



# Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens

Noa Liscovitch-Brauer<sup>1,2,4</sup>, Antonino Montalbano<sup>1,2,4</sup>, Jiale Deng<sup>1,2</sup>, Alejandro Méndez-Mancilla <sup>1,2</sup>, Hans-Hermann Wessels <sup>1,2</sup>, Nicholas G. Moss<sup>1,2</sup>, Chia-Yu Kung<sup>1,2</sup>, Akash Sookdeo<sup>1,2</sup>, Xinyi Guo<sup>1,2</sup>, Evan Geller<sup>1,2</sup>, Suma Jaini<sup>1,3</sup>, Peter Smibert <sup>1,3</sup> and Neville E. Sanjana <sup>1,2</sup> ✉

**CRISPR screens have been used to connect genetic perturbations with changes in gene expression and phenotypes. Here we describe a CRISPR-based, single-cell combinatorial indexing assay for transposase-accessible chromatin (CRISPR-sciATAC) to link genetic perturbations to genome-wide chromatin accessibility in a large number of cells. In human myelogenous leukemia cells, we apply CRISPR-sciATAC to target 105 chromatin-related genes, generating chromatin accessibility data for ~30,000 single cells. We correlate the loss of specific chromatin remodelers with changes in accessibility globally and at the binding sites of individual transcription factors (TFs). For example, we show that loss of the H3K27 methyltransferase EZH2 increases accessibility at heterochromatic regions involved in embryonic development and triggers expression of genes in the HOXA and HOXD clusters. At a subset of regulatory sites, we also analyze changes in nucleosome spacing following the loss of chromatin remodelers. CRISPR-sciATAC is a high-throughput, single-cell method for studying the effect of genetic perturbations on chromatin in normal and disease states.**

Chromatin accessibility orchestrates *cis*- and *trans*-regulatory interactions to control gene expression, and is dynamically regulated in cell differentiation and homeostasis. Alterations in chromatin state have been associated with many diseases, including several cancers<sup>1</sup>. In recent years, CRISPR screens have been combined with single-cell RNA-sequencing (RNA-seq) to measure the effects of genetic perturbations on gene expression across the transcriptome<sup>2–5</sup>. However, methods to capture changes in the epigenome following CRISPR perturbations have limited throughput<sup>6</sup>. To study how genetic perturbations affect chromatin states, we developed a platform for scalable pooled CRISPR screens with single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) profiles, called CRISPR-sciATAC. In CRISPR-sciATAC, we simultaneously capture transcripts encoding a Cas9 guide RNA (gRNA) and perform single-cell combinatorial indexing ATAC-seq<sup>7</sup> (Fig. 1a and Supplementary Fig. 1). Following cell fixation and lysis, nuclei are recovered and the open chromatin regions of the genomic DNA undergo barcoded tagmentation in a multi-well plate using a unique, easy-to-purify transposase from *Vibrio parahaemolyticus* (ViPar; Fig. 1b and Supplementary Fig. 2). Next, gRNA sequences are barcoded with the same barcode sequence as the ATAC fragments, using *in situ* reverse transcription. The nuclei are pooled and split again into a new multi-well plate, and both the ATAC fragments and gRNA sequences are tagged with a second barcode in two consecutive PCR steps. At the end of this process, a unique combination of barcodes ('cell barcode') tags both the gRNA and ATAC fragments from each cell (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1).

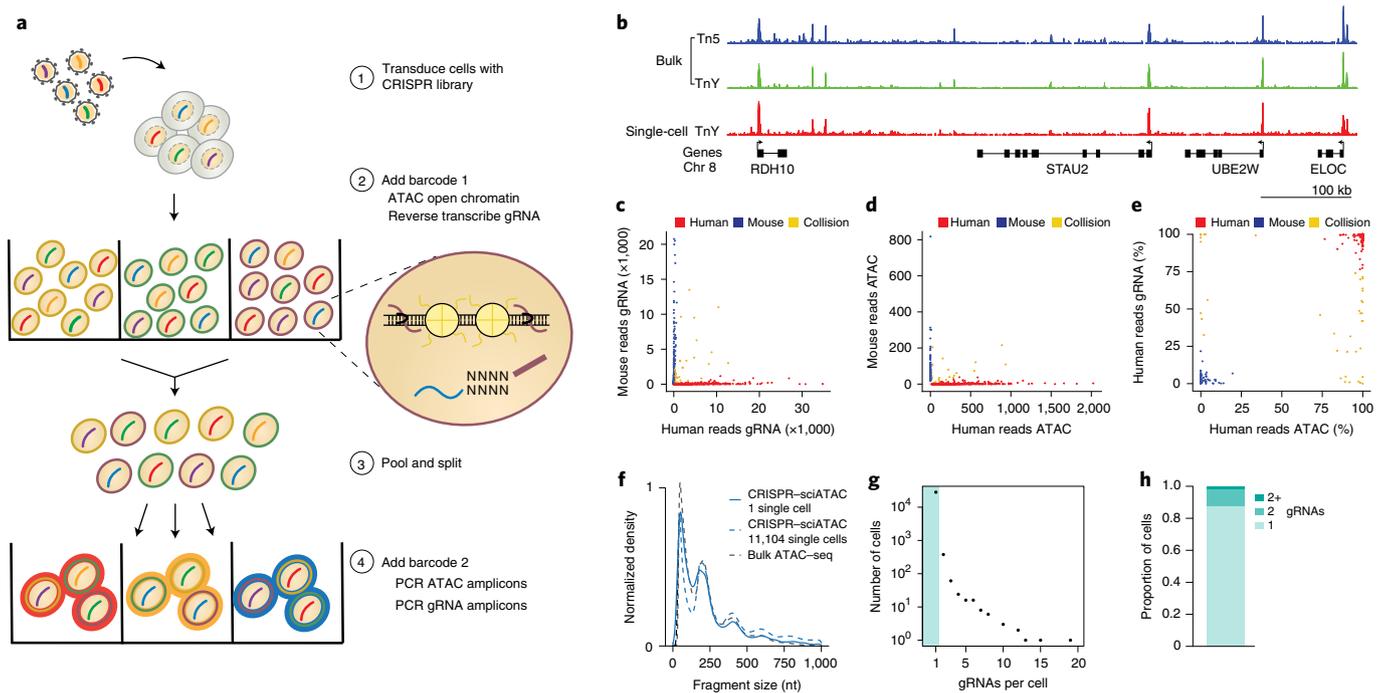
## Results

**CRISPR-sciATAC captures both chromatin accessibility profiles and CRISPR perturbations in single cells.** To quantify capture and barcoding of single cells, we performed CRISPR-sciATAC on a

mix of human (HEK293) and mouse (NIH-3T3) cells. Human and mouse cells were each transduced with a small library of ten distinct nontargeting gRNAs, with no overlapping gRNAs between the two pools. The gRNA sequences were cloned into a lentiviral vector that includes the Pol3-driven gRNA within a longer Pol2 transcript for perturbation readout (CROP-seq vector)<sup>2</sup>. We found that 93% of cell barcodes had gRNA-containing reads that could uniquely be assigned to either human or mouse gRNAs (Fig. 1c), and 96% of cell barcodes had ATAC-seq reads mapping to either the human or mouse genome, indicating that the majority of cell barcodes were correctly assigned to single cells (Fig. 1d). As an additional verification of single-cell separation, we also measured species concordance between ATAC-seq and gRNA reads. We found that for 92% of the captured cell barcodes, both ATAC-seq and gRNA reads aligned to either human or mouse reference genomic and gRNA sequences, respectively. In 4.4% of cells, the ATAC-seq and/or gRNA reads could not be exclusively assigned to one species. ATAC-seq and gRNA reads were assigned to different species (species collision) in 3.6% of cells (Fig. 1e). The low rates of these two failure modes suggest that CRISPR-sciATAC can simultaneously identify accessible chromatin and CRISPR gRNAs in single cells.

To test the ability of CRISPR-sciATAC in capturing biologically meaningful changes in chromatin accessibility, we targeted 21 chromatin modifiers that are highly mutated in cancer (Supplementary Fig. 3a,b). Using the Catalog of Somatic Mutations in Cancer (COSMIC) database<sup>8</sup>, we selected 21 chromatin-related genes that carry the highest mutational load across all cancers, including nine chromatin remodelers (*ARID1A*, *ATRX*, *CHD4*, *CHD5*, *CHD8*, *MBD1*, *PBRM1*, *SMARCA4* and *SMARCB1*), two DNA methylation regulators (*DNMT3A* and *TET2*), three histone methyltransferases (*EZH2*, *PRDM9* and *SETD2*), one histone demethylase (*KDM6A*), one histone deacetylase (*HDAC9*), three histone

<sup>1</sup>New York Genome Center, New York, NY, USA. <sup>2</sup>Department of Biology, New York University, New York, NY, USA. <sup>3</sup>Technology Innovation Lab, New York Genome Center, New York, NY, USA. <sup>4</sup>These authors contributed equally: Noa Liscovitch-Brauer, Antonino Montalbano. ✉e-mail: [neville@sanjanalab.org](mailto:neville@sanjanalab.org)



**Fig. 1 | CRISPR screens with CRISPR-sciATAC enable the joint capture of chromatin accessibility profiles and CRISPR perturbations.** **a**, CRISPR-sciATAC workflow with initial barcoding, nucleus pooling and resplitting followed by second-round barcoding. **b**, Comparison of bulk ATAC-seq chromatin accessibility profiles from K562 cells using Tn5 and TnY transposases and aggregated CRISPR-sciATAC single-cell profiles from 11,104 cells. **c**, gRNA reads mapping to human or mouse CRISPR libraries ( $n=1,986$  cells). **d**, ATAC reads mapping to human or mouse genomes ( $n=721$  cells). For display purposes, we removed one cell that had >tenfold average number of ATAC reads. **e**, Concordance between percentage of ATAC and gRNA reads mapping to human and mouse genomes and human and mouse gRNA libraries, respectively, for each cell ( $n=496$  cells). **f**, ATAC-seq fragment size distribution from K562 cells of bulk ATAC-seq data, aggregated CRISPR-sciATAC single-cell profiles from 11,104 cells and one representative single cell from CRISPR-sciATAC. **g**, Number of CRISPR gRNAs detected per cell. **h**, Proportion of cells with one, two or more than two gRNAs.

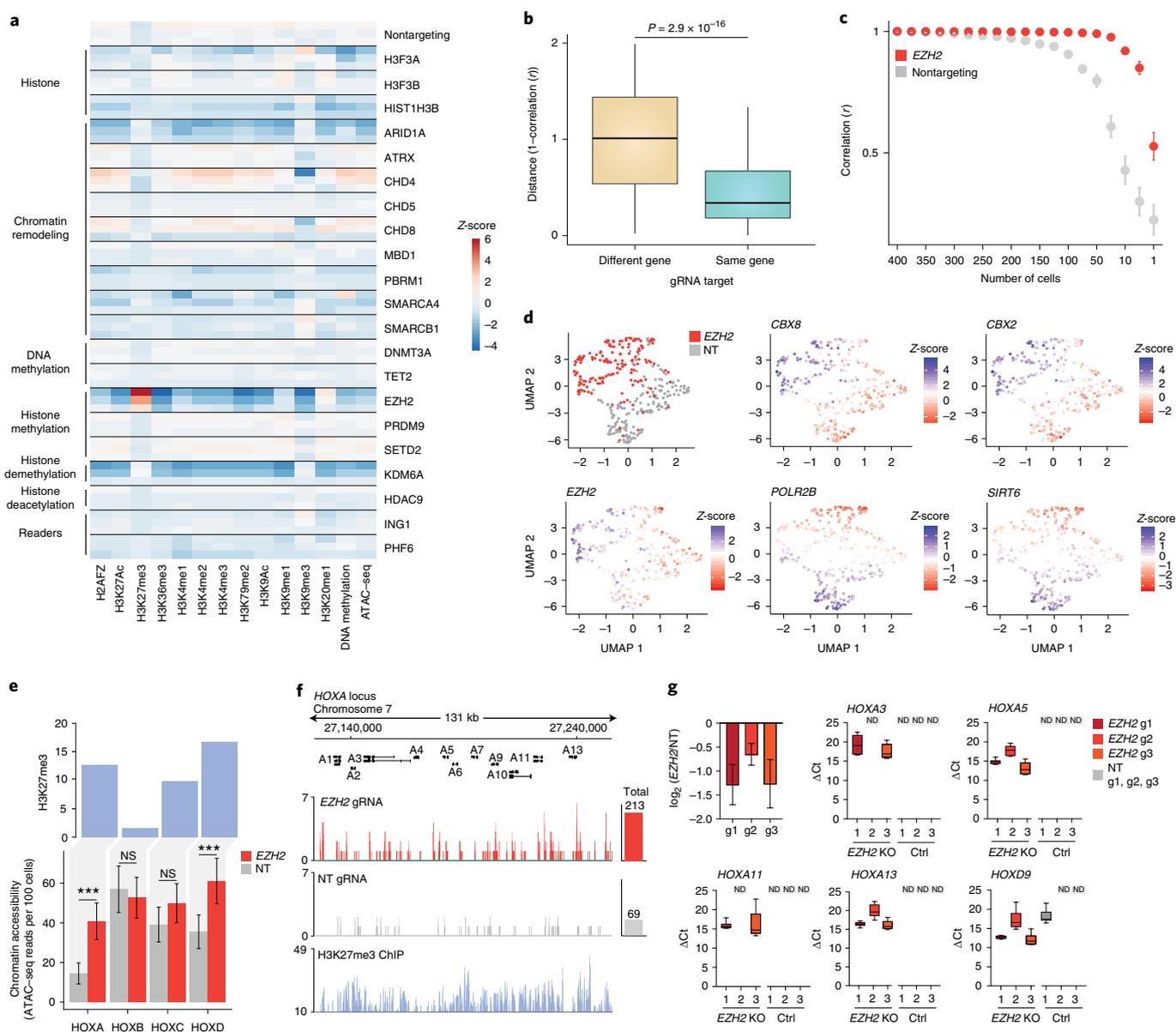
subunits (*H3F3A*, *H3F3B* and *HIST1H3B*) and two readers (*ING1* and *PHF6*). We designed three gRNAs per gene and also included three nontargeting gRNAs in our library (Supplementary Table 2). We transduced Cas9-expressing human myelogenous leukemia K562 cells with this lentiviral gRNA library at a low multiplicity of infection (MOI), and selected with puromycin for transduced cells. After 1 week of selection, we collected single-cell, paired-end ATAC-seq data. Having filtered for cells with  $\geq 500$  unique ATAC-seq fragments and  $\geq 100$  gRNA reads (Supplementary Fig. 3c–h), we obtained 11,104 cells with a median of 1,977 unique ATAC-seq fragments mapping to the human genome. Aggregated ATAC-seq profiles for these cells correlate well with bulk data from K562 cells (Fig. 1b and Supplementary Fig. 3i). Single cells retained an ATAC fragment length distribution similar to cells tagged in bulk (Fig. 1f). The majority of cell barcodes (83%) had one gRNA (Fig. 1g,h) and, for 90% of cell barcodes, a single gRNA represented  $\geq 99\%$  of reads (Supplementary Fig. 3j).

We recovered all 66 gRNAs with a median of 148 single cells per gRNA and 468 single cells per gene. Following closer examination, we noticed that not all gene targets resulted in the same number of single cells captured, suggesting that some of our targets might be essential genes whose targeting leads to dropout of those cells. To distinguish gRNA depletion of essential genes from inability to capture gRNAs using CRISPR-sciATAC, we separately amplified gRNAs from the bulk population at an early time point and at 1 and 2 weeks post selection (Supplementary Fig. 4a). We found high correlation between all samples across three independent transduction replicates (Supplementary Fig. 4b,c). For several genes, multiple, distinct gRNAs targeting the same gene were consistently depleted or enriched: *H3F3A*, *CHD4*, *SMARCA4* and

*SMARCB1* were depleted, while targeting *KDM6A* accelerated cell growth (Supplementary Fig. 4d). Using robust rank aggregation to measure consistent enrichment across multiple gRNAs<sup>9</sup>, we computed gene-level enrichment scores (Supplementary Fig. 4e and Supplementary Table 3), which were highly correlated with a previous genome-wide CRISPR screen in K562 cells<sup>10</sup> ( $r=0.85$ ) (Supplementary Fig. 4f). Reassuringly, enrichment of individual gRNAs was positively correlated with cell numbers estimated from CRISPR-sciATAC cell barcodes ( $r=0.73$ ; Supplementary Fig. 4g). Different gRNAs targeting the same gene tend to result in similar numbers of single cells, highlighting consistent proliferation phenotypes between different genetic perturbations targeting the same gene (Supplementary Fig. 4h,i). We did not observe changes in the number of ATAC fragments per cell between the different perturbed genes (Supplementary Fig. 5a,b), and gene enrichment was not correlated with the number of ATAC fragments, peaks or differential peaks obtained from gRNAs targeting the same gene (Supplementary Fig. 5c–e).

#### Loss of chromatin modifiers alters chromatin accessibility at regulatory elements in single cells.

We examined how loss of these chromatin modifiers impacts accessibility within known chromatin marks (primarily histone post-translation modifications) using ENCODE chromatin immunoprecipitation (ChIP)-seq data from K562 cells (Fig. 2a and Supplementary Tables 4 and 5). We found similar accessibility changes between different gRNAs targeting the same genes, further highlighting the consistency between distinct genetic perturbations targeting the same gene (Fig. 2b). Targeting of the Polycomb repressive complex 2 (PRC2) subunit *EZH2* resulted in an increase in chromatin accessibility at regions with H3K27me3,



**Fig. 2 | CRISPR-sciATAC reveals changes in accessibility at *HOX* genes following loss of EZH2.** **a**, Heatmap of chromatin accessibility z-scores at histone and DNA modifications for different CRISPR perturbations ( $n = 3$  gRNAs per gene). We converted the fraction of accessible regions for each modification into z-scores (using all cells in the screen). For visualization, we show the average z-score for all cells receiving a particular gRNA. **b**, Distances in the histone and DNA modification accessibility profiles (shown in **a**) between gRNAs targeting different genes and those targeting the same gene. The distance metric is  $1 - (\text{Pearson correlation of z-scores})$ . **c**, Pearson correlation between averaged accessibility z-scores at histone and DNA modifications of the indicated number of single cells and the average profile of 400 single cells, for cells with either *EZH2*-targeting or nontargeting gRNAs. **d**, UMAP representation of chromatin accessibility z-scores at histone and DNA modifications from single cells receiving either *EZH2* or nontargeting (NT) gRNAs. Also shown is the same UMAP representation with single cells colored by TFBS accessibility enrichment scores for *CBX2*, *CBX8*, *EZH2*, *POLR2B* and *SIRT6*. **e**, H3K27me3 ChIP-seq coverage at *HOXA-D* loci (top). Changes in accessibility at *HOXA-D* loci in cells transduced with *EZH2*-targeting or NT gRNAs (bottom). \*\*\* $P < 0.001$ . **f**, CRISPR-sciATAC fragments mapping to the *HOXA* locus in cells transduced with *EZH2*-targeting or NT gRNAs ( $n = 510$  cells per condition). The sum of all ATAC fragments over the entire *HOXA* locus in cells transduced with *EZH2*-targeting and NT gRNAs is shown on the right. K562 H3K27me3 ChIP-seq coverage is shown at the bottom. **g**, Gene expression (qPCR) of *EZH2*, *HOXA3*, *HOXA5*, *HOXA11*, *HOXA13* and *HOXD9* for cells transduced with either *EZH2*-targeting or NT gRNAs. *HOX* gene expression in cells targeted by the two more effective *EZH2*-targeting gRNAs (g1 and g3, as defined by decrease in *EZH2* expression as compared to NT gRNAs) is greater than in cells targeted by the less effective gRNA (g2) (Student's *t*-test,  $P < 0.05$  for *HOXA3*, *HOXA5*, *HOXA11* and *HOXA13*,  $P = 0.09$  for *HOXD9*). ND, not detected. NT, nontargeting. KO, knockout. Ctrl, control. Error bars represent SD over biological replicates.  $\Delta Ct$  is defined as  $Ct(\text{HOX}) - Ct(\text{ACTB})$ .

a marker of heterochromatin (Fig. 2a). EZH2 catalyzes nucleosome compaction via H3K27 trimethylation<sup>11</sup>, and thus loss of EZH2 increases accessibility in these regions. Differential accessibility at known chromatin marks can be highly specific: a downsampling

analysis reveals that, for some target genes, such as *EZH2* and *ARID1A*, a small number of cells correlates well (Pearson's  $r \geq 0.75$ ) with an aggregated (pseudobulk) cell population (five single cells for *EZH2*, 25 single cells for *ARID1A*) (Fig. 2c and Supplementary

Fig. 6a,b). For cells receiving a nontargeting gRNA, we find that 75 cells correlate well with the respective pseudobulk populations (Supplementary Fig. 6c). Over all perturbations, we find that the median cell number to represent the pseudobulk is also 75 cells. For certain CRISPR perturbations, more cells are needed to accurately represent the pseudobulk (for example, 225 cells for *TET2*) (Supplementary Fig. 6d), indicating that disruption of these genes creates more variable chromatin accessibility than nontargeting controls.

Hierarchical clustering of single cells transduced with *EZH2*-targeting or nontargeting gRNAs reveals a clear separation (Supplementary Fig. 7a,b), which can also be observed in a uniform manifold projection (UMAP) (Fig. 2d). We verified that this separation is not due to differences in library complexity in cells with *EZH2*-targeting gRNAs (Supplementary Fig. 7c), and found that increased accessibility in Polycomb repressive complex 1 (PRC1) components CBX2 and CBX8 binding sites has the highest predictive power in differentiating *EZH2*-targeted cells. Similarly, differential accessibility of POLR2B and SIRT6 binding sites can also be used to differentiate between cells with *EZH2*-targeting and nontargeting gRNAs, although in the opposite direction, where loss of *EZH2* leads to decreased accessibility in their binding sites. As expected, we find an increase in accessibility at *EZH2* binding sites, which is expected given the role of *EZH2* in repression through heterochromatin formation<sup>12</sup>.

Using Gene Ontology (GO) analysis of differentially accessible regions in *EZH2*-targeted cells, we found an enrichment in genes involved in embryonic development and cell differentiation (Supplementary Fig. 8 and Supplementary Table 6). Indeed, *EZH2* is known to play important roles in embryonic development and cell- and tissue-specific differentiation<sup>11</sup>, and we found large changes in chromatin accessibility at several of the homeobox (*HOX*) genes (Fig. 2e). In K562 cells, the *HOXA* and *HOXD* gene clusters contain the most H3K27me3 repressive heterochromatin among the *HOX* gene clusters. In the *HOXA* gene cluster, we found a near-threefold increase in accessibility (Fig. 2f). A similar increase in accessibility was also seen at the *HOXD* gene cluster (Supplementary Fig. 8d). To understand the functional consequences of these changes, we measured the expression of *EZH2* and several *HOX* genes (*HOXA3*, *HOXA5*, *HOXA11*, *HOXA13* and *HOXD9*) (Fig. 2g). After *EZH2* loss, we found that these previously silenced genes become highly expressed. Among the three gRNAs targeting *EZH2*, we noticed that the least effective gRNA resulted in less *HOX* gene expression, further reinforcing the role of *EZH2* in maintenance of *HOX* gene repression (Fig. 2g and Supplementary Fig. 8e). Taken together, these results suggest that loss-of-function mutations in *EZH2* lead to aberrant expression of genes from the *HOXA* and *HOXD* clusters.

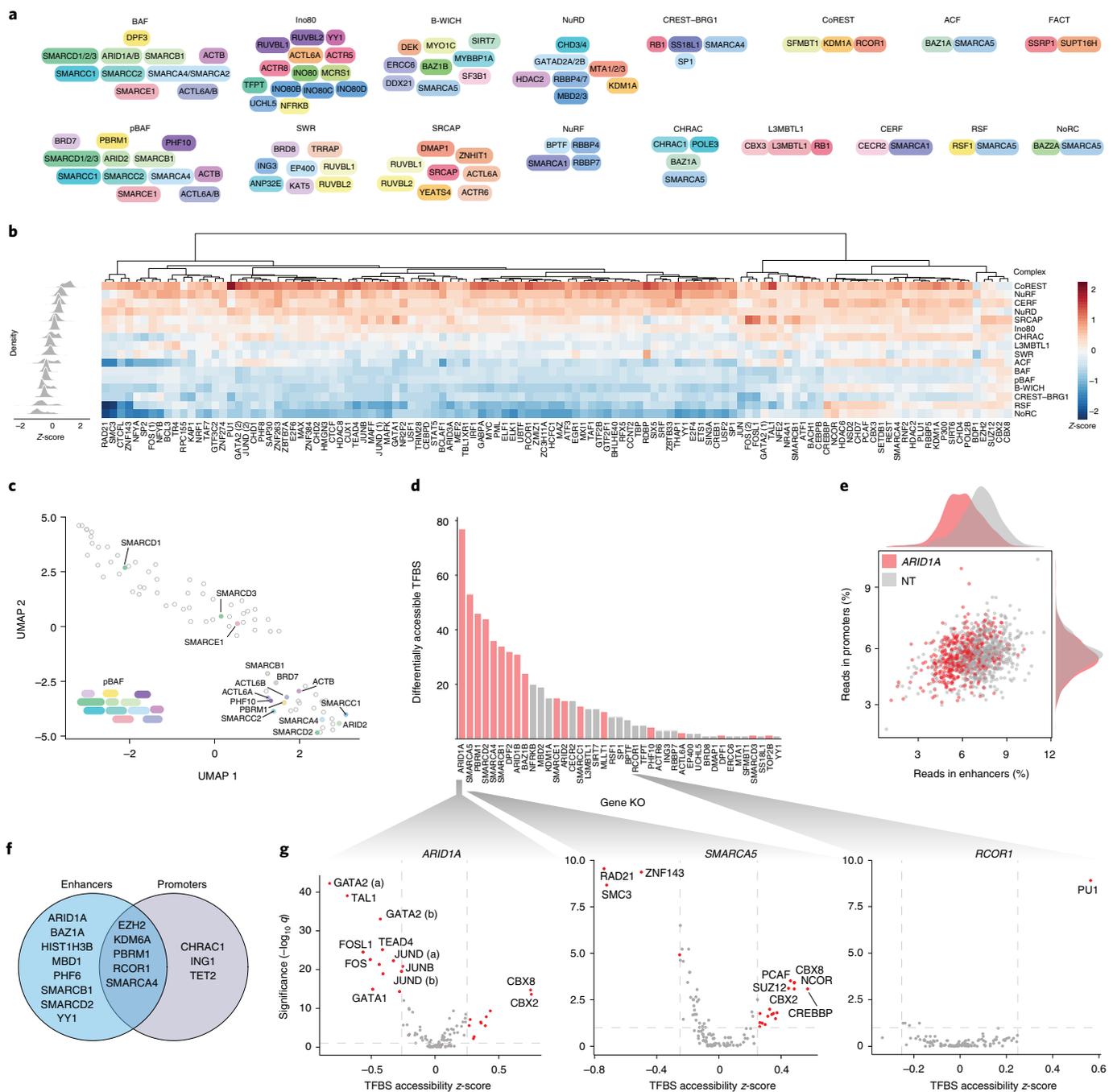
Beyond *EZH2*, we found that changes in accessibility in single cells at TF binding sites (TFBS) are consistent between gRNAs targeting the same gene (Supplementary Fig. 9a,b). We also noticed that large changes in TF accessibility correlate with decreased cell proliferation (Supplementary Fig. 9c), suggesting that perturbation of chromatin modifiers that broadly disrupt TFBS can impact cell viability. We found similar changes in TFBS accessibility using either ENCODE ChIP-seq data from K562 or predicted TF binding sites (JASPAR TF motifs<sup>13</sup> and chromVAR<sup>14</sup>) (Supplementary Fig. 9d,e). To determine whether chromatin accessibility is modified at single-nucleotide polymorphisms (SNPs) that regulate gene expression, we measured overlap with *cis*-regulatory expression quantitative trait loci (*cis*-eQTLs). For two of our targets—*KDM6A* and *ARID1A*—we found a reduction in accessibility at tissue-matched (blood) *cis*-eQTLs in cells after perturbation of these genes (Supplementary Fig. 10a). *KDM6A*-targeted cells had the largest reduction of *cis*-eQTL accessibility with eQTL genes (eGenes) involved in DNA condensation and chemokine receptor activity (Supplementary Fig. 10b–e).

**A loss-of-function screen of chromatin remodeling complexes using CRISPR-sciATAC.** To further demonstrate the scalability of CRISPR-sciATAC, we designed a CRISPR library to target all human chromatin remodeling complexes in the EpiFactors database<sup>15</sup> (Fig. 3a). In total, we targeted 17 chromatin remodeling complexes that each include between two and 14 subunits. As before, we targeted the coding exons of each subunit with three gRNAs and also included gRNAs designed not to target anywhere in the human genome. Over the 17 chromatin remodeling complexes we captured paired CRISPR perturbation and single-cell ATAC-seq data from 16,676 cells. As in the previous screen, the number of cells recovered for each CRISPR perturbation correlated with gene essentiality scores<sup>10</sup> (Supplementary Fig. 11a). We recovered particularly low numbers of cells for the two subunits of the FACT complex, which are known to be highly essential<sup>16</sup> (Supplementary Fig. 11a,b).

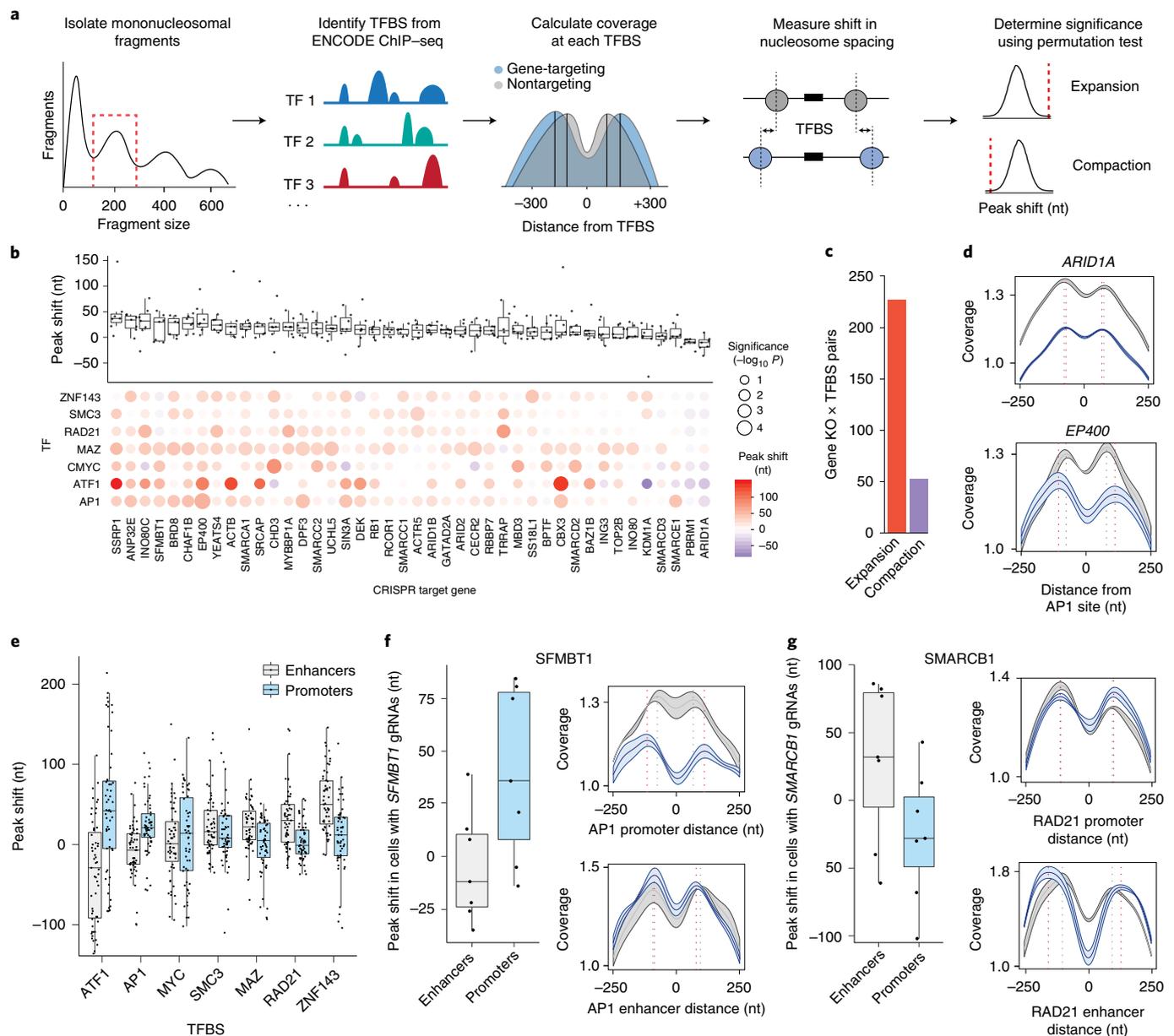
Given the larger scale of this CRISPR-sciATAC screen, we initially analyzed changes in accessibility at the level of different chromatin remodeling complexes as opposed to individual proteins/subunits (Fig. 3b). Examination of differential accessibility in TFBS revealed two major groups: complexes where loss of subunits generally results in increased accessibility, such as the CoRepressor for Element-1-Silencing Transcription factor (CoREST) and the Nucleosome Remodeling Factor (NuRF) complexes, and another group where loss of subunits leads to decreased accessibility, such as the complexes Calcium REsponsive Transactivator-BRG1 (CREST-BRG1) and SWItch/sucrose nonfermentable (SWI/SNF)-B (pBAF). However, loss of individual subunits within these complexes displays tremendous heterogeneity: nearly all complexes have subunits where loss triggers increased accessibility, and other subunits with the opposite effect (Supplementary Fig. 12 and Supplementary Table 7). A two-dimensional UMAP projection of TFBS accessibility profiles reveals a cluster enriched in SWI/SNF components and, in particular, pBAF components (hypergeometric  $P = 4 \times 10^{-4}$ ) (Fig. 3c). Loss of SWI/SNF subunits tends to alter accessibility at many TFBS, with the greatest number of disrupted TFBS from ARID1A loss (Fig. 3d). Previously, ARID1A loss has been shown to impair enhancer-mediated gene regulation<sup>17</sup> and, indeed, we find that loss of ARID1A dramatically reduced accessibility at enhancers but not at promoters (Fig. 3e).

Combining data from both CRISPR-sciATAC experiments, we found that the chromatin modifiers targeted in our two screens resulted in a greater number of accessibility changes at enhancers than at promoters (Fig. 3f), supporting a gene regulatory model with more dynamic chromatin accessibility at distal regulatory elements compared to promoters<sup>18</sup>. Loss of either SWI/SNF-ATPase subunit ARID1A or ISWI-ATPase subunit SMARCA5 results in many changes in TFBS accessibility (Fig. 3g). For ARID1A, some of these changes include a reduction in accessibility at binding sites JUN and FOS, which are subunits of the AP-1 TF that cooperate with the SWI/SNF complex to regulate enhancer activity<sup>19</sup>. Loss of SMARCA5, which helps load cohesin onto chromosomes<sup>20</sup>, triggered a reduction in accessibility in binding sites of cohesin subunits RAD21 and SMC3, along with cohesin cofactor ZNF143 (ref. <sup>21</sup>). In contrast to these gene perturbations that affect a wide range of TFBS, other perturbations result in accessibility changes at only one or a few TFBS. For example, we observed an increase in accessibility at only PU.1 binding sites following loss of RCOR1 (Fig. 3g). RCOR1 has previously been shown to promote erythroid differentiation via repression of myeloid genes such as PU.1, and thus may have a focused role in lineage specification<sup>22</sup>.

**Quantification of nucleosome movement at TF binding sites following chromatin remodeler loss.** In addition to changes in accessibility, chromatin remodeling complexes can regulate gene expression by changing specific nucleosome positions around regulatory sequences<sup>23</sup>. We developed a computational framework



**Fig. 3 | CRISPR-sciATAC screen targeting 17 chromatin remodeling complexes uncovers widespread disruptions in accessibility following SWI/SNF disruption.** **a**, Chromatin remodeling complex subunits and cofactors targeted in the CRISPR library. **b**, Heatmap of chromatin accessibility z-scores at TFBS for the various chromatin remodeling complexes targeted in the screen. We converted the fraction of accessible regions for each TFBS into z-scores (using all cells in the screen). To enhance visualization, we first average over all cells for a particular target gene and then average over all genes in the complex. Histograms (left) show the distribution of z-scores for each complex. The FACT complex is not shown due to a low number of single cells ( $n = 75$  cells). **c**, UMAP representation of genes perturbed in the screen based on TFBS differential accessibility z-score profiles. Subunits of the SWI/SNF pBAF complex are labeled with filled circles and gene names. **d**, Number of TFBS with significant differential accessibility for cells receiving a specific gene-targeting CRISPR perturbation as compared to cells receiving a nontargeting (NT) control gRNA ( $FDR q \leq 0.1$ ). SWI/SNF components and cofactors are highlighted in red. **e**, Percentage of ATAC fragments in enhancers and promoters in cells transduced with *ARID1A*-targeting and NT gRNAs. Each point is a single cell. K562 enhancer and promoter genome segmentation derived from ENCODE (Methods). **f**, CRISPR-targeted chromatin complex genes with significant differential accessibility at enhancers and/or promoters. **g**, Volcano plots showing significant changes in accessibility at TFBS in cells transduced with *ARID1A*- (left), *SMARCA5*- (middle) and *RCOR1*-targeting gRNAs (right). Standardized z-scores are averaged over single cells. Points in red represent TFBS with significant change in accessibility ( $FDR q \leq 0.1$  and  $|z\text{-score}| > 0.25$ ).



**Fig. 4 | Nucleosome dynamics around TFBS following CRISPR targeting of chromatin remodelers.** **a**, Schematic depicting the computational approach to identifying changes in nucleosome positions around TFBS. **b**, Difference in nucleosomal distances in gene-targeted and nontargeting cells (peak shift) across seven TFBS following CRISPR targeting of chromatin remodelers (top). Bubble plot of peak shifts for individual TFBS (bottom). Bubble color corresponds to peak shift (nt), and bubble size represents the empirical  $P$  value calculated by a label permutation test. **c**, The number of nucleosome expansion and compaction events around TFBS following CRISPR targeting of chromatin remodelers. **d**, Coverage profiles of mononucleosomal fragments around AP-1 binding sites in cells transduced with *ARID1A*-targeting (blue) and nontargeting (gray) gRNAs (top) and in cells transduced with *EP400*-targeting (blue) and nontargeting (gray) gRNAs (bottom). Dashed lines represent the most highly covered base in each peak. **e**, Peak shifts in TFBS located in enhancers and promoters. Each point is a CRISPR-perturbed gene (average of all gRNAs for that gene). **f**, Peak shifts in TFBS located in enhancers and promoters in *SFMBT1*-targeted cells (left). Coverage profiles of mononucleosomal fragments in cells transduced with *SFMBT1*-targeting (blue) and nontargeting (gray) gRNAs around AP-1 binding sites in promoters (top right) and in enhancers (bottom right). **g**, Peak shifts in TFBS located in enhancers and promoters in *SMARCB1*-targeted cells (left). Coverage profiles of mononucleosomal fragments in cells transduced with *SMARCB1*-targeting (blue) and nontargeting (gray) gRNAs around RAD21 binding sites in promoters (top right) and in enhancers (bottom right). **d, f, g**, Shaded regions represent s.e.m. ( $n=3$  gRNAs).

to measure changes in nucleosome position in CRISPR-sciATAC, focusing on seven TFs with a bimodal coverage profile around their binding sites, suggesting a symmetric positioning of nucleosomes around them<sup>24</sup> (Fig. 4a,b). Using this pipeline, we found that loss of chromatin remodelers generally results in expansion

of nucleosomes around TFBS (Fig. 4c), with the exception of BAF/pBAF (SWI/SNF) subunits *ARID1A* and *PBRM1*, where knockout led to compaction of nucleosomes around the TFBS studied (Fig. 4b). At specific TFBS, loss of different chromatin remodelers can have opposing effects: for example, *ARID1A* loss results in a

20-nt nucleosome compaction at AP-1 binding sites ( $P=0.03$ ), which was also demonstrated in a recent study, suggesting that the BAF complex controls occupancy of AP-1 (ref. 25). In contrast, loss of EP400, which is part of the Sick With Rat8ts (SWR) complex, causes a large (56-nt) expansion of nucleosomes around AP-1 binding sites ( $P=10^{-4}$ ) (Fig. 4d).

We further asked whether there are specific differences in nucleosome dynamics surrounding TFBS residing in enhancers versus promoters. We found that changes in nucleosome peak positions occur typically in either enhancers or promoters, depending on the specific TFBS. For example, across all CRISPR perturbations, the expansion of nucleosome spacing around AP-1 binding sites occurs mostly at sites located in promoters (Fig. 4e). In contrast, expansion of nucleosome spacing around ZNF143 binding sites occurs mostly at sites located in enhancers. An exception to this trend is ATF1—knockout of chromatin remodelers results in nucleosome expansion around ATF1 binding sites in promoters, but compaction in ATF1 binding sites in enhancers (Supplementary Fig. 13a,b). For specific chromatin modifiers, we often observed greater expansion in either enhancers or promoters (Supplementary Fig. 13c). Knockout of CoREST subunit SFMBT1 tends to cause nucleosome expansion around TFBS in promoters but not in enhancers, for the seven TFs analyzed—for example, an 85-nt expansion around AP-1 binding sites in promoters and no change in nucleosomal positions around AP-1 binding sites in enhancers (Fig. 4f). In contrast, knockout of BAF/pBAF subunit SMARCB1 tends to cause nucleosome expansion around TFBS in enhancers but not in promoters (for example, at RAD21 sites) (Fig. 4g).

## Discussion

In this work we developed CRISPR–sciATAC, a platform for pooled forward genetic screens that jointly captures CRISPR perturbations and ATAC profiles in single cells. Pooled CRISPR screens have been used extensively to identify genes responsible for therapeutic resistance, cell proliferation and Mendelian disorders<sup>26</sup>. While several methods that combine CRISPR screens with single-cell RNA-seq have been developed<sup>2–5</sup>, the ability to capture changes in chromatin accessibility following CRISPR perturbations has been limited. Rubin and collaborators<sup>6</sup> published a related method (Perturb–ATAC) that uses a programmable microfluidic device to physically isolate single cells into small chambers. This method delivers high-depth, single-cell ATAC–seq data ( $\sim 10^4$  fragments per cell), but the throughput per experiment is limited to the 96 chambers of the microfluidic device. CRISPR–sciATAC offers an alternative approach that takes advantage of two-step combinatorial indexing to label DNA molecules with unique cell barcodes and requires no specialized equipment. When compared with Perturb–ATAC, CRISPR–sciATAC can generate thousands of single cells at  $\sim 20\times$  less reagent cost and requires  $\sim 14\times$  less time (Supplementary Tables 8 and 9). In this work we analyzed 28,510 cells, which is  $\sim 7$ -fold more than in the Perturb–ATAC dataset (Supplementary Fig. 14a). Using a library of 318 gRNAs targeting 105 genes, we investigated differential accessibility at histone and DNA modifications and at TFBS following loss of chromatin modifiers. By perturbation of chromatin remodeling complexes in a high-throughput and uniform setting, we reduce batch effects and generate data for a large number of different chromatin complexes. Since it is based on combinatorial indexing ATAC–seq, CRISPR–sciATAC has comparable yield to published sciATAC datasets, which do not have the additional modality of gRNA capture (Supplementary Fig. 14b). As we demonstrate with gRNAs targeting *EZH2*, one important caveat of CRISPR nuclease-driven perturbation is that knockout can be incomplete due to in-frame repair, and efficiency can vary depending on the gRNA. However, CRISPR–sciATAC is not critically dependent on any specific gRNA, but rather searches for consistent effects between gRNAs. To more completely address these issues,

future computational methods used to discern perturbed cells from unperturbed cells, as recently developed for single-cell RNA-seq<sup>27</sup>, will also be useful for single-cell ATAC–seq.

Future CRISPR–sciATAC studies could profile the effects of genetic perturbations on chromatin accessibility in diverse tissues and organoids, and in biological processes where dramatic changes in chromatin accessibility occur, such as differentiation or oncogenic transformation. Although we primarily targeted chromatin modifiers whose loss might be expected to result in large changes in accessibility, perturbations resulting in more subtle accessibility changes can also be detected but may require greater cell numbers. Cell throughput could be increased by uniting the combinatorial approach of CRISPR–sciATAC and droplet-based, single-cell methods. This hybrid strategy could yield high-complexity, single-cell ATAC–seq libraries with the ability to profile many more perturbations, including potentially genome-scale CRISPR libraries. Overall, CRISPR–sciATAC can be applied to the study of diverse phenotypes and diseases, and to understand the interaction between genetic changes and genome-wide chromatin accessibility.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-021-00902-x>.

Received: 22 April 2019; Accepted: 17 March 2021;

Published online: 29 April 2021

## References

- Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
- Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
- Adams, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
- Dixit, A. et al. Perturb–seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
- Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896 (2016).
- Rubin, A. J. et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376 (2019).
- Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
- Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
- Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012).
- Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
- Margueron, R. et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
- Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).
- Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. ChromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
- Medvedeva, Y. A. et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* **2015**, bav067 (2015).
- Lejeune, E. et al. The chromatin-remodeling factor FACT contributes to centromeric heterochromatin independently of RNAi. *Curr. Biol.* **17**, 1219–1224 (2007).
- Mathur, R. et al. ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* **49**, 296–302 (2017).
- Nord, A. S. et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).

19. Vierbuchen, T. et al. AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol. Cell* **68**, 1067–1082 (2017).
20. Hakimi, M. A. et al. A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* **123**, 3175–3184 (2002).
21. Wen, Z., Huang, Z. T., Zhang, R. & Peng, C. ZNF143 is a regulator of chromatin loop. *Cell Biol. Toxicol.* **34**, 471–478 (2018).
22. Swiers, G., Patient, R. & Loose, M. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* **294**, 525–540 (2006).
23. Li, M. et al. Dynamic regulation of transcription factors by nucleosome remodeling. *Elife* **4**, e06249 (2015).
24. Kundaje, A. et al. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* **22**, 1735–1747 (2012).
25. Kelso, T. W. R. et al. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *Elife* **6**, e30506 (2017).
26. Guo, X., Chitale, P. & Sanjana, N. E. Target discovery for precision medicine using high-throughput genome engineering. *Adv. Exp. Med. Biol.* **1016**, 123–145 (2017).
27. Papalexis, E. et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Cell culture and monoclonal K562-Cas9 cell line.** NIH-3T3 and K562 cells were acquired from ATCC (nos. CRL-1658 and CCL-243, respectively). HEK293FT cells were acquired from Thermo Fisher (no. R70007). NIH-3T3 (mouse) and HEK293FT (human) cells were maintained at 37°C and 5% CO<sub>2</sub> in D10 medium (DMEM with high glucose and stabilized L-glutamine (Caisson, no. DML23) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher, no. 16000044)). K562 cells were maintained at 37°C and 5% CO<sub>2</sub> in R10 medium (RPMI with stabilized L-glutamine (Thermo Fisher, no. 11875119) supplemented with 10% FBS). To generate monoclonal K562 cells expressing Cas9, K562 cells were transduced with lentiCas9-Blast (Addgene, no. 52962) at MOI = 0.1, and selected and maintained in R10 with 5 µg ml<sup>-1</sup> blasticidin. Monoclonal K562-Cas9 cells were isolated and expanded by limiting dilution. Expression of Cas9 was confirmed by immunoblot using an anti-2A peptide antibody (Millipore Sigma, no. MABS2005).

**Lentiviral CRISPR libraries.** To generate NIH-3T3 and HEK293FT cells expressing gRNAs for the human/mouse experiment, ten human and ten mouse nontargeting gRNAs (Supplementary Table 2) were individually synthesized and cloned into the lentiviral transfer vector CROPseq-Guide-Puro<sup>+</sup> (Addgene, no. 86708), resulting in the synthesis of an RNA Pol3 transcript of Cas9 gRNA and an RNA Pol2 polyadenylated transcript containing the puromycin resistance gene, a U6 promoter and the gRNA. The RNA Pol2 transcript facilitates the selection of transduced cells (via puromycin) and detection of the gRNA targeting sequence via reverse transcription and PCR (Supplementary Fig. 1c). Equal amounts of each gRNA plasmid were mixed and then, with packaging plasmids pMD2.G (Addgene, no. 12259) and psPAX2 (Addgene, no. 12260), transfected into HEK293FT cells<sup>28</sup>. NIH-3T3 and HEK293FT cells were transduced at MOI = ~0.1 and selected and maintained in D10 with 1 µg ml<sup>-1</sup> puromycin. The gRNA library coverage was 1,500× on average for the species-mixing experiment, the chromatin modifier screen and the chromatin remodeling complex subunit screen.

For the chromatin modifier pooled CRISPR screen, we identified 21 frequently mutated chromatin modifiers across all cancers in the COSMIC database<sup>8</sup> (Supplementary Fig. 3a,b) and designed three targeting gRNAs per gene. An important issue to consider when designing gRNAs is that not all CRISPR nuclease-driven modifications will result in loss of function, because some genome modifications (nonhomologous end joining) will result in in-frame repair that may preserve gene function. However, it was previously demonstrated that even in-frame mutations can be disruptive when targeting functional domains of proteins<sup>29</sup>. To capitalize on this discovery, we chose a CRISPR library design algorithm that uses protein functional domains (from the Pfam database) to target our gRNAs<sup>30</sup>. The final library comprised 63 targeting and three nontargeting gRNAs that were individually synthesized (IDT) and annealed (Supplementary Table 2). For the chromatin remodeling complex subunit pooled CRISPR screen, we designed a CRISPR library to target all chromatin remodeling complexes in the human genome, as defined by the EpiFactors database<sup>15</sup> (Fig. 3a). The library comprised 252 targeting and three nontargeting gRNAs that were individually synthesized (IDT) and annealed (Supplementary Table 2). Annealed oligos were pooled in an equimolar ratio and cloned as a pool into the CROPseq-Guide-Puro lentiviral transfer vector. K562-Cas9 cells were transduced at MOI = ~0.1, and selected and maintained in 1 µg ml<sup>-1</sup> puromycin and 5 µg ml<sup>-1</sup> blasticidin. The CRISPR-sciATAC protocol was performed on these cells at 1 week post selection.

**Transposase identification and isolation.** We were motivated to use a transposase other than Tn5 due to the difficulty in obtaining sufficient yields of Tn5 (ref. <sup>31</sup>). To identify new transposases, sequences were aligned using ClustalW<sup>32</sup> (v.2.1). We found a range of transposon sequences related to the Tn5 sequence and selected a transposon from ViPar for further analysis. The inside and outside ends (IE and OE, respectively) of the ViPar transposon utilize the same sequence as IE and OE of the Tn5 transposon, giving us confidence that the ViPar transposon would be compatible with existing Tn5-based workflows (Supplementary Fig. 2a,b). The identified ViPar transposase was synthesized (Twist Bioscience) and cloned into the vector pTXB1 (NEB, no. N6707S). Two mutations were introduced: (1) P50K, equivalent to the mutation E54K in Tn5, which is predicted to render the transposon hyperactive<sup>33</sup> and (2) M53Q, which changes the residue that interacts with nucleotide 9 (a thymine) on the nontransferred strand of the mosaic end (ME) similar to Tn5 Q57, predicted to increase binding to the Tn5 ME. The ViPar transposase with P50K and M53Q mutations, henceforth referred to as TnY, showed Tn5 ME loading and tagmentation activity (Supplementary Fig. 2c–f). Finally, we characterized the insertion site preference of TnY by performing tagmentation on NA12878 DNA and sequencing on a MiSeq Instrument (Illumina); we found that TnY has insertion site preferences distinct from, but of a similar magnitude to, those of Tn5 (Supplementary Fig. 2g,h). The chromatin accessibility profiles resulting from TnY and Tn5 are highly correlated (Supplementary Fig. 2i,j).

**TnY transposase production.** The pTXB1-TnY vector was transformed into BL21(DE3)-competent *Escherichia coli* cells (NEB, no. C2527), and TnY was produced via intein purification with an affinity chitin-binding tag<sup>31</sup>. One liter of Luria broth (LB) culture was grown at 37°C to optical density (OD<sub>600</sub>) = 0.6. TnY

expression was then induced with isopropyl-β-d-thiogalactopyranoside (IPTG) 0.5 mM at 18°C overnight. After induction, cells were pelleted and then frozen at –80°C overnight. Cells were then lysed by sonication in 100 ml pf HEGX (20 mM HEPES-KOH pH 7.5, 0.8 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100) with a protease inhibitor cocktail (Roche, no. 04693132001). The lysate was pelleted at 30,000g for 20 min at 4°C. The supernatant was transferred to a new tube, and 3 µl of neutralized 8.5% polyethylenimine (Sigma-Aldrich, P3143) was added dropwise to each 100 µl of bacterial extract, gently mixed and centrifuged at 30,000g for 30 min at 4°C to precipitate DNA. The supernatant was loaded on four 1-ml chitin columns (NEB, no. S6651S). Columns were washed with 10 ml of HEGX, then 1.5 ml of HEGX containing 100 mM DTT was added to the column with incubation for 48 h at 4°C to allow cleavage of TnY from the intein tag. TnY was eluted directly into two 30-kDa molecular-weight cutoff (MWCO) spin columns (Millipore, no. UFC903008) by the addition of 2 ml of HEGX. Protein was dialyzed in five dialysis steps using 15 ml of 2× dialysis buffer (100 HEPES-KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 20% glycerol) and concentrated to 1 ml by centrifugation at 5,000g. The protein concentrate was transferred to a new tube and mixed with an equal volume of 100% glycerol. Then, we added Triton X-100 (0.04% final concentration). TnY aliquots were stored at –80°C.

**Transposome assembly.** To produce mosaic-end, double-stranded (MEDS) oligos, we annealed the single T5 tagmentation oligo with the pMENT common oligo (100 µM each; Supplementary Table 2) as follows, in TE buffer: 95°C for 5 min then cooling at 0.2°C s<sup>-1</sup> to 4°C (MEDSA). The same process was used to anneal each barcoded T7 tagment sciATAC oligo with the pMENT common oligo (MEDS B; Supplementary Table 2). MEDSA and MEDSB were mixed, diluted 1:6 in TE buffer then 2 µl was transferred to a new tube and mixed with 3 µl of TnY enzyme. After 30 min at room temperature to allow for transposome assembly, we added 45 µl of dilution buffer, mixed by pipetting up and down, and stored at –20°C until ready for tagmentation. The dilution buffer consists of 2× dialysis buffer (TnY transposase production) diluted 1:1 by volume with 100% glycerol. We observed optimal tagmentation when transposome assembly was carried out on the same day as CRISPR-sciATAC tagmentation.

**PfuX7 polymerase production.** For CRISPR-sciATAC we used a purified PfuX7 DNA polymerase<sup>34</sup>. First, we transformed BL21(DE3)-competent *E. coli* cells (NEB, no. C2527) with pET-PfuX7 and grew them in 1 l of LB culture at 37°C to OD<sub>600</sub> = 0.6. PfuX7 expression was then induced with IPTG (0.5 mM final concentration) at 30°C overnight. After induction, cells were pelleted and resuspended in 20 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM PMSE, 10 µg ml<sup>-1</sup> EDTA-free protease inhibitor (Sigma, no. 11873580001)) and sonicated in an ice slurry. Sonication was performed at 20% amplitude for ten cycles of 1-min duration with a 30-s pause between cycles (Branson Ultrasonics, Model 450 Digital Sonifier). The lysate was pelleted at 30,000g for 15 min at 4°C. The supernatant was transferred to a new tube and incubated with DNA digestion buffer (20 µl of DNase I (NEB, no. M0303), 0.5 mM CaCl<sub>2</sub>, 2.5 mM MgCl<sub>2</sub>) for 30 min at 37°C. DNase I was then inactivated by incubation for 30 min at 85°C. After inactivation, the lysate was placed on ice for 20 min then centrifuged at 50,000g for 20 min at 4°C. The supernatant was loaded on two 1-ml Ni-NTA columns (Qiagen, no. 30210) and washed twice with wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl). PfuX7 enzyme was eluted in 5 ml of elution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.25 M imidazole) and desalted in storage buffer (100 mM Tris-HCl pH 8.0, 0.2 mM EDTA, 2 mM DTT) by performing buffer exchange three times using one Amicon 30-kDa MWCO spin column (Millipore, no. UFC903008). The purified protein was then transferred to a new tube, combined with an equal volume of 100% glycerol and adjusted with Tween-20 (0.1% final concentration) and IGEPAL CA630 (0.1% final concentration). Aliquots were stored at –20°C.

**Bulk ATAC-seq.** To perform bulk ATAC-seq<sup>35</sup> we resuspended 500,000 cells in 1 ml of PBS with gentle lysis, by the addition of 10 ml of resuspension buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>) with 0.1% Tween-20. Cells were then centrifuged at 500g for 10 min at 4°C to pellet nuclei. Pelleted nuclei were resuspended in 600 µl of 1× tagmentation buffer (10 mM TAPS-NaOH pH 8.5, 5 mM MgCl<sub>2</sub>, 10% dimethylformamide), from which 30 µl (~25,000 nuclei) was then transferred to 1.5-ml tubes and 20 µl of TnY transposomes was added. Tagmentation was performed at 37°C for 30 min. Samples were then purified using the DNA Clean & Concentrator kit (Zymo Research, no. D4014) and eluted in 10 µl of Tris and EDTA (TE) buffer. Eluted DNA was thermocycled with PfuX7 in Phusion GC Buffer (Thermo Fisher, no. F519L) as follows: 72°C for 5 min and (98°C for 30 s, 98°C for 10 s, 63°C for 30 s, 72°C for 3 min) × ten cycles, 4°C hold. Samples were purified using the DNA Clean & Concentrator kit, eluted in 6 µl of TE buffer and size selected using a 0.9× volume of Ampure XP Beads (Beckman Coulter, no. A63882) to remove excess oligos.

**CRISPR-sciATAC species-mixing experiment in HEK293 and NIH-3T3 cells.** HEK293FT (human) and NIH-3T3 (mouse) cells transduced with nontargeting gRNAs libraries were grown separately. On the day of the experiment, cells were counted and 500,000 were resuspended in 1 ml of PBS per cell line (1:1 ratio of

human and mouse cells). Cells were then pelleted, resuspended in fixation buffer and fixed for 7 min at room temperature. Fixation buffer consists of 2.8 ml of H<sub>2</sub>O, 790 µl of 100% ethanol, 310 µl of 40% glyoxal (Sigma, no. 128465) and 30 µl of glacial acetic acid (Sigma, no. A6283). After preparation of fixation buffer, the pH is adjusted to 5.0 by the addition of NaOH and kept ice-cold until immediately before use. In line with a previous study<sup>36</sup>, we found that glyoxal fixation resulted in better preservation of intact nuclei than the more commonly used fixative, paraformaldehyde.

After fixation, cells were washed three times with 1 ml of PBS and gently lysed by addition and resuspension in 10 ml of resuspension buffer (Bulk ATAC-seq) with 0.1% Tween-20 and 0.1% Igepal CA630. Cells were then incubated on ice for 3 min and pelleted at 500g for 10 min at 4°C to obtain nuclei. Nuclei were washed in 1 ml of tagmentation buffer (Bulk ATAC-seq) with 5 µl of RiboLock RNase Inhibitor (Thermo Fisher, no. EO0381) and centrifuged at 500g for 5 min at 4°C. Human and mouse nuclei were resuspended and mixed in a final volume of 3.2 ml of tagmentation buffer with 28 µl of RiboLock RNase Inhibitor. Nuclei (30 µl, ~20,000) were distributed among each well of a 96-well plate containing 20 µl of TnY assembled with MEDSA and 96barcoded MEDSB (see Supplementary Table 2 for MEDS sequences). Tagmentation was performed for 30 min at 37°C and then stopped by the addition of 2 µl of 500 mM EDTA to each well. After incubation for 15 min at 37°C, EDTA was quenched before reverse transcription by the addition of 2 µl of 50 mM MgCl<sub>2</sub> to each well.

For reverse transcription, 5 µl of the nucleus solution (~2,000 nuclei) was transferred to a new 96-well plate containing barcoded reverse transcription primers. Reverse transcription primers contain the same barcode as the MEDSB oligos (see Supplementary Table 2 for reverse transcription oligos). Nuclei were transferred, maintaining plate orientation to match tagmentation and reverse transcription barcodes. The reverse transcription master mix (RTMM) consisted of 1 ml of 5× RT buffer, 270 µl of dNTPs, 1.6 ml of water, 262 µl of RevertAid reverse transcriptase and 27 µl of RiboLock RNase Inhibitor (all components: Thermo Fisher, no. EP0442). We distributed 15 µl of RTMM to each well, with mixing and incubation for 30 min at 37°C.

Reverse transcription was stopped by the addition of 2 µl of Stop and Stain buffer (1 ml of 500 mM EDTA, 2 µl of 5 mg ml<sup>-1</sup> DAPI) and incubation for 5 min on ice. Nuclei were pooled and pelleted at 500g for 5 min at 4°C. The supernatant was carefully removed, taking care to not disturb the pellet. Nuclei were gently resuspended in 250 µl of PBS and counted using a hemocytometer. PBS was added to obtain a final concentration of ten nuclei µl<sup>-1</sup>. Next, 2 µl of the nuclear solution (~20 nuclei) was transferred to a new 96-well plate with DNA extraction and digestion buffer in each well. Specifically, each well contained 24.5 µl of DNA rapid extract buffer (1 mM CaCl<sub>2</sub>, 3 mM MgCl<sub>2</sub>, 1% Triton X-100, 10 mM Tris-HCl pH 7.5) and 2 µl of digestion buffer (1 µl of H<sub>2</sub>O, 0.5 µl of SDS 5.8%, 0.5 µl of Proteinase K 20 mg ml<sup>-1</sup> (Sigma, no. P2308)). Nuclei were digested for 5 min at 65°C, and digestion was stopped by the addition of 3 µl of PMSF (Sigma, no. 93482) and incubation for 30 min at room temperature.

For the first PCR, ATAC-seq and gRNA-PCR1 primers were added at a final concentration of 0.5 and 0.1 µM, respectively. Amplification for ATAC-seq/gRNA-PCR1 was performed with PfuX7 in Phusion GC Buffer as follows: 72°C for 5 min, 98°C for 30 s (98°C for 10 s, 63°C for 30 s, 72°C for 3 min) × 14–18 cycles, 4°C hold. For the second PCR, 2 µl of PCR product was transferred to a new 96-well plate while maintaining plate orientation to match ATAC-seq and gRNA barcodes. gRNA-PCR2 primers were added to a final concentration of 0.5 µM. Amplification for gRNA-PCR2 using PfuX7 in Phusion GC buffer was 98°C for 30 s (98°C for 10 s, 55°C for 10 s, 72°C for 20 s) × 20 cycles, 72°C for 5 min, 4°C hold.

We then purified ATAC-seq and gRNA amplicons. The ATAC-seq/gRNA-PCR1 PCR plate was purified using four columns of the DNA Clean & Concentrator kit, eluted in 10 µl of elution buffer and size selected using a 0.9× volume of Ampure XP Beads. The gRNA-PCR2 PCR plate was purified using ten columns of the DNA Clean & Concentrator kit, eluted in 20 µl of elution buffer. Eluted samples were run on E-gel 2% (Thermo Fisher, no. G402002) and the expected band (~250 base pairs (bp)) gel extracted, purified using one column of the ZymoClean Gel DNA Recovery Kit (Zymo Research, no. D4008) and eluted in 20 µl. Libraries were separately sequenced on the MiSeq Sequencer (Illumina) using the read lengths shown in Supplementary Fig. 1d,<sup>37,38</sup>.

**CRISPR-sciATAC for chromatin modifiers in K562 cells.** The CRISPR-sciATAC protocol for the chromatin modifier library in K562 cells was performed similarly to the human/mouse experiment described above. K562-Cas9 cells transduced with the pool of 63 chromatin modifier gRNAs and three nontargeting gRNAs (library 1), or with the pool of 252 chromatin modifier gRNAs and three nontargeting gRNAs (library 2), were cultured for 1 week after selection. We prepared either 12 (library 1) or 41 (library 2) 96-well plates and pooled amplicons. ATAC-seq amplicons were sequenced on a HiSeq 2500 (Illumina) and gRNA amplicons were sequenced on a MiSeq.

**Gene essentiality screen and analyses.** K562-Cas9 cells were transduced with the chromatin modifier pooled CRISPR screen at MOI = ~0.1 and selected and maintained in 1 µg ml<sup>-1</sup> puromycin and 5 µg ml<sup>-1</sup> blasticidin. Genomic DNA was

extracted at 3 days (early time point), 1 week and 2 weeks post selection. The gRNA cassette was PCR amplified<sup>3</sup>. Libraries were sequenced on a MiSeq sequencer (Illumina). In addition to the CRISPR-sciATAC experiment, two independent transduction replicates were also analyzed. To identify essential genes, a *P* value per gRNA was calculated using the MAGeCK algorithm, with *P* values for the three gRNAs targeting one gene aggregated into a gene-level *P* value using a robust rank aggregation approach followed by a Bonferroni correction<sup>9,39</sup>.

**Read alignment.** CRISPR-sciATAC gRNA and ATAC datasets were demultiplexed based on cellular barcodes using the snATAC\_mat.py script in an established sciATAC-seq pipeline (<https://github.com/r3fang/snATAC>)<sup>40</sup>. The processed gRNA sequences were aligned to a custom guide reference using bowtie<sup>41</sup> (v.1.1.2) using the command bowtie -v 1 -m 1. For the human/mouse experiment we show data for cells with at least 20 mapped reads. Cells with >90% of gRNA reads that mapped exclusively to human or mouse gRNAs were considered species specific. Cells where one gRNA represented at least 90% of the total reads were kept for further analyses, while the remaining cells were considered collisions and/or the result of multiple infections. For downstream analysis of the K562 data, we required each cell to have at least 100 aligned gRNA reads with ≥99% of reads assigned to one gRNA sequence for the chromatin modifier screen and at least ten aligned gRNA reads with ≥90% of the reads assigned to one gRNA sequence for the chromatin remodeling complex subunit screen. Gene knockouts with at least 50 identified single cells were considered for further analysis (98/105 targeted genes).

The processed ATAC sequences were aligned to the reference genome using bowtie2 (ref. <sup>42</sup>) (v.2.2.8) using the command bowtie2 -D 15 -R 2 -L 22 -i S,1,1.15 -p 5 -t -X2000 -e 75 --no-mixed --no-discordant. The reference genome was a chimeric human hg19 and mouse mm10 genome for the human/mouse experiment and a human hg19 for the K562 datasets. Improperly paired and nonuniquely mapped alignments and reads mapping to mitochondrial DNA were removed. Reads overlapping ENCODE blacklist regions were removed (<https://www.encodeproject.org/annotations/ENCSR636HFF/>). Reads were then deduplicated using Picard (v.2.16.0) (<http://broadinstitute.github.io/picard/>). For the human/mouse experiment we show data for cells with at least 20 unique ATAC-seq reads; for the K562 datasets we require at least 500 unique ATAC-seq reads.

**Differential accessibility at genomic regions with specific chromatin and DNA modifications.** To assess changes in accessibility, we downloaded from ENCODE ChIP-seq files covering post-translational histone modifications and DNA methylation (Supplementary Table 4). For each ChIP-seq track, we considered the fraction of fragments in each single cell overlapping ChIP-seq peaks. We standardized the averaged fractions over all single cells into *z*-scores and then averaged those obtained for each ChIP-seq file over cells that received the same gRNA for the visualization shown in Fig. 2a. To determine significant deviations in accessibility per gene knockout and per modification, we performed a two-tailed *t*-test on the *z*-scores, of all cells for one gene knockout and of all nontargeting cells for each modification. *P* values were adjusted for multiple-hypothesis testing using a Benjamini-Hochberg false-discovery rate (FDR) correction ( $q \leq 0.1$ ).

For correlation of downsampled cell populations with the aggregated (pseudobulk) data, we randomly sampled cells (from a total of 400 single cells) without replacement. We performed this resampling procedure 200 times for each cell number. For each cell sample, we average the accessibility *z*-scores and then compute the Pearson correlation with pseudobulk. For this analysis, we included only target genes with at least 400 single cells.

**Differential accessibility in TFBS using ENCODE ChIP-seq.** To identify enrichment or depletion in accessibility of TFBS following chromatin modifier knockout, we downloaded 116 TF K562 ChIP-seq peak files from ENCODE (Supplementary Table 4) and considered the fraction of fragments in each single cell overlapping ChIP-seq peaks. We standardized the averaged fractions over all single cells into *z*-scores and then averaged the *z*-scores obtained for each ChIP-seq file over cells that received the same gRNA for the visualization shown in Supplementary Fig. 9a. For dimensionality reduction, we used the function umap (from the R package umap) and, to predict cell perturbation, we fit TFBS *z*-scores with a generalized linear model using the function glm (from the R package stats). To determine significant deviations in accessibility per gene knockout and per TF, we performed a two-tailed *t*-test on the *z*-scores, of all cells for one gene knockout and of all nontargeting cells for each TF. *P* values were adjusted for multiple-hypothesis testing using a Benjamini-Hochberg FDR correction ( $q \leq 0.1$ ). For genes with multiple ENCODE ChIP-seq datasets, we denote these with either (1) ENCODE ChIP-seq profiles obtained using an antibody that directly recognizes the protein of interest, or (2) ENCODE ChIP-seq profiles obtained using an antibody directed against an EGFP-tag.

**Differential accessibility in TFBS using JASPAR motifs.** As an orthogonal method to ENCODE ChIP data, we also utilized predicted TFBS from the JASPAR database (386 motifs from JASPAR 2016, human CORE dataset)<sup>13</sup>. Transcription factor motif enrichment and depletion scores were calculated using chromVAR<sup>14</sup>. Briefly, *z*-scores quantifying deviations in the frequency of each motif in each

of the single cells were calculated based on the frequency of the motif in the collection of peaks existing in each cell, out of all 358,028 peaks called on the aggregated single-cell alignment files (pseudobulk). This frequency was compared to that of the motif in peaks found in the entire aggregated single-cell dataset<sup>44</sup>. We considered cells with a minimum of 2,000 fragments per cell and a minimum of 10% of total fragments in peaks. To avoid bias from recovery of different numbers of cells for each gRNA, we subsampled all gRNA cell populations to 12 cells (the lowest number of cells for a single gRNA in our K562 dataset), calculated the deviation  $z$ -scores and repeated this resampling process 1,000 times to obtain deviation  $z$ -scores for each gRNA.

**GO analysis of differential EZH2 chromatin accessibility sites.** To identify and annotate genomic regions differentially accessible in cells with *EZH2*-targeting gRNAs, we aggregated equal numbers of single cells ( $n = 170$  cells per gRNA) for each of the three *EZH2* and nontargeting gRNAs. We next binned the genome into 150-nt regions and identified all bins covered by all three *EZH2* gRNAs but not covered by any of the three nontargeting gRNAs. These bins were then mapped to the transcription start site of the closest genes. We used this (unranked) gene list ( $n = 3,740$ ) as input for GO enrichment analysis, with all human genes as a background set<sup>43</sup>.

**Differential accessibility at HOX loci and gene expression.** To measure accessibility at *HOX* loci, *EZH2*-targeted and nontargeting single cells were downsampled to 100 cells, aggregated and fragments overlapping the *HOXA-D* loci were counted. Empirical  $P$  values were calculated over 1,000 bootstrap iterations. To select *HOX* genes for expression profiling, we compared CRISPR-sciATAC coverage in *EZH2* knockout and NT cells. We computed the number of reads in each *HOX* gene body (including a 500-nt flanking sequence on each side). We then selected the top five *HOX* genes with the most significant change in CRISPR-sciATAC coverage (Student's  $t$ -test). Gene expression of *HOX* genes (*HOXA3*, *HOXA5*, *HOXA11*, *HOXA13* and *HOXD9*) and *EZH2* following *EZH2* knockout was quantified using quantitative PCR with reverse transcription (RT-qPCR). Briefly, 1 million K562-Cas9 cells were infected with either *EZH2* gRNA 1–3 or NT gRNA 1–3 at MOI = ~0.1 for each of the six gRNAs and grown in six-well plates. At 24 h post infection, cells were selected in  $1 \mu\text{g ml}^{-1}$  puromycin. Cells were harvested 10 days after transduction and lysed using TRIzol (Life Technologies) and RNA was purified using Direct-zol (Zymo Research). We reverse transcribed  $1 \mu\text{g}$  of total RNA using random hexamer primers and RevertAid Reverse Transcriptase (Thermo Fisher) at 25 °C for 10 min, 37 °C for 60 min and 95 °C for 5 min. After complementary DNA synthesis, qPCR reactions were performed using Luna Universal Probe qPCR Master Mix (NEB); custom primers and probes (IDT) were designed to detect each target gene, and normalized to  $\beta$ -actin (*ACTB*) (see Supplementary Table 2 for primer and probe sequences). All qPCRs were thermocycled on a ViiA 7 Real-Time PCR System (Applied Biosystems) as follows: initial denaturation at 95 °C for 1 min, then 40 cycles at 95 °C for 15 s and 60 °C for 30 s. Quantification was performed via  $\Delta\Delta\text{Ct}$  using three biological replicates and four technical (qPCR) replicates for each biological replicate.  $\Delta\Delta\text{Ct}$  was calculated as  $\Delta\text{Ct}(\text{KO}) - \Delta\text{Ct}(\text{NT})$ , where  $\Delta\text{Ct}(\text{NT})$  represented an average over all NT gRNAs and biological replicates.

**eQTL enrichment.** To test whether targeting chromatin modifiers resulted in changes in accessibility at SNPs associated with regulatory function through eQTL association testing, we utilized *cis*-eQTLs (SNP-gene combinations within 1 Mbp) from the eQTLGen Consortium. The Consortium performed association testing for 19,960 genes expressed in blood in 31,684 samples<sup>44</sup>. We considered the fractions of fragments in each single cell overlapping *cis*-eQTLs and compared these for each population of single cells that received gRNAs targeting a gene to the fractions in nontargeting cells, using a Wilcoxon signed-rank test followed by a Benjamini–Hochberg multiple-hypothesis correction. To identify specific *cis*-eQTLs with altered accessibility, we downsampled KDM6A single and nontargeting single cells to the same number of cells ( $n = 737$  cells) and focused on a subset of 7,829 highly covered ( $\geq 50$  reads) *cis*-eQTLs in the two cell populations combined. For each of these 7,829 *cis*-eQTLs we considered the proportion of cells with a read covering the *cis*-eQTL in both the KDM6A cell population ( $n = 921$  cells) and the nontargeting cell population ( $n = 737$  cells) and performed a  $\chi^2$  test of proportion. Allelic effects are from the Genotype-Tissue Expression database<sup>45</sup>. For each *cis*-eQTL, we show allele-specific expression for the two closest genes in whole-blood samples. In cases where allele-specific expression was not available in whole-blood samples, we show the most significant association.

**Differential accessibility in enhancers and promoters.** For each single cell, we calculated the fraction of reads intersecting with promoters and with enhancers, as defined by ENCODE (wgEncodeAwgSegmentationCombinedK562.bed, <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeAwgSegmentation/>). We then compared the fractions in each gene knockout cell population with those in the nontargeting cell population. To determine significant differences in reads in promoters/enhancers (versus nontargeting), we performed a two-sample Wilcoxon test and the  $P$  values were adjusted for multiple-hypothesis testing using a Benjamini–Hochberg FDR  $q \leq 0.1$ .

**Nucleosome dynamics at TFBS, promoters and enhancers.** To investigate nucleosome dynamics around TFBS, we first subset the ATAC fragments into those putatively spanning one nucleosome (mononucleosome fragments, 147–280 bp<sup>33</sup>). We next calculated coverage profiles around TFBS sites with BEDTools<sup>46</sup> (v.2.25.0). We focused on those TFBS with two nucleosomes, spanning them symmetrically according to our data. The TFBS selected for this analysis thus reflect those whose nucleosome positions are strongly bimodal. We chose these sites by calculating a nucleosome-free region (NFR) score for each, a metric to assess bimodality<sup>47</sup>. Specifically, we take the difference in average base-pair coverage between flanking regions (50–150 bp upstream and downstream of site) and the central region (50 bp across the site center). We focused on a small subset of seven TFs with NFR > 0 and, indeed, for TF binding sites we observed strong bimodality. These TFs had also previously been shown to have a bimodal profile<sup>24</sup>.

ATAC-seq fragment coverage plots were smoothed using the smooth.spline function from the R package stats, with smoothing parameter spar = 0.8. Next, positions of maximum coverage upstream and downstream of motif centers were used to estimate nucleosome location. To determine expansion, we first calculated the distance between the upstream and downstream nucleosomes at a particular TFBS; this distance was then compared to nontargeting cells, to obtain a positive (expansion) or negative (compaction) score. Empirical  $P$  values for each score were generated using a label permutation test, where nontargeting and knockout labels were randomly shuffled while keeping group size constant to avoid biases stemming from different numbers of cells in the nontargeting and knockout cell populations. Labels were shuffled 10,000 times and, in each iteration, the distance between upstream and downstream nucleosomes was measured to create a null distribution to which the true distance was compared.

For each TF, we calculate mononucleosomal coverage profiles separately for sites located in promoters or enhancers, as defined by the University of California, Santa Cruz (wgEncodeAwgSegmentationCombinedK562.bed, <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeAwgSegmentation/>).

**Statistical analysis.** Data between two groups were analyzed using either a two-tailed unpaired  $t$ -test or nonparametric Wilcoxon signed-rank test.  $P$  values and statistical significance were estimated for all analyses. Corrections for multiple-hypothesis testing were performed using the Benjamini–Hochberg approach<sup>48</sup>. In all box plots, the middle line is the median, and the lower and upper hinges correspond to the first and third quartiles. The upper whisker extends from the hinge to the largest value no further than  $1.5 \times$  interquartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most  $1.5 \times$  IQR of the hinge. Data beyond the end of the whiskers are outlying points and are plotted individually. All statistical analyses were performed in R/RStudio.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Processed and raw data can be downloaded from NCBI GEO (PRJNA674902, GSE1161002).

## Code availability

The scripts and pipeline for the analysis can be found at <https://gitlab.com/sanjanaalab/crispr-sciatac>.

## References

- Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Shi, J. et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
- Meier, J. A., Zhang, F. & Sanjana, N. E. GUIDES: sgRNA design for loss-of-function screens. *Nat. Methods* **14**, 831–832 (2017).
- Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
- Norholm, M. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* **10**, 21 (2010).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Richter, K. N. et al. Glyoxal as an alternative fixative to formaldehyde in immunostaining and super-resolution microscopy. *EMBO J.* **37**, 139–159 (2017).

37. Adey, A. et al. In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res.* **24**, 2041–2049 (2014).
38. Amini, S. et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014).
39. Liu, X. S. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
40. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
41. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
43. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
44. Vösa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/447367> (2018).
45. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
47. Duttke, S. et al. Identification and dynamic quantification of regulatory elements using total RNA. *Genome Res.* **29**, 1836–1846 (2019).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).

## Acknowledgements

We thank the entire Sanjana laboratory for support and advice. We thank J. Morris for help with eQTL resources, M. Zaran and R. Satija for computational resources and the NYGC Sequencing Platform and NYU Biology Genomics Core for sequencing resources. BL21(DE3) cells transformed with pET-PfuX7 were kindly provided by J. Gregory. N.L.-B. is supported by a postdoctoral fellowship from the Human Frontier Science Program Organization (no. LT000672/2019-L), an EMBO long-term fellowship (no. ALTF 826-2018) and the Weizmann Institute of Science National Postdoctoral Award Program for Advancing Women in Science. N.E.S. is supported by NYU and NYGC startup funds, NIH/NHGRI (nos. R00HG008171 and DP2HG010099), NIH/NCI (no. R01CA218668), DARPA (no. D18AP00053), the Sidney Kimmel Foundation, the Melanoma Research Alliance and the Brain and Behavior Foundation.

## Author contributions

N.E.S. conceived and supervised the project. N.E.S., A.M. and N.L.-B. designed the experiments. A.M., N.L.-B., J.D., A.M.-M., C.-Y.K. and A.S. performed the experiments. N.L.-B., A.M., J.D., N.E.S., H.-H.W. and N.G.M. analyzed the data. P.S. isolated TnY. S.J. purified PhuX7. A.M., J.D., C.-Y.K., A.S., P.S. and S.J. purified TnY. N.L.-B., A.M. and N.E.S. wrote the manuscript with input from all authors.

## Competing interests

The New York Genome Center and New York University have applied for patents relating to the work in this article. N.E.S. is an adviser to Vertex.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-021-00902-x>.

**Correspondence and requests for materials** should be addressed to N.E.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |     |           |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

- Transposase sequences were aligned using ClustalW (version 2.1).
- CRISPR-sciATAC gRNA and ATAC datasets were demultiplexed based on cellular barcodes using the snATAC\_mat.py script (version 12.24.2017, <https://github.com/r3fang/snATAC>).
- The processed gRNA sequences were aligned to a custom guide reference using bowtie (version 1.1.2).
- The processed ATAC sequences were aligned to the reference genome using bowtie2 (version 2.2.8).
- Reads were deduplicated using Picard (version 2.16.0).
- p-value per gRNA was calculated using the MAGeCK algorithm (version 0.5.7).
- For dimensionality reduction, we used the function umap (from the R package umap, version 0.2.6.0).
- To predict cell perturbation, we fit TFBS Z-scores with a generalized linear model using the function glm (from the R package stats, version 3.5.2).
- Transcription factor motif enrichment and depletion scores were calculated using chromVAR (version 1.4.1).
- Gene Ontology enrichment analysis was performed using GOrilla (version 03.8.2013).
- We calculated coverage profiles around TFBS sites with BEDTools (version 2.25.0).
- ATAC-seq fragment coverage plots were smoothed using the smooth.spline function (from the R package stats, version 3.5.2).

#### Data analysis

The scripts and pipeline for the analysis can be found at <https://gitlab.com/sanjanalab/crispr-sciatac>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Processed and raw data generated in this study can be downloaded from NCBI GEO (PRJNA674902, GSE161002). The pTXB1-TnY plasmid generated in this study has been deposited with Addgene (Plasmid #163975).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for each sgRNA was determined by the number of cell surviving the CRISPR targeting of the gene. In every comparison made in the analyses, the number of cells per sgRNA/gene-KO was taken into consideration either by normalization or by down-sampling as described in the figure legends and in the Methods section.
Data exclusions	Human/mouse mixture CRISPR-sciATAC pooled screen: Cells were selected for further analyses if they had at least 20 sgRNA reads and/or at least 20 unique ATAC-seq fragments.  K562 CRISPR-sciATAC pooled screens: For both screens, we required at least 500 unique ATAC-seq reads for each single cell. For the chromatin modifier screen (21 genes), we required each cell to have at least 100 aligned sgRNA reads with at least 99% of the reads assigned to one sgRNA. For the chromatin remodeling complex subunit screen (84 genes), we required each cell to have at least 10 aligned sgRNA reads with at least 90% of the reads assigned to one sgRNA. Gene perturbations with $\geq 50$ identified single cells were considered for further perturbed-gene based analyses (98/105 targeted genes).  For all ATAC-seq data, fragments overlapping ENCODE Blacklist regions (ENCSR636HFF) were removed.
Replication	The cell enrichment/depletion screen was performed with three independent biological replicates.
Randomization	Downsampling/resampling was performed using the R function "base::sample".
Blinding	Blinding is not relevant to our study because it is not a subjective trial and the results presented here are purely based on objective description of our experiments.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

---

Antibodies used	Anti-2A monoclonal antibody, Millipore Sigma MABS2005, clone 3H4, lot 3192032
Validation	Anti-2A peptide, clone 3H4, Cat. No. MABS2005 was validated by the manufacturer (immunofluorescence, immunoprecipitation, and western blot).

## Eukaryotic cell lines

---

Policy information about [cell lines](#)

Cell line source(s)	NIH-3T3 and K562 cells were acquired from ATCC (CRL-1658 and CCL-243). HEK293FT cells were acquired from Thermo Fisher (R70007).
Authentication	None of the cells used were authenticated independent of the labeling from ATCC and ThermoFisher.
Mycoplasma contamination	All cell lines were tested as mycoplasma-free using Lonza MycoAlert (#LT07-518).
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used.