**A reusable single cell for epigenomic analysis**

Hidetaka Ohnuki1\*, David Venzon2, Alexei Lovanov3,4, and Giovanna Tosato1

**1** Laboratory of Cellular Oncology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

**2** Biostatistics and Data Management Section, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

**3** CCR Collaborative Bioinformatics Resource (CCBR), Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

**4** Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research sponsored by the National Cancer Institute, Frederick, MD, USA.

\*Correspondence to: hidetaka.ohnuki@nih.gov

**Abstract**

Single-cell methods for epigenomic analysis are changing current understanding of cell populations. Existing experimental approaches cleave genomic DNA and discard the single cells after single use hampering re-analysis of the same single cell to confirm the results and collect data of additional epigenetic marks. Here, we describe a new single-cell method for epigenomic analysis using a “reusable” single cell. The results indicate that the “reusable” single cell is reusable at least 5 times for epigenomic analysis. A “reusable” single cell for epigenomic analysis (RscEpi-seq) supports two types of experiments. One type of experiment allows repeating the same experiment using the same single cell. Repeated experiments reveal signal patterns from epigenetic antibodies that are distinct from the random patterns of controls. Statistical analysis of accumulated single-cell signals permits signal-noise separation and identification of specific epigenetic marks. Another type of experiment permits collecting data of additional epigenetic marks in the same single cell. These features of RscEpi-seq overcome unsurmountable limitations in the current single-cell epigenomic analyses.

(164 words)

**Introduction**

Genomic, epigenomic, transcriptomic and proteomic heterogeneity characterizes cencer cells within a tumor 1-3. Intratumor heterogeneity can endows cancer cells flexibility for developing resistance to treatment through clonal evolution. Although cancer is a genetic disease, genetic mechanisms are often not clear, suggesting the importance of non-genetic mechanisms 4-6. Recent advances in single-cell epigenomic analysis are changing substantially our understanding of intratumor heterogeneity 7-12. These new technologies allow analysis of epigenetic heterogeneity and identify cell subsets within cell populations.

Currently, single-cell epigenomic analysis, presents an unsolved crucial limitation. The nucleosome, the basic unit of chromatin structure and epigenetic signaling module, only possesses one double-stranded DNA per nucleosome and one binding transcription factor (TF). This limits the number of attainable epigenomic signals per nucleosome resulting in digital/binary-like low signal patterns. Digital-like, low signal patterns require data aggregation from other single cells to distinguish signal from noise. This aggregation-based approach results in the identification of common signals within the cell population. However, it hampers identification of epigenetic signals unique to one cell, such as the clinically important cancer founder and treatment-resistant cells.

To solve this issue, we developed a new single-cell method for epigenomic analysis stemming from our development of a “reusable” single cell. We show that “reusable” single cells afford repeating epigenomic characterization of the same cell. Repeated experiments allow application of statistical analysis to distinguish signal from noise, without data aggregation of other single cells, which was not previously possible. In addition, reusable single cells permit analysis of different epigenetic markers, such as H3K27ac, H3K27me3, Med1 and 5hmC in the same single cells. Such broader epigenetic characterization is useful to classify cis-regulatory elements, including promoters and enhancers of a single cell, without depending on commonality with other single cells, identifying previously undetected single-cell distinguishing features.

**Results**

**Method design: A “reusable” single cell for epigenomic analysis (RscEpi-seq)**

The design of the new method was conceived from principles applied in the tissue clearing method CLARITY 13, protein-retention expansion microscopy (ExM) 14, single-cell whole genome amplification method Multiple Annealing and Looping Based Amplification Cycles (MALBAC)15 and proximity ligation assay 16. The new method consists of two sequential steps. The first step (**Fig. 1a**) generates “reusable” single cells. Cellular proteins, including nuclear proteins, are modified with monomer acrylamide using a paraformaldehyde (PFA)/acrylamide mixture; the monomer acrylamide on the proteins is incorporated into a polyacrylamide scaffold by polymerizing the acrylamide (**Fig. 1a**). Individual cells are then embedded in a polyacrylamide gel-bead (**Supplementary Fig. 1a**). The second step acquires locational information of individual antibodies on the genome through a series of biochemical reactions (**Fig. 1b**). Random primers annealed to the genomic DNA are extended with a DNA polymerase to acquire locational information on the genome. A DNA polymerase, which does not have exonuclease activity, is used to protect genomic DNA. Antibodies are conjugated to a DNA probe containing a unique barcode and a ligation sequence. The antibodies are incubated with the “reusable” single cell. After washing the cell, proximity ligation is performed between the antibody-DNA-probe-conjugate and the random primers using a ligation adapter with a T4 DNA ligase. The ligated products are amplified by MALBAC with a second set of random primers containing a cell-specific barcode. The antibody-DNA-probe-conjugate also contains a T7 promoter sequence. In vitro transcription, RNA purification and reverse transcription are performed to reduce MALBAC byproducts, i.e. genome sequences lacking the antibody-DNA probe (**Supplementary Fig. 2**).

**A single cell anchored to polyacrylamide scaffold is reusable in RscEpi-seq**

We first evaluated whether “reusable” single cells can retain cellular proteins over repeated heating and cooling cycles, which are used for acquiring locational information of antibodies (**Fig. 1b**). The results (**Supplementary Fig. 1b**) indicated that cells treated with 28% acrylamide+4% PFA before embedding into the polyacrylamide scaffold preserve the highest amount of cellular protein over 100 annealing cycles compared to other solutions.

Next, we tested whether RscEpi-seq can generate the desired products. We performed the series of reactions (**Fig. 1b**) using 8 single cells anchored to the polyacrylamide scaffold, a single cell embedded into the gel without protein anchoring to the gel and blank gel (no cell). Antibodies (anti-H3K27ac and anti-H3K27me3) and control IgG were used in the experiment. Final DNA products are shown in **Supplementary Figure 3a**.Single cells anchored to polyacrylamide generated more DNA products than either blank gel (no cell) or a single cell without anchoring, suggesting that anchoring cellular proteins and genomic DNA to polyacrylamide increases output of desired products. Shallow sequencing using MiSeq (**Supplementary Fig. 3b**) indicated that a single cell anchored to polyacrylamide generates a higher number of desired DNA products containing antibody barcode, ligated sequence and genomic sequences than single cells without anchoring.

Cellular proteins were retained over repeated experiments (**Supplementary Fig. 1b**). Next, we evaluated whether RscEpi-seq maintains the position of nuclear proteins and genomic DNA over repeated experiments. We repeated the same 3 experiments using the same 8 single cells with anti-H3K27ac, anti-H3K27me3 and control IgG. Amplification duplicates were computationally removed based on combinations of antibody barcode, 8 nt random sequence of the 1st random primer and mapped position of the 5’ end on the genome. After removal of the amplification duplicates, read distribution was analyzed in the first, second and third experiments (**Fig. 2a** and **2b**). We identified the signal distributions in the genome detected in the first experiment (green). Subsequently, we identified the signal distribution detected in the second (blue) and the third (red) experiment in the same genomic regions detected in the first experiment. Read distribution patterns in the second and third experiment showed similar patterns to those found in the first experiment. These results indicated that the positions of H3K27ac and H3K27me3 are closely maintained over the experiments. These results support the conclusion that the polyacrylamide-anchored single cells are reusable at least 3 times.

**Putative signals are identified in a cell by repeated experiments without relying on other cells**

We further analyzed data from the 3 repeated experiments using the same reusable single cells to answer the question of whether signals from specific epigenetic antibodies differ from those of control IgG and stochastic background. To this end, we applied the bootstrap statistical test (see Online Methods) to each 500 bp genomic region from each of the triplicate experiments with each single cell, to identify genomic regions where antibody-derived signal are enriched compared to control IgG and stochastic noise. The test results identified, on average, 125,502 regions containing H3K27ac signals (**Fig. 2c**) and 136,387 regions containing H3K27me3 signals (**Fig. 2e**) that were statistically significantly different from controls (p < 0.01). These signal-enriched regions contained, on average, 556,432 H3K27ac signals per cell (**Fig. 2d**) and 375,756 H3K27me3 signals per cell (**Fig. 2f**). Single-cell #3 was an outliner for H3K27me3 (**Fig. 2e** and **2f)**. However, single-cell #3 showed an inverse trend for H3K27ac (**Fig. 2c** and **2e)**, suggesting that results from single-cell #3 are not reflective of experimental error. Signals in the enriched 500 bp regions, on average, amounted to 4.91 H3K27ac signals and 3.81 H3K27me3 signals. Since a nucleosome includes 146 bp of wrapping DNA and the length of linker DNA is up to 80 bp, a 500 bp genome region may contain 2 to 3 nucleosomes. Since a nucleosome contains 2 copies of histone H3, 4 to 6 copies of histone H3 may be present per 500 bp of genome. Therefore, the average frequencies of 4.91 H3K27ac and 3.81 H3K27me3 signals per 500 bp region, detected here likely reflect high-density signal regions. These results also indicate that repeated experiments using the same single cells accumulate signals in specific regions compared to controls. In a single experiment, one nucleosome of a cell can generate only a few signals because one nucleosome has only one dsDNA. In repeated experiments, the results revealed signal patters from epigenetic antibodies that are distinct from the random pattern of controls. Statistically significant (p < 0.01 by the bootstrap test) signals were considered putative signals in the subsequent analyses.

**Evaluation of putative signals at promoters and enhancers**

Since histone H3K27ac is enriched at promoter and enhancer regions 17, we evaluated the location of putative single-cell H3K27ac signals in promoters and enhancers by comparing with bulk ChIP-seq results (**Fig. 3a-3d**). For this analysis, we used the Human ACtive Enhancers to interpret Regulatory variants (HACER) atlas 18, which defines cell-specific, active enhancers based on bidirectional enhancer RNA expression in 265 human cell types. HACER also contains data experimentally validated, active enhancers-promoter interaction. In HACER, enhancers and promoters were classified into 2 groups, K562-cell-type specific, active enhancer/promoter (referred here as K562-active enhancer/promoter) and K562-cell-type non-specific, active enhancers/promoter (referred here as other active enhancer/promoterl) enhancer/promoter. RscEpi-seq detected 9,974 promoters with significant putative signal enrichment (*p* < 0.05), 62.93% of which were confirmed by bulk ChIP-seq (**Fig. 3a**). Among K562-active promoter (**Fig. 3b**), RscEpi-seq detected 4,212 promoters, 77.28% of which were confirmed by bulk ChIP-seq. Among other active enhancers (**Fig. 3c**), RscEpi-seq detected 78,874 enhancers with significant signal enrichment (p < 0.05), 25.77% of which were confirmed by bulk ChIP-seq. Among K562-active enhancers detected by RscEpi-seq (**Fig. 3d**), 73.89% were confirmed by bulk ChIP-seq. The 73.89% confirmation rate of K562-active enhancers is higher than the 25.77% confirmation rate of other active enhancers. These results indicate that putative signals identified by RscEpi-seq and the bootstrap statistical test detect established epigenetic signals of K562 cells, which can recognize cell-specific, active enhancers.

**Enhancer classification using four epigenetic marks**

The presence of H3K27ac uniquely marks active enhancers that are linked to active genes 17,19. Poised or inactive enhancers that are linked to inactive genes are distinguished by the absence of H3K27ac and enrichment of H3K27me3 19. Ratios of H3K27ac/H3K27me3, but not simply by presence or absence of H3K27ac or H3K27me3, explain changes in gene expression during hematopoietic lineage differentiation 20,21. Here, we used relative ratios Log2(H3K27ac/H3K27me3) (shown in the X-axis, **Fig. 3e-3h**) to classify the enhancers. Results are expressed as average number of H3K27ac (Y-axis, **Fig. 3e**) and H3K27me3 (Y-axis, **Fig. 3f**) putative signals per kb/enhancer. Consistent with previous reports 17,19, enhancers with Log2(H3K27ac/H3K27me3) values above 6, were exclusively H3K27ac-positive and H3K27me3-negative. In the transitional -5 to +5 range, H3K27ac signals were dominant in the 0 to +5 range and H3K27me3 signals were dominant in the 0 to -5 range. In the less than -5 range, H3K27me3 was exclusively positive in enhancers. These results indicated that signals from RscEpi-seq could classify enhancers into at least 3 groups; poised enhancers (exclusively H3K27me3-positive), intermediate enhancers (transitional) and active enhancers (exclusively H3K27ac-positive).

Super-enhancers are clusters of highly active enhancers marked by high level H3K27ac, high density master transcription factors and mediator complexes 22. Therefore, if the enhancer classification by RscEpi-seq (**Fig. 3e** and **3f**) is correct, mediator complex subunit1 (Med1) should be enriched in the active enhancers. To examine this, we re-tested the same 8 single cells with anti-Med1 antibody and control IgG (4th experiment). Med1 putative signals were identified by bootstrap test statistics using aggregated datasets from 8 single cells (p < 0.01, **Supplementary Fig. 4a** and **4b**). This aggregated data identified putative common Med1 signals among the 8 single cells based on signal-enriched regions over control IgG and stochastic noise. The results from counting Med1 signals in the enhancers classified by RscEpi-seq (**Fig. 3g**) show that Med1 signals were enriched in the active enhancers, especially in highly active enhancers. These results indicate that RscEpi-seq data can correctly identify the epigenetic status of enhancers as defined by the relative ratio of Log2(H3K27ac/H3K27me3).

We evaluated further the enhancer classification using the additional epigenetic mark, 5-hydroxymethylcytosine (5hmC). Stroud et al. 23 reported that 5hmC is significantly enriched in enhancers with H3K27ac. In addition, 5hmC is generated by the enzyme Ten-eleven translocation 1 (TET1) 24,25 that interacts with Med1 super-enhancer complex 26. If the enhancer classification by RscEpi-seq was correct, 5hmC should be enriched in active enhancers. For this evaluation, we re-tested the same 8 single cells with anti-5hmC and control IgG (5th experiment). Putative signals of 5hmC were identified based on statistical significance (bootstrap test; p < 0.01) using aggregated datasets from 8 single cells (**Supplementary Fig. 4c** and **4d**). Next, we calculated the average number of putative 5hmC signals in each classified enhancer range (**Fig. 3h**). Consistent with previous reports23, 26, the results indicated that the 5hmC putative signals were enriched in the active enhancers. These results indicate that the H3K27ac and H3K27me3 putative signals from RscEpi-seq can correctly recognize the epigenetic status of enhancers. These results also indicate that the reusable single cells are reusable at least 5 times in RscEpi-seq. Furthermore, these results indicate that RscEpi-seq can analyze at least 4 epigenetic marks in the same single cells.

**Signal evaluation in enhancers based on gene targets**

The results of RscEpi-seq testing support that H3K27ac and H3K27me3, combined, can correctly identify active enhancers in K562 single cells. We now assessed the function of these active enhancers by focusing on interacting genes. The HACER dataset 18 contains experimentally validated interactions between enhancers and genes from 4DGenome 27 and chromatin interaction studies from Hi-C, ChIA-PET, HiChIP and Capture HiC. K562-active enhancers detected by RscEpi-seq (2,363 enhancers) and by bulk ChIP-seq (3,938 enhancers) interact with 3,154 and 5,002 genes, respectively (**Fig. 4a**). Interacting genes identified by RscEpi-seq were mostly (93.18%) common with identified genes by bulk ChIP-seq. Among other active enhancers (**Fig. 4b**), overlap between RscEpi-seq and bulk ChIP-seq (78.44%) was smaller than the K562-active enhancers (93.18%). Since enhancer-promoter/gene interactions transmit enhancer activity to target genes 28, these results suggest that signals from RscEpi-seq can correctly capture sets of enhancers for functions of K562 single cells.

**Signal evaluation in enhancers and promoters based on canonical signaling pathways**

We further evaluated signals from RscEpi-seq by pathway enrichment analysis 29 through analysis of genes proximal to active enhancers and promoters detected by RscEpi-seq and bulk ChIP-seq. Scince enhancer-promoter/gene interactions transmit enhancer activity to target genes28, we mimicked these transmission in silico. We calculated the combined relative scores {Log2(H3K27ac/H3K27me3)} of enhancers and promoters and transmitted this score to the proximal gene (**Fig. 4c**). The transmitted scores to genes were used in the pathway enrichment analysis. By comparing the 5,000 genes with the highest scores (putatively active genes) identified by RscEpi-seq and bulk ChIP-seq, we found that 88.21% of the enriched canonical pathways (referred here as pathways) from RscEpi-seq were common to the enriched pathways from bulk ChIP-seq (**Fig. 4d**). Activation *z*-scores of the enriched pathways from RscEpi-seq and bulk ChIP-seq showed a strong correlation (Pearson correlation coefficient, *r*=0.875). We further evaluated the results of the enriched pathways from RscEpi-seq with the results of enriched pathways from RNA-seq of bulk K562 cells. The RNA-seq data from the embryonic stem cell line H1 were used as a reference control. The top 5,000 transcripts from bulk K562 cells and bulk H1 cells were used in pathway enrichment analysis (**Fig. 4e**). Most (89.15%) of the enriched pathways from RscEpi-seq were confirmed as enriched pathways from bulk RNA-seq of K562 cells (**Fig. 4f**). Activation *z*-scores of the enriched pathways from RNA-seq and RscEpi-seq showed a strong correlation (Pearson correlation coefficient, *r* = 0.864). However, only 17.92% of the enriched pathways from RscEpi-seq overlapped with the enriched pathways of bulk H1 RNA-seq, and the overlapped pathways were poorly correlated (Pearson correlation coefficient =0.365). These results suggest that signals from RscEpi-seq are useful to infer downstream functions of epigenetically active enhancers.

**Cell identity identified by RscEpi-seq**

The results in **Figure 4g** suggested that signals from RscEpi-seq can define cell identity. To confirm this possibility, we compared different bulk cell types to K562 single cells based on epigenetic status [Log2(H3K27ac/H3K27me3)] of enhancers and promoters. The results are visualized by t-stochastic neighbor embedding (t-SNE, **Fig. 5a**). K562 single cells (red dots) and K562 bulk cells (blue dots) differed from endothelial cells (yellow), fibroblasts (green) and myoblasts (light blue). K562 single and bulk cells bordered hematopoietic malignant cell lines and primary hematopoietic cells. K562 cell line is a myelogenous leukemia cell line 30 with multi-differentiation potential along erythroid, macrophage and megakaryocytic lineages 31-33. Counter plots of K562 single cells (red lines) and K562 bulk cells (blue lines) indicated epigenetic similarities between datasets from K562 single and bulk cells. These data support the conclusion that signals from RscEpi-seq can recognize cell identity.

**Active enhancers usage and Med1 association among single cells**

In **Figure 3c** and **3d**, the overlap between genes from RscEpi-seq and bulk ChIP-seq was lower for other active enhancers (**Fig. 3c**) than for K562-acive enhancers (**Fig. 3d**), suggesting high epigenetic heterogeneity in other active enhancers. When visualized by t-SNE, the 8 K562 single cells localized at somewhat different positions (**Fig.** **5a**), suggesting epigenetic heterogeneity among these single cells.We examined the basis of this heterogeneity identified by RscEpi-seq.K562 cells have features shared by common myeloid progenitors (CMPs), megakaryocyte-erythroid progenitor (MEPs) and/or granulocyte-monocyte progenitor (GMPs) and display differentiation potential along the erythroid, myeloid and megakaryocytic lineages 34, 35, 36, suggesting a basis for epigenetic heterogeneity. CD34 is expressed in CMP and GMP, but not in MEP 37. Stem-like chronic myeloid leukemia cells are marked by CD33 and CD47 38. The epigenetic status of the CD34 locus, the CD33 locus and the CD47 locus in the 8 K562 single cells is shown in **Figure 5b-5d**. K562-active enhancers (green) and other active enhancers (red) were detected uniquely in some of the single K562 cells by Rsc-Epi-seq. In addition, Med1 and 5hmC marked active enhancers and other regions in most cells, suggesting the occurrence of interaction between some of the active enhancers and mediator complex. These results indicate that the CD34, CD33 and CD47 genes could be epigenetically active in selected K562 single cells.

Megakaryocytic differentiation is induced in K562 cells by the homeobox transcription factor DLX4 whereas erythroid differentiation is repressed. Other active enhancers were detected upstream of the DLX4 locus, some of which were marked with Med1 (**Fig. 5e**), suggesting an epigenetic bias toward megakaryocytic differentiation in some of the K562 cells.

K562 cells have been used as a Human Leukocyte Antigen (HLA)-deficient cell line to measure cytotoxicity of natural killer cells 39. Expression of HLA genes can be induced in K562 cells by sodium butyrate, which inhibits histone deacetylation and increases histone acetylation 40. Our analysis of HLA-A, HLA-B and HLA-C loci (**Supplementary Fig. 5a-5c**) is consistent with previous reports 39,40 in showing that these regions are mostly epigenetically inactive, exept for upstream regions of HLA-B in single-cells #5 and #6. Overall, the results in **Figure 5b-5e** and **Supplementary Fig. 5a-5c** indicate that RscEpi-seq can capture two types of epigetic heterogeneity. One type is cell-to-cell heterogeneity. The other is epigenetic heterogeneity in the usage of cell-type specific and other cell-type specific enhancers.

**Discussion**

Here, we provide evidence that the reusable single cell is amenable to re-analysis at least 5 times. Reusable single cells can preserve nuclear proteins and genomic DNA location over repeated experiments. The reusability of single cells in epigenomic analysis opens unprecedented new opportunities for epigenomic analysis.

One such opportunity is that the same experiment can be repeated with the same single cells using the same antibodies and control IgG. Repeated experiments indicated that noise is relatively random, whereas antibody signals accumulate at specific positions on the genome. This affords the possibility to distinguish signal from noise at the single-cell level. Current single-cell epigenomic analyses 7-12 recognize signals from noise based on enrichment/overlap of signals among other single cells. Therefore, identified signals are biased toward common signals among cells in a cell population. Data from repeated experiments permits identification of signals in a cell without relying on other single cells. This is important, for example, when analyzing stem cells and clonal evolution in cancer. During cancer evolution 41, a few cancer cells 42 can be essential effectors of cancer relapse, metastasis and the emergence of treatment-resistant tumors. Therefore, identifying epigenetic signals in single cancer cells without depending on other single cells could be a fundamental step for successful cancer treatment.

The second opportunity afforded by reusable single cells is that epigenetic characterization can be expanded by re-analyzing the same single cells with different antibodies for distinct epigenetic marks. At least 4 epigenetic marks could be defined in the same single cell. The epigenome is a multi-layered apparatus that includes histone modifications, DNA modifications, transcription factors and other genome associated proteins 43. Analyzing different epigenetic marks in the same single cells would be helpful to decode epigenetic mechanisms at the single-cell level.

**Online Methods**

**Cell culture**

K562 cells from the American Type Culture Collection (ATCC, CCL-243) were cultured in Iscove's Modified Dulbecco's Medium (12440053, Thermo Fisher Scientific) containing 10% fetal bovine serum (F2442-500ML, Sigma-Aldrich). Cells tested mycoplasma-negative by a PCR-based method 44.

**Antibody conjugation to DNA probe**

Anti-H3K27ac (ENCODE ID: ENCAB000ADS, 39133, Active Motif), anti-H3K27me3 (ENCODE ID: ENCAB000ADT, 39155, Active Motif), anti-Med1 (ab60950, Abcam), anti-5hmC (A-1018, Epigentek) antibodies and control IgG (I5006, Sigma-Aldrich) were conjugated with the DNA probes (Integrated DNA technologies) listed in **Table S1**. Glycerol and sodium azide were removed from all antibodies using Zeba Spin Desalting Columns according to the manufacturer's protocol (89682, Thermo Fisher Scientific).

The antibodies were activated by introducing the heterobifunctional crosslinker Succinimidyl-6-hydrazino-nicotinamide (S-HyNic, S-1002-105, TriLink Biotechnologies) according to the following procedures: S-HyNic (1 mg) was dissolved with 100 l of anhydrous Dimethylformamide (DMF, S-4001-005, TriLink Biotechnologies), and 0.6 l of S-HyNic/DMF was added to 100 l of antibody solution (1 mg/ml). The antibody was incubated for 2 hours at room temperature. Extra S-HyNic was removed using the Zeba Spin Desalting Column.

To conjugate the DNA probe to the activated antibody, the heterobifunctional crosslinker N-succinimidyl-4-formyl benzamide (S-4FB, S-1004-105, TriLink Biotechnologies) was introduced into the DNA probe according to the procedure below. The DNA probes listed in **Table S1** were synthesized with an amino group at 5’ end with a C12 spacer (Integrated DNA Technologies). Twenty nanomoles of the DNA probe was dissolved into 20 l of 100 mM sodium phosphate/150 mM NaCl, pH8.0. S-4FB (1 mg) was dissolved into 50 l of anhydrous DMF. Ten l of S-4FB/DMF was added to the DNA probe and incubated for 2 hours at room temperature. Extra S-4FB was removed using the Zeba Spin Desalting Column.

The activated antibody and the DNA probe were mixed and incubated overnight at room temperature. Unbound DNA probe was removed by Ultrafiltration using Amicon Ultra (Molecular Cut Off 100 kDa, UFC5010, Millipore), and the buffer was exchanged to TBS containing 50% glycerol, 5 mM EDTA and 5% BSA. Antibody concentration is measured by Sandwich ELISA 45 using a standard curve of rabbit IgG, and antibody concentration was adjusted to 1 mg/ml. The antibody conjugated to the DNA probe was stored at -20°C until use.

**Preparation of reusable single cells**

K562 cells were harvested by centrifuging at 1,200 x g for 5 min and washed with PBS. The cells were treated with one of the following solutions: PBS only, PBS containing 4% paraformaldehyde (PFA, 15713, Electron Microscopy Sciences), PBS containing 5% acrylamide and 4% PFA, PBS containing 20% acrylamide and 4% PFA, or PBS containing 28% acrylamide and 4% PFA. After 1-hour incubation, the cells were washed with TBS containing 2.5% BSA. Individual cells were transferred into a 0.2 ml tube containing 3 l/tube of the acrylamide solution [PBS containing 3.88% (w/v) Acrylamide (01697-500ML, Sigma-Aldrich), 0.12% (w/v) Bis-acrylamide (M1533-25ML, Sigma-Aldrich) and 1% (w/v) ammonium persulfate (A3678-25G, Sigma-Aldrich)] using ALS CellCelector (ALS Automated Lab Solutions) or PicoPipet with a micromanipulator (Nepa Gene) under microscopy. Fifty l of mineral oil containing 0.2% TEMED was added to the 0.2 ml tube. Polymerization was performed at room temperature for 1 hour. A cell was embedded into the outside layer of the two-layered acrylamide bead. The cell was stained with SYBR Gold (S11494, Thermo Fisher Scientific), and the presence of a single cell was confirmed using fluorescent microscopy (BZ-X710, Keyence).

**Cellular protein retention assay**

K562 cells were harvested and washed in PBS. Cells were suspended in PBS and 0.33 x105 cells were aliquoted into 0.5 ml Protein LoBind tubes (0030108094, Eppendorf). The tubes were centrifuged at 1,200 x g for 5 min. As input controls, 1x105 cells were lysed with 100 l/tube of 1 x lithium dodecyl sulfate buffer (LDS buffer, NP0007, Thermo Fisher) and stored at 4 °C until measurement. The cells were treated with PBS only, PBS containing 4% PFA, PBS containing 5% acrylamide and 4% PFA, PBS containing 20% acrylamide and 4% PFA, or PBS containing 28% acrylamide and 4% PFA. After 1-hour incubation, cells were washed with TBS containing 10% fetal bovine serum. Cells were suspended with 90 l of the acrylamide solution and the cell suspension was transferred into 9 PCR tubes containing 50 l/tube mineral oil containing 0.2% TEMED. After 1-hour incubation, polymerized beads (3 beads, approximately 1 x 105 cells) were transferred into a PCR tube. ThermoPol Mg(-) buffer (100 l/tube) was added and the following treatments were performed. One group was stored at 4 °C. Other groups were placed in a thermal cycler, heated at 94°C for 3 min and cooled at 4°C for 5min (number of cycles 1, 2, 5, 10 and 100). The supernatant was collected from each condition. Protein concentration was measured by Micro BCA Protein Assay kit according to the manufacturer’s protocol (23235, Thermo Fisher). Solubilized cellular proteins from non-treated cells in LDS buffer were used as 100% in calculating eluted cellular proteins from the polyacrylamide beads.

**Agarose gel electrophoresis**

Agarose gel electrophoresis was performed through 2% agarose gel, E-Gel (G521802, Thermo Fisher Scientific) and DNA products are visualized by the imager LAS 4000 (FujiFilm).

**Acquisition of locational information of individual antibodies on the genome**

Cell membranes and nuclear membranes were permeabilized using the following protocol. Single cells were treated with 1 x TBS containing 1% Triton X-100, 1 mM EDTA and 10% glycerol (TBS-Triton) for 15 min. The cells were then serially treated with 25%, 50%, 75% and 100% methanol for 5 min each, then treated with 75%, 50% and 25% methanol. The single cell was washed with the TBS-Triton.

In the “Annealing” step (**Fig. 1b**), the single cell was washed with “ThermoPol Mg(-) buffer” (20 mM Tri-HCl, pH8.8, 10 mM (NH4)2SO4, 10 mM KCl and 0.1% Triton X-100) and the supernatant was removed. ThermoPol Mg (-) buffer (15 l/tube) containing 17.5 M of the 1st Random Primer (5’-/5Phos/CGACGCTNNNNNNNN-3’, Integrated DNA Technologies) was added to the single cell and incubated for 1 hour to deliver the random primer to the nucleus. The tube was incubated at 94 °C for 3 min and incubated on ice for 2 min or longer.

In the “Extension” step (**Fig. 1b**), a solution containing MgSO4 (final concentration 2 mM), NaCl (final concentration 300 mM) and dNTPs (final concentration 1.47 mM) was added followed by addition of the DNA polymerase DeepVent Exo(-) (1 unit, M0259L, New England BioLabs) and incubation for 4 hours on ice, 2 hours at 4°C, 2 hours at 10 °C, 2 hours at 20°C and 4 hours at 25°C.

In the “Antibody binding” step (**Fig. 1b**), the cell was washed and incubated with the ThermoPol Mg(-) buffer containing 300 mM NaCl, 10 mM EDTA and 1% BSA and 10% glycerol for 1 hour. Antibodies conjugated with the DNA probe were added (0.5 g IgG/ml each) to the cell and incubated overnight on ice. The cell was washed with the ThermoPol Mg (-) buffer containing 300 mM NaCl.

In the “Proximity joining” and “Proximity ligation” steps (**Fig. 1b**), the cell was incubated with ThermoPol Mg (-) buffer containing 300 mM NaCl and 0.5 M Ligation Adapter Probe (5’-AGCGTCGTGTAGGGAA-3’, Integrated DNA Technologies) for 1 hour at 25°C. The cell was washed once with 1x Quick Ligation Reaction buffer (M0202L, New England BioLabs). Quick Ligase (1 l) and 1 x Quick Ligation Reaction buffer (19 l) were added and incubated for 4 hours at 15 °C and 30 min at 25°C.

The ligated product of Antibody-DNA probe + 1st Random primer was fully extended using the polymerase Bst 3.0 (M0374L, New England BioLabs). The cell was washed with ThermoPol Mg (-) buffer containing 300 mM NaCl. The polymerase solution was then added (320 unit/ml Bst 3.0 polymerase, 1.4 mM dNTPs and 2.5 mM MgSO4 in ThermoPol buffer). The cell was incubated for 4 hours at 4°C, 1 hour at 10°C, 1 hour at 20°C, 1 hour at 30°C, 1 hour at 40°C, 1 hour at 50°C and 1 hour at 65°C.

The extended products were amplified by a modified Multiple Annealing and Looping Based Amplification Cycles (MALBAC) 15,46. Two M of the 2nd Random primer (**Table S1**) was added and incubated for 2 hours. The cell and the extended products were heated at 94°C for 3 min and incubated on ice for 2 min. Bst 3.0 polymerase (1 l/tube) was added and incubated for 4 hours at 4°C, 30 min at 10°C, 30 min at 20°C, 30 min at 30°C, 30 min at 40°C, 30 min at 50°C, 60 min at 65°C and 3 min at 94°C. The supernatant was collected, and 20 l of buffer containing 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA and 0.05% Tween 20 was added. After overnight incubation at 4°C, the supernatant was collected; supernatant collection was repeated for a total of 3 times. The collected supernatant was combined into a 0.2 ml PCR tube. Four l of 10 x ThermoPol buffer (New England BioLabs) and dNTPs (total 1.4 mM) were added, and then the combined solution was heated at 94°C for 3 min and placed on ice for 2 min. Bst large fragment (2 ul/tube, M0275L, New England BioLabs) was added and incubated at 10°C for 45secs, 20°C for 45secs, 30°C for 45secs, 40°C for 45secs, 50°C for 45secs, 65°C for 2mins and 94°C for 20s. The tubes were then quickly quenched on ice. After quenching on ice, Bst large fragment (2 ul/tube) was added and then incubated at 10°C for 45secs, 20°C for 45secs, 30°C for 45secs, 40°C for 45secs, 50°C for 45secs, 65°C for 2mins and 94°C for 20s. The tubes were then quickly quenched on ice. The above cycles were repeated for a total of 8 times.

The products were further amplified by PCR using a primer (5'-ATCCATAGTGTCAGCAGGCT-3’) with DeepVent exo(-) polymerase (step1: 95°C-5 min, step2: 95°C-30 sec, step3: 60°C-30 sec, step4: 72°C-30 sec, step5: repeat step2-4 20 times, step6: 72°C-5 min and step 7: 4°C-forever). DNA was purified using UltraPure Phenol:Chloroform:Isoamyl Alcohol according to the manufacturer’s protocol (15593031, Thermo Fisher Scientific). Extra primers were removed by size-selective precipitation 47 using polyethylene glycol; 200 g/tube linear acrylamide (AM9520, Thermo Fisher Scientific) and 2 M MgCl2 (final concentration 20 mM) were added, and then 50%(w/v) PEG8000 was added (final concentration 14%). After 20 min incubation, the tube was centrifuged at 1,3400 x g for 10 min. The pellet was washed with 80% ethanol 3 times and dried. After dissolving the pellet with 0.1x TE buffer, the amount of double-stranded DNA (dsDNA) was measured by Quant-iT PicoGreen dsDNA Assay kit (P11496, Thermo Fisher Scientific) and concentration of dsDNA was adjusted for *in vitro* transcription. The reusable single cell was stored at -20°C in TBS buffer containing 50% glycerol, 0.1% Triton X-100, 0.5% BSA and 1 mM EDTA) until the next round of experiments.

MALBAC, developed to amplify genomic DNA from single cells15,46, can amplify all types of DNA. In our system, MALBAC could amplify desired products (antibody-DNA probe + 1st random primer + genome sequence + 2nd random primer) and genome-derived byproducts. Genome-derived byproducts were removed by the following steps. The antibody-DNA probe contains a T7 promoter sequence (**Table 1** and **Supplementary Fig. 2**). We converted the ligated products into RNA by *in vitro* transcription using HiScribe T7 High Yield RNA Synthesis Kit according to the manufacturer’s protocol (E2040S, New England BioLabs). DNA was digested using RNase-free DNase I according to the manufacturer’s protocol (EN0521, Thermo Fisher Scientific). RNA was purified using TRIzol LS Reagent according to the manufacturer’s protocol (10296028, Thermo Fisher Scientific). The purified RNA was converted into DNA using SuperScript IV Reverse Transcriptase according to the manufacturer’s protocol (18090010, Thermo Fisher Scientific) with a primer (5’-ATCCATAGTGTCAGCAGGCT-3’, RNase-free HPLC purified oligo, Integrated DNA Technologies) specific for the sequence in the 2nd random primer (see “Reverse transcription” part in **Supplementary Fig. 2d**). Second strand synthesis was performed using the DNA polymerase, DeepVent Exo (-) (1 unit/50 l) with a 3.68 M primer (5’-TAGCTAAGGTATCCTCCAGG-3’), 200 M dNTPs and 1 x ThermoPol Reaction buffer. Extra sequences in the DNA products were removed by restriction-enzyme digestion with BciVI according to the manufacturer’s protocol (R0596L, New England BioLabs, see “Digestion with a restriction enzyme” in **Supplementary Fig. 2f**). Fragments larger than 49 bp were selected using E-gel 2% agarose (Thermo Fisher Scientific) and were extracted using Freeze ‘N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad). The extracted DNA was purified using UltraPure Phenol:Chloroform:Isoamy Alcohol according to the manufacturer’s protocol (Thermo Fisher Scientific). The purified DNA was used for library construction using the Illumina TruSeq PCR free kit (Illumina) with TruSeq DNA index kit plate (96 samples, Illumina). Individual samples were labeled with unique Illumina indexes. The constructed libraries were sequenced by MiSeq, NextSeq 550 or NovaSeq 6000.

**Preparation of reads before mapping to the genome**

Illumina adapter sequences were trimmed using the trimming software TrimGalore (version 0.4.5, http://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/) with options [--paired --illumina -e 0.2 --length]. Trimmed reads were demultiplexed using antibody barcodes shown in Supplementary Table 2 by software FlexBar (version 3.0.3) 48. Reads containing reverse a complement sequence of antibody barcodes were inverted using FastX tool kit (version 0.0.14, <http://hannonlab.cshl.edu/fastx_toolkit/index.html> ). Left side sequence of reads from 5’ end to an antibody barcode-ligated sequence was trimmed by FlexBar. Right side sequence of reads from 3’ end to a cell barcode was also trimmed by FlexBar. After the trimming, An 8 nucleotide sequence of the 1st random primer was removed and add to read ID as a unique molecular identifier by software UMI tools (version 1.0.1)49. All used codes here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/tree/master/001_Data_Prep>. Outline of the Shell script are also shown as drawing on page 2 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/blob/master/000_Workflow_page1-11.pdf>.

**Mapping to the genome and removing amplification duplicates**

The trimmed reads are mapped to the human genome, GRCh38 using software Bowtie2 (version 2.3.5.1) 50. Amplification duplicates were removed by UMI tools based on the unique molecular identifier derived from 8 nt random sequences and mapped position on the genome. The unique mapped reads were used for the subsequent data analysis. All used codes here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/001_Data_Prep> . Outline of the Shell script are explained are also shown as drawing on page 2 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/blob/master/000_Workflow_page1-11.pdf> .

**Analysis of locational preservation in H3K27ac and H3K27me3 marks over the repeated experiments**

The unique mapped reads of 8 single cells were combined. The combined datasets of the first experiment were converted into a BED file and used as a peak file in the next step. Signal distributions in genomic regions where the signal was detected in the first experiment were analyzed by Homer (version 4.10.4) 51. The signal distribution pattern was calculated by the Homer in the first, second and third experiments. The output data were visualized as a line graph by Microsoft EXCEL. All used codes here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/002_Figure2a-2h>. Outline of the Shell script are also shown as drawing on page 3 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf>

**Identification of putative signals without data aggregation of other single cells**

Datasets of the first, second and third experiments were combined in reads derived from antibody or control IgG, separately. Datasets of individual single cells remain as separate files. Randomized controls were generated from the combined antibody reads or control IgG using software BED tools (version 2.29.0) 52. The randomized controls of each single cell were used to estimate stochastic background levels in the following bootstrap statistical test. Genomic regions were subdivided into every 500 bp bin and the bin is shifted every 250 bp. Signals in the 500 bp bins were counted by BED tools with an option “intersect”. To estimate signal enrichment over control IgG, counts of control-IgG reads were subtracted from counts of antibody reads in each bin (*AbIgG*). To estimate accidental enrichment, counts of randomized control-IgG reads were subtracted from counts of randomized antibody reads (*Random AbRandom IgG*). Upperside of 99% confidence interval in the *Random AbRandom IgG* were calculated by bootstrap statistical test using R package “boot” (version 1.3-23). Regions, which have greater values in *AbIgG* than the upper side of the 99% confidence interval of *Random AbRandom IgG* were considered as significantly signal-enriched regions compare to the background. The number of significantly signal-enriched regions were counted and shown in Figure 2c and 2d. Signals in the significantly signal-enriched regions were counted and shown in Figures 2d and 2f considered as putative signals for further evaluation in the subsequent analysis. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/003_Fiugre2c-2f>. Outline of the Shell script in combination with R script is also shown on page 4 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Identification of putative signals with data aggregation of other single cells**

For Med1 and 5hmC, the conventional approach was used to identify putative signals based on the commonality of signals among cells using data aggregation of other single cells. The conventional approach is useful to increase the number of epigenetic marks in the same single cell and to save the time to repeat the experiments. The unique reads of 8 single cells were combined into one BAM file using SAM tools (version 1.9) 53. Each randomized control was generated from reads of antibody or control IgG by BED tools. Number of reads in each bin was counted by BED tools with an option “intersect”. To estimate signal enrichment over control IgG, counts of control-IgG reads were subtracted from counts of antibody reads in each bin (*AbIgG*). To estimate accidental enrichment, counts of randomized control-IgG reads were subtracted from counts of randomized antibody reads (*Random AbRandom IgG*). Upperside of 99% confidence interval in the *Random AbRandom IgG* were calculated by bootstrap statistical test using R package “boot”. Regions, which have greater values in *AbIgG* than the upper side of 99% confidence interval of *Random AbRandom IgG* were considered as significantly signal-enriched regions compare to the background. The number of significantly signal-enriched regions were counted and shown in Figure 2c and 2d. Signals in the significantly signal-enriched regions were counted and shown in Figures 2d and 2f considered as putative signals for further evaluation in the subsequent analysis. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/004_Supplementary_Figure4> . Outline of the Shell script is also shown on page 5 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Signal-enriched promoters and enhancers**

Putative H3K27ac signals identified by the bootstrap test were used to determine signal-enriched promoters and enhancers. A list of human promoters was downloaded from Eukaryotic Promoter Database 54 of the Swiss Institute of Bioinformatics as a [BED file](https://epd.vital-it.ch/wwwtmp/human_epdnew_ouMA3.bed) (version 006, assembly: hg38). A list of human enhancers was downloaded from the Human ACtive Enhancers to interpret Regulatory variants (HACER) atlas 18. To compare the results of RscEpi-seq with standard data of bulk cell analysis, bulk ChIP-seq dataset of K562 cells was used (from ENCODE, Data set ID: ENCFF301TVL). To estimate the frequency of accidental detection, the putative H3K27ac signals in RscEpi-seq and bulk ChIP-seq were randomized using BED tools, and the generated data were used as controls. Numbers of putative H3K27ac signals and random control in promoters and enhancers ware counted by BED tools with an option “intersect”. Signal counts were normalized based on the length of promoters or enhancers (per kilobase) and the number of input signals (per million input). Bootstrap statistical test was performed for the random controls to determine upperside of 95% confidence interval. The value of 95% confidence interval was used to select significantly signal-enriched promoters and enhancers were determined, and then the numbers were counted in bulk ChIP-seq and RscEpi-seq. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/tree/master/005_Figure3a-3d> . For the sake of clarity, workflow and data flow of the Shell script was shown on page 6 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Enhancer classification**

Enhancers in the HACER datasets were classified based on relative ratios Log2(H3K27ac/H3K27me3). Putative signals of H3K27ac, H3K27me3, Med1 and 5hmC in enhancers were counted using BED tools with option “map”. The signal counts were normalized based on the length of enhancers (per kilobase) and the number of input signals (per million input). Enhancers were separated based on Log2(H3K27ac/H3K27me3). Average signal counts were calculated in the classified enhancers. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/006_Figure3e-3h>. For the sake of clarity, workflow and data flow of the Shell script was shown on page 7 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Genes interacting with active enhancers**

Active enhancers were determined based on relative ratios Log2(H3K27ac/H3K27me3) calculated using the putative signals, which have a p-value of less than 0.01. To compare the results of RscEpi-seq with bulk ChIP-seq, H3K27ac and H3K27me3 data sets of bulk ChIP-seq were used (ENCODE, dataset ID: ENCFF301TVL and ENCFF330YFF). Number of putative H3K27ac or H3K27me3 signals in enhancers was counted by BED tools. The signal counts were normalized based on the length of enhancers (per kilobase) and the number of input signals (per million input). Relative ratio Log2(H3K27ac/H3K27me3) were calculated from the normalized signal counts, and enhancers having a value greater than 0 were selected as putative active enhancers. The putative active enhancers were separated based on cell-type specificity defined by the HACER. A list of genes, which interact with the identified putative active enhancers were generated based on experimentally validated information in the HACER. The number of genes were counted and also overlapping genes between bulk ChIP-seq and RscEpi-seq were shown in Figure 4a and 4b. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/007_Figure4a-4b> . For the sake of clarity, workflow and data flow of the Shell script was shown on page 8 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Pathway enrichment analysis**

Functions of genes proximal to the putative active enhancers were analyzed by Pathway Enrichment Analysis 29. A BED file was generated by combining BED files of enhancers (HACER) and promoters (Eukaryotic Promoter Database 16 of Swiss Institute of Bioinformatics). H3K27ac and H3K27me3 signals from RscEpi-seq and bulk ChIP-seq (ENCFF301TVL and ENCFF330YFF) in enhancers and promoters were counted by BED tools. The signal counts ware normalized based on length of enhancers (per kilobase) and a number of input signals (per million input). Relative ratio Log2(H3K27ac/H3K27me3) were calculated from the normalized signal counts. A list of genes was generated form HACER and the promoter datasets. An epigenetic score of a gene was calculated by totalizing epigenetic scores of proximal enhancers and promoters. The epigenetic scores of genes were used as input data for Ingenuity Pathway Analysis (Qiagen). For RNA-seq data from bulk K562 and H1 cells, datasets were downloaded from Cancer Cell Line Encyclopedia (Broad Institute) and ENCODE (dataset ID: ENCFF093NEQ). Read per kilobase per million (RPKM) was converted to Log2, and the difference between K562 and H1 cells were calculated. The difference was converted into a ranking metric, and the data were converted into a text file. The text file was used as input data for Ingenuity Pathway Analysis. The top 5,000 transcripts in K562 and H1 cells were used in the Pathway Enrichment Analysis. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/008_Figure4d>. For the sake of clarity, workflow and data flow of the Shell script was shown on page 9 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**t-Stochastic Neighbor Embedding (t-SNE)**

t-SNE plot was generated from the epigenetic status of enhancers and promoters. Bulk ChIP-seq data of H3K27ac and H3K27me3 were downloaded from ENCODE (258 datasets, the list is shown in Table S4). Signals of the bulk ChIP-seq in enhancers (HACER) and promoters (Eukaryotic Promoter Database 16 of Swiss Institute of Bioinformatics) were counted by BED tools. Also, putative signals of each single cell were counted by the BED tools. The counts were normalized based on the length of enhancers or promoters (per kilobase) and the number of inputs (per million inputs). Relative ratios Log2(H3K27ac/H3K27me3) were calculated from the normalized signal counts. The Log2(H3K27ac/H3K27me3) values of enhancers and promoters were used as input data for software SeqGeq (FlowJo) and t-SNE plot was genetrated. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/009_Figure5a> . For the sake of clarity, workflow and data flow of the Shell script was shown on page 10 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**Visualization of K562-cell-type-specific and non-specific, active enhancers with Med1 and 5hmC signals**

Putative signals of Med1 and 5hmC in individual cells were extracted from unique reads using BED files generated in Supplementary Figure 4 by BED tools. The extracted reads (BAM files) of individual cells were converted into BED files by BED tools with option “bamtobed”. Data sets of significantly active enhancers generated in Figure 3a-3d were converted into TDF files by software IGVtools (version 2.7.2) 55. The generated TDF files and BED files were visualized by Integrated Genome Viewer, and data were shown in Figure 5b-5e and Supplementary Figure 5a-5c. All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/tree/master/010_Figure5b-5e%2BSupplementary_Figure5a-5c> . For the sake of clarity, workflow and data flow of the Shell script was shown on page 10 of a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**General reagents**

In the experiments above, the following reagents were used; 10 x TBS, pH7.4 (351-086-101, Quality Biological), glycerol (15514-011, Thermo Fisher Scientific), 0.5 M EDTA, pH8.0 (15575-038, Thermo Fisher), BSA (B6917-100MG, Sigma-Aldrich), 10 x PBS, pH7.4 (70011-044, Thermo Fisher Scientific), Tween 20 (P1379-500ML, Sigma-Aldrich), 0.2 ml PCR tubes (A30588, Thermo Fisher Scientific), Tri-HCl, pH8.8 (T1588, Teknova), Ammonium sulfate (A4418-100G, Sigma-Aldrich), 2 M KCl (AM9640G, Thermo Fisher Scientific), Triton X-100 (T8787-50ML, Sigma-Aldrich), MgSO4(M3409-10x1ML, Sigma-Aldrich), 5 M NaCl (AM9760G, Thermo Fisher Scientific), 10 mM dNTP mix (R0192, Thermo Fisher Scientific), polyethylene glycol 8000 (P5413-500G, Sigma-Aldrich), 2 M MgCl2 (340-034-721, Quality Biological), TE buffer (351-010-131, Quality Biological) and UltraPure Water (10977-015, Thermo Fisher Scientific).

**Data availability**

All sequencing data will be available from the NIH National Center for Biotechnology Information Sequence Read Archive (PRJNA522467).

**Code availability**

All codes used here are available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq>. For the sake of clarity, workflows and data flow of the Shell scripts were shown in a PDF file available at <https://github.com/HidetakaOhnuki-NCI/RscEpi-seq/000_Workflow_page1-11.pdf> .

**References**

1 Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81-94, doi:10.1038/nrclinonc.2017.166 (2018).

2 Mazor, T., Pankov, A., Song, J. S. & Costello, J. F. Intratumoral Heterogeneity of the Epigenome. *Cancer Cell* **29**, 440-451, doi:10.1016/j.ccell.2016.03.009 (2016).

3 McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613-628, doi:10.1016/j.cell.2017.01.018 (2017).

4 Gardner, E. E. *et al.* Chemosensitive Relapse in Small Cell Lung Cancer Proceeds through an EZH2-SLFN11 Axis. *Cancer Cell* **31**, 286-299, doi:10.1016/j.ccell.2017.01.006 (2017).

5 Salgia, R. & Kulkarni, P. The Genetic/Non-genetic Duality of Drug 'Resistance' in Cancer. *Trends Cancer* **4**, 110-118, doi:10.1016/j.trecan.2018.01.001 (2018).

6 Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431-435, doi:10.1038/nature22794 (2017).

7 Carter, B. *et al.* Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat Commun* **10**, 3747, doi:10.1038/s41467-019-11559-1 (2019).

8 Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060-1066, doi:10.1038/s41588-019-0424-9 (2019).

9 Harada, A. *et al.* A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat Cell Biol* **21**, 287-296, doi:10.1038/s41556-018-0248-3 (2019).

10 Ku, W. L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325, doi:10.1038/s41592-019-0361-7 (2019).

11 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-1172, doi:10.1038/nbt.3383 (2015).

12 Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol Cell* **76**, 206-216 e207, doi:10.1016/j.molcel.2019.07.015 (2019).

13 Chung, K. *et al.* Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332-337, doi:10.1038/nature12107 (2013).

14 Tillberg, P. W. *et al.* Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nat Biotechnol* **34**, 987-992, doi:10.1038/nbt.3625 (2016).

15 Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622-1626, doi:10.1126/science.1229164 (2012).

16 Gullberg, M. *et al.* Cytokine detection by antibody-based proximity ligation. *Proc Natl Acad Sci U S A* **101**, 8420-8424, doi:10.1073/pnas.0400552101 (2004).

17 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

18 Wang, J. *et al.* HACER: an atlas of human active enhancers to interpret regulatory variants. *Nucleic Acids Res* **47**, D106-D112, doi:10.1093/nar/gky864 (2019).

19 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).

20 Gray, S. M., Amezquita, R. A., Guan, T., Kleinstein, S. H. & Kaech, S. M. Polycomb Repressive Complex 2-Mediated Chromatin Repression Guides Effector CD8(+) T Cell Terminal Differentiation and Loss of Multipotency. *Immunity* **46**, 596-608, doi:10.1016/j.immuni.2017.03.012 (2017).

21 Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83, doi:10.1016/j.cell.2015.08.036 (2015).

22 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

23 Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* **12**, R54, doi:10.1186/gb-2011-12-6-r54 (2011).

24 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).

25 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348, doi:10.1038/nature10066 (2011).

26 Ding, J. *et al.* Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. *Cell Stem Cell* **16**, 653-668, doi:10.1016/j.stem.2015.04.001 (2015).

27 Teng, L., He, B., Wang, J. & Tan, K. 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics* **31**, 2560-2564, doi:10.1093/bioinformatics/btv158 (2015).

28 Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110-1122, doi:10.1016/j.molcel.2019.05.032 (2019).

29 Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).

30 Lozzio, C. B. & Lozzio, B. B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**, 321-334 (1975).

31 Chylicki, K. *et al.* p53-mediated differentiation of the erythroleukemia cell line K562. *Cell Growth Differ* **11**, 315-324 (2000).

32 Duncan, M. T. *et al.* SIRT1 is a critical regulator of K562 cell growth, survival, and differentiation. *Exp Cell Res* **344**, 40-52, doi:10.1016/j.yexcr.2016.04.010 (2016).

33 Yang, C., Cai, H. & Meng, X. Polyphyllin D induces apoptosis and differentiation in K562 human leukemia cells. *Int Immunopharmacol* **36**, 17-22, doi:10.1016/j.intimp.2016.04.011 (2016).

34 Andersson, L. C., Nilsson, K. & Gahmberg, C. G. K562--a human erythroleukemic cell line. *Int J Cancer* **23**, 143-147, doi:10.1002/ijc.2910230202 (1979).

35 Klein, E. *et al.* Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *Int J Cancer* **18**, 421-431, doi:10.1002/ijc.2910180405 (1976).

36 Lozzio, B. B., Lozzio, C. B., Bamberger, E. G. & Feliu, A. S. A multipotential leukemia cell line (K-562) of human origin. *Proc Soc Exp Biol Med* **166**, 546-550, doi:10.3181/00379727-166-41106 (1981).

37 Warren, L., Bryder, D., Weissman, I. L. & Quake, S. R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc Natl Acad Sci U S A* **103**, 17807-17812, doi:10.1073/pnas.0608512103 (2006).

38 Valent, P. *et al.* DPPIV (CD26) as a novel stem cell marker in Ph+ chronic myeloid leukaemia. *Eur J Clin Invest* **44**, 1239-1245, doi:10.1111/eci.12368 (2014).

39 Nagel, J. E., Collins, G. D. & Adler, W. H. Spontaneous or natural killer cytotoxicity of K562 erythroleukemic cells in normal patients. *Cancer Res* **41**, 2284-2288 (1981).

40 Sutherland, J. *et al.* Induction of the expression of HLA class I antigens on K562 by interferons and sodium butyrate. *Hum Immunol* **12**, 65-73, doi:10.1016/0198-8859(85)90344-1 (1985).

41 Faltas, B. M. *et al.* Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat Genet* **48**, 1490-1499, doi:10.1038/ng.3692 (2016).

42 Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314-322, doi:10.1038/nature09781 (2011).

43 Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**, 69-75, doi:10.1126/science.aan6826 (2017).

44 Young, L., Sung, J., Stacey, G. & Masters, J. R. Detection of Mycoplasma in cell cultures. *Nat Protoc* **5**, 929-934, doi:10.1038/nprot.2010.43 (2010).

45 Weiland, G. & zu Dohna, H. G. [The suitability of the enzyme-linked immunosorbent assay (ELISA) for the demonstration of babesia antibodies in hyperimmune serums with the use of ectoantigens]. *Berl Munch Tierarztl Wochenschr* **91**, 33-35 (1978).

46 Lu, S. *et al.* Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **338**, 1627-1630, doi:10.1126/science.1229112 (2012).

47 Paithankar, K. R. & Prasad, K. S. Precipitation of DNA by polyethylene glycol and ethanol. *Nucleic Acids Res* **19**, 1346 (1991).

48 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* **1**, 895-905, doi:10.3390/biology1030895 (2012).

49 Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491-499, doi:10.1101/gr.209601.116 (2017).

50 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

51 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

52 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

53 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

54 Dreos, R., Ambrosini, G., Groux, R., Cavin Perier, R. & Bucher, P. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res* **45**, D51-D55, doi:10.1093/nar/gkw1069 (2017).

55 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

**Acknowledgments**

We thank Drs. David Sanchez-Martin and Christopher B. Buck for comments during the conceptualization stage of the project. We also thank the Genomics Core, Center for Cancer Research, National Cancer Institute, National Institutes of Health for help in preliminary experiments, and the Collaborative Bioinformatics Resource, CCR, NCI, NIH for advice in computational analysis. We thank Ms. Anna Word for helping with the optimization of DNA polymerases used in the method. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). This project is supported by the Intramural Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health, the NCI Director’s Innovation Award (#397172) and Federal funds from the National Cancer Institute under Contract No. HHSN261200800001E. We thank Drs. Tom Misteli, Douglas R. Lowy and all members of Laboratory of Cellular Oncology for productive comments. We also thank Drs. Christopher B. Buck, Robert Yarchoan, and Douglas R. Lowy for reviewing the manuscript.