavage efficiency was examined by transfecting the plasmid in 293T cells, followed by the SURVEYOR assay. Briefly, genomic DNA was extracted using a DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer’s instructions, and CRISPR-targeted sites in genomic regions were amplified with PrimeSTAR Max DNA polymerase (Takara). PCR products were subjected to the SURVEYOR assay following the manufacturer’s protocol (Transgenomics). The primers used are listed in Supplemental Table S7. To facilitate HDR, we purchased a single-stranded DNA oligo donor (ssODN) consisting of a homology sequence, HA tag sequence, and an EcoRI restriction site for screening purposes from Integrated DNA Technologies. The ssODN was designed to include the removal of a single nucleotide within the protospacer adjacent motif (PAM) to eliminate multiple genome editing events at the same loci.

To generate HA knock-in NE cells, sgRNA and Cas9 expression plasmids were electroporated together with a ssODN at a ratio of 1:1 into AF22 cells using a Nucleofector device (Amaxa) following the manufacturer’s guidelines. The electroporated cells were cultured and maintained before FACS sorting. The Single-cell sorting was performed using a flow cytometer SH800 (Sony), and the cells were subsequently cultured in RHB-A medium (Takara) supplemented with EGF (Peprotech) and FGF2 beads (Stem Cultures) in a Matrigel (Corning)-coated 96-well plate (Greiner Bio-One). Each clone derived from a single cell was expanded for further analysis.

**Preparation of nuclear extract**

Cells were scraped and collected from 10 cm culture dishes in 300 μl of ice-cold NP-40 lysis buffer with 1X protease inhibitor cocktail (Roche). After the cells were incubated at 4°C for 10 min, they were homogenized by being drawn through a 23-gauge needle
to rupture the plasma membrane. The soluble cytoplasmic fraction was separated from
the nuclear fraction by centrifugation at 800 \( \times g \) for 10 min at 4°C. The remaining pellet
was washed once with ice-cold lysis buffer without NP-40, followed by centrifugation.
Subsequently, the pellet was resuspended in ice-cold RIPA lysis buffer with 1X
protease inhibitor cocktail. The lysate was sonicated three times with 30-s pulses (30-s
interval between pulses) and cleared of cell debris by centrifugation at 20,000 \( \times g \) for 10
min. The supernatant was collected as a nuclear fraction. The protein concentration was
estimated using the BCA method with a BCA protein assay kit (Pierce) and iMark
microplate absorbance reader (Bio-Rad).

**Quartz-seq**

Control and CHD7-knockdown NE cells (day 8 after transduction) were sorted into a
96-well plate with a flow cytometer SH800 (Sony) and dissolved with single-cell lysis
buffer (0.5% NP40). Whole-transcript amplification was performed as previously
described (Sasagawa et al. 2013). Amplified cDNA from single cells was processed for
library preparation using a Nextera XT Library Prep Kit (Illumina). The DNA
sequencing library was analyzed with a massively parallel sequencer Hiseq 2500. Raw
reads were trimmed based on read quality and read length using Trimmomatic (v0.33)
software. Trimmed reads were aligned to reference genome hg19 using Sailfish (v0.7.6)
with default settings. Samples were filtered with the following parameters and used for
analysis: read number > 1 million, aligned rate > 70%, and detected gene number >
5000. The details of each sample are listed in Supplemental Table S8.

Correlations of differentially expressed genes of individual cells following CHD7
knockdown versus iPSC-derived NPCs and MSCs were evaluated using ExAtlas
(https://lgsun.irp.nia.nih.gov/exatlas/). The criteria of significance for the correlation
analysis were adjusted with FDR < 0.05 and change > 2-fold. To plot the correlation in
a heat map, datasets were sorted with hierarchical clustering. Published gene expression
profiles of iPSC-derived NPCs and iPSC-derived NCCs were loaded from the GEO.
Chromatin immunoprecipitation assay

Cells grown on 10 cm dishes (≈9,000,000 cells; 3 × 10 cm dishes) were used for each ChIP reaction. For single fixation with formaldehyde, cells were fixed with 1% formaldehyde-containing medium for 10 min. For double fixation, the cells were first fixed with 2 mM EGS for 30 min followed by with 1% formaldehyde-containing medium for 10 min. All fixation steps were performed at room temperature. Cells were washed and incubated with NP-40 buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 0.5% NP-40) for 10 min at RT with agitation. The buffer was aspirated and cells were scraped, collected in fresh NP-40 buffer and centrifuged at 1000 ×g for 5 min at 4°C. The pellet was resuspended in SDS lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and protease inhibitor cocktail) and topped up with 400 µl of ChIP dilution buffer (50 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.1% Triton X-100, 0.11% sodium deoxycholate and protease inhibitor cocktail). The lysates were sonicated with either a Bioruptor (COSMO BIO) (treatment time: 15 min; 30 sec ON, 60 sec OFF, 10 cycles, power high) or a Covaris S2 focused-ultrasonicator (intensity: 4; duty cycle: 5%; cycles per burst: 200; treatment time: 15 min) followed by centrifugation at 20,000 ×g at 4°C for 10 min. For each ChIP reaction, 50 µl of Dynabeads M-280 sheep anti-mouse IgG, Dynabeads M-280 sheep anti-rabbit IgG, or Dynabeads sheep anti-rat IgG (all from Life Technologies) were washed with PBS twice. Beads were collected by magnet and conjugated to 5 µg of desired primary antibody in 500 µl of ice-cold RIPA buffer I (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100 and 0.1% sodium deoxycholate) with 1x protease inhibitor cocktail. The suspensions were rotated at 4°C overnight. The next day, beads were collected by
magnet and washed twice with 500 µl of ice-cold RIPA buffer I. Following the last wash, sheared chromatin was added to the magnetic beads and allowed to react overnight at 4°C with rotation. The immunoprecipitated samples were washed once with RIPA buffer I, once with RIPA buffer II (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100 and 0.1% sodium deoxycholate), and twice with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Each wash was accomplished with rotation at 4°C for 5 min. Immunoprecipitated DNAs or input DNAs were eluted by adding 200 µl of ChIP elution buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM EDTA, 0.5% SDS) and were incubated at 65°C overnight to reverse the crosslinks. Samples were then treated with RNase A at 37°C for 30 min and with Proteinase K at 55°C for 3 h. DNA was purified using a ChIP DNA Clean & Concentrator Kit (Zymo Research). Antibodies used for ChIP-seq experiments are listed in Supplemental Table S6.

**ChIP-Seq peak calling and data analysis**

Sequenced raw reads were trimmed based on read length and read quality using Trimmomatic (v0.33). The trimmed reads were aligned to the reference genome (UCSC hg19) using Bowtie2 (v2.1.0) with the default parameters, and only uniquely mapped reads were used for downstream analyses. The resulting SMA files were converted to the BAM format using SAMtools (v0.1.19). Peak calling was performed using Model-based Analysis of ChIP-seq (MACS2) version 2.1.0.20150731 with default setting (Zhang et al., 2008). Overlaps for each peak were calculated using bedtools (v2.17.0) with default parameters. The genome-wide peak distribution from TSS was calculated using the ChIPseeker R package, and functional analysis (GO) of CHD7 peaks was performed using the GREAT website (http://bejerano.stanford.edu/great/public/html/index.php). Heatmaps for each sample against the CHD7 binding region were made using deepTools (v1.5.11). The H3K27ac density plot at CHD7 binding regions, CNS-specific enhancers, TEs and SEs were
created using deepTools. The average H3K27ac read density at each region and the corresponding flanking region were calculated (bin size=50). The length of typical and SE regions (between Start and End) was scaled relative to its median length. The enhancer category was calculated based on overlap of the CHD7 binding site and the enhancer region; these overlaps were validated in the VISTA enhancer database. The genome-wide peak distribution from TSS was calculated using the ChIPseeker R package, and the functional analysis (GO) of CHD7 peaks was performed using the GREAT website (http://bejerano.stanford.edu/great/public/html/index.php).

**Identification of super-enhancers and associated genes**

SEs were identified using the ROSE algorithm with default parameters based on H3K27ac intensity in which enhancer peaks located within 12.5 kb were stitched together and ranked based on their input-subtracted signal of H3K27ac (Loven et al. 2013; Whyte et al. 2013). We used a promoter exclusion zone of 5,000 bp to exclude any enhancer contained within a window of ± 2,500 bp around an annotated transcription start site from being stitched. Enhancer-associated genes were defined on the basis of the calculated distance from the center of the super-enhancer to the nearest TSS of each gene. The assignment of enhancers to the closest genes in human NE cells is shown in Supplemental Table S2.

**Computational analysis of microarray data**

Raw probe intensity data were exported from Illumina GenomeStudio gene expression software (v 1.9.0) and loaded onto R (v3.2.2) statistical computing software for statistical analysis. Limma software (v 3.26.20) was used for background correction, quantile normalization and log (base 2) conversion. Gene sets with no detectable expression in all samples were excluded after normalization. For the evaluation of the overall gene expression profiles, genes differentially expressed by 1.5-fold between in control versus CHD7-knockdown samples and healthy control versus CHARGE-NE
samples were extracted and used for GO analysis using DAVID Bioinformatics Resources (http://david.ncifcrf.gov).

Correlations of differentially expressed genes following CHD7 knockdown versus iPSC-derived NPCs and NCCs were evaluated using ExAtlas (https://lgsun.irp.nia.nih.gov/exatlas/). The criteria of significance for the correlation analysis were adjusted with FDR < 0.05 and change > 2-fold. To plot the correlation in a heat map, datasets were sorted with hierarchical clustering. Published gene expression profiles of iPSC-derived NPCs and iPSC-derived NCCs were loaded from the GEO database. The datasets used in study for comparison were as follows: iPSC-derived NPCs (GSM1553290, GSM1553289, GSM1553291, GSM1538561, GSM1538558, GSM1538560) and iPSC-derived NCCs (GSM1470884, GSM1470883, GSM1470885, GSM1538548, GSM1538547, GSM1538546).

References
Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.
Higuchi R, Krummel B, Saiki RK. 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* **16**: 7351-7367.
Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**: 320-334.
Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281-2308.
Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, Ueda HR. 2013. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression
heterogeneity. *Genome Biol* 14: R31.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307-319.

Yusa K, Zhou L, Li MA, Bradley A, Craig NL. 2011. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* 108: 1531-1536.