

# Genome-wide ATAC-seq screening identifies TFDP1 as a modulator of global chromatin accessibility

Received: 5 February 2020

Accepted: 8 January 2024

Published online: 15 February 2024

 Check for updates

Satoko Ishii<sup>1,2,3</sup>, Taishi Kakizuka<sup>4</sup>, Sung-Joon Park<sup>1</sup>, Ayako Tagawa<sup>6</sup>, Chiaki Sanbo<sup>2,3</sup>, Hideyuki Tanabe<sup>7</sup>, Yasuyuki Ohkawa<sup>8</sup>, Mahito Nakanishi<sup>9</sup>, Kenta Nakai<sup>5</sup> & Yusuke Miyanari<sup>1,6,10</sup> 

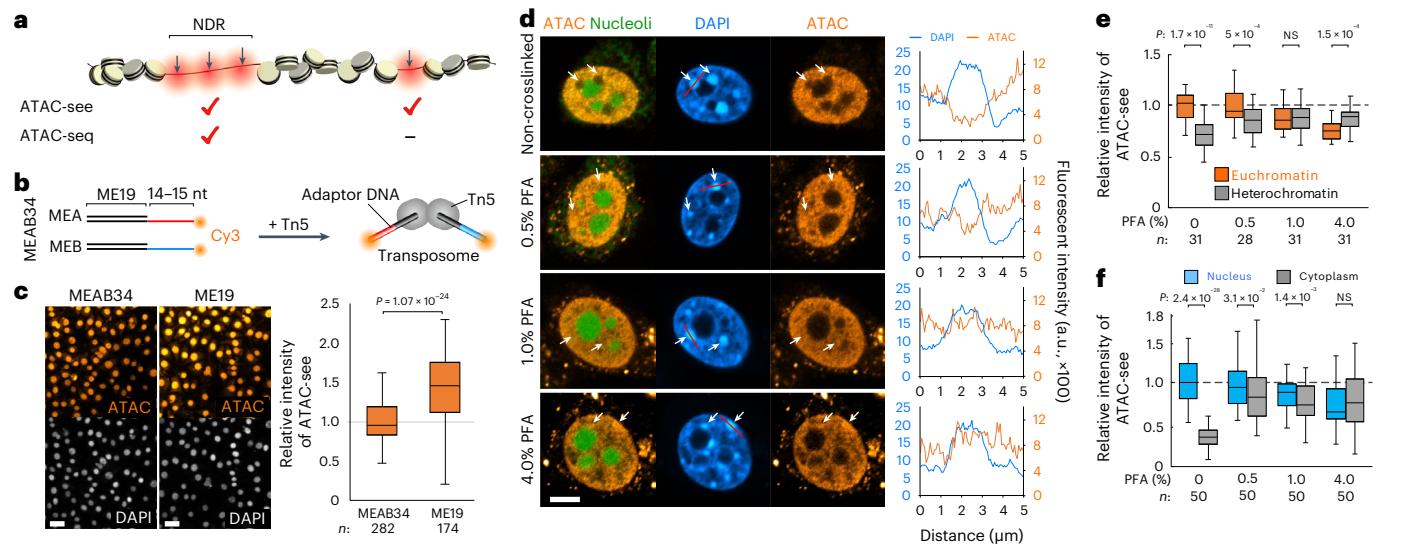
Chromatin accessibility is a hallmark of active regulatory regions and is functionally linked to transcriptional networks and cell identity. However, the molecular mechanisms and networks that govern chromatin accessibility have not been thoroughly studied. Here we conducted a genome-wide CRISPR screening combined with an optimized ATAC-seq protocol to identify genes that modulate global chromatin accessibility. In addition to known chromatin regulators like *CREBBP* and *EP400*, we discovered a number of previously unrecognized proteins that modulate chromatin accessibility, including *TFDP1*, *HNRNPU*, *EIF3D* and *THAP11* belonging to diverse biological pathways. ATAC-seq analysis upon their knockouts revealed their distinct and specific effects on chromatin accessibility. Remarkably, we found that *TFDP1*, a transcription factor, modulates global chromatin accessibility through transcriptional regulation of canonical histones. In addition, our findings highlight the manipulation of chromatin accessibility as an approach to enhance various cell engineering applications, including genome editing and induced pluripotent stem cell reprogramming.

Active regulatory regions such as promoters and enhancers contain nucleosome-depleted regions (NDRs), establishing chromatin accessible for DNA-binding proteins (Fig. 1a). In contrast, outside of these NDRs, nucleosomes occupy most parts of genomic DNA, resulting in a less accessible state. Therefore, the core and linker histones, fundamental components of chromatin, are key determinants of accessibility<sup>1–3</sup>. Accessible chromatin provides scaffolds for various DNA-dependent processes including transcription, DNA repair and replication. Consequently, multiple cellular processes such as differentiation,

reprogramming and disease development are accompanied by dynamic changes in chromatin accessibility<sup>4–6</sup>. Therefore, targeting chromatin accessibility holds promise for manipulating chromatin plasticity and cellular phenotype.

Chromatin accessibility is regulated at local level by chromatin remodelers, histone chaperones, histone modifying enzymes and pioneer transcription factors<sup>7</sup>. However, the molecular mechanisms that shape chromatin accessibility on a global scale remain poorly understood. Transposase-accessible chromatin with visualization

<sup>1</sup>The Graduate University for Advanced Studies, SOKENDAI, Okazaki, Japan. <sup>2</sup>National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Japan. <sup>3</sup>Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Okazaki, Japan. <sup>4</sup>Graduate School of Frontier Biosciences, Osaka University, Suita, Japan. <sup>5</sup>Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. <sup>6</sup>WPI Nano Life Science Institute, Kanazawa University, Kanazawa, Japan. <sup>7</sup>Research Center for Integrative Evolutionary Science, SOKENDAI, Hayama, Japan. <sup>8</sup>Division of Transcriptomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan. <sup>9</sup>TOKIWA-Bio Inc., Tsukuba, Japan. <sup>10</sup>Cancer Research Institute, Kanazawa University, Kanazawa, Japan.  e-mail: miyanari@staff.kanazawa-u.ac.jp



**Fig. 1 | Optimization of ATAC-see assay.** **a**, Schematic diagram of the distinct specificity of ATAC-see and ATAC-seq. Arrows indicate transposed sites. **b**, Schematic illustration of the Tn5 transposome and its adapters. **c**, Representative images of ATAC-see with the indicated transposome on NIH3T3 (left; scale bars, 30 μm) and the relative intensity of ATAC-see per nucleus (right). *n*, number of cells analyzed. The box plot represents the interquartile range (IQR; the midline indicates the median) with whiskers that extend 1.5 times the IQR. Statistics include a two-tailed unpaired Student's *t*-test. **d**, Representative images of ATAC-see (orange), DAPI (cyan) and nucleoli visualized by EGFP-H2B (21–35) (green) in NIH3T3 cells crosslinked with the indicated concentration of PFA (the left panels). Intensities of ATAC-see and DAPI along red lines drawn in images are visualized in the plots. Arrows indicate DAPI-enriched heterochromatins; scale bar, 5 μm. **e,f**, Relative quantification of ATAC-see

signals in constitutive heterochromatin and euchromatin per nucleus (**e**) or in the nucleus and cytoplasm (**f**) under the indicated crosslinking conditions. *n*, number of cells analyzed; NS, non-significant. **g**, Alignment of adapter DNA sequences recognized by Tn5. Non-conserved bases between IE and OE are highlighted in red. **h**, A representative genomic browser track displaying ATAC-seq with mES cells using the indicated transposomes (left) with the proportion of reads mapped on ATAC-seq peaks or on the other genomic regions (right). Error bars refer to the standard deviation. **i**, A box plot showing relative chromatin accessibility upon indicated gene knockouts relative to non-targeting control. The box plot represents the IQR (the midline indicates the median) with whiskers (the 10th and 90th percentiles). Statistics include a two-tailed unpaired Student's *t*-test using the median from two or three independent biological replicates to compare between control and indicated samples. *n*, number of cells analyzed.

(ATAC-see) is a method that uses Tn5 DNA transposase assembled with fluorescently labeled adapter DNAs to selectively insert the adapters into accessible chromatin<sup>8</sup>. This method enables us to visualize and quantify chromatin accessibility at the single-cell level, using microscopy or fluorescence-activated cell sorting (FACS). Importantly, the specificity of ATAC-see is distinct from that of ATAC-seq (Fig. 1a). ATAC-seq principally requires two adjacent tagmentation events for subsequent polymerase chain reaction (PCR) amplification and short-read deep sequencing<sup>9</sup>. This leads to a preferential detection of short DNA fragments, primarily on NDRs, while being less efficient at capturing sparse tagmentation over 1,000 bp that mainly occur outside of NDRs. Consequently, ATAC-seq predominantly identifies 'local' accessible regions as distinct peaks. In contrast, ATAC-see captures all the tagmentation events including the sparse ones distributed across the genome, thus providing a 'global' view of accessibility. In this Article, we performed a genome-wide clustered regularly interspaced short palindromic repeats (CRISPR) screening combined with ATAC-see and discovered key proteins modulating chromatin accessibility. Notably, we showed that TFDP1 is a global modulator of chromatin accessibility by controlling histone transcription. In addition, we proposed that manipulation of chromatin accessibility could be applied

to improve cell engineering applications, including genome editing and somatic cell reprogramming.

## Results

### Optimization of ATAC-see for a genome-wide screening

We first optimized the ATAC-see protocol to improve its sensitivity and specificity for successful screening. In the initial publication reporting on ATAC-see<sup>8</sup>, -33–34-nucleotide (nt) adapter DNAs (MEAB34) were employed to prepare Tn5 transposome. MEAB34 is composed of a 19 bp mosaic end sequence (ME19), essential for assembly with Tn5, and additional -14–15-nt flanking sequences (A or B), which were not required for the ATAC-see assay (Fig. 1b). We found that Tn5 transposome with ME19 displayed relatively higher activity than that with MEAB34 in both ATAC-see on NIH3T3 cells (Fig. 1c) and in vitro tagmentation assay (Extended Data Fig. 1a). Thus, we used ME19 for subsequent ATAC-see assays.

We next investigated the impact of crosslinking on the ATAC-see signals. DAPI-enriched heterochromatin displays lower ATAC-see signals than surrounding euchromatic regions (Fig. 1d and Extended Data Fig. 1b) without crosslinking, which is consistent with the relatively less accessible states of heterochromatic regions<sup>1</sup>.

However, formaldehyde crosslinking led to a dose-dependent increase in ATAC-seq signals in heterochromatin, with comparable or even higher levels than euchromatic regions (Fig. 1d,e). Similar effects of crosslinking were observed on the staining pattern of mitotic chromosomes and the inactive X chromosome, both of which exhibit condensed chromatin and reduced accessibility (Extended Data Fig. 1c-h). These results suggest that crosslinking of cells with higher concentration of formaldehyde could hamper the detection of accessible chromatin by ATAC-seq. Additionally, non-crosslinked conditions significantly reduced cytoplasmic background signals, presumably originating from mitochondrial DNAs, resulting in higher specificity for nuclear accessible chromatin (Fig. 1d,f).

The Tn5 transposome preferentially targets NDRs, but it still inserts its adapter within the accessible linker DNA between adjacent nucleosomes to some extent (Fig. 1a)<sup>9</sup>. We conducted ATAC-seq to investigate the specificity of the optimized ATAC-seq using short adapter DNAs under the non-crosslinked condition. In contrast to conventional ATAC-seq adapters (MEAB34), ME19 lacks primer annealing sites for PCR amplification. Therefore, we used alternative 19 bp adapter DNAs (IE and OE) containing 7-nt differences between them<sup>10</sup>, which allowed specific PCR amplification without affecting Tn5 activity (Fig. 1g and Extended Data Fig. 2a,b). ATAC-seq analysis with mouse embryonic stem (mES) cells showed that the transposome with IEOE19 or Cy3-conjugated IEOE19 (IEOE19-CY3) had a high specificity to NDRs, resulting in approximately 50% of reads enriched in the ATAC-seq peaks, which is slightly higher than that achieved with MEAB34 (Fig. 1h and Extended Data Fig. 2c-h). Overall, the optimized ATAC-seq assay using the short adapter DNAs and non-crosslinked condition allowed sensitive and specific detection of ‘global’ chromatin accessibility at the single-cell level. This approach is particularly suitable for CRISPR screening using FACS.

To evaluate the utility of ATAC-seq as a robust readout for a CRISPR screening, we first analyzed the effects of disrupting 16 chromatin regulators expected to alter chromatin accessibility, including chromatin remodelers and histone acetyltransferases, as potential positive controls. We used CRISPR-associated protein 9 (Cas9)-expressing eHAP, a fully haploid human cell line that enables efficient and complete disruption of gene function by targeting a single allele, making it ideal for genome-wide knockout screening<sup>11</sup>. We confirmed high knockout efficiency of the target proteins by transient transfection of a plasmid DNA encoding single guide RNAs, achieving 90–98% for all genes tested (Extended Data Fig. 3). To ensure the reproducibility and accurate quantification of ATAC-seq signals by FACS, we introduced spike-in cells to calibrate the signals across different samples. Depletion of the chromatin remodeler *EP400*, or the histone acetyltransferase *CREBBP* (also known as *CBP*), resulted in significant increase or decrease in ATAC-seq fluorescence, respectively (Fig. 1i). These results demonstrate that our optimized ATAC-seq can serve as a reliable phenotypic readout for the genome-wide screening to identify modulators of chromatin accessibility. While some of the other regulators analyzed have been shown to regulate local chromatin accessibility<sup>12,13</sup>, individual depletion of these regulators may not be sufficient to alter the global accessibility measured by ATAC-seq.

### Novel modulators of chromatin accessibility

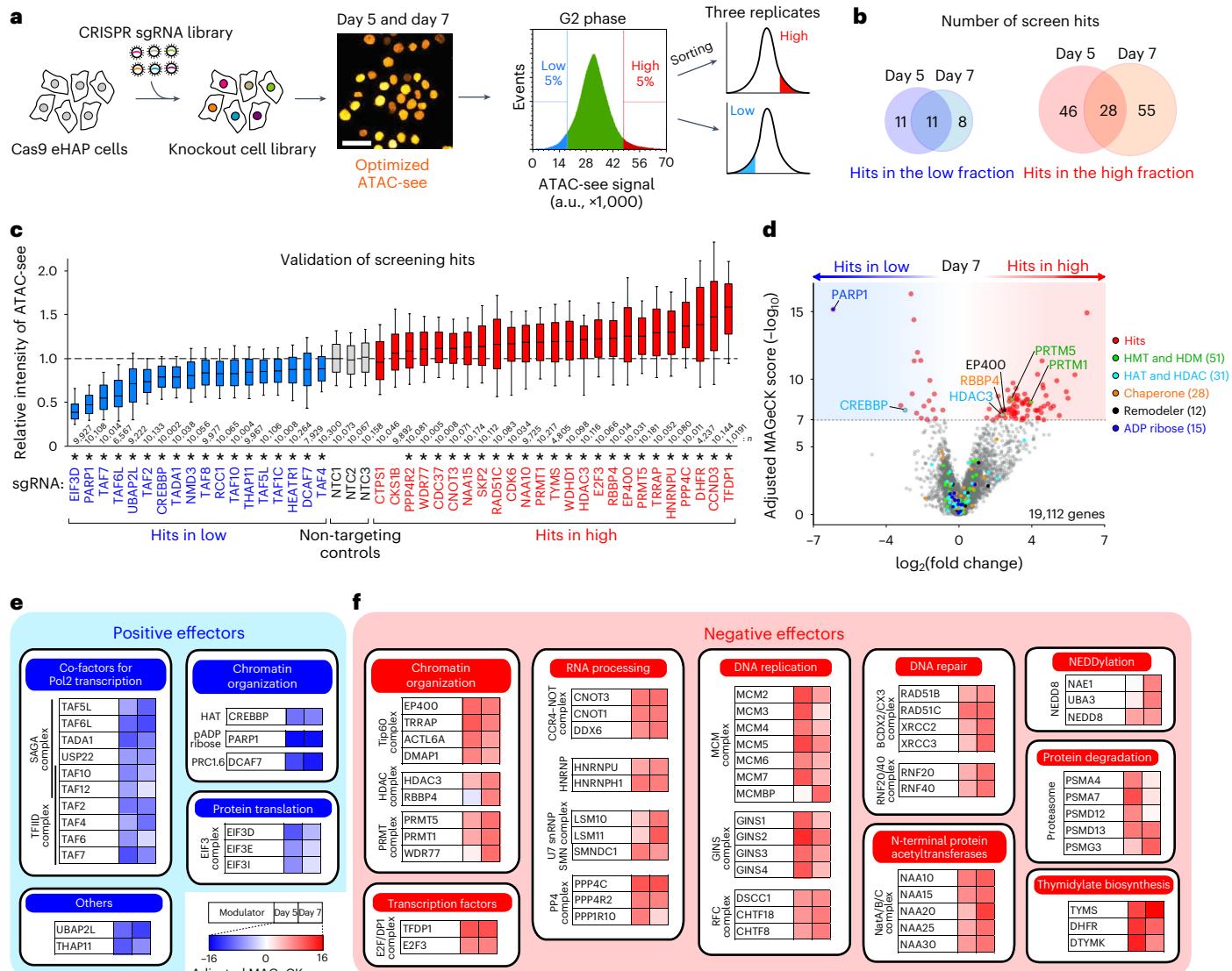
To identify genes that modulate chromatin accessibility, a genome-wide sgRNA lentivirus library that covered a broad range of human genes (19,114 genes) was transduced to Cas9-expressing eHAP cells<sup>14</sup>. The resulting pool of knockout cells was subjected to the optimized ATAC-seq assay on day 5 and day 7. Because chromatin accessibility varies greatly throughout the cell cycle (Extended Data Fig. 1g,h), we gated cells in the G2 phase and specifically collected those within the top or bottom 5% of the distribution of the ATAC-seq signals (referred to as ‘high’ or ‘low’ population in Fig. 2a and Extended Data Fig. 4a). The sgRNA enrichment relative to each population was then analyzed by high-throughput sequencing and the MAGeCK program<sup>15</sup>.

We obtained highly reproducible datasets between three replicates (Extended Data Fig. 4b). Hit selection based on stringent thresholds with MAGeCK significance score below  $10^{-7}$  identified 74 and 83 genes in the high population, and 22 and 19 genes in the low population on days 5 and 7, respectively (Fig. 2b and Supplementary Table 2). To validate the screening results, we individually knocked out 43 hit genes randomly selected from the high and low populations and evaluated their effects on chromatin accessibility by ATAC-seq assay and FACS. A total of 41 out of 43 cases (95.3%) showed significantly altered chromatin accessibility (Fig. 2c). Consistent with our pilot experiments (Fig. 1i), *CREBBP* and *EP400* were included in screen hits (Fig. 2d and Extended Data Fig. 4c,d). These results collectively demonstrate high reproducibility and reliability of screening outcomes. Furthermore, we confirmed the similar results with the same set of 43 genes in both ‘diploidized’ eHAP cells and human primary diploid fibroblast cell (TIG3) (Extended Data Fig. 5a), indicating the robustness of screening results, which is not limited to haploid karyotype and cellular immortality. While 84% of the screen hits are essential genes, further validations comprehensively confirmed minimal false positives, with negligible influence driven by cellular defects such as apoptosis accompanied with essential gene knockout (Extended Data Fig. 5b-e and Supplementary Note 1).

Notably, most of the known catalytic enzymes for maintaining chromatin integrity, including histone chaperones, chromatin remodelers and histone modifiers, were not identified as screen hits (Fig. 2d and Extended Data Fig. 4d). However, specific enzymes, such as *CREBBP*, *PARP1*, *EP400*, *PRMT1*, *PRMT5*, *HDAC3* and *RBBP4*, were identified, suggesting that these enzymes have prominent functions in controlling global chromatin accessibility. The opposite effects upon *CREBBP* and *HDAC3* knockouts indicate involvement of histone acetylation in regulating chromatin accessibility, as reported previously<sup>16</sup> (Extended Data Fig. 5f and Supplementary Note 2). Interestingly, knockout of *PARP1*, a poly(ADP-ribose) polymerase, remarkably reduced the global chromatin accessibility (Fig. 2c,d), which is in line with previous studies showing that PARP1-mediated poly(ADP)-ribosylation of chromatin facilitated nucleosome remodeling and the dissociation of linker histone H1 (refs. 17–20). Importantly, the chromatin remodeler *SMARCA4* was not identified in our screen hits, which aligns with the result in Fig. 1i. However, ATAC-seq analysis revealed a significant reduction in accessibility in NDRs upon *SMARCA4* knockout (Extended Data Fig. 5g-i), consistent with previous reports<sup>13,21</sup>. This result suggests the possible presence of false negatives in our screen, presumably due to distinctive features of ATAC-seq. Notably, we also discovered a number of previously unrecognized genes in the screen hits (red spots in Fig. 2d) and some of these hits displayed even more significant effects than known chromatin regulators.

We classified the screen hits into positive and negative effectors, based on the impact of their depletion—either decreased or increased accessibility, respectively (Fig. 2e,f). The positive effectors included several components of SAGA and TFIID complexes, which are general cofactors of RNA polymerase II and participate in transcriptional initiation<sup>22</sup>. Our findings reinforce the notion that the initiation of transcription is a critical factor in the regulation of chromatin accessibility<sup>23</sup>. *DCAF7*, a scaffold protein associated with several complexes, including the PRC1–AUTS2 complex<sup>24,25</sup>, was identified as a positive regulator. We also identified other novel positive effectors, including three components of EIF3 translational initiation factors (*EIF3D*, *EIF3F* and *EIF3I*), *THAP11* (a transcription factor also known as *RONIN*) and *UBAP2L* (a component of stress granules).

The negative effectors included components of specific complexes regulating chromatin integrity, such as the Tip60–EP400 remodeling complex, histone deacetylase (HDAC) complex and protein-arginine methyltransferase (PRMT) complex (Fig. 2f). Additionally, transcription factors, *TFDP1* and its heterodimerization partner *E2F3*, were identified as negative effectors. Notably, depletion of *TFDP1* led to the strongest



**Fig. 2 | CRISPR screening for identifying effectors of chromatin accessibility.** **a**, Schematic of the screening workflow. Scale bar, 30  $\mu\text{m}$ . **b**, Venn diagrams representing the overlap of screen hits in the high and low fractions on days 5 and 7. **c**, Validation of the screening results by single-gene CRISPR knockout in eHAP cells. Cells transfected with sgRNA against the indicated genes were analyzed by ATAC-see. ATAC-see signal intensity relative to that of non-targeting control (NTC1) is plotted. Hits in the high and low fractions are highlighted in red and blue, respectively. Statistics and definition of box plots are the same as Fig. 1i. \* $P < 0.05$ .  $n$ , number of cells analyzed. **d**, A volcano plot for gene hits from the screening result on day 7. The x axis shows the median of log fold change for all sgRNAs of each gene. The y axis shows the adjusted significance score computed by MAGeCK ( $-\log_{10}$ ). Each dot corresponds to a single gene, and significant hits

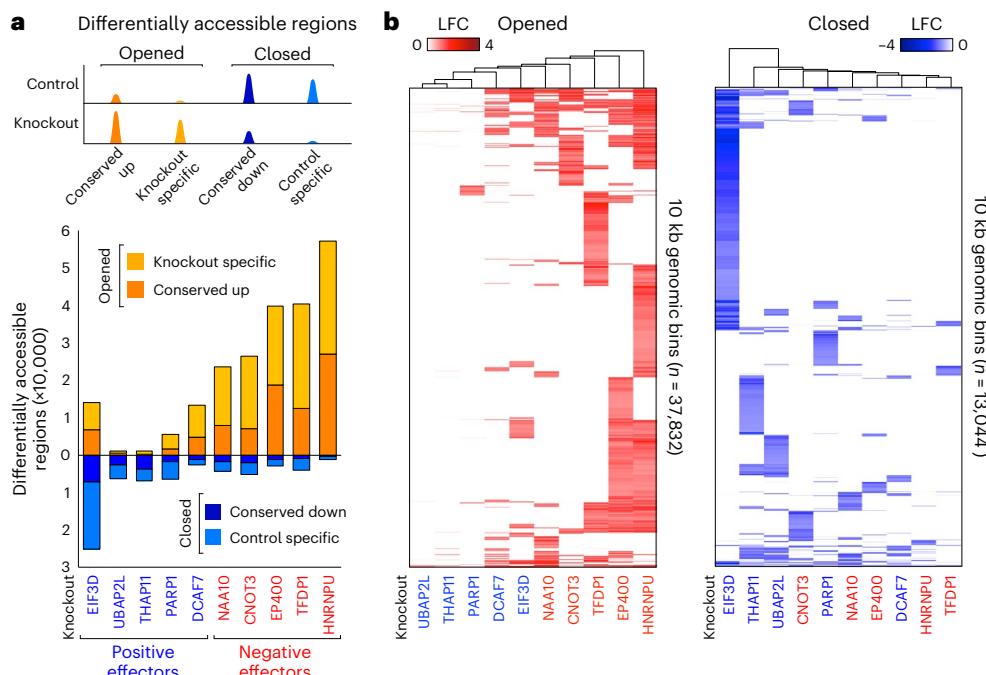
are highlighted in red. Red and blue shaded regions indicate hits in the high and low fractions, respectively. Genes encoding proteins with enzymatic activity for chromatin regulation, including histone methyltransferase and demethylase (HMT and HDM, green), histone acetyltransferase and deacetylase (HAT and HDAC, light blue), histone chaperone (orange), chromatin remodeler (black), and ADP ribosyltransferase (blue). The number of the genes in each category is shown on the right side. Only genes expressed but not essential in eHAP cells are included in these categories (transcripts per million (TPM)  $\geq 10$ , essential FDR  $\geq 0.001$ ; Benjamini–Hochberg procedure). **e,f**, Modulators of chromatin accessibility identified by the screening. Selected positive (**e**) and negative (**f**) modulators categorized on the basis of functional pathways and complexes are shown with adjusted significance MAGeCK scores for days 5 and 7.

increase in the ATAC-seq signals amongst all the validated hits (Fig. 2c). We also identified previously unrecognized pathways such as DNA repair, RNA processing and DNA replication (Extended Data Fig. 5j). Taken together, our genome-wide screen provides insights into specific biological pathways that modulate global chromatin accessibility and identified candidate effectors that involved in its regulation.

## Distinct impacts of modulators on the accessibility

To investigate whether the identified global effectors also play a role in controlling the local accessibility on NDRs, we conducted ATAC-seq analysis on knockout for selected eight novel effectors representative

from different pathways (four positive: *UBAP2L*, *THAP11*, *DCAF7* and *EIF3D*, and four negative effectors: *TFDP1*, *CNOT3*, *NAA10* and *HNRNPU*) and two known chromatin regulators (*PARP1* and *EP400*). To ensure robust and quantitative analysis, we utilized external spike-in controls and obtained highly consistent datasets among biological replicates (Extended Data Fig. 6a–g). Pairwise comparisons between the control and each knockout enabled the identification of differentially accessible regions (DARs). These DARs were then classified into one of four groups: conserved-up (common peaks upregulated upon knockout), conserved-down (common peaks downregulated upon knockout), knockout-specific (de novo peaks upon knockout) and control-specific



**Fig. 3 | ATAC-seq profiling on knockout of effector genes in eHAP cells.** **a**, Top: schematic illustration shows classification of DARs: conserved-up (orange), knockout-specific (light orange), conserved-down (dark blue) and control-specific (light blue). Bottom: the number of the DARs upon the indicated

knockout. **b**, Heatmap of hierarchical clustering shows normalized read depth of conserved DARs within a 10 kb genomic bin for opened (left) and closed (right) regions. LFC, log fold change.

(lost peaks upon knockout) DARs (Fig. 3a). The varying accessibility of the DARs was validated by ATAC–quantitative PCR analysis (Extended Data Fig. 6h), confirming the reliability of our ATAC-seq datasets. We observed variable numbers of DARs in each knockout (Fig. 3a). The trend for the opened and closed DARs was consistent with that of the ATAC-seq data except for DCAF7 KO. In addition, hierarchical clustering of signal enrichment on ATAC-seq peaks within 10 kb genomic bins revealed that depletion of each hit resulted in specific effects on the distinct genomic loci (Fig. 3b). Consistently, comprehensive annotations of DARs revealed that these hits exhibit distinct and variable effects on accessibility at NDRs with diverse genomic features (Extended Data Fig. 7 and Supplementary Note 3), which suggests their specific functions in regulating local chromatin accessibility.

#### TFDP1/E2Fs: modulators of the global accessibility

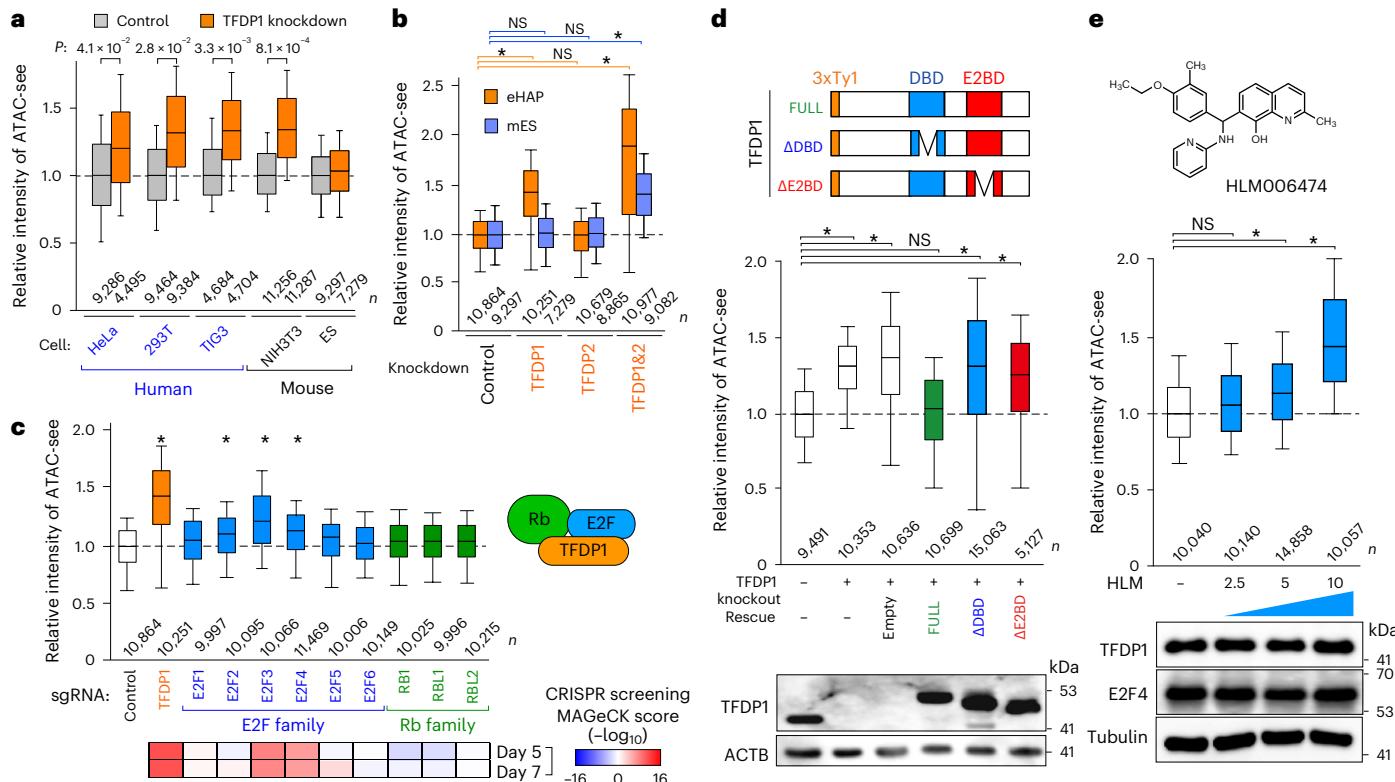
Remarkably, the transcription factor TFDP1 had a marked impact on chromatin accessibility. As *TFDP1* was not previously implicated in chromatin regulation, we sought to understand its mechanistic role. Consistent with *TFDP1* being categorized as a non-essential gene<sup>14,15</sup>, knockout of *TFDP1* had a limited effect on cell proliferation (Extended Data Fig. 8a), indicating that the alteration of chromatin accessibility caused by *TFDP1* depletion was not due to defects in cell proliferation. Depletion of *TFDP1* in other cell lines, including human (HeLa, HEK293T and TIG3) and mouse (NIH3T3) cells, induced significant increase in ATAC-seq signals, except for mES cells (Fig. 4a and Extended Data Fig. 8b). This result suggests that *TFDP1* plays a robust and conserved role in regulating chromatin accessibility. *TFDP1* has a paralog called *TFDP2*, which is expressed at low levels in eHAP cells (fragments per kilobase of transcript per million mapped reads: *TFDP1* with 89 and *TFDP2* with 3). While sgRNAs against *TFDP2* had no effect on ATAC-seq signals, *TFDP1* and *TFDP2* double knockout induced a synergistic increase in the ATAC-seq signal compared with *TFDP1* depletion alone in eHAP cells (Fig. 4b). Consistently, the double knockdown showed elevated ATAC-seq signals in mES cells. However, depletion of both *TFDP1* and *TFDP2* resulted in a substantial cell-growth

retardation (Extended Data Fig. 8c). These findings suggest that both *TFDP1* and *TFDP2* redundantly regulate chromatin accessibility and that a minimal expression level of these genes is required for proper cell growth. Relatively higher expression level of *TFDP2* in mES cells (fragments per kilobase of transcript per million mapped reads: *TFDP1* with 75 and *TFDP2* with 9) than in eHAP cells may provide an explanation for the lack of change in ATAC-seq signals upon *TFDP1* depletion alone, as *TFDP2* compensates for the function of *TFDP1*.

To determine if known *TFDP1* interacting partners, including E2F1–6 and Rb family proteins (RB1, RBL1 and RBL2)<sup>26</sup>, were also involved in the regulation of chromatin accessibility, we evaluated the effects of these genes using ATAC-seq. Knockout for *E2F2/E2F3/E2F4* induced a modest increase in ATAC-seq signals (Fig. 4c), in line with our screening results, indicating that E2F proteins also play a role in regulating accessibility. The effect of *TFDP1* knockout was rescued by expression of full-length *TFDP1*, but not by its mutants lacking either DNA- ( $\Delta$ DBD) or E2F-binding domain ( $\Delta$ E2BD)<sup>27</sup> (Fig. 4d), suggesting that both activities are crucial for regulating chromatin accessibility. Additionally, the treatment with a pan-E2F inhibitor (HLM006474), which blocks chromatin-binding activity of E2F<sup>28,29</sup>, caused a dose-dependent increase in chromatin accessibility (Fig. 4e and Extended Data Fig. 8d). As reported previously<sup>29</sup>, cell-growth retardation was observed at the higher concentration of the E2F inhibitor. These results collectively suggest that *TFDP1* and E2F proteins cooperatively regulate chromatin accessibility.

#### TFDP1 modulates the accessibility via histone transcription

Surprisingly, knockout of *TFDP1* resulted in a global elevation of the NDR accessibility (Fig. 5a,b). To assess its effect on nucleosome occupancy, we conducted a quantitative MNase-seq analysis with spike-in calibration. Remarkably, we observed a widespread reduction in nucleosome occupancy throughout the genome upon *TFDP1* depletion (Fig. 5c,d). The reduction was equally observed in both genic and intergenic regions (Extended Data Fig. 8e). However, the effect on ‘well-positioned’



**Fig. 4 | The TFDP1/E2F heterodimer plays a role in regulating chromatin accessibility.** **a**, Relative intensity of ATAC-see in the indicated cells treated with non-targeting (control) and *TFDP1* sgRNAs (for HeLa and HEK293T cells), or *TFDP1* siRNAs (for TIG3 cells, NIH3T3 cells and mES cells). **b**, eHAP cells and mES cells with TFDP1 and/or TFDP2 depletion were analyzed by ATAC-see. **c**, Results from ATAC-see analysis with eHAP cells transfected with sgRNAs targeting the indicated TFDP1-interaction partners. MAGeCK significance score for each gene is shown at the bottom of the plot. **d**, Top: schematic of full-length (FULL) and truncated mutants ( $\Delta$ DBD and  $\Delta$ E2BD) of TFDP1. eHAP cells treated with

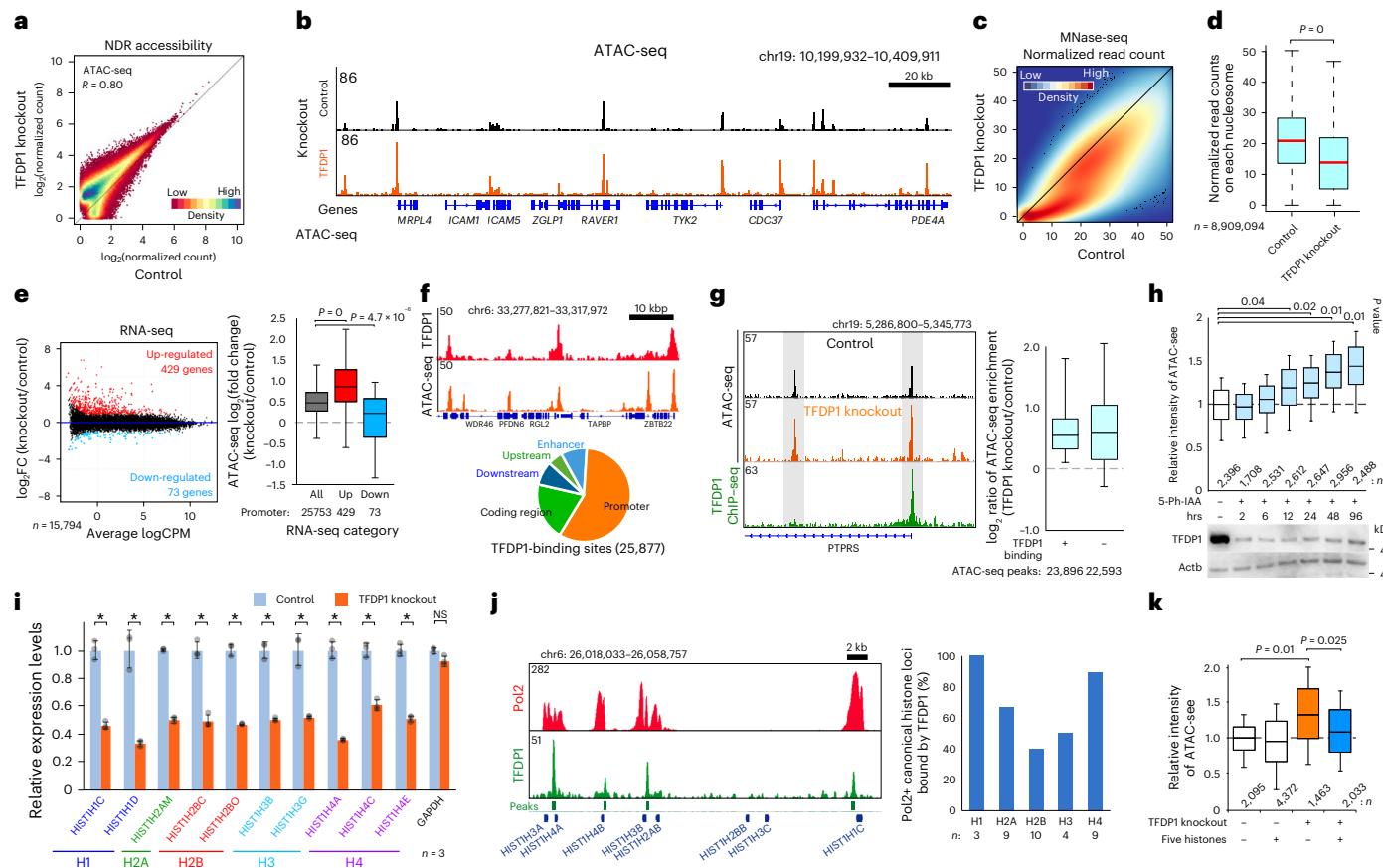
non-targeting (knockout –) or *TFDP1* sgRNA (knockout +) were transfected with an empty vector or plasmids expressing the *TFDP1* gene (FULL,  $\Delta$ DBD or  $\Delta$ E2BD) and analyzed by ATAC-see. Bottom: western blots for TFDP1 and ACTB. **e**, Top: the chemical structure of the E2F inhibitor, HLM006474. Relative intensity of ATAC-see signals in the cells treated with vehicle or 2.5  $\mu$ M, 5  $\mu$ M or 10  $\mu$ M of HLM006474 (HLM). Bottom: western blotting images of TFDP1,  $\alpha$ -tubulin and E2F4. All the statistics and definition of box plots in **a–e** are the same as Fig. 1*i*. \* $P < 0.05$ . NS, not significant; n, number of cells analyzed. All the ATAC-see signals were measured by FACS.

nucleosomes (-1–3) downstream of transcription start sites (TSSs) of expressed genes was limited compared with that of non-expressed genes, indicating preferential nucleosome loading to transcribing gene loci. Nucleosome phasing and nucleosome repeat length (NRL) remained unchanged (Extended Data Fig. 8f), suggesting that TFDP1 depletion mainly affects nucleosome occupancy, without altering nucleosome positioning. Given that the genome-wide reduction of nucleosomes upon TFDP1 depletion could potentially affect transcriptional activity, we conducted a quantitative RNA-seq with spike-in normalization. While we did not observe a global increase in gene expression, the majority of differentially expressed genes were upregulated in TFDP1 knockout (429 upregulated and 73 downregulated genes with  $|log$  fold change  $>1$  and false discovery rate (FDR)  $<0.005$ ), accompanied by increased accessibility on their promoters (Fig. 5e, Extended Data Fig. 8g and Supplementary Table 3). This limited effect on the transcription profile could be attributed to the transcription-dependent nucleosome loading downstream of TSSs that we observed in MNase-seq results. These results collectively suggest that depletion of TFDP1 induced a global increase in chromatin accessibility, as well as a genome-wide reduction in nucleosome occupancy.

To address the mechanism by which TFDP1 modulates chromatin accessibility, we profiled TFDP1-binding sites by chromatin immunoprecipitation followed by sequencing (ChIP-seq). To achieve this, we prepared eHAP cells in which 3 $\times$ FLAG tag was knocked in at the N-terminus of TFDP1 as a fusion protein. We confirmed that FLAG-tagged TFDP1 was functional, as the upregulation of chromatin

accessibility in TFDP1 knockout cells was efficiently reversed by expression of FLAG-TFDP1 (Extended Data Fig. 9a–c). ChIP-seq with anti-FLAG antibody showed significant enrichment of known TFDP1-binding motifs in the ChIP-seq peaks (Extended Data Fig. 9d). Meta-analysis revealed that TFDP1 preferentially binds to accessible regions, especially to active promoters (Fig. 5f and Extended Data Fig. 9e,f), implying a role of TFDP1 in regulating chromatin accessibility. However, as TFDP1 knockout induced elevated chromatin accessibility, not only at TFDP1-binding sites but also at non-binding sites to a comparable extent (Fig. 5g), we investigated whether TFDP1 directly regulates global chromatin accessibility. We carried out acute depletion of TFDP1 using auxin-induced degradation 2 (AID2) system<sup>30</sup> (Extended Data Fig. 9g) and assessed its effect by ATAC-see. While TFDP1 was efficiently degraded within 2 h upon the addition of 5-phenyl-indole-3-acetic acid (5-Ph-IAA, a ligand for AID2 system), the ATAC-see signals were initially unaffected (Fig. 5h). However, the signals were gradually elevated over time until 96 h, suggesting that TFDP1 plays an indirect role in regulating chromatin accessibility, at least in a global context.

We next investigated whether downstream genes of *TFDP1* are involved in the regulation of global chromatin accessibility. RNA-seq analysis revealed that expression levels for screen hit genes were not significantly altered upon TFDP1 depletion (Extended Data Fig. 8h), suggesting TFDP1 regulated chromatin accessibility in a distinct mechanism compared with other screen hits. Therefore, we came up with a hypothesis that TFDP1 may regulate the expression levels of histones, which are basic components of the nucleosome. Quantitative



**Fig. 5 | TFDP1 modulates global chromatin accessibility through direct regulation of histones.** **a**, A scatter plot showing normalized ATAC-seq read counts for the control and TFDP1 knockout on 135,791 ATAC-seq peaks. **R**, Pearson correlation coefficient. **b**, Genome browser tracks displaying ATAC-seq with and without TFDP1 knockout. **c,d**, A density dot plot (**c**) and a box plot (**d**) displaying normalized MNase-seq read counts for each nucleosome position in control and TFDP1 knockout cells. Statistics include a two-sided Mann–Whitney *U* test. *n*, number of nucleosomes analyzed. **e**, Left: an MA plot of log fold change (FC) against log CPM (counts per million mapped reads) showing RNA-seq data representing TFDP1 knockouts versus control eHAP cells. Right: changes of ATAC-seq enrichments at promoter associated peaks. ‘All’ indicates all the promoters. Up, upregulated promoters categorized on the basis of RNA-seq results; Down, downregulated promoters categorized on the basis of RNA-seq results. Dashed line,  $\log_2 FC = 0$ . Statistics include a two-sided Mann–Whitney *U* test. **f**, Genome browser images showing ATAC-seq and ChIP-seq for TFDP1 in eHAP cells. A pie chart represents the proportion of genomic features within TFDP1 ChIP-seq peaks. **g**, Left: genome browser tracks displaying ATAC-seq and TFDP1 ChIP-seq. NDRs are highlighted in gray. Right: the enrichment of ATAC-seq signals (TFDP1 knockout/control) that overlapped with (+) or without (−) TFDP1 ChIP-seq peaks. Dashed line,  $\log_{10}$  ratio = 0. **h**, Top: relative intensity of ATAC-seq signals analyzed by FACS in the cells treated with 5-Ph-IAA for the indicated time. Bottom: western blotting images of TFDP1 and  $\beta$ -actin. *n*, number of cells analyzed. **i**, Relative histone mRNA levels normalized to  $ATCB$  in the control or TFDP1 knockout. Statistics include a two-tailed unpaired Student’s *t*-test. Data are presented as mean values  $\pm$  standard deviation. *n*, independent experiments. **j**, Left: genomic browser images displaying ChIP-seq data for RNA polymerase II (Pol2) and TFDP1 at the HIST1 locus. Right: percentage of Pol2-positive histone gene loci associated with TFDP1 ChIP-seq peaks. *n*, number of genes. **k**, ATAC-seq analysis for TFDP1 knockouts transfected with plasmids expressing five canonical histones (H1, H2A, H2B, H3 and H4). Dashed line, relative intensity = 1; *n*, number of cells analyzed. The statistics and definition of box plots in **h** and **k** are the same as in Fig. 1*i*. chr, chromosome. \**P* < 0.05. The box plots in **d**, **e** and **g** are shown as Fig. 1*c*.

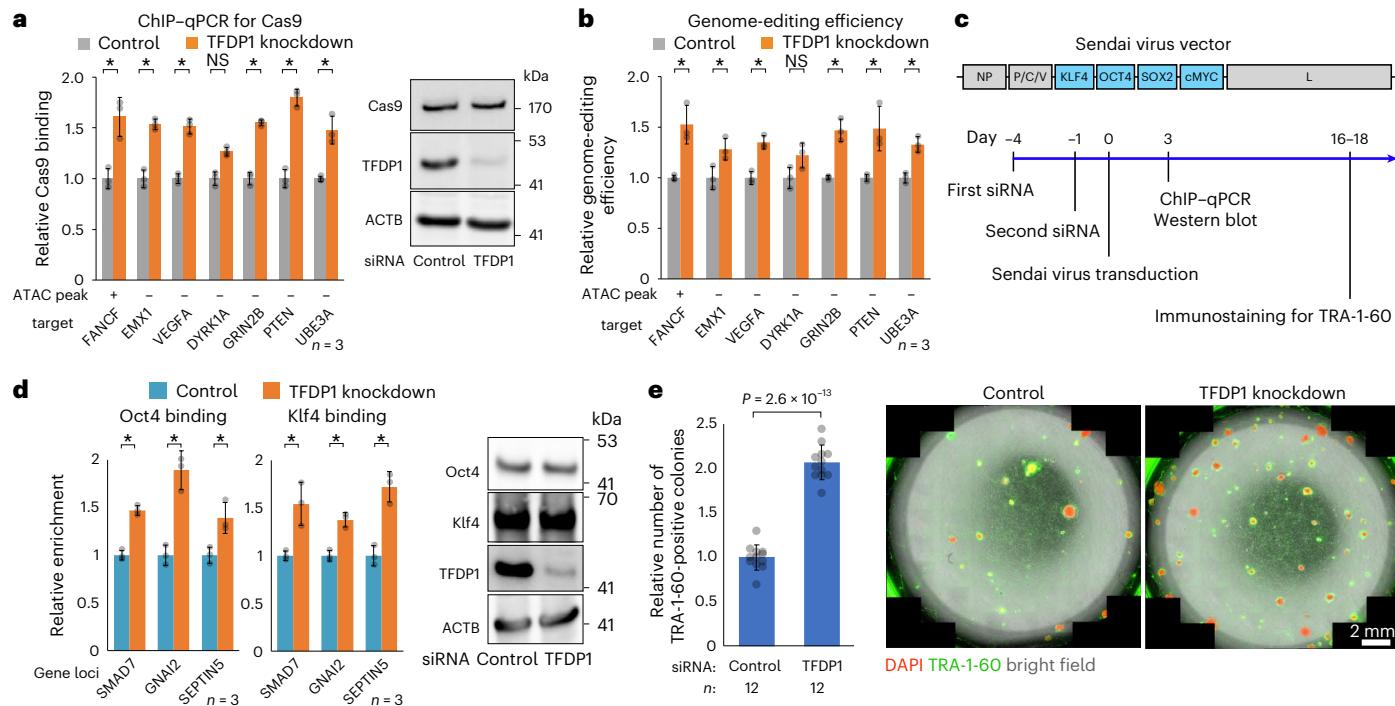
reverse transcription PCR analysis, as well as RNA-seq, revealed that the expression levels of these canonical histone genes (H1, H2A, H2B, H3 and H4) were significantly downregulated by TFDP1 depletion (Fig. 5*i* and Extended Data Fig. 10*a*). The reduction in the canonical histones was also observed at the protein level, as determined by quantitative western blotting and Coomassie brilliant blue (CBB) staining with whole cell extracts and chromatin fractions (Extended Data Fig. 10*b*). In contrast, the effects on non-canonical histone variants, including H3.3, H2A.X and H2A.Z, were limited (Extended Data Fig. 10*c*). Remarkably, ChIP-seq analysis revealed that TFDP1 binds to proximal promoters of multiple copies of canonical histone loci, which are occupied by RNA polymerase II (Fig. 5*j*). In addition, publicly available ChIP-seq data from other cell types also showed binding of TFDP1 at histone gene loci (Extended Data Fig. 10*d,e*). Collectively, these results suggest that TFDP1 transcriptionally regulates the expression levels of canonical histones, which is in line with previous reports implicating E2F1 and E2F4

ATAC-seq and TFDP1 ChIP-seq. NDRs are highlighted in gray. Right: the enrichment of ATAC-seq signals (TFDP1 knockout/control) that overlapped with (+) or without (−) TFDP1 ChIP-seq peaks. Dashed line,  $\log_{10}$  ratio = 0. **h**, Top: relative intensity of ATAC-seq signals analyzed by FACS in the cells treated with 5-Ph-IAA for the indicated time. Bottom: western blotting images of TFDP1 and  $\beta$ -actin. *n*, number of cells analyzed. **i**, Relative histone mRNA levels normalized to  $ATCB$  in the control or TFDP1 knockout. Statistics include a two-tailed unpaired Student’s *t*-test. Data are presented as mean values  $\pm$  standard deviation. *n*, independent experiments. **j**, Left: genomic browser images displaying ChIP-seq data for RNA polymerase II (Pol2) and TFDP1 at the HIST1 locus. Right: percentage of Pol2-positive histone gene loci associated with TFDP1 ChIP-seq peaks. *n*, number of genes. **k**, ATAC-seq analysis for TFDP1 knockouts transfected with plasmids expressing five canonical histones (H1, H2A, H2B, H3 and H4). Dashed line, relative intensity = 1; *n*, number of cells analyzed. The statistics and definition of box plots in **h** and **k** are the same as in Fig. 1*i*. chr, chromosome. \**P* < 0.05. The box plots in **d**, **e** and **g** are shown as Fig. 1*c*.

in histone gene regulation<sup>31–33</sup>. Importantly, the ectopic expression of five canonical histones partially reversed the de-regulated chromatin accessibility in TFDP1 knockout cells (Fig. 5*k*). Taken together, we conclude that TFDP1 regulates global chromatin accessibility through transcriptional regulation of canonical histone genes.

### Manipulating accessibility for DNA-related applications

Finally, we examined whether the actual protein binding to chromatin was changed upon transient depletion of TFDP1 by using small interfering RNA. We evaluated the effect by monitoring Cas9, since chromatin accessibility is a major determinant of Cas9 binding to the target sites<sup>34,35</sup>. We transfected Cas9-expressing eHAP cells with in vitro transcribed sgRNAs targeting seven distinct genomic loci. ChIP-qPCR assay revealed that TFDP1 depletion significantly elevated binding of Cas9 to the target sites, without affecting the expression level of Cas9 (Fig. 6*a*). Importantly, these effects were observed not only in



**Fig. 6 | TFDP1 depletion boosts DNA-related biological applications.**

**a, b**, siRNA against *TFDP1* or non-targeting control was transfected to Cas9-expressing eHAP cells. The cells were subsequently transfected with in vitro transcribed sgRNAs targeting the indicated gene loci and analyzed by ChIP-qPCR analysis for Cas9 binding (**a**, left), western blotting (**a**, right) and genome-editing assay (**b**). Target regions overlapping with or without ATAC-seq peaks are shown as + or -, respectively. Statistics include a two-tailed unpaired Student's *t*-test. \*P < 0.05. NS, not significant. **c**, Top: structure of Sendai virus vector used for the somatic cell reprogramming. Blue boxes, reprogramming transcription factors. Bottom: flow chart showing the timeline of the reprogramming experiment.

**d, e**, ChIP-qPCR analysis for Oct4 or Klf4 (left), and western blotting for indicated proteins (right) performed 3 days after the Sendai virus transduction. Statistics include a two-tailed unpaired Student's *t*-test. \*P < 0.05. **e**, Left: relative reprogramming efficiency evaluated by counting TRA-1-60-positive colonies when using control or *TFDP1*-knockdown iPS3 cells. Statistics include a two-tailed unpaired Student's *t*-test. Right: the representative images of iPS cell colonies for the indicated conditions. The images were captured using three different channels: DAPI (red), TRA-1-60 (green) and bright field (white). Data in **a**, **b** and **d** are presented as mean values ± standard deviation. *n*, independent experiments.

highly accessible regions associated with ATAC-seq peaks, but also in regions without peaks. These results suggest that chromatin accessibility was globally elevated in the *TFDP1*-depleted cells, resulting in greater binding of Cas9. It is noteworthy that *TFDP1* depletion also enhanced efficiency of genome editing (Fig. 6b), indicating that manipulating global chromatin accessibility could offer a promising approach to control chromatin plasticity and possibly to improve various cell engineering applications using DNA-binding proteins.

To further explore the potential of chromatin accessibility manipulation, we employed it in the context of induced pluripotent stem (iPS) cell reprogramming. The conversion of somatic cells into iPS cells relies on the binding of pluripotent transcription factors (Oct4, Sox2, Klf4 and cMyc) to their respective target sites. However, in somatic cells, these target sites are often inaccessible due to the closed chromatin conformation<sup>36,37</sup>. Therefore, chromatin accessibility represents a significant obstacle to the reprogramming process. To investigate the impact of chromatin accessibility manipulation on iPS cell reprogramming, we transduced *TFDP1*-depleted TIG3 cells with Sendai virus-based RNA vector encoding the pluripotent factors<sup>38</sup> (Fig. 6c). As anticipated, the transient knockdown of *TFDP1* led to an enhanced binding of both Oct4 and Klf4 to their putative target sites in the early phase of reprogramming, without affecting their expression levels (Fig. 6d), thus confirming the elevated chromatin accessibility by *TFDP1* depletion. Remarkably, we observed an increase in the number of colonies positive for TRA-1-60, a critical marker for the fully reprogrammed state<sup>39</sup>, when using *TFDP1*-knockdown TIG3 cells (Fig. 6e). These results suggest that *TFDP1* depletion promotes the chromatin-binding of the pluripotent factors, resulting in enhanced efficiency of iPS cell

reprogramming. This is in line with a previous study that *TFDP1* was identified as a barrier gene for the iPS cell reprogramming through genome-wide RNA interference (RNAi) screening<sup>40</sup>. Taken these results together, our study shows the potential of manipulating chromatin accessibility as a strategy to enhance DNA-templated biological applications, including genome editing and cellular reprogramming.

## Discussion

In this study, the unique characteristics of ATAC-seq in detecting global chromatin accessibility allowed us to identify a range of modulators. Some of these modulators are involved in DNA-templated reactions, including transcription, DNA repair and DNA replication. Given that chromatin accessibility provides physical scaffolds to recruit factors involved in these reactions, our findings suggest a strong interdependence between the establishment of chromatin accessibility and these chromatin-templated reactions. We also discovered several pathways involved in the process, which are not directly related to chromatin integrity but rather have roles in RNA processing, N-terminal protein acetylation and protein translation/degradation. Further analysis of these identified effectors is expected to advance our understanding of the mechanisms that regulate chromatin accessibility.

We showed that global elevation of chromatin accessibility upon *TFDP1* depletion enhances DNA-associated applications, including iPS cell reprogramming. Consistent with these results, chromatin decondensation in somatic cells by treatment with chemical inhibitors or depletion of chromatin regulators has been shown to increase the efficiency of iPS cell reprogramming<sup>41–47</sup>. Notably, a previous study showed that overexpression of PARP1, which is discovered as a positive

effector in our screening, facilitates iPS cell generation<sup>48</sup>. These findings collectively support the concept that manipulation of chromatin accessibility is a promising strategy for enhancing the efficiency of DNA-templated reactions. However, it is important to note that the extensive manipulation of the accessibility may lead to the misregulation of gene expression through mechanisms such as off-target protein binding or result in other phenotypes such as cell-growth retardation, as observed in both cases of double knockdown for TFDP1/TFDP2, and treatment with high dose of the E2F inhibitor. Therefore, fine-tuning of chromatin accessibility, either locally or globally, is crucial for controlling chromatin plasticity.

Our improved ATAC-seq approach enabled specific detection of nuclear accessible chromatin with high sensitivity and high signal-to-noise ratio. The staining pattern of ATAC-seq was profoundly affected by crosslinking with formaldehyde, although there is no clear molecular explanation. Several studies have reported that the crosslinking of chromatin could induce artifacts in imaging, ChIP-seq and ATAC-seq analyses<sup>49–56</sup>. We recommend that the crosslinking condition for ATAC-seq assays should be carefully optimized for each experimental setup. We also found that use of short adapter DNA substantially improve Tn5 activity. Interestingly, the small molecule inhibitor Pitstop2 has been shown to enhance nuclear permeability of Tn5, thereby improving efficiency of chromatin fragmentation<sup>57</sup>. Since Tn5 transposase is widely used for versatile sequencing applications, these improvements are potentially applicable to enhance the performance of Tn5-based techniques, including bulk and single-cell approaches such as CUT&TAG<sup>58,59</sup>, ATAC-seq<sup>60,61</sup> and RNA-seq<sup>62</sup>.

Taken together, our study uncovers the involvement of multiple biological pathways in establishing the chromatin accessibility landscape and highlights the contribution of TFDP1 to this process through the direct transcriptional regulation of canonical histones. Additionally, we propose that manipulating chromatin accessibility has a potential to improve various cell engineering applications.

## Online content

Any methods, additional references, Nature Portfolio 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/s41588-024-01658-1>.

## References

1. Boskovic, A., Bing, X. Y., Kaymak, E. & Rando, O. J. Control of noncoding RNA production and histone levels by a 5' tRNA fragment. *Genes Dev.* **34**, 118–131 (2020).
2. Celona, B. et al. Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biol.* **9**, e1001086 (2011).
3. Pal, S. & Tyler, J. K. Epigenetics and aging. *Sci. Adv.* **2**, e1600584 (2016).
4. Klein, H.-U. et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
5. Li, D. et al. Chromatin accessibility dynamics during iPSC reprogramming. *Cell Stem Cell* **21**, 819–833.e6 (2017).
6. Liu, L. et al. An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat. Commun.* **10**, 364 (2019).
7. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
8. Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
9. 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).
10. Zhou, M., Bhasin, A. & Reznikoff, W. S. Molecular genetic analysis of transposase-end DNA sequence recognition: cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase. *J. Mol. Biol.* **276**, 913–925 (1998).
11. Essletzbichler, P. et al. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).
12. de Dieuleveult, M. et al. Genome-wide nucleosome specificity and function of chromatin remodelers in ES cells. *Nature* **530**, 113–116 (2016).
13. Schick, S. et al. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethaliies in human cancers. *Nat. Genet.* **51**, 1399–1410 (2019).
14. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
15. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
16. Görisch, S. M., Wachsmuth, M., Tóth, K. F., Lichter, P. & Rippe, K. Histone acetylation increases chromatin accessibility. *J. Cell Sci.* **118**, 5825–5834 (2005).
17. Ryu, K., Kim, D.-S. & Kraus, L. W. New facets in the regulation of gene expression by ADP-ribosylation and poly(ADP-ribose) polymerases. *Chem. Rev.* **115**, 2453–2481 (2015).
18. Azad, G. et al. PARP1-dependent eviction of the linker histone H1 mediates immediate early gene expression during neuronal activation. *J. Cell Biol.* **217**, 473–481 (2018).
19. Kim, M. Y., Mauro, S., Gévrí, N., Lis, J. T. & Kraus, W. NAD<sup>+</sup>-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803–814 (2004).
20. Krishnakumar, R. et al. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* **319**, 819–821 (2008).
21. Schick, S. et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nat. Genet.* **53**, 269–278 (2021).
22. Fischer, V., Schumacher, K., Tora, L. & Devys, D. Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* **10**, 1–8 (2018).
23. Stewart-Morgan, K. R., Nazaret, R.-G. & Groth, A. Transcription restart establishes chromatin accessibility after DNA replication. *Mol. Cell* **75**, 284–297.e6 (2019).
24. Hauri, S. et al. A high-density map for navigating the human polycomb complexome. *Cell Rep.* **17**, 583–595 (2016).
25. Wang, Q. et al. WDR68 is essential for the transcriptional activation of the PRC1-AUTS2 complex and neuronal differentiation of mouse embryonic stem cells. *Stem Cell Res.* **33**, 206–214 (2018).
26. Bracken, A. P. et al. E2F target genes: unraveling the biology. *Trends Biochem. Sci.* **29**, 409–417 (2004).
27. Wu, C. L., Classon, M., Dyson, N. & Harlow, E. Expression of dominant-negative mutant DP-1 blocks cell cycle progression in G1. *Mol. Cell. Biol.* **16**, 3698–3706 (1996).
28. Lan, W. et al. E2F signature is predictive for the pancreatic adenocarcinoma clinical outcome and sensitivity to E2F inhibitors, but not for the response to cytotoxic-based treatments. *Sci. Rep.* **8**, 8330 (2018).
29. Ma, Y. et al. A small-molecule E2F inhibitor blocks growth in a melanoma culture model. *Cancer Res.* **68**, 6292–6299 (2008).

30. Yesbolatova, A. et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* **11**, 5701 (2020).
31. Gokhman, D., Livyatan, I., Sailaja, B. S., Melcer, S. & Meshorer, E. Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones. *Nat. Struct. Mol. Biol.* **20**, 119–126 (2013).
32. Oswald, F., Dobner, T. & Lipp, M. The E2F transcription factor activates a replication-dependent human H2A gene in early S phase of the cell cycle. *Mol. Cell. Biol.* **16**, 1889–1895 (1996).
33. Rabinovich, A., Jin, V. X., Rabinovich, R., Xu, X. & Farnham, P. J. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res.* **18**, 1763–1777 (2008).
34. Singh, R., Kuscu, C., Quinlan, A., Qi, Y. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* **43**, e118 (2015).
35. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).
36. Knaupp, A. S. et al. Transient and permanent reconfiguration of chromatin and transcription factor occupancy drive reprogramming. *Cell Stem Cell* **21**, 834–845.e6 (2017).
37. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
38. Nishimura, K. et al. Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302. *Stem Cell Res.* **23**, 13–19 (2017).
39. Chan, E. M. et al. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat. Biotechnol.* **27**, 1033–1037 (2009).
40. Yang, C.-S., Chang, K.-Y. & Rana, T. M. Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. *Cell Rep.* **8**, 327–337 (2014).
41. Chen, J. et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
42. Schmidt, R. & Plath, K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol.* **13**, 251 (2012).
43. Xu, Y. et al. Transcriptional control of somatic cell reprogramming. *Trends Cell Biol.* **26**, 272–288 (2016).
44. Huangfu, D. et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26**, 795–797 (2008).
45. Mikkelsen, T. S. et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
46. Shi, Y. et al. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525–528 (2008).
47. Liu, B. et al. Inhibition of histone deacetylase 1 (HDAC1) and HDAC2 enhances CRISPR/Cas9 genome editing. *Nucleic Acids Res.* **48**, 517–532 (2020).
48. Doege, C. A. et al. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
49. Baranello, L., Kouzine, F., Sanford, S. & Levens, D. ChIP bias as a function of cross-linking time. *Chromosome Res.* **24**, 175–181 (2016).
50. Birk, U. J. Super-resolution microscopy of chromatin. *Genes* **10**, 493 (2019).
51. Nagaich, A. K., Walker, D. A., Wolford, R. & Hager, G. L. Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling. *Mol. Cell* **14**, 163–174 (2004).
52. Schmiedeberg, L., Skene, P., Deaton, A. & Bird, A. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS ONE* **4**, e4636 (2009).
53. Teytelman, L. et al. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS ONE* **4**, e6700 (2009).
54. Whelan, D. R. & Bell, T. D. M. Image artifacts in single molecule localization microscopy: why optimization of sample preparation protocols matters. *Sci. Rep.* **5**, 1–10 (2015).
55. Oh, K.-S., Ha, J., Baek, S. & Sung, M.-H. XL-DNase-seq: improved footprinting of dynamic transcription factors. *Epigenetics Chromatin* **12**, 30 (2019).
56. Zhang, H. et al. Extensive evaluation of ATAC-seq protocols for native or formaldehyde-fixed nuclei. *BMC Genomics* **23**, 214 (2022).
57. Mulqueen, R. et al. Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis. Preprint at *bioRxiv* <https://doi.org/10.1101/637256> (2019).
58. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
59. Meers, M. P., Llagas, G., Janssens, D. H., Codomo, C. A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using Multi-Tag. *Nat. Biotechnol.* **41**, 708–716 (2023).
60. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
61. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
62. Di, L. et al. RNA sequencing by direct fragmentation of RNA/DNA hybrids. *Proc. Natl Acad. Sci. USA* **117**, 2886–2893 (2020).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

## Methods

Our study complies with all relevant ethical regulations. National Institute for Basic Biology and Kanazawa University approved this work.

### Cell culture

eHAP cells (Horizon Discovery)<sup>11</sup> and TIG3 cells (human primary embryonic fibroblast, JCRB Cell Bank) were grown in Iscove's modified Dulbecco's medium (Wako) supplemented with 10% fetal bovine serum (FBS; Sigma) and penicillin/streptomycin (Wako). MEF, NIH3T3, HeLa and HEK293T cells were cultured in Dulbecco's modified Eagle medium (Wako) supplemented with 10% FBS and penicillin/streptomycin. mES cells (E14) were cultured in Dulbecco's modified Eagle medium containing 20% FBS, leukemia inhibitory factor, 3 μM CHIR99021 (a GSK3b inhibitor; Sigma), 1 μM PD0325901 (a MEK inhibitor; Sigma), 1 mM sodium pyruvate, penicillin/streptomycin, non-essential amino acids and 0.1 mM 2-mercaptoethanol. All the mammalian cells were grown at 37 °C with 5% CO<sub>2</sub>. *Drosophila* S2 cells were grown in Schneider's *Drosophila* S2 medium (Thermo Fisher), supplemented with 10% FBS and penicillin/streptomycin at 25 °C.

### Generation of Cas9-expressing cells

eHAP cells were transfected with pSpCas9B vector expressing SpCas9 and blasticidin-resistance gene, followed by selection with 10 μg ml<sup>-1</sup> of blasticidin. Single colonies were isolated and a clone expressing Cas9 (A6) was used for further assays. HEK293T, HeLa and TIG3 cells were infected with lentivirus (lentiCRISPR-BSD2) expressing SpCas9 and blasticidin-resistance gene, and subsequently selected with 10–20 μg ml<sup>-1</sup> blasticidin.

### Preparation of knockout cells

Cas9-expressing eHAP (A6) and diploidized A6 (D-A6) cells were transfected with a plasmid expressing sgRNAs and puromycin-resistance gene (psgR1.OP) using either FUGENE HD (Promega) or HilyMax (Dojindo). Meanwhile, HeLa and HEK293T cells were transfected with HilyMax (Dojindo), and TIG3 cells with Lipofectamine 2000 (Thermo). Cells were subsequently selected with 1 μg ml<sup>-1</sup> puromycin from 24 h to 72 h post-transfection. On day 5 or day 7 post-transfection, cells were used for further assays. To investigate the effect of apoptosis induced by knockout, A6 cells were cultured with 2 μM QVQ-OPh (Abcam) for 3 days after the puromycin selection and subjected to ATAC-see assay at day 5. Unless otherwise stated, mixture of two plasmids encoding different sgRNAs against single gene were transfected to cells for efficient knockout. Knockout of each gene was confirmed by western blotting or immunostaining. Sequences of sgRNAs are listed in Supplementary Table 1.

### Generation of diploidized eHAP cells

Since haploid eHAP cells spontaneously turn to diploid state over passage, Cas9-expressing eHAP cells (A6) were cultured for 5 weeks. We confirmed that the ploidy of resulting eHAP cells (D-A6) reached a pure diploid state by FACS analysis.

### Generation of FLAG, Ty1 or mini-Auxin-inducible degron-tagged TFDP1 knockin cells

eHAP cells were co-transfected with a plasmid expressing Cas9 and sgRNA targeting N-terminus of TFDP1 and a targeting plasmid (pGb-TFDP-PflxFI3, pGb-TFDP-PflxTy3 or pGb-BAflox-TFDP1\_v2) carrying the desired tag. The transfected cells were selected with either 1 μg ml<sup>-1</sup> puromycin or 10 μg ml<sup>-1</sup> blasticidin, and single colonies were isolated and expanded in culture. The genotypes of the colonies were confirmed by PCR and western blotting. The selection cassette (puromycin or blasticidin-resistance gene) was then removed by transfection of pCAG-Cre-iBSD or pCAG-Cre-iPuro-expressing Cre recombinase. In the case of AID-tagged TFDP1 knockin, cells were further infected with lentivirus encoding TiR1mutant (F74G) and a puromycin-resistance

gene (pLv\_fTiR1\_F74G-Puro), and subsequently selected with 1 μg ml<sup>-1</sup> puromycin. The resulting cells were named DP1AID2-15. Plasmid DNAs used for gene targeting are listed in Supplementary Table 1.

### Treatment of eHAP cells with apoptotic inducing reagents

As eHAP cells are prone to detach from a culture dish during the early stage of apoptosis, eHAP cells were first dissociated by treatment with trypsin and transferred to a 1.5 ml tube with culture medium at a concentration of 1 × 10<sup>6</sup> cells ml<sup>-1</sup>. The cells were then treated with either 100 μM H<sub>2</sub>O<sub>2</sub>, 5 μM Raptinal or 1 μM staurosporine for 1 h or 2 h at 37 °C with 5% CO<sub>2</sub>. After treatment, the cells including detached cells were subjected to either ATAC-see assay or staining with annexin V fluorescein isothiocyanate (MBL).

### Preparation of Tn5 complex for ATAC-see

Recombinant Tn5 was produced in *Escherichia coli* as previously described<sup>63</sup>. Concentration of the purified Tn5 was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by CBB staining with different amounts of bovine serum albumin (BSA) as a standard. Tn5 complex was assembled by mixing the same molecular amount of purified Tn5 and pre-annealed adapter DNA for 1 h at room temperature. Since the complex tends to aggregate in the ATAC-see reaction buffer, we removed the aggregates by centrifugal filtration before use, otherwise it would lead to strong background signals in the cytoplasm or periphery of nucleus. After suspending fluorescent Tn5 complex in TAPS-DMF buffer (10 mM TAPS-NaOH pH 8.5, 5 mM MgCl<sub>2</sub> and 10% dimethylformamide), the solution was sonicated by Bioruptor (30 s on and 30 s off at medium power for five cycles; Cosmobio), and passed through a centrifugal 0.45 μm filter (Ultrafree-MC; Millipore) to remove aggregated Tn5 complex. Alternatively, the aggregates were precipitated by centrifugation at 20,400g for 10 min at room temperature. Sequences of adapter DNAs are listed in Supplementary Table 1.

### ATAC-see for microscopic imaging

NIH3T3 or MEF cells were seeded in wells of a four-well micro-insert (80406, Ibidi) placed on a glass bottom dish coated with iMatrix-511 (Wako), which supported cell adhesion and was effective for keeping cells on a dish during the ATAC-see reaction even without crosslinking. ATAC-see was modified from the previous report<sup>8</sup>. Briefly, after washing with phosphate-buffered saline (PBS) twice, cells were fixed with 0.5%, 1% or 4% formaldehyde in PBS for 10 min at room temperature. This step was skipped for the non-crosslinked condition. Cells were subsequently transposed in TAPS-DMF buffer containing fluorescent Tn5 complex and 0.1% IGEPAL CA-630 at 37 °C for 10 min. The cells were washed with PBS containing 0.05% Triton X-100 twice and crosslinked in 4% paraformaldehyde (PFA) for 10 min at room temperature. After immunostaining with specific antibodies or DNA fluorescent in situ hybridization, cells were mounted in Vectashield (VECTOR). Images were acquired with Nikon A1 confocal microscopy or Cell Voyager CV1000 (Yokogawa). Unless stated, the ATAC-see reaction was performed with non-crosslinked cells. For the validation of effects of crosslinking, NIH3T3 cells were transfected with pEGFP-H2B (21–35) to label nucleoli<sup>64</sup>. Cells were then fixed with indicated concentration of PFA for 10 min at room temperature and subjected to ATAC-see reaction. Antibodies used were listed in Supplementary Table 1.

### Optimized ATAC-see for FACS with spike-in control cells

For analyses of cells by ATAC-see using flow cytometry, we introduced a spike-in control to calibrate technical variability between samples, which enabled precise quantification and high reproducibility. eHAP cells expressing H2B-Kusabira-Orange were used as the spike-in control that could be easily gated away from knockout cells using signals of Kusabira-Orange. Knockout cells were first mixed with the spike-in cells in a tube and subsequently processed for ATAC-see analyses. The

mixed cells were washed with PBS, permeabilized with 0.1% IGEPAL CA-630 in TAPS-DMF buffer containing 1% BSA, 1 mg ml<sup>-1</sup> polyvinylpyrrolidone (PVP) for 5 min on ice, and centrifuged at 526g for 5 min at 1 °C. Then, cells were sequentially washed with the same buffer and with TAPS-DMF buffer containing 1% BSA and 1 mg ml<sup>-1</sup> PVP. A total of 1 × 10<sup>6</sup> cells were transposed for 30 min at 37 °C using Tn5 assembled with ME19-Cy5 in TAPS-DMF containing 1% BSA. After the reaction, cells were washed twice with PBS containing 1% BSA, 1 mg ml<sup>-1</sup> PVP and 0.05% Triton X-100 and stained with DAPI, followed by analyses with flow cytometry (SH800 Cell Sorter or SH3800 Cell Analyzer; Sony). The mean fluorescent intensity of ATAC-see signals of the spike-in cells was used to normalize between samples.

### In vitro tagmentation with Tn5 complex

A total of 300 ng of mouse genomic DNA was tagmented using Tn5 complex with Cy3-labeled adapter DNA in 10 μl of TAPS-DMF buffer at 37 °C for 10 min. The reaction was stopped by adding 1 μl of STOP buffer (0.5% SDS and 100 mM ethylenediaminetetraacetic acid (EDTA)) and incubating at 70 °C for 5 min. DNA was analyzed by agarose gel electrophoresis and stained with ethidium bromide. Fluorescent images for ethidium bromide and Cy3 were acquired with Typhoon 6000 (GE Healthcare).

### Genome-wide CRISPR screening

We used a pooled CRISPR library called Brunello consisting of 76,441 unique sgRNAs targeting 19,114 human genes (~4 sgRNAs per gene) along with 1,000 non-targeting controls (Addgene, 73178)<sup>14</sup>. A total of 1.74 × 10<sup>8</sup> Cas9-expressing eHAP (A6) cells were transduced with pooled lentiviral library encoding puromycin-resistance gene and sgRNAs at a multiplicity of infection of 0.25–0.5 in the presence of 4 μg ml<sup>-1</sup> of polybrene (Sigma). The cells were selected with 1 μg ml<sup>-1</sup> of puromycin for 2 days, and then used for ATAC-see at 5 and 7 days post-transduction, respectively. Since knockout of essential genes often leads to cell death or growth arrest, these genes are progressively eliminated during the screening process before reaching day 7 post-transduction. However, the highly efficient knockout in eHAP cells enabled us to analyze even essential genes in early phase of the screening. Therefore, we conducted the ATAC-see assay on both day 5 and day 7 to ensure comprehensive detection of essential and non-essential genes. To eliminate the contamination of apoptotic or dead cells from the culture dish, cells were washed with PBS twice and culture dishes were tapped during the wash. This tapping procedure is inspired by the concept of ‘mitotic shake off’, to further detach apoptotic cells from the culture dishes. This tapping step effectively remove apoptotic cells from the dish, reducing their presence in the screening assay. After trypsinization, 5 × 10<sup>7</sup> cells were transposed with Cy5-labeled Tn5 complex in 2 ml of TAPS-DMF buffer containing 500 nM okadaic acid, an inhibitor of phosphatase, for 30 min at 37 °C. After the ATAC-see reaction, cells were processed for immunostaining to specifically label mitotic cells. Cells were blocked with PBS containing 2% BSA, 5 mM NaF, phosphatase inhibitor cocktail (Nacalai Tesque), 0.05% IGEPAL CA-630 and 1 mg ml<sup>-1</sup> PVP for 5 min on ice. Cells were subsequently incubated with primary antibodies (anti-histone H3S10P and anti-histone H3S28P antibodies, MAB Institute) in Can Get Signal Immunostain Solution A (Toyobo) containing 5 mM NaF and phosphatase inhibitor cocktail for 30 min on ice. After washing the cells with wash buffer (PBS containing 1% BSA, 1 mg ml<sup>-1</sup> PVP, 0.05% Triton X-100 and 5 mM NaF), cells were incubated with secondary antibody (anti-mouse IgG Alexa Fluor 555, Thermo Fisher, 1:1,000 dilution) on ice for 10 min. Cells were washed again and stained with DAPI for 30 min, followed by cell sorting with SH800 Cell Sorter. Since the signal of ATAC-see depends on cell cycle, we specifically analyzed cells at G2 phase by gating G2/M phase with DAPI staining and eliminating cells at M phase based on signals of H3S10P and H3S28P. We sorted the populations with the highest and lowest 5% of ATAC-see signal in gated G2 cells, respectively, for the subsequent analysis. Cells in parental G2 population were also collected

as a control representing all knockout cells. In one batch screening, a series of these processes was repeated five times to collect ~2.2 × 10<sup>6</sup> cells in total for each population. Three batch screenings were independently performed to collect three biological replicates. The sorted cells were treated with 0.5 mg ml<sup>-1</sup> proteinase K in genome lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 0.1% SDS) at 50 °C overnight followed by a 10 min incubation at 80 °C. Subsequently, the lysate was treated with 10 μg ml<sup>-1</sup> RNase A at 37 °C for 30 min. Genomic DNA was then extracted by phenol-chloroform extraction and precipitated with ethanol. The integrated lentivirus genome encoding sgRNA was amplified by PCR using KOD FxNeo (TOYOBO). PCR cycling conditions included an initial 2 min at 94 °C, followed by 15 s at 94 °C, 30 s at 56 °C and 30 s at 68 °C for 20 cycles and a final 3 min extension at 68 °C. Primer sequences used for the PCR are shown in Supplementary Table 1. The PCR products were fractionated by electrophoresis using a 2% agarose gel and subsequently purified using QIAquick Gel Extraction kit (QIAGEN). The libraries were sequenced with Illumina HiSeq2500 for 50 bp single-end reads. sgRNAs enriched in cell fractions with the highest or lowest 5% of ATAC-see signal were identified by comparing each other using MAGeCK-RRA in MAGeCK-VISPR v0.5.3 (ref. 65). Three biological replicates were implemented for the analysis, and data from days 5 and 7 post-transduction were analyzed independently. Diversity of sgRNAs in original lentivirus pool used for the infection and total fraction of G2 cells at day 7 were used to calculate an essential score (essential FDR) for each gene. Data were further analyzed with MAGeCKFlute<sup>66</sup>. Genes with MAGeCK score <10<sup>-7</sup> were judged as screen hits. Functional enrichment analysis was performed by using g:Profiler<sup>67</sup>. All the screen hits are shown in Supplementary Table 2.

### The rescue experiments of TFDP1 knockouts by expression of TFDP1 mutants or histones

eHAP cells were transfected with a vector expressing Cas9 and sgRNA targeting TFDP1 and selected with 1 μg ml<sup>-1</sup> puromycin for 2 days to knockout TFDP1. For the rescue with TFDP1 mutants, cells were transfected with a plasmid expressing full-length or truncated TFDP1 mutants, including ΔDBD and ΔE2BD lacking amino acids 103–126 and amino acids 233–272, respectively. All the TFDP1 expression plasmids are resistant to Cas9 cleavage due to point mutations at sgRNA targeting sites. For the rescue with canonical histones, three plasmids encoding histone H1, H2A and H2B, H3 and H4 (pLvE-FLH1, pLvE-FLH2AB and pLvE-FLH3H4) were co-transfected to the TFDP1 knockout eHAP cells using HilyMax (Dojindo). Cells were subsequently selected with 10 μg ml<sup>-1</sup> blasticidin for 2 days and assayed for ATAC-see 3 days after the transfection.

### Quantification of histone proteins

To analyze amount of histone proteins in whole cell extracts, cell lysates containing equal number of cells were applied on a SDS-PAGE gel and subsequently analyzed by CBB staining or western blotting. Chromatin fractions were extracted as described previously<sup>68</sup>. In brief, cells were washed with PBS twice and suspended in lysis buffer A (20 mM HEPES pH 7.5, 0.25 M sucrose, 3 mM MgCl<sub>2</sub>, 0.5% IGEPAL CA-630, 3 mM 2-mercaptoethanol, 0.4 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitors) for 10 min at 4 °C. The cell lysates were homogenized with 15 strokes of a type B pestle of a Dounce homogenizer. After centrifugation at 3,000g for 15 min at 4 °C, the pellet was suspended with buffer B (20 mM HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 0.2 mM egta-zinc acid, 3 mM 2-mercaptoethanol, 0.4 mM PMSF and protease inhibitors) and added to an equal volume of buffer B containing 0.6 M KCl and 10% glycerol in a dropwise fashion followed by gentle agitation for 10 min at 4 °C. The chromatin fractions were pelleted by centrifugation at 20,400g for 30 min at 4 °C, and subsequently lysed in SDS-PAGE sample buffer. Amount of genomic DNA in the lysate was measured using Qubit 2.0 with dsDNA BR assay kit (Thermo Fisher) and used for the calibration across samples. For quantitative western blotting, NuPAGE

SDS-PAGE gel was used for fractionation of proteins. Histone proteins were detected with specific primary antibodies and fluorescent secondary antibodies (anti-mouse IgG IRDye 800CW or anti-rabbit IgG IRDye 680RD, LI-COR). All the images were acquired with Odyssey CLx (LI-COR) and quantified using Image Studio software (LI-COR). The concentration of each antibody and amount of proteins loaded on an SDS-PAGE gel were titrated within a linear detection range for precise quantification. Tricine SDS-PAGE gel<sup>69</sup> was used for CBB staining.

### Calibrated ATAC-seq for analysis of knockout eHAP cells with spike-in controls

ATAC-seq was performed as described previously<sup>9,70</sup> with some modifications. To quantitatively analyze ATAC-seq data between experimental conditions, we used a fixed number of *Drosophila* S2 cells as a spike-in control for each sample. The read counts for the spike-in were used to internally calibrate ATAC-seq data, as described previously<sup>71–73</sup>. In brief,  $1 \times 10^4$  (for TFDP1 knockouts and PARP1 knockouts) or  $2.5 \times 10^3$  S2 cells (for the other knockouts) were first mixed with  $1 \times 10^5$  eHAP cells in a tube, washed with cold PBS, and centrifuged at 935g for 5 min. Cells were permeabilized with cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.1% IGEPAL CA-630). Then cells were transposed using Tn5 complex with MEAB34 adapters in TAPS-DMF buffer at 37 °C for 30 min. Genomic DNA was purified with Monarch PCR and DNA Cleanup Kit (NEB). After amplification of transposed DNA using Q5 High-Fidelity 2× Master Mix (NEB), DNA was size-selected using solid-phase reversible immobilization beads to produce libraries with fragments between 150 bp and 600 bp. To reduce a contamination of mitochondrial DNA, the libraries were treated with an anti-mitochondrial DNA CRISPR/Cas9 library to specifically digest mitochondrial DNA as described previously<sup>74</sup>. The libraries were sequenced on HiSeq3000 or HiSeq2500 for 100 bp or 50 bp paired-end reads, respectively. Two biological replicates were analyzed for each experimental condition.

### Preparation of libraries for ATAC-seq with IEOE or MEAB34 adapters

A total of  $1 \times 10^5$  mES cells were transposed using Tn5 complex with 19 bp of IEOE adapter DNAs. Since sequences of IE and OE display low melting temperature ( $T_m$ ) as 56 °C and 54 °C, respectively, custom primers with locked nucleic acid (IDT) with  $T_m$  at 63 °C and 61 °C were used for the first PCR amplification. Transposed DNA was first amplified for four cycles with KOD FxNeo polymerase (TOYOBO) using the following PCR conditions: 72 °C for 5 min, 98 °C for 30 s and thermal cycling at 98 °C for 10 s, 63 °C for 30 s and 68 °C for 1 min. After the removal of the locked nucleic acid primers using Monarch PCR and DNA Cleanup Kit, the second PCR for indexing was performed using Q5 High-Fidelity 2× Master Mix for five cycles with primers designed to amplify the first PCR products. The libraries were sequenced on HiSeq2500 for 50 bp paired-end reads using custom sequencing primers designed for IE and OE sequences. All the primer sequences are listed in Supplementary Table 1. ATAC-seq with MEAB34 adapters was performed as described previously<sup>9</sup>. Three biological replicates were analyzed for each experimental condition.

### Assay for CRISPR-Cas9 cleavage efficiency and ChIP-qPCR for Cas9 with TFDP1-knockdown cells

Since genome-editing efficiency by ectopic expression of sgRNAs from plasmids was extremely high (around 90%) in eHAP cells as shown in Extended Data Fig. 3c, we introduced in vitro transcribed sgRNAs prepared using T7 CUGA transcription kit (Nippon Gene) to control the genome-editing efficiency around 50% so that we could evaluate the effect of TFDP1 depletion on Cas9 binding and the genome editing. Cas9-expressing eHAP cells were transfected with non-targeting or TFDP1 siRNA using Lipofectamine RNAiMAX. At 2 days post-transfection, cells were transfected with in vitro transcribed sgRNAs using Lipofectamine RNAiMAX Reagent. At 2 days after the sgRNA transfection, the

genomic DNA was isolated using NucleoSpin Tissue (Marcheray-Nagel) or DirectPCR Lysis Reagent (Viagen Biotech) and genomic DNAs around the sgRNA targeting sites were amplified by PCR using KOD FX Neo polymerase (Toyobo) or Q5 High-Fidelity DNA polymerase (NEB) and subjected to gel purification. The PCR amplicons were directly processed for Sanger sequencing. The editing efficiency was quantified by TIDE program<sup>75</sup>. For ChIP-qPCR analysis, cells were crosslinked 1 day after sgRNA transfection as described above. Chromatin fractions were prepared by sonication using BioRuptor (30 s on, 30 s off, at high power for 10 cycles). The sgRNA-target regions immunoprecipitated with 1 µg of Cas9 antibody were analyzed relative to input DNA by qPCR. Antibodies and primers used for ChIP are listed in Supplementary Table 1.

### Calibrated MNase-seq with spike-in controls

To analyze nucleosome occupancy quantitatively in control and TFDP1 knockout cells, we included *Drosophila* S2 cells as a spike-in control. After mixing  $1 \times 10^6$  eHAP cells and  $5 \times 10^4$  S2 cells, cells were washed with PBS containing 0.1% BSA and subsequently permeabilized with 0.5% IGEPAL CA-630 in reticulocyte standard buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 1 mM CaCl<sub>2</sub>, 3 mM MgCl<sub>2</sub> and 0.5 mM PMSF) for 20 min on ice. Nuclei were digested in reticulocyte standard buffer containing 2.5 U MNase (Takara) at 37 °C for 10 min. The reaction was terminated by adding EDTA, egtazic acid and proteinase K. The extracted DNA was fractionated by agarose gel electrophoresis, and DNA fragments between 125 bp and 175 bp representing mono-nucleosome was purified with Monarch PCR and DNA Cleanup Kit followed by treatment with RNase. Libraries for sequencing were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) and sequenced on Illumina HiSeq3000 for 100 base paired-end reads. Two biological replicates were analyzed for each experimental condition.

### Calibrated RNA-seq with spike-in controls

A total of  $5 \times 10^5$  eHAP cells and  $1 \times 10^4$  S2 cells were mixed in a tube. RNA was extracted from  $5 \times 10^5$  eHAP cells using RNeasy Mini Kit (Qiagen). Genomic DNA was removed by TURBO DNase (Thermo). The quality of the extracted RNA was evaluated by using Bioanalyzer RNA kit (Agilent). For polyA-based RNA-seq, messenger RNA was purified using polyA magnetic beads (NEB). For Ribo-zero RNA-seq, total RNA was treated with NEBNext rRNA Depletion Kit (NEB). Libraries for sequencing were prepared using NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB) and sequenced on HiSeq4000 for 50 base single-end reads. Three biological replicates were analyzed for each experimental condition.

### Bioinformatic analysis for deep sequencing data

The details of bioinformatic analyses for ATAC-seq, ChIP-seq, MNase-seq and CRISPR screen are described in Supplementary Information.

### Statistics and reproducibility

In this study, no statistical method was used to predetermine sample size, no data were excluded from the analyses and the experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. The outcomes of all statistical tests, including *P* values and number of samples, are included in the figure panels or the corresponding figure legends. For all the ATAC-seq analysis analyzed by FACS in Figs. 1*i*, 2*c* and 4*a–e* and Extended Data Fig. 5*a,c,f*, two-tailed unpaired Student's *t*-test using the median from independent biological replicates to compare corresponding control and indicated samples were used to calculate *P* values. Significance was defined as any statistical outcome that resulted in a *P* value of less than 0.05.

### Reporting summary

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

## Data availability

The RNA-seq, ChIP-seq, ATAC-seq, MNase-seq and CRISPR-screening datasets in this research are available from the NCBI GEO repository: [GSE144454](#). Data used in, but not generated in, this study include ChIP-seq datasets: [GSE72800](#) (POLR2A), [GSE94992](#) (SMC1A and CTCF), [GSE108390](#) (H3K36me3, H3K4me1, H3K9me3 and H2AK119ub), PRJEB8671 (TFDP1 in LoVo cells), [GSE105217](#) (TFDP1 in K562 cells), [GSE80661](#) (TFDP1 in U266 cells) and [GSE80661](#) (TFDP1 in MM1.S cells). Source data are provided with this paper.

## Code availability

The publicly available softwares used are indicated in the Reporting Summary. The custom codes to analyze ATAC-seq data in Fig. 3 and Extended Data Fig. 7 are available in GitHub (<https://github.com/Park-Sung-Joon/ATACprofWS>, <https://doi.org/10.5281/zenodo.10417228>)<sup>76</sup>.

## References

63. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
64. Musinova, Y. R. et al. Nucleolar localization/retention signal is responsible for transient accumulation of histone H2B in the nucleolus through electrostatic interactions. *Biochim. Biophys. Acta* **1813**, 27–38 (2011).
65. Li, W. et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* **16**, 281 (2015).
66. Wang, B. et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat. Protoc.* **14**, 756–780 (2019).
67. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
68. Schnitzler, G. R. Isolation of histones and nucleosome cores from mammalian cells. *Curr. Protoc. Mol. Biol.* **50**, 21.5.1–21.5.12 (2000).
69. Schägger, H. Tricine–SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
70. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
71. Bonhoure, N. et al. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res.* **24**, 1157–1168 (2014).
72. Hu, B. et al. Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res.* **43**, e132 (2015).
73. Orlando, D. A. et al. Quantitative ChIP-seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).
74. Montefiori, L. et al. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci. Rep.* **7**, 1–9 (2017).
75. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
76. Park, S. J. ATACprofWS.vR2. Zenodo <https://doi.org/10.5281/zenodo.10417228> (2023).

## Acknowledgements

We appreciate all members of Miyanari lab for their kind discussion, M. Hirabayashi and T. Kobayashi (National Institute for Physiological Sciences) for help with FACS analysis, T. Hiyama and M. Noda (National

Institute for Basic Biology, NIBB) for providing HEK293T cells, T. Nishiyama (Nagoya University) and T. Kuraishi (Kanazawa University) for providing S2 cells, E. V. Sheval (Moscow State University) for providing pEGFP-H2B (21–35), K. Yamaguchi, A. Akita, S. Shigenobu and M. Matsumoto (NIBB) for technical support for deep sequencing, K. Naruse and A. Kato (NIBB) for Sanger sequencing, M. E. Torres-Padilla (Helmholtz Munich), T. Ishiuchi (Yamanashi University), M. A. Koyanagi, T. Aoi (Kobe University) and K. Nishimura (Tsukuba University) for comments on the manuscript, the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, NYU Genome Technology Center for their deep sequencing service, NIBB Data Integration and Analysis Facility, the NIG supercomputer at ROIS National Institute of Genetics for the computational resources, and NIBB Bioimaging Facility for their imaging supports. S.I. and T.K. are supported by JSPS Grant-in-Aid for Young Scientists. S.-J.P. is supported by JSPS KAKENHI (18H04710 and 20H05940). Y.O. is supported by JSPS KAKENHI (JP23H00372, JP22H04676 and JP22K19275), AMED BINDS (JP22ama121017j0001), the Medical Research Center Initiative for High Depth Omics and the Cooperative Research Project Program of the Medical Institute of Bioregulation at Kyushu University. Y.M. is supported by JSPS KAKENHI (16H06279 (PAGS), 22H04925 (PAGS), 18H04722, 22H04688 and 21H04765), World Premier International Research Center Initiative (WPI, MEXT), Okazaki Orion Project, Astellas Foundation, JST/PRESTO, Takeda Science Foundation, the Cell Science Research Foundation, Mochida Memorial Foundation for Medical and Pharmaceutical Research, the Naito Foundation, the Cannon Foundation and Ohsumi Frontier Science Foundation.

## Author contributions

S.I. and Y.M. designed and performed experiments and wrote the paper. T.K. set up the optimized ATAC-seq. A.T. supported cell sorting. C.S. performed the construction of plasmids. S.-J.P. and Y.M. conducted bioinformatic analyses. K.N. supported the bioinformatic analyses. H.T. provided a resource and technical advice for DNA FISH analysis. Y.O. contributed to the initial setup for deep sequencing. M.N. provided a resource for iPSC cell reprogramming.

## Competing interests

M.N. is the founder and Chief Technology Officer of Tokiwa-Bio, Inc. The other authors declare no competing interests.

## Additional information

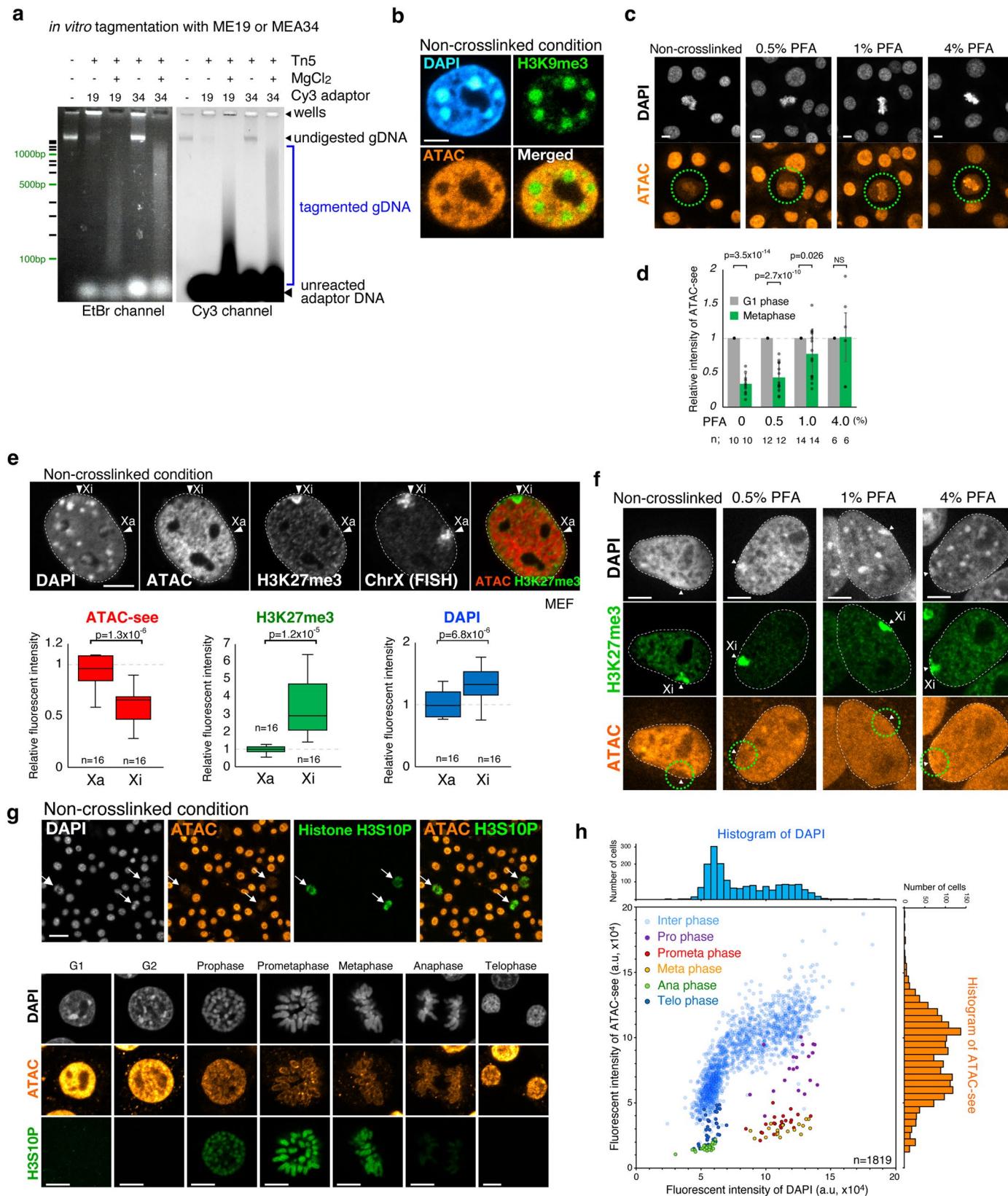
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-024-01658-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-024-01658-1>.

**Correspondence and requests for materials** should be addressed to Yusuke Miyanari.

**Peer review information** *Nature Genetics* thanks Andrey Krokhin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

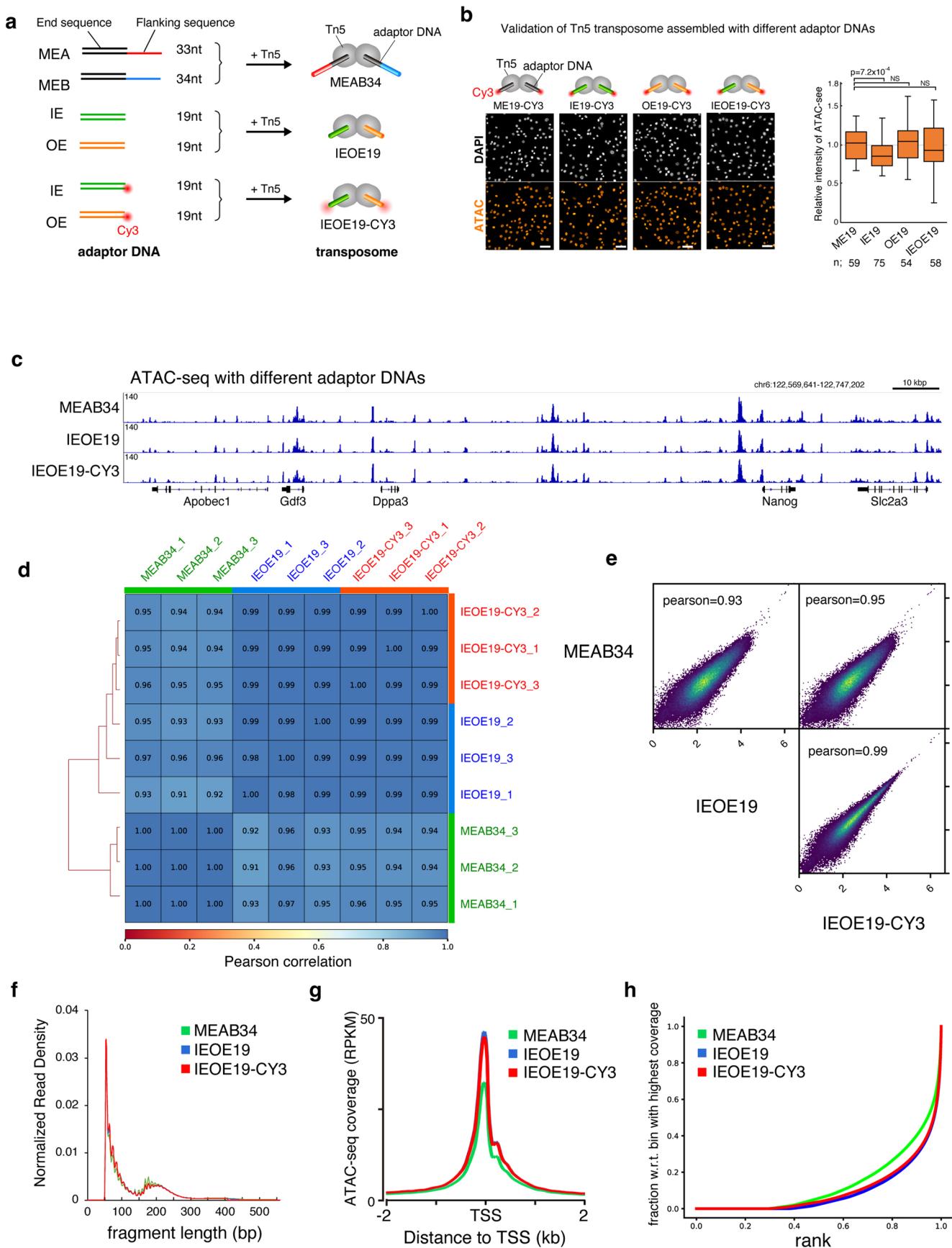
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Characterization of optimized ATAC-see.** **a**, Mouse genomic DNA (gDNA) were fragmented using Tn5 transposome with Cy3-adaptor DNAs (ME19 or MEA34) with or without MgCl<sub>2</sub>. Agarose gel images for ethidium bromide (EtBr) and Cy3 channels are displayed. **b**, Representative images of NIH3T3 cells stained for ATAC-see, DAPI and H3K9me3 under the non-crosslinked condition. scale bar, 2  $\mu$ m. **c**, Representative images of NIH3T3 cells stained for ATAC-see and DAPI under indicated crosslink conditions (the left panels). Green dashed circles, mitotic cells. scale bars, 5  $\mu$ m. **d**, A bar plot shows intensity of ATAC-see on cells at metaphase relative to that at G2 phase. Data are presented as mean values  $\pm$  SD. n, number of cells analyzed. Statistics: a two-tailed unpaired Student's t-test. **e**, Representative images of female mouse embryonic fibroblasts (MEFs) stained for ATAC-see, H3K27me3, DAPI, and X chromosome (Chr X, DNA-FISH) without cell-crosslinking (the top panels). Xi and Xa indicate inactive

and active Chr X, respectively. Relative fluorescent intensities for indicated marks in Chr X territories are displayed at bottom panels. Box plots (the Bottom panels) are shown as Fig. 1c. n, number of cells analyzed. Scale bars, 5  $\mu$ m. Statistics: a two-tailed unpaired Student's t-test. **f**, Images for ATAC-see and H3K27me3 on female MEFs crosslinked with indicated concentration of PFA. Arrowheads and green dashed circles, inactive Chr X. Scale bars, 5  $\mu$ m. **g**, Representative images of NIH3T3 cells stained for ATAC-see, DAPI, phospho histone H3 Ser10 (H3S10P) under the non-crosslinked condition (the upper panels, scale bar, 30  $\mu$ m). Arrows, H3S10P-positive mitotic cells. Lower panels show higher magnification images of indicated cell-cycle stages. Scale bars, 5  $\mu$ m. **h**, Scatter plot and histograms representing intensities for DAPI and ATAC-see in cells at i n, number of cells analyzed. Experiments in a, b, c, and g were reproduced three times.

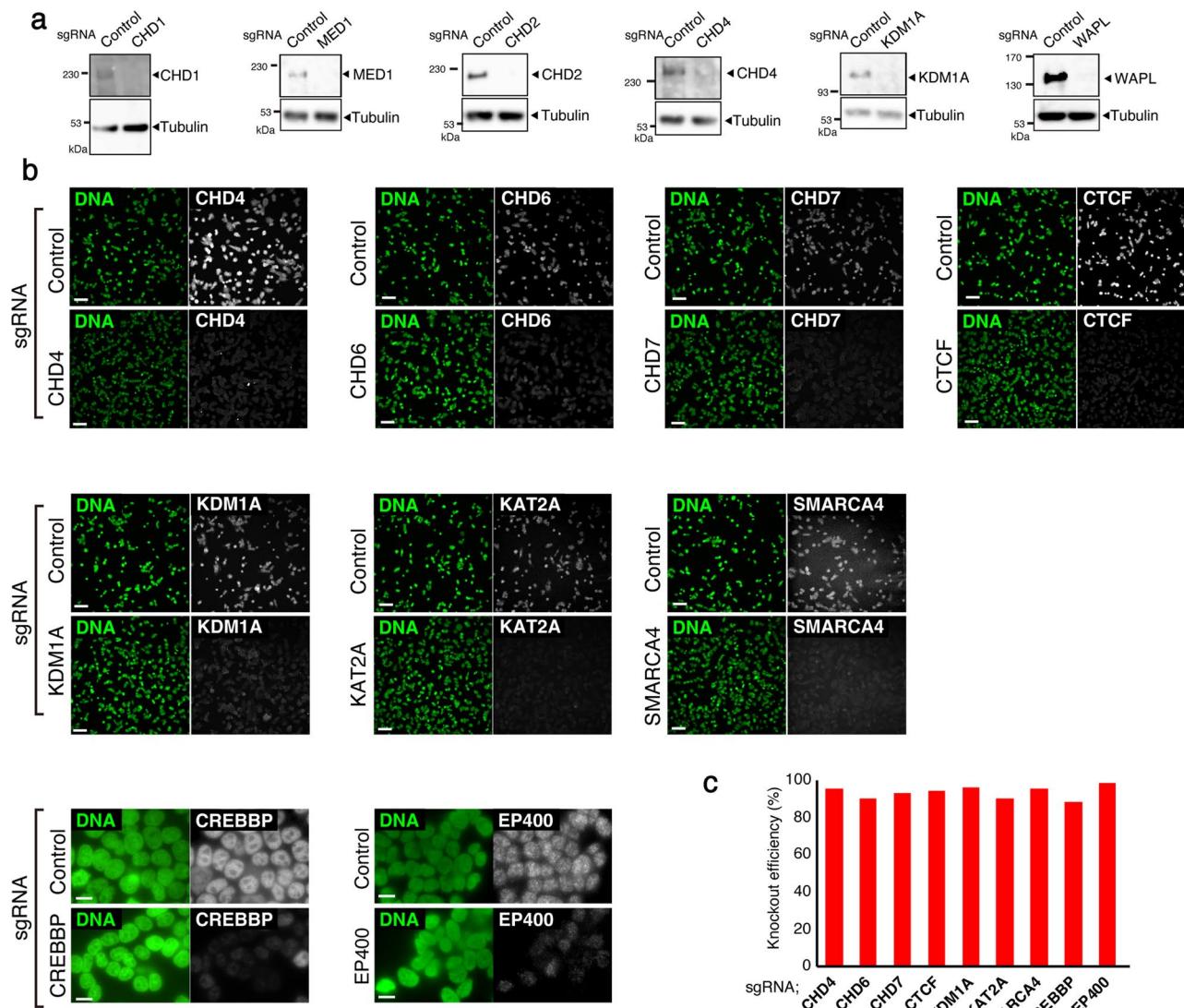


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Validation of optimized ATAC-seq by ATAC-seq analysis.**

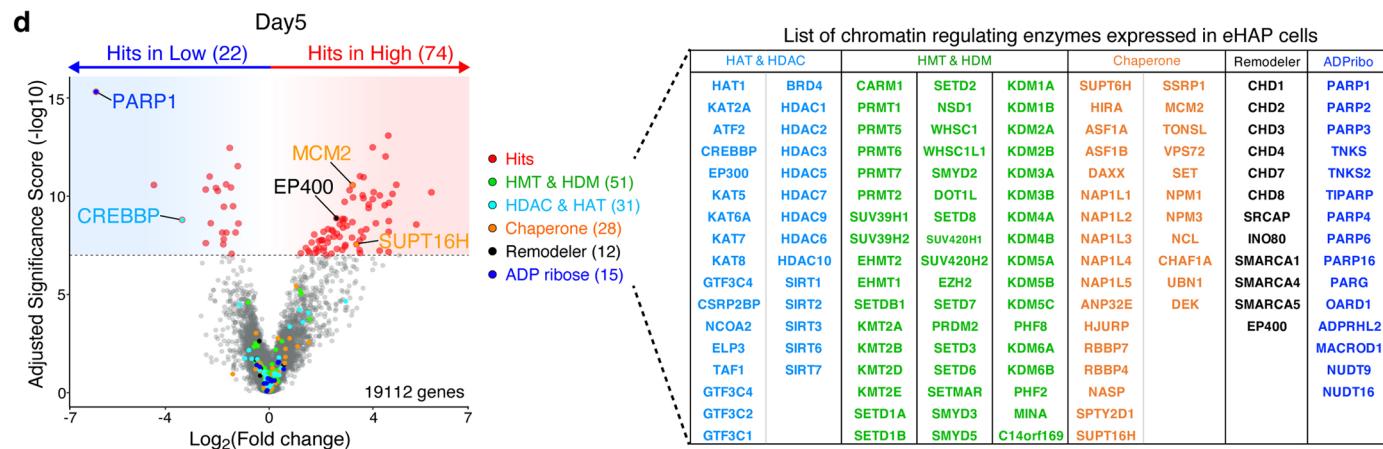
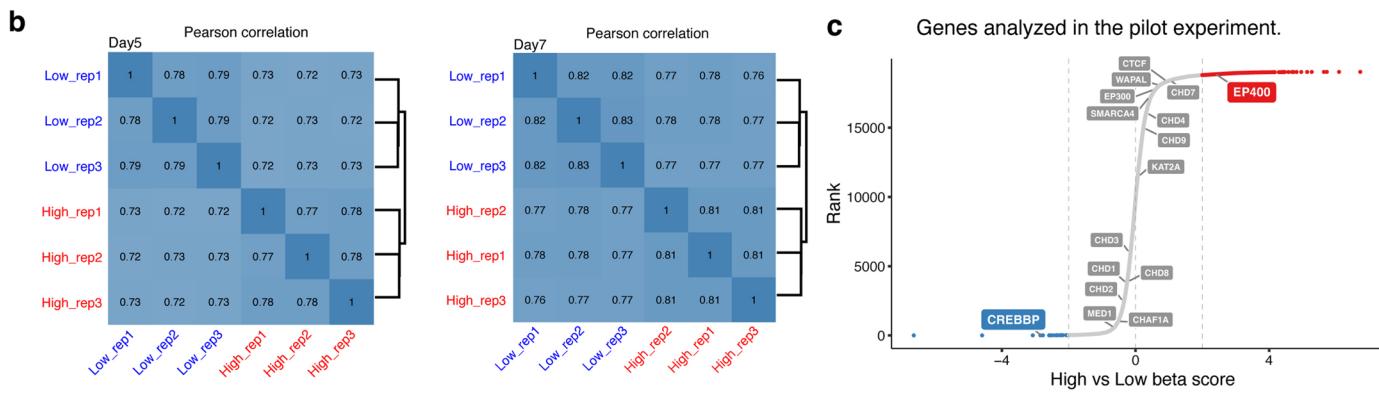
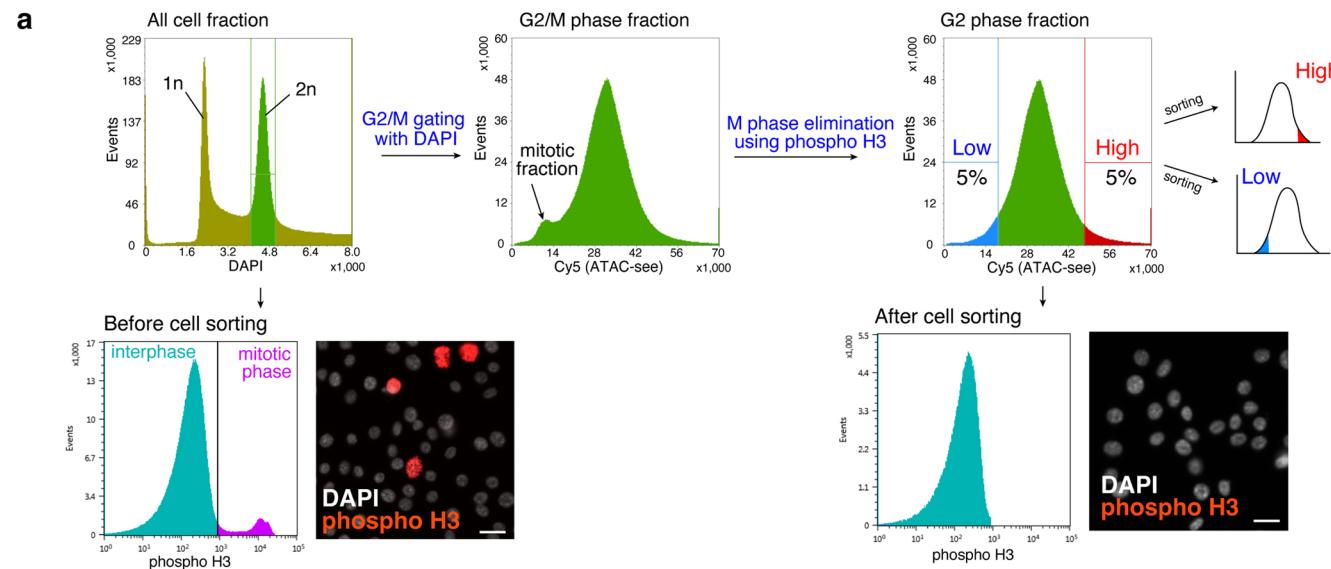
**a**, Schematic illustrations displaying Tn5 transposome assembled with indicated adaptor DNAs. **b**, Representative images of NIH3T3 cells analyzed by ATAC-seq using indicated transposomes (the left panels, scale bars, 30  $\mu$ m). n, number of cells analyzed. A boxplot in the right panel shows relative intensity of ATAC-seq. Statistics and definition of the box plot are the same as Fig. 1*i*. Experiments were reproduced three times. **c**, Representative genome browser tracks of ATAC-seq analysis with mES cells using indicated transposomes. **d**, A correlation matrix between ATAC-seq using indicated transposomes. A heat map represents

clustering of three independent biological replicates for each experimental condition based on read enrichments in merged ATAC-seq peaks. **e**, Pairwise scatter plots of signals in the merged peaks. Values of correlation coefficients (Pearson) for each comparison are indicated. **f**, Insert-size distributions of ATAC-seq using indicated adaptor DNAs. **g**, Metagene profiles for read enrichments around the transcriptional start sites (TSS). **h**, Signal extraction scaling (SES) computed by using deeptools plotFingerprint represents cumulative read coverages.



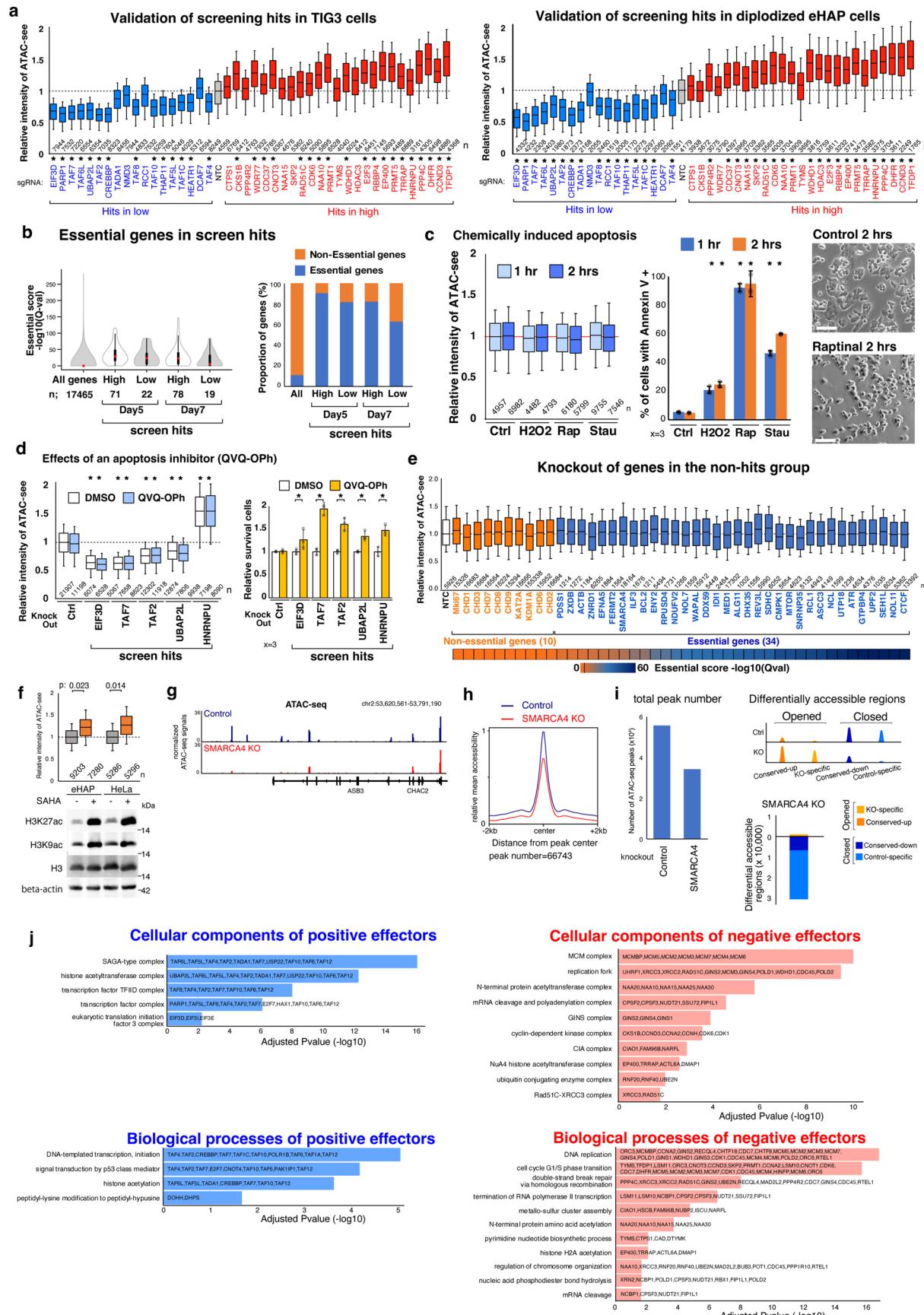
**Extended Data Fig. 3 | CRISPR-mediated depletion of chromatin related genes.** eHAP cells expressing Cas9 were transfected with a plasmid encoding two pairs of sgRNAs against indicated genes and analyzed by either western blotting (**a**) or immunostaining (**b**) on day 4–5 post-transfection. **a**, Specific bands using indicated antibodies are shown with arrowheads. Size markers (kDa)

are indicated in left side of each panel. These experiments were reproduced twice. **b**, Immunofluorescence images with indicated antibodies (gray) and DNA staining (green). scale bars, 30  $\mu$ m. **c**, Knockout efficiency calculated based on immunostaining data. n; number of cells analyzed. Scale bars, 10  $\mu$ m.



**Extended Data Fig. 4 | Extended results of the CRISPR screening.** **a**, A gating strategy for flow cytometry to collect cells used for the CRISPR screening. Cells stained for ATAC-see, DAPI, and phospho Histone H3 were analyzed for a DAPI channel to gate cells at G2/M phase. The cells were subsequently gated out based on the signals of phospho Histone H3 to eliminate the mitotic fraction. A cell population displaying the highest or lowest 5% of ATAC-see signals was collected, respectively. Immunostaining images for phospho H3 and DAPI before and after the cell-sorting (the bottom panels) confirm the high purity of G2 phase. Scale bars, 15  $\mu$ m. **b**, Pearson correlations of CRISPR screen data between

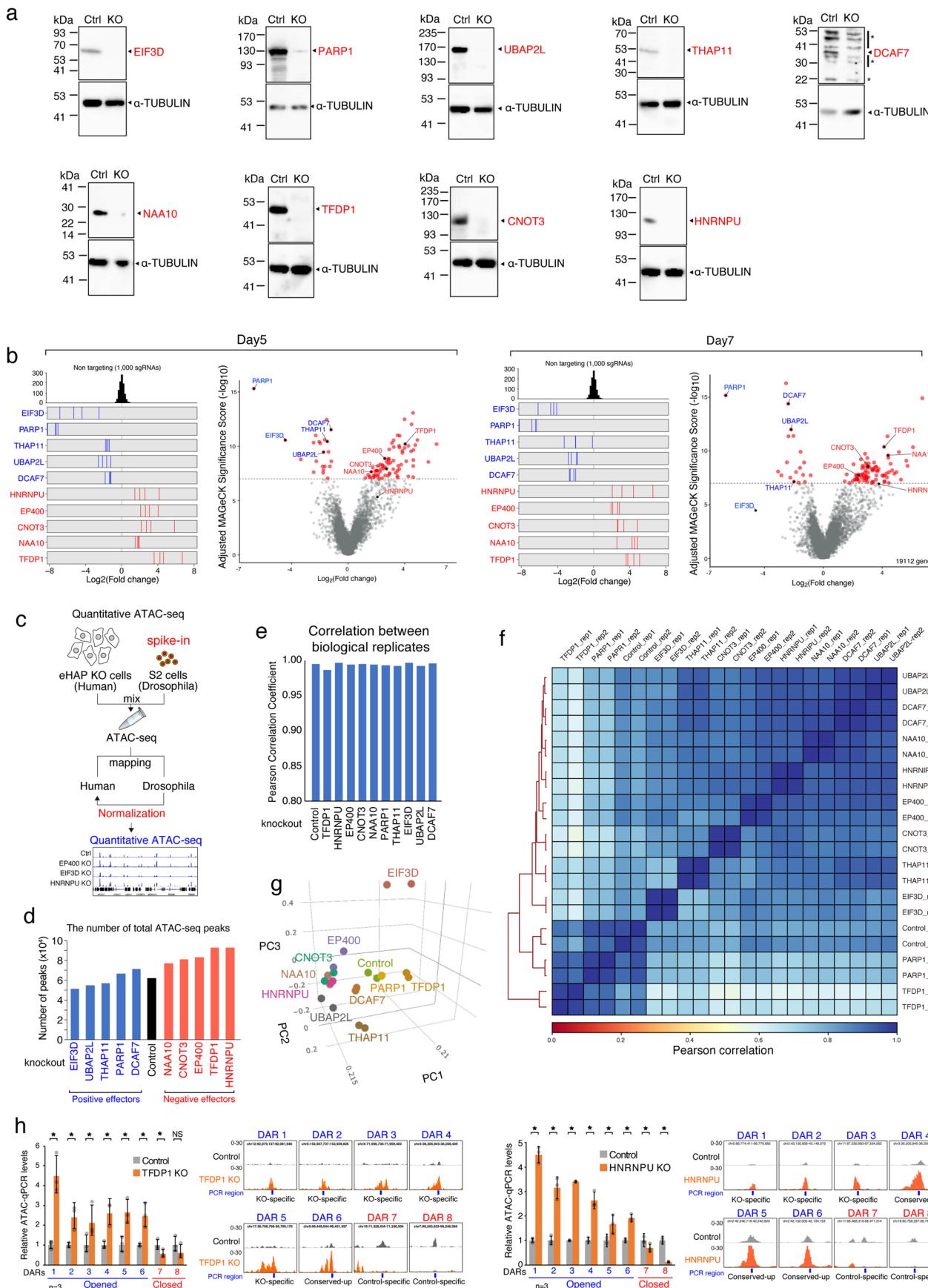
biological replicates from the high and low fractions. Data from day 5 or day 7 after transduction were analyzed by MAGeCK-VISPR. **c**, The beta score and rank of genes analyzed in Fig. 1i. Only the data from day 7 is shown here. **d**, A volcano plot for gene hits from the screening result at day 5. Data is shown as in Fig. 2d (the left panel). Lists of chromatin regulating enzymes highlighted in the plot (the right panel). Genes are included in these categories only in the case where they are expressed in eHAP cells (TPM  $\geq$ 10) and corresponding sgRNAs are not depleted at day 7.



Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Validation of screen hits.** **a**, ATAC-seq analysis with indicated cell lines as shown in Fig. 2c. **b**, Distribution of essential scores and proportion of the essential genes (the left and right panels) for all genes and indicated screen hits. **c**, ATAC-seq analysis measured by FACS (the left panel), annexin V staining (the middle panel), and bright field images with 30  $\mu$ m scale bars (the right panel) under indicated conditions. Ctrl: control, H2O2: Hydrogen peroxide, Rap: Raptinal, Stau: Staurosporine. **d**, eHAP cells were cultured with or without QVQ-OPh during the knockout, and subsequently assayed for ATAC-seq and survival rates (the left and right panels). **e**, ATAC-seq analysis measured by FACS on knockouts for indicated non-essential and essential genes selected from the non-hit group. Essential score for each gene is displayed at bottom. All the samples showed statistically non-significant ( $P > 0.05$ ). **f**, eHAP or HeLa cells were treated with 1  $\mu$ M SAHA, for 8 hrs and subjected to ATAC-seq measured by FACS (the top panel) and western blotting with indicated antibodies (the bottom panel).

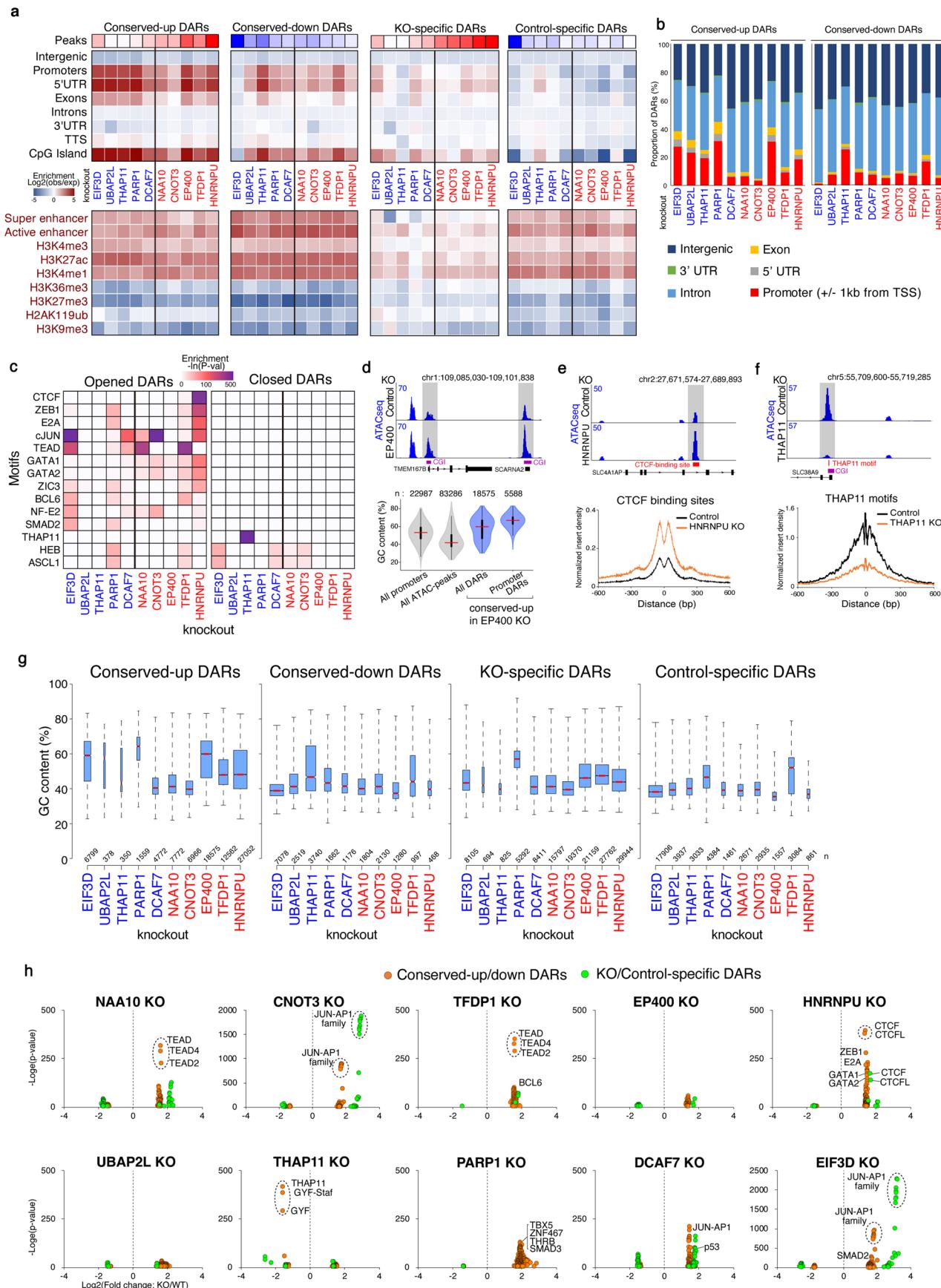
**g**, Representative genome browser tracks of spike-in calibrated ATAC-seq on control and *SMARCA4* knockout eHAP cells. **h**, Relative mean ATAC-seq signals around merged ATAC-seq peaks between control and *SMARCA4* KO. **i**, The left panel, the total number of ATAC-seq peaks detected in indicated knockouts. The right panels, schematic illustration showing classification of differentially accessible regions (DARs, top) and the number of DARs in *SMARCA4* knockout eHAP cells (bottom). **j**, Gene enrichment analyses with gene hits. Statistics, hypergeometric test computed using g:Profiler. n in a, c-f indicates number of cells analysed. x in c and d indicates number of independent experiments. Statistics in the right panels of c and d: a two-tailed unpaired Student's t-test. Statistics and the definition of box plots for ATAC-seq assays in a, c-f, and i are the same as Fig. 1i. Data in the right panels of c and d are presented as mean values  $\pm$  SD. \*:  $P < 0.05$ .



Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Validation of ATAC-seq analyses for knockout of the selected screen hits.** **a**, Knockouts of indicated genes (colored in red) were confirmed by western blotting. Arrowheads indicate specific bands and asterisks indicate non-specific bands. These experiments were reproduced twice. **b**, Results of CRISPR screen for the selected screen hits at day 5 and day 7. Left panels for each day show log<sub>2</sub> fold changes of 4 sgRNAs designed to the indicated positive (blue) and negative regulators (red) analyzed by ATAC-seq in Fig. 3. Volcano plots for these genes are shown as Fig. 2d (the right panels). **c**, Schematic illustrations displaying the quantitative ATAC-seq with spike-in control. **d**, Numbers of total ATAC-seq peaks in the indicated knockout conditions. **e**, Pearson correlation between two biological replicates. Normalized

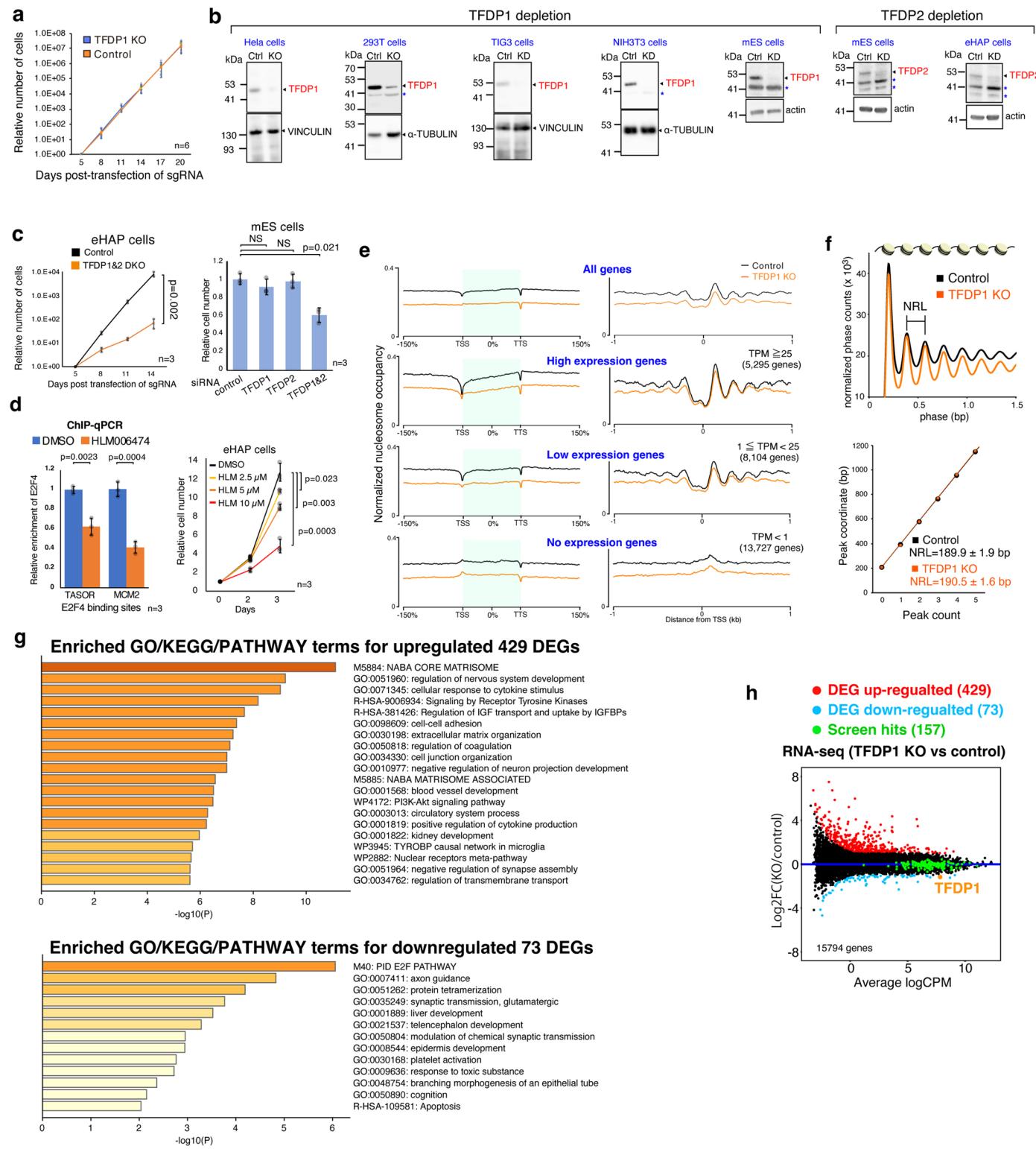
read enrichments within TSS proximal regions (TSS +/− 1 kb, 22987 RefSeq genes) were analyzed. **f**, Pairwise Pearson correlation of ATAC-seq samples based on normalized read enrichments in the consensus peak set ( $n = 256,617$ ). **g**, PCA clustering analysis of ATAC-seq data. **h**, ATAC-qPCR analysis for TFDPI KO (the left panels) and *HNRNPU* KOs (the right panels) to confirm differential accessibility at DARs identified by ATAC-seq analysis. Genome browser tracks for ATAC-seq of control and the indicated knockout cells at corresponding DARs are shown on the right side. The DAR categories are shown at the bottom of corresponding browser tracks. Statistics: a one-side unpaired Student's t-test. \*:  $P < 0.05$ . NS, not significant. n, independent experiments. Data of ATAC-qPCR are presented as mean values  $\pm$  SD.



Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Extended ATAC-seq analyses for DARs of each knockout.** **a**, Enrichments of genomic elements (upper half) and chromatin signatures (lower half) at DARs categorized as Fig. 3a in indicated knockouts. **b**, Proportion of genomic features in conserved-up and conserved-down DARs in indicated knockouts. **c**, Enrichments of transcription factor motifs at the opened and closed DARs in indicated knockouts. **d-f**, Representative genome browser tracks of ATAC-seq analysis of the control and the indicated knockouts. The DARs are highlighted in grey. The GC contents of the indicated regions are shown in a violin plot (the bottom of d). Red bars show the medians. The box plots are shown

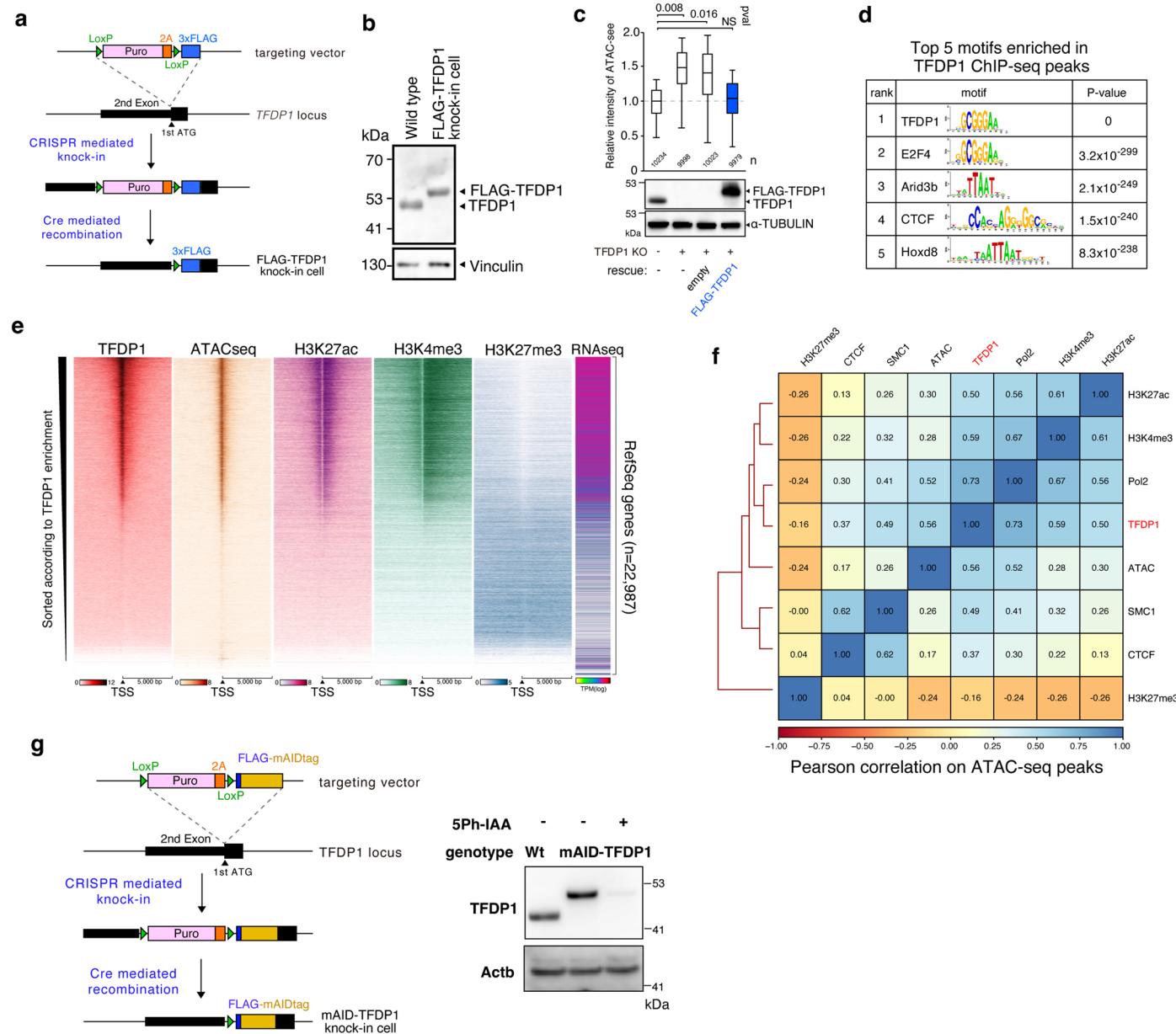
as Fig. 1c. Aggregation plots at the bottom of e and f indicate normalized density of insertion sites at the indicated genomic sites. **g**, Box plots show proportion of GC content of indicated DARs in each knockout. The width of the boxes indicates a relative number of DARs. The box plots are shown as Fig. 1c. **n**, number of DARs. **h**, Motif enrichment analysis of the DARs. Each dot represents significantly enriched motif. x-axis, logFC of ATAC-seq signals at the DARs associated with enriched motif. y-axis, significance of motif enrichment. Motifs with similar sequences are marked with dashed circles.



Extended Data Fig. 8 | See next page for caption.

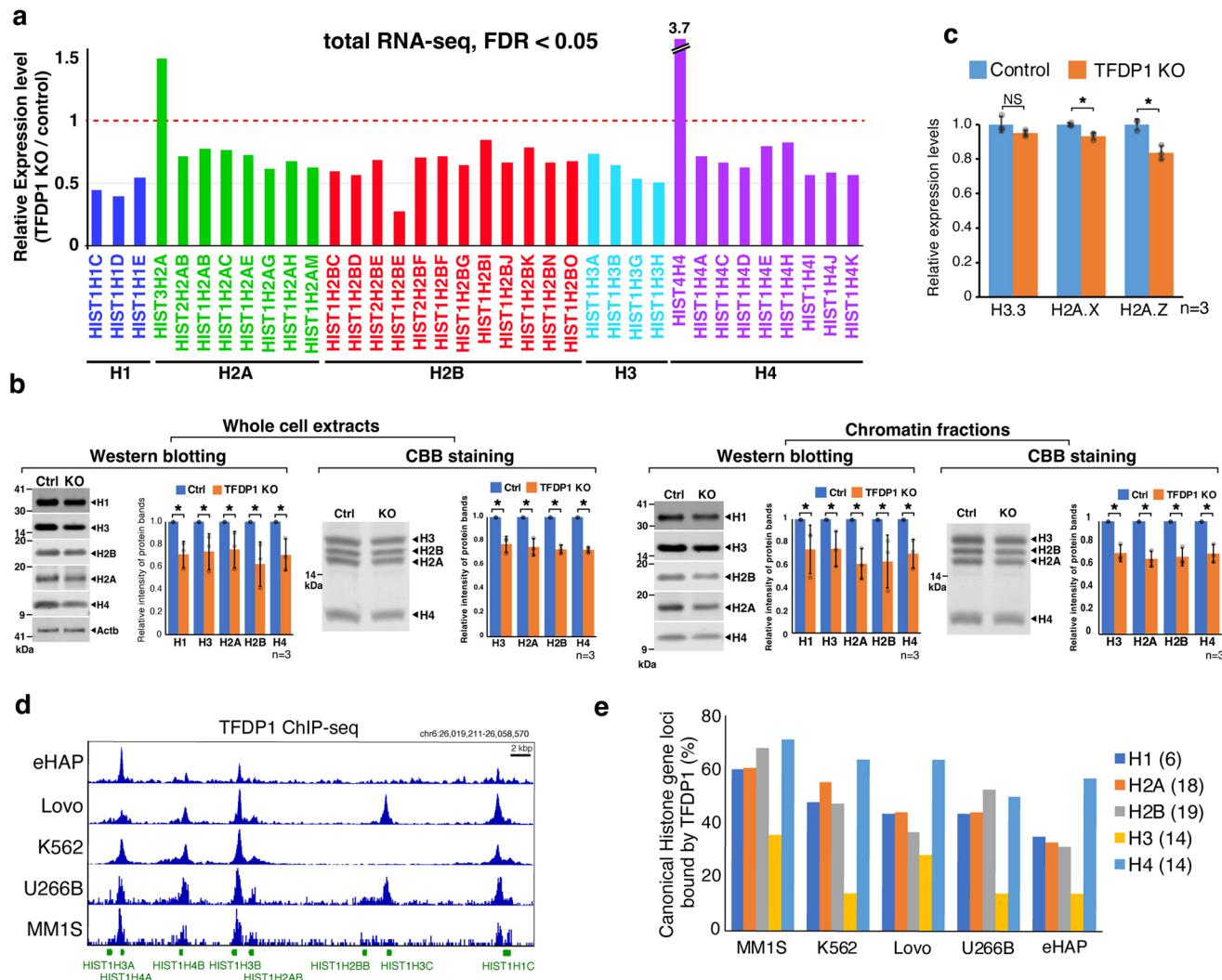
**Extended Data Fig. 8 | Characterization of *TFDP1* knockout cells.** **a**, Growth curves of eHAP cells transfected with either non-targeting (control) or *TFDP1* sgRNAs. **b**, Western blotting with the indicated cell lines transfected with sgRNA (for eHAP, HeLa and 293 T cells) and siRNA (for TIG3, NIH3T3, and mES cells) against non-targeting control (Ctrl), *TFDP1* (*TFDP1*-depletion) or *TFDP2* (*TFDP2*-depletion). Arrowheads indicate specific bands and asterisks indicate non-specific bands. These experiments were reproduced twice. **c**, Growth curves of eHAP cells transfected with either non-targeting (control) or both *TFDP1* and *TFDP2* sgRNAs (the left panel). Relative cell number of mES cells 4 days after the transfection with indicated siRNA (the right panel). Statistics: a two-tailed unpaired Student's t-test. NS: non-significant. **d**, ChIP-qPCR analysis for the putative E2F4 binding sites of indicated gene loci. eHAP cells were cultured with or without 5 μM HLM006474. (the left panel). Growth curves of eHAP cells treated with the indicated concentration of HLM006474

(the right panel). Statistics: a two-tailed unpaired Student's t-test. **e**, Averaged nucleosome occupancy of the control and *TFDP1* KO at positions across 300% of coding regions (the left panels) or within 2 kb of TSS of all genes, high-, low-, and non-expressed genes (the right panels). The expression status (High, Low, or no expression) is categorized based on transcripts per kilobase million (TPM). **f**, Normalized MNase-seq read counts of distances between neighboring nucleosomes (the top panel). Linear fits to the phase peaks calculated from the data in the top are shown with resulting NRLs (the bottom panel). **g**, Gene enrichment analysis on RNA-seq data for differentially expressed genes upon *TFDP1* KO. P-values were computed by representation analysis using Metascape. **h**, A MA plot for the comparison between *TFDP1* KO and control cells with gene spots for up- (red spots), down-regulated genes (blue spots), screen hit genes (green spots), and *TFDP1* (an orange spot). Data in a, c-d are presented as mean values ± SD. n in a, c and d, independent experiments.



**Extended Data Fig. 9 | Characterizations of FLAG or mAID knock-in TFDP1 eHAP cells.** **a**, Schematic diagram showing the generation of FLAG-TFDP1 knock-in eHAP cells. The targeting cassette was knocked-in to *TFDP1* locus by CRISPR-Cas9 mediated homologous recombination. Representative clone was further transfected with a plasmid expressing Cre recombinase to flip out Puromycin resistance cassette, resulting in FLAG-TFDP1 knock-in cells. **b**, Western blotting for indicated proteins in wild-type and FLAG-TFDP1 eHAP cells. **c**, *TFDP1* knockout eHAP cells (*TFDP1* KO) were transfected with either an empty or FLAG-TFDP1 expressing plasmid. 3 days after the transfection, cells were analyzed by ATAC-seq (the top panel). The statistical test and definition of the box plot are the same as in Fig. 1i. \*:  $P < 0.05$ . Western blotting for TFDP1 and  $\alpha$ -Tubulin

are shown in the bottom panels. **d**, Top 5 transcription factor-binding motifs enriched in TFDP1 binding sites. Statistics, Hypergeometric test computed using Pscan-ChIP. **e**, Heat maps of the indicated chromatin profiles aligned at  $\pm 5,000$  bp of TSSs in 22,987 individual RefSeq genes, sorted by TFDP1 ChIP-seq enrichment levels. Expression levels of corresponding genes analysed by RNA-seq are shown in the right side. **f**, Pearson correlation between enrichments of indicated factors on ATAC-seq peaks. **g**, Schematic diagram representing the generation of mAID-TFDP1 knock-in cells (the left panel). Acute depletion of TFDP1 upon adding 5Ph-IAA for 6 hrs was confirmed by western blotting (the right panel). Experiments in b and g were reproduced three times.



**Extended Data Fig. 10 | TFDP1 transcriptionally regulates canonical histones.** **a**, Relative expression levels of canonical histones calculated from total RNA-seq of wild-type and *TFDP1* KO eHAP cells. Only expressed histone genes (LogCPM>2) are shown here. **b**, Quantitative fluorescent western blotting analyses (left) and CBB staining quantification (right) in either whole cell extracts or chromatin fractions of wild-type and *TFDP1* KO eHAP cells. Amount of genomic DNA was used as an internal control for calibrating sample variations. Statistical comparison: a two-side unpaired Student's t-test. \*: P < 0.05. n, independent

experiments. Data are presented as mean values +/- SD. **c**, Relative expression levels of mRNA of the indicated non-canonical histone variants in control and *TFDP1* KO cells. Statistics: a two-tailed unpaired Student's t-test. NS: non-significant. \*: P < 0.05. n, independent experiments. Data are presented as mean values +/- SD. **d**, Genome browser track of *TFDP1* ChIP-seq at a histone cluster *HIST1* locus in the indicated human cell lines. **e**, Proportion of the indicated canonical histone gene loci occupied by *TFDP1*. The total numbers of analyzed genes for each canonical histone are shown in the right side.

## Reporting Summary

Nature Portfolio 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 Portfolio 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

Libraries were sequenced on a NovaSeq6000, HiSeq2500, and HiSeq300 (illumina). Fluorescent images were acquired using CellVoyager CV1000 software or Nikon A1 NIS-elements and analyzed using ImageJ (2.9.0). Western blots or CBS stained gels were analyzed by LI-COR Odyssey Image Studio or Bio-Rad Image Lab. FACS data was analyzed by SH800 and FlowJo 10.1.

Data analysis

fastqc v0.11.9, bowtie2 v.2.2.5.1, macs2 2.2.8, bedtools 2.31.0, deeptools 3.5.2, picardtools 3.0.0, IDR 2.0.0, annoPeakR, IGV 2.12.2, EaSeq, HOMER 4.11, PScan-ChIP, NuMap, trimgalore 0.6.5, sambamba 1.0.0, samtools 1.17, ROSE, featureCounts 2.0.6, Trimmomatic, HISAT2 2.2.0, StringTie 2.2.0, NuMap. Custom codes used in this manuscript can be found at <https://doi.org/10.5281/zenodo.1041722>.

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Deep sequencing data in this study is available from GSE144454 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144454>). Data used in, but not generated in, this study includes ChIP-seq datasets: GSE72800 (ROLR2A), GSE94992 (SMC1A, CTCF), GSE108390 (H3K36me3, H3K4me1, H3K9me3, H2AK119ub), PRJEB8671 (TFDP1 in LoVo cells), GSE105217 (TFDP1 in K562 cells), GSE80661 (TFDP1 in U266 cells), and GSE80661 (TFDP1 in MM1.S cells). Reference genomes are hg38, mm10, or dm6.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<a href="#">Not applicable</a>
Reporting on race, ethnicity, or other socially relevant groupings	<a href="#">Not applicable</a>
Population characteristics	<a href="#">Not applicable</a>
Recruitment	<a href="#">Not applicable</a>
Ethics oversight	<a href="#">Not applicable</a>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## 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 ChIP-seq, ATAC-seq, and MNase-seq was determined according to our experience as well as literatures reporting in terms of specific experiments (Kurihara M. et al, Molecular Cell, 78, 3, 493-505, 2020, Landt SG., Genome Research, 2012. 22(9):1813-1831)
	CRISPR screening, RNA-seq, quantitative fluorescent western blotting, and CBB staining for histones were repeated at 3 times. Experiments for iPS reprogramming were independently performed 12 times.
Data exclusions	No data were excluded from any of our analyses
Replication	Typically, three independent biological replicates (minimum two replicates) were used for all experiments, unless otherwise stated. All trends reported in this study were confirmed across replicates

Randomization	Choosing the gene sets to validate the screening results was performed randomly.
Blinding	There was no blinding performed in this study. Our study did not include any subjective measurements that could be influenced by an unconscious bias to the analysis.

## 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"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

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

## Antibodies

### Antibodies used

All the antibodies used in this study are provided in supplementary table 1.

aTUBULIN Calbio CP06, WB (1:1000)  
 CHD1 Active Motif 39990, WB (1:1000)  
 CHD2 GeneTex GTX112359, WB (1:1000)  
 CHD4 GeneTex GTX124186, WB (1:1000)  
 CHD6 Santa Cruz sc-393445, IF (1:1000)  
 CHD7 Santa Cruz sc-390742, IF (1:1000)  
 MED1 IGBMC 1TRA-1A10, WB (1:1000)  
 KDM1A Abcam ab17721, WB (1:1000)  
 WAPL Santa Cruz sc-365189, WB (1:1000)  
 CTCF Abcam ab70303, IF (1:1000)  
 KAT2A Abcam ab153903, IF (1:1000)  
 SMARCA4 Active Motif 61626, IF (1:1000)  
 CREBBP Cell Signaling Technology 7389, IF (1:1000)  
 EP400 Abcam ab5201, IF (1:1000)  
 H3S10P MAB Institute 305-35193, FACS (1:1000)  
 H3S28P MAB Institute 383-09143, FACS (1:1000)  
 EIF3D Santa Cruz sc-271515, WB (1:1000)  
 PARP1 Abcam ab191217, WB (1:1000)  
 UBAP2L Bethyl Laboratories A300-534A-T, WB (1:1000)  
 THAP11 R&D Systems MAB5727-SP, WB (1:1000)  
 DCAF7 Novus Biologicals NBP1-74157, WB (1:1000)  
 NAA10 Santa Cruz sc-373920, WB (1:1000)  
 TFPD1 Abcam ab124678, WB (1:1000)  
 TFPD1 MBL K0095-3, WB (1:1000)  
 CNOT3 proteintech 11135-1-AP, WB (1:1000)  
 HNRNPU Santa Cruz sc-32315, WB (1:1000)  
 ACTB sigma A1978, WB (1:1000)  
 E2F4 Abcam ab150360, WB (1:1000)  
 VINCLIN MBL PM088, WB (1:1000)  
 FLAG Sigma F1802, WB (1:1000), ChIP (1ug/reaction)  
 H3K27ac MAB Institute MABI0309, ChIP (1ug/reaction)  
 H3K4me3 MAB Institute MABI0304, ChIP (1ug/reaction)  
 H3K27me3 Cell Signaling Technology 9733, WB (1:1000), ChIP (1ug/reaction)  
 H3K9ac MAB Institute CMA305, WB (1:1000)  
 H1 Active Motif 39708, WB (1:1000)  
 H2A Active Motif 39112, WB (1:1000)  
 H2B Active Motif 39126, WB (1:1000)  
 H3 CST 4499, WB (1:1000)  
 H4 Millipore 07-108, WB (1:1000)  
 Cas9 Abcam ab191468, WB (1:1000)  
 Oct4 BD bioscience 611202, WB (1:1000), ChIP (1ug/reaction)  
 Klf4 R&D AF3640, WB (1:1000), WB (1:1000), ChIP (1ug/reaction)  
 TFPD2 proteintech 11500-1-AP, WB (1:1000)  
 TRA-1-60 eBioscience 14-8863-80, IF (1:1000)  
 anti-mouse IgG Alexa Fluor 555, IF (1:1000)  
 anti-mouse IgG IRDye 800CW, WB (1:2000)  
 anti-rabbit IgG IRDye 680RD, LI-COR, WB (1:2000)

### Validation

aTUBULIN manufacturer validated  
 CHD1 manufacturer validated and confirmed by WB on knockout eHAP cells  
 CHD2 manufacturer validated and confirmed by WB on knockout eHAP cells  
 CHD4 manufacturer validated and confirmed by WB on knockout eHAP cells  
 CHD6 manufacturer validated and confirmed by IF on knockout eHAP cells  
 CHD7 manufacturer validated and confirmed by IF on knockout eHAP cells

MED1 manufacturer validated and confirmed by WB on knockout eHAP cells  
 KDM1A manufacturer validated and confirmed by WB on knockout eHAP cells  
 WAPL manufacturer validated and confirmed by WB on knockout eHAP cells  
 CTCF manufacturer validated and confirmed by IF on knockout eHAP cells  
 KAT2A manufacturer validated and confirmed by IF on knockout eHAP cells  
 SMARCA4 manufacturer validated and confirmed by IF on knockout eHAP cells  
 CREBBP manufacturer validated and confirmed by IF on knockout eHAP cells  
 EP400 manufacturer validated and confirmed by IF on knockout eHAP cells  
 H3S10P manufacturer validated  
 H3S28P manufacturer validated  
 EIF3D manufacturer validated and confirmed by WB on knockout in eHAP cells  
 PARP1 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 UBAP2L manufacturer validated and confirmed by WB on knockout in eHAP cells  
 THAP11 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 DCAF7 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 NAA10 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 TFPD1 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 TFPD1 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 CNOT3 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 HNRNPU manufacturer validated and confirmed by WB on knockout in eHAP cells  
 ACTB manufacturer validated  
 E2F4 manufacturer validated and confirmed by WB on knockout in eHAP cells  
 VINCULIN manufacturer validated  
 FLAG manufacturer validated  
 H3K27ac manufacturer validated  
 H3K4me3 manufacturer validated  
 H3K27me3 manufacturer validated  
 H3K9ac manufacturer validated  
 H1 manufacturer validated  
 H2A manufacturer validated  
 H2B manufacturer validated  
 H3 manufacturer validated  
 H4 manufacturer validated  
 Cas9 manufacturer validated  
 Oct4 manufacturer validated  
 Klf4 manufacturer validated  
 TFPD2 manufacturer validated  
 TRA-1-60 manufacturer validated  
 anti-mouse IgG Alexa Fluor 555, manufacturer validated  
 anti-mouse IgG IRDye 800CW, manufacturer validated  
 anti-rabbit IgG IRDye 680RD, LI-COR, manufacturer validated

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

All the cell lines used in this study are provided in supplementary table 1.  
 NIH3T3 lab stock  
 mESC (E14) lab stock  
 eHAP cells from Horizon Discovery  
 Mouse embryonic fibroblast Lab made  
 HeLa cell, A gift from Dr. Noda (ATCC CCL2)  
 HEK293T cell, A gift from Dr. Noda (ATCC CRL-3216)  
 TIG3 cell from JCRB cell bank  
 Drosophila S2 cells, a gift from Dr. Nishiyama (ATCC CRL-1963)

Authentication

All cell lines used in this study have previously been validated by PCR, sequencing, and western blotting, and microscopic observation where appropriate.

Mycoplasma contamination

All cell lines used were regularly tested for mycoplasma contamination and found to be negative.

Commonly misidentified lines  
 (See [ICLAC](#) register)

No commonly misidentified cell lines were used.

## Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a

# ChIP-seq

## Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

### Data access links

*May remain private before publication.*

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144451>

### Files in database submission

GSM4288482 ChIP-seq\_H3K27ac-1  
 GSM4288483 ChIP-seq\_H3K27ac-2  
 GSM4288484 ChIP-seq\_H3K27me3-1  
 GSM4288485 ChIP-seq\_H3K27me3-2  
 GSM4288486 ChIP-seq\_H3K4me3-1  
 GSM4288487 ChIP-seq\_H3K4me3-2  
 GSM4288488 ChIP-seq\_TFDP1-FL1  
 GSM4288489 ChIP-seq\_TFDP1-FL2  
 GSM4288490 ChIP-seq\_eHAP input gDNA

### Genome browser session (e.g. [UCSC](#))

*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

## Methodology

### Replicates

ChIP-seq analyses for TFDP1, H3K27ac, H3K4me3, and H3K27me3 were performed with 2 biological replicates.

### Sequencing depth

H3K27ac-1 total\_reads\_35,967,315 50 PE  
 H3K27ac-2 total\_reads\_41,451,666 50 PE  
 H3K27me3-1 total\_reads\_47,927,768 50 PE  
 H3K27me3-2 total\_reads\_56,661,732 50 PE  
 H3K4me3-1 total\_reads\_41,070,028 50 PE  
 H3K4me3-2 total\_reads\_40,342,050 50 PE  
 TFDP1-1 total\_reads\_55,314,858 50 PE  
 TFDP1-2 total\_reads\_55,916,780 50 PE

### Antibodies

FLAG Sigma F1802, 1 ug per reaction  
 H3K27ac MAB Institute MABI0309, 1 ug per reaction  
 H3K4me3 MAB Institute MABI0304, 1 ug per reaction  
 H3K27me3 Cell Signaling Technology 9733, 1 ug per reaction

### Peak calling parameters

Peak calling was performed using MACS2. Narrowpeak or broadpeak mode was used for TFDP1 or histone modifications, respectively. The corresponding input library or normal IgG control was used as background signals for the estimation of the significance of peak enrichment.

### Data quality

Quality of ChIP-seq data for histone modifications was assessed by comparison to published datasets in the same cell lines. Quality check for TFDP1 is assessed by enrichment of TFDP1 binding sites in detected peaks.

### Software

Paired end reads were aligned to human (hg38) using Bowtie2. After filtering out duplicated reads, reads on ChIPseq blacklist, coverage (bigwig files), metaplot, and heatmaps were produced using deeptools.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

### Sample preparation

For analyses of cells by ATAC-seq using flow cytometry, we introduced a spike-in control to calibrate technical variability between samples, which enabled precise quantification and high reproducibility. eHAP cells expressing H2B-Kusabira-Orange were used as the spike-in control which could be easily gated away from knockout cells using signals of Kusabira-Orange. Knockout cells were first mixed with the spike-in cells in a tube and subsequently processed for ATAC-seq analyses. The mixed cells were washed with PBS, permeabilized with 0.1% IGEPAL CA-630 in TAPS-DMF buffer containing 1% BSA, 1 mg/ml

polyvinylpyrrolidone (PVP) for 5 min on ice, and centrifuged at 526 x g for 5 min at 1 °C. Then, cells were sequentially washed with the same buffer, and with TAPS-DMF buffer containing 1% BSA and 1 mg/ml PVP. 1x 106 cells were transposed for 30 min at 37 °C using Tn5 assembled with ME19-Cy5 in TAPS-DMF containing 1% BSA. After the reaction, cells were washed twice with PBS containing 1% BSA, 1 mg/ml PVP, and 0.05 % TritonX-100 and stained with DAPI, followed by analyses with flow cytometry (SH800 Cell Sorter, or SH3800 Cell Analyzer, Sony). The mean fluorescent intensity of ATAC-seq signals of the spike-in cells was used to normalize between samples.

Instrument	SONY SH800 or SA3800
Software	SH800 software, SA3800 software, FlowJo 10.1
Cell population abundance	The purity of sorted cells were checked by fluorescent-microscopic observations for ATAC-seq, anti-phospho H3 antibody, and DAPI as shown in Extended Fig. 4a.
Gating strategy	For all analysis, initial gate from forward scatter (FCS-A) vs side-scatter (SSC-A) plot was used to select cells. Single-cell gating from FSC-A vs FSC-H plots was used to exclude doublet and debris. The cells at G2/M cell-cycle was gated based on DAPI signals. For CRISPR screening, the cells were further gated for the G2 fraction by excluding cells at M phase (anti-phospho H3 +). We sorted the populations with the highest and lowest 5% of ATAC-seq signal in the gated G2 cells. Signals for tdTomato were used for gating spike-in cells (H2B-tdTomato+).

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.