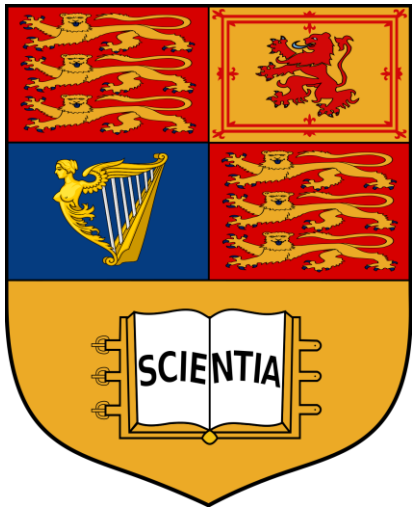
Benchmarking of CUT&RUN to profile histone post-translational modifications in K562 cell line



Yi Yang

Department of Medical Biosciences

School of Medicine, Imperial College London

CID number: 01728962

Supervisor: Dr Nathan Skene

UKDRI, Department of Brain Sciences,

Faculty of Medicine, Imperial College London

This research report is submitted in partial fulfilment of the requirements for the degree of *BSc in Medical Biosciences*

April 2022

Benchmarking of CUT&RUN to profile histone post-translational modifications in K562 cell line

Key abbreviations
PTMs, post-translational modifications; ChIP-seq, chromatin immunoprecipitation coupled with next-generation sequencing; CUT&RUN, Cleavage Under Targets and Release Using Nuclease; ENCODE, The Encyclopedia of DNA Elements; MNase, micrococcal nuclease; CUT&Tag Cleavage Under Targets & Tagmentation; FANS, fluorescence-activated nuclei sorting; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology; FRiP, fraction of reads in peaks.

ABSTRACT

Epigenetic regulation in eukaryotic cells is defined as mechanisms that control gene expression through chromatin dynamics. One typical example is protein-DNA interactions including histone post-translational modifications (PTMs) and binding of transcription factors. Recent studies found that most biological traits were associated with non-coding regulatory elements of human genome, making epigenetic regulation the key to understand mechanisms underlying various disease. However, detailed annotations are lacking because the current standard technique chromatin immunoprecipitation sequencing (ChIP-seq) is not suitable to generate cell type-specific epigenomic profiles due to enormous cells demanded. Cleavage Under Targets and Release Using Nuclease (CUT&RUN) is a potential alternative which profiles PTMs and transcription factors *in situ* at high resolution with low input. Nevertheless, its application is limited by sample types and absence of benchmarking against ChIP-seq. In this study, a modified CUT&RUN protocol was introduced to improve compatibility to post-mortem tissues; benchmarking was conducted by performing CUT&RUN of 11 histone PTMs in K562 cell line and comparing the data to one of the most authoritative ChIP-seq datasets The Encyclopedia of DNA Elements (ENCODE). The results showed that profiles derived from CUT&RUN and ENCODE ChIP were largely different in peaks and functional annotations, and the overlap varied by PTMs. CUT&RUN could potentially replace ChIP-seq for heterochromatic marks but the validity for transcriptionally active marks and transcription factors needs further evidence. This study is a cornerstone for future epigenomic profiling of various biological samples, with an attempt to decode epigenetic abnormality in major human disease.

KEYWORDS

Epigenetics, epigenomics, CUT&RUN, ChIP-seq, histone modifications, ENCODE

INTRODUCTION

Epigenetics is the study of heritable phenotypic changes that happen without alternations in DNA sequence. Eukaryotic DNAs are wrapped around an octamer of histone proteins to form nucleosomes for getting packaged into more compact chromatin structures. Highly condensed and transcriptionally silent regions define heterochromatin whereas loosely compacted and transcriptionally active regions are euchromatin. Expression of a specific gene is strictly under the control of chromatin dynamics such as histone post-translational modifications (PTMs), transcription factor binding, nucleosome positioning, DNA methylation, etc, near regulatory elements of the gene. (1) In particular, accessibility of DNA segments to transcriptional machineries is largely dependent on histone PTMs including methylation, acetylation, phosphorylation, etc. (2) For instance, H3K27ac (histone H3 lysine 27 acetylation) is commonly associated with open chromatin and active transcription while H3K27me3 (histone H3 lysine 27 tri-methylation) with gene silencing. (3,4)

Over the past few years epigenetics has become a promising approach to understand mechanisms underlying human diseases, especially complex traits of brain such as Alzheimer's disease, Parkinson's disease, and autism spectrum disorders. (5–8) Compared to transcriptomics which examines levels of RNAs, epigenetics derives more reliable results in many cases. It was found that post-mortem tissues, a prevalent source of biological samples, had very different transcriptomic features in contrast to fresh tissues within just a few hours after death while the epigenomes, i.e., histone PTMs, tended to be stable over a relatively long period. (9,10) Moreover, transcriptomic studies on brain cells mostly used isolated nuclei for RNA -sequencing due to technical difficulties in obtaining cells from frozen tissues, which means that RNAs that remain in nuclei after sorting are not a good representation of the complete transcriptome. (11) Epigenomics avoids this problem naturally since all chromatin dynamics happen in nuclei.

In addition to the technical advantages, epigenetics provides more insights into gene expression regulation than simply looking at RNA levels since some regulatory elements are not transcribed into RNAs. As revealed by a number of genome-wide association studies, the majority of human disease risk variants are located in non-coding areas of the genome and show preferential enrichment in regulatory elements, indicating that they are likely to affect gene expression rather than protein function. (12–16) Plenty of studies suggested that different patterns of PTMs and transcription factors were correlated with neurodegenerative disorders, possibly caused by dysregulated expression of disease-associated genes. (3, 17-20)

Despite risk variants being identified, their functional link to disease and mechanisms of action remain largely elusive due to absence of epigenomic annotations in normal and pathological tissues. Moreover, epigenetic features were proved to be dynamic during various cellular processes such as ageing, cell fate determination, carcinogenesis, etc. (21–23) They are also highly dependent on cell types. Patterns of histone PTMs are predictive of gene expression across different cells. (24-26) The fact that some cells are more vulnerable in disease emphasises the need to assess cell type-specific epigenomic profiles. For example, many risk variants of Alzheimer's disease were found to be particularly enriched in microglia compared to other brain cells. (14,16,27,28) Given that the currently available epigenomic data was limited and mostly from sequencing of bulk tissues that consisted of heterogeneous cells, it is important to develop advanced profiling techniques to understand how epigenetic abnormality causes disease and hopefully identify new therapeutic targets.

The standard technique to profile DNA-protein interactions in the past few decades was chromatin immunoprecipitation (ChIP), coupled with next generation sequencing (ChIP-seq). One of the most extensive epigenomic databases The Encyclopedia of DNA Elements (ENCODE) was generated using ChIP-seq. In ChIP-seq, DNA is cross-linked to target proteins with formaldehyde and fragmented by sonication. The resulting pieces of DNA-protein complexes are immunoprecipitated using antibodies against the targets, then extracted, amplified, and sequenced. **(Fig.1A)** Despite the wide use, ChIP-seq has many limitations (for detailed discussion see Meyer and Liu *et al.* (29)). Various adaptations of the method were developed to overcome problems such as low resolution (i.e., ChIP-exo), *in vivo* applications (i.e., ChIP-nexus), epitope masking (i.e., native ChIP-seq), etc. (30-32)

However, none of the currently available ChIP-seq approaches seem suitable to cells sorted for cell type-specific analysis (especially in post-mortem human samples). The major difficulty of standard ChIP-seq is that it requires a great number of cells, typically over 5 million, while some cell types are rare. For example, microglia account for up to 16.6% of total brain cells in human, (33) it is hard to obtain enough cells for ChIP-seq. This problem could be overcome using Cleavage Under Targets and Release Using Nuclease (CUT&RUN), a recently developed technique which profiles PTMs *in situ*. (34) In CUT&RUN, permeabilised cells or nuclei are bound to magnetic beads and incubated with antibodies against target proteins. Micrococcal nuclease (MNase) is then tethered to the antibodies via fusion protein pA/G which, when activated by calcium ions, can cleave DNA on both proximal sides. Only short target DNA fragments will be released into supernatant, amplified, and sequenced **(Fig.1B)**. Compared to ChIP-seq, CUT&RUN requires far less cells/nuclei for a nearly base-pair resolution and is therefore especially suitable for profiling rare biological samples. Other

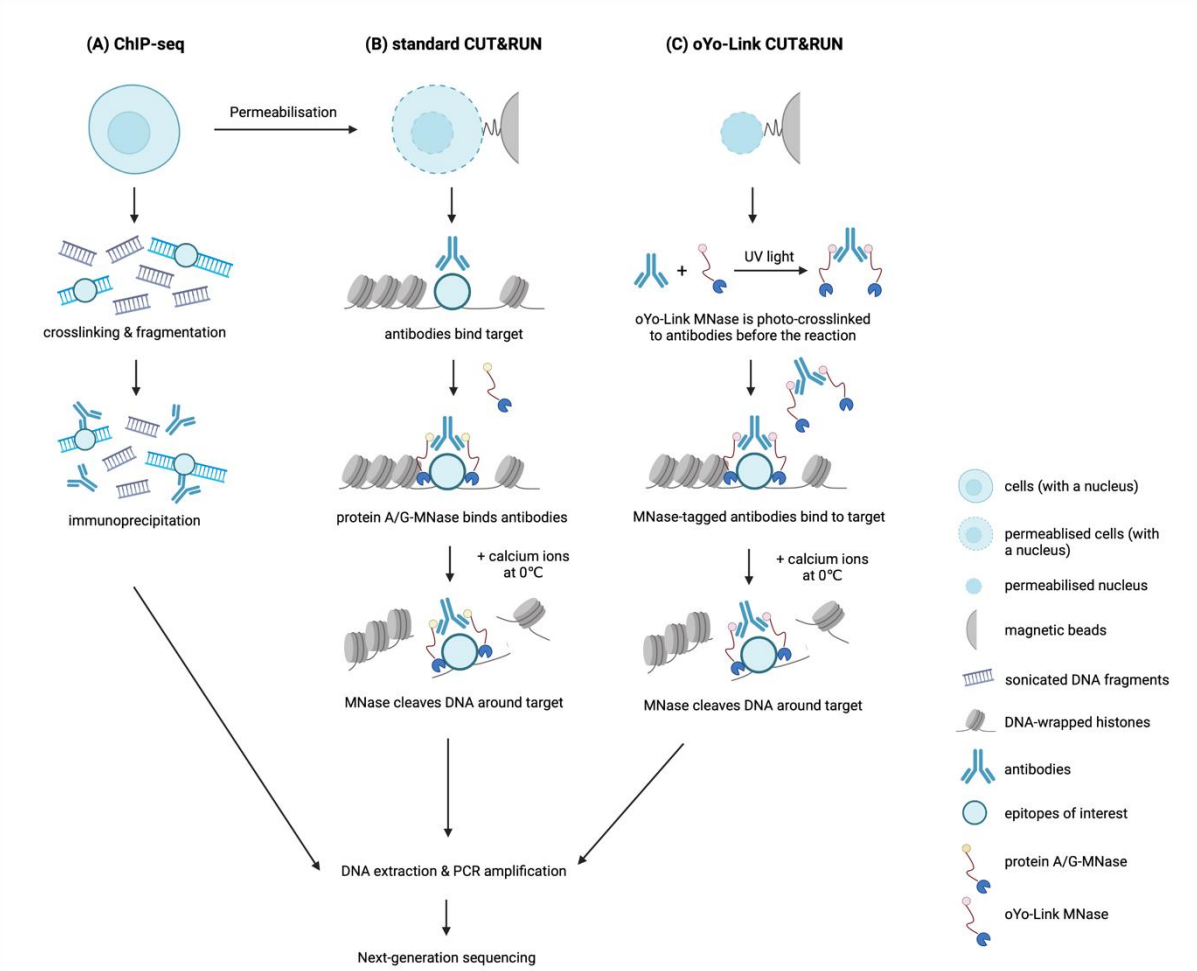


Fig.1: Schematic diagrams of the techniques (A) ChIP-seq. Proteins of interest are profiled in a crosslinking condition. Low resolution is resulted from sequencing sonicated fragments of variable lengths. **(B)** The standard CUT&RUN. Proteins of interest are profiled *in situ*. Cleavage of MNase generates fragments of nucleosomal length (or multiples of the length) and therefore achieves better resolution. **(C)** CUT&RUN with the oYo-Link modification for compatibility to more biological samples. Fusion protein A/G-MNase is replaced by oYo-Link MNase, avoiding mixing of antibodies. Created with BioRender.com

enzyme-tethering approaches for low inputs, e.g., Cleavage Under Targets & Tagmentation (CUT&Tag), do exist but CUT&RUN was proved to outperform them in terms of resolution, transcription factor profiling, washing conditions, etc, due to the use of MNase. (35)

Before applying CUT&RUN to sorted cells from human post-mortem samples, two problems must be solved. First, one of the promising methods to isolate particular nuclei types from frozen human tissues is fluorescence-activated nuclei sorting (FANS), which relies on immunostaining of marks specific to the cell type (i.e., interferon regulatory factor 8, a transcription factor, to microglia). (36) With standard CUT&RUN, the antibodies retained from FANS will mix with the ones of CUT&RUN, causing MNase to cleave excessive genomic areas.

One possible solution is to directly attaches MNase to CUT&RUN antibodies beforehand (e.g. oYo-Link method), (37) rather than using fusion protein pA/G-MNase from the original protocol **(Figure.1C)**.

Second, despite that CUT&RUN replaced ChIP-seq in many studies, its actual performance has not been systematically evaluated; data is limited to a few PTMs and transcription factors. It is still in question whether CUT&RUN-derived data can yield similar, better, or less findings than ChIP-seq. Notably, many studies observed more peaks being identified using CUT&RUN compared to ChIP-seq (38–41) while a few reported the opposite. (42,43) Furthermore, recovery rates of ChIP-seq peaks in CUT&RUN studies tended to vary considerably depending on given experimental conditions. (38–43) Thus, it is important to benchmark CUT&RUN before use. By comparing CUT&RUN to ChIP-seq data from an established epigenomic database, i.e., ENCODE, the performance of CUT&RUN could be verified. Given the limitations of ChIP-seq, the comparison also provides a unique opportunity to examine if ENCODE ChIP is indeed the ground truth in epigenetics.

As provision for profiling epigenetic features, i.e., PTMs, this study aims to **1)** introduce a modified CUT&RUN protocol compatible to more biological samples; **2)** benchmark CUT&RUN by performing modified CUT&RUN of 11 histone PTMs (i.e., H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K9me3, H3K27ac, H3K27me3, H3K36me3, H3K79me2, H4K20me1; see **Supplementary Table 1** for full names and descriptions) in human K562 cell line and comparing the results to the corresponding ChIP-seq data from the extensive epigenomic database ENCODE. PTMs of different kinds were profiled for relatively comprehensive evaluation. The specific cell line was selected because both CUT&RUN and ENCODE ChIP data are widely available, which is ideal for benchmarking. The hypothesis is that CUT&RUN, with the oYo-Link modification, provides an efficient and reliable alternative to traditional ChIP-seq in PTM profiling of rare cell types because it is expected to recall most ENCODE ChIP features accurately with low input.

MATERIALS AND METHODS

Biological materials

Human K562 cells were purchased from ATCC (Manassas, VA, CCL-243), cultured following the supplier's protocol, and cryopreserved in 10% dimethyl sulfoxide (DMSO) in foetal bovine serum (Signa, F3885). Antibody information is detailed in **Table 1**. MNase-

conjugated antibodies were made by mixing antibodies and oYo-Link MNase (AlphaThera, #AT6001) at 1:1.6 ratio under 365 nm UV light for 2 hours on ice.

Table 1: CUT&RUN antibodies Antibodies of 11 histone PTMs and a negative control were photo-crosslinked to oYo-Link MNase (AlphaThera, #AT6001) to target epitopes of interest during CUT&RUN. These antibodies were the same ones used in ENCODE ChIP-seq data in K562 cell line (**Supplementary Table 1**). (44,45) They were selected on purpose for consistency in CUT&RUN benchmarking. *The only exception was the H3K27me3 antibody from Cell Signaling, which had been optimised by CUT&RUN's developers. (34) It was used as a positive control to mark proper CUT&RUN reactions. ** The polyclonal Guinea pig anti-rabbit IgG antibody targets an absent epitope in K562 cells. It was used as a negative control to assess the level of background cutting of MNase.

Name	Dilution used	Consistent to ENCONDE ChIP antibodies
H3K4me1 (Abcam; ab8895; Lot GR3407156-1)	1:200	Yes
H3K4me2 (Abcam; ab7766; Lot GR3397853-1)	1:200	Yes
H3K4me3 (Abcam; ab8580; Lot GR3362386-3)	1:200	Yes
H3K9ac (Abcam; ab4441; Lot GR3418494-1)	1:200	Yes
H3K9me1 (Abcam; ab8896; Lot GR3416786-1)	1:200	Yes
H3K9me3 (Abcam; ab8898; GR3414325-1)	1:200	Yes
H3K27ac (Abcam; ab4729; GR3394626-1)	1:200	Yes
H3K27me3 (positive control) (Cell Signaling; #9733, Lot 14)	1:500	No*
H3K36me3 (Abcam; ab9050; Lot GR3382010-2)	1:200	Yes
H3K79me2 (Active Motif; #39923; Lot 2771001)	1:200	Yes
H4K20me1 (Abcam; ab9051; Lot GR3344957-2)	1:200	Yes
Polyclonal Guinea Pig Anti-Rabbit IgG Antibody (negative control) (Antibodies online; #ABIN6923140)	1:100	N/A**

CUT&RUN

The CUT&RUN protocol was based on Meers et al. (46) and adapted for compatibility to more sample types. Buffer recipes see **Supplementary Table 2**. Quickly thawed K562 cells (50,000 per sample) were washed three times in 1 mL wash buffer to remove DMSO and incubated with 10 µL activated concanavalin A-coated magnetic beads (Bangs Laboratories, #BP531) at RT with agitation. (47) After 10 minutes, tubes were placed on a magnetic stand until the solution was clear. Supernatant was discarded to remove unbound beads. 100 µL antibody buffer and MNase-conjugated antibodies were added per sample, followed by overnight incubation at 4℃ with agitation (see **Table 1** for dilutions). Reactions were placed on a magnetic stand to clear and liquid was pulled off to remove unbound antibodies. Beads-bound cells were washed in 200 µL wash buffer twice and 200 µL high-calcium/low-salt rinse buffer once to reduce premature DNA release as chromatin would shrink in low-salt conditions. (46) 150 µL ice-cold incubation buffer was added, followed by gentle vortexing to dislodge all beads. Tubes were incubated at 0℃ with agitation for 30 minutes to allow antibody-guided cleavage by MNase. After being placed on a magnetic stand to clear and liquid being withdrawn, reactions were stopped by addition of 33 µL stop buffer to chelate away remaining calcium ions and incubated at 37℃ for 30 minutes to release DNA fragments from insoluble nuclear chromatin. While placed in a magnetic stand to clear, 30 µL supernatant containing digested DNA was transferred to new PCR tubes for library preparation.

Library preparation and sequencing

10 µL per sample CUT&RUN fragments were diluted in 40 µL 0.1x TE buffer (0.1 mM Tris-HCl, pH 8.0; 0.1 mM ethylenediaminetetraacetic acid) and PCR-amplified using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB, #E7103L) and NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) (NEB, #7335L) following the manufacturer's protocol. Adaptors were diluted in Tris/NaCl, pH 8.0 for a concentration of 0.6 µM during ligation to reduce excessive primer dimers. PCR conditions were optimised for CUT&RUN fragments: initial denaturation at 98℃ for 45 s; 14 cycles of 98℃ for 15 s and 60℃ for 10 s; final extension at 72℃ for 4 mins. Primer dimers were cleaned up from PCR products by adding 1.2x sample purification beads provided in the kit. Before sequencing, fragment size distributions of CUT&RUN characteristics were verified for quality control using Agilent 2100 Bioanalyser and High Sensitivity DNA Kit (Agilent, 5067-4627). Libraries were pooled together at 10 nM equimolar. Paired-end Illumina sequencing (NovaSeq PE150) was conducted on indexed libraries by Novogene, Cambridge.

Data processing

The workflow for CUT&RUN data processing was set up according to the tutorial created by Henikoff's lab, with some modifications. (48)

The quality of raw sequencing data was checked using FastQC (version 0.11.9) before adaptor removal by Cutadapt (version 3.7) and TrimGalore (version 0.6.6). Trimmed paired-end reads were aligned to human reference genome GRCh38 using Bowtie2 (version 2.2.5; --local --very-sensitive --no-discordant --no-mixed --phred33 -I 10 -X 700). They were additionally aligned to *E.coli* genome Ensembl EB1 for spike-in calibration with two more parameters (--no-dovetail --no-overlap) to avoid cross-mapping. Potential PCR duplicates were removed from name-sorted mapped reads using Picard (version 2.26.11), the rationale of which was explained in the lab's previous work. (49) BAM files after de-duplication were converted into fragment BED files to select for fragment length less than 1000 bp using samtools (version 1.9) and bedtools (version 2.30.0). No extra filtering was needed because of high alignment quality scores across all samples. To normalise the effect of different epitope abundance, (50) scaling factor S (S = 10,000 / fragments aligned to *E. coli* genome) was calculated and used as a bedtools genomecov parameter to prepare genomic coverage bedgraph files for peak calling.

Peaks were called using SEACR (version 1.4; -c 0.01 -n non -m stringent), a software specifically designed for data with low background. The cut-off threshold was set at false discovery rate = 0.01 as optimised in the lab's previous study. (49)

ENCODE benchmarking and functional analysis

Comparison of CUT&RUN peaks to ENCODE ChIP peaks (**Supplementary Table 3**) (44,45) and all functional analysis were completed using EpiCompare (version 0.99.13; genome_build = "hg38" blacklist = hg38_blacklist chrmHMM_annotation = "K562"), an R package developed by members in the lab for comparing epigenetic datasets. (51) Peaks in non-standard chromosomes and hg38 blacklisted regions (ENCODE ID: ENCFF356LFX) were removed to eliminate the effect of areas with anomalous signals across many cell lines and experiments. (52) Adjusted p-values ≤ 0.01 denoted statistical significance in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and gene ontology (GO) enrichment analysis.

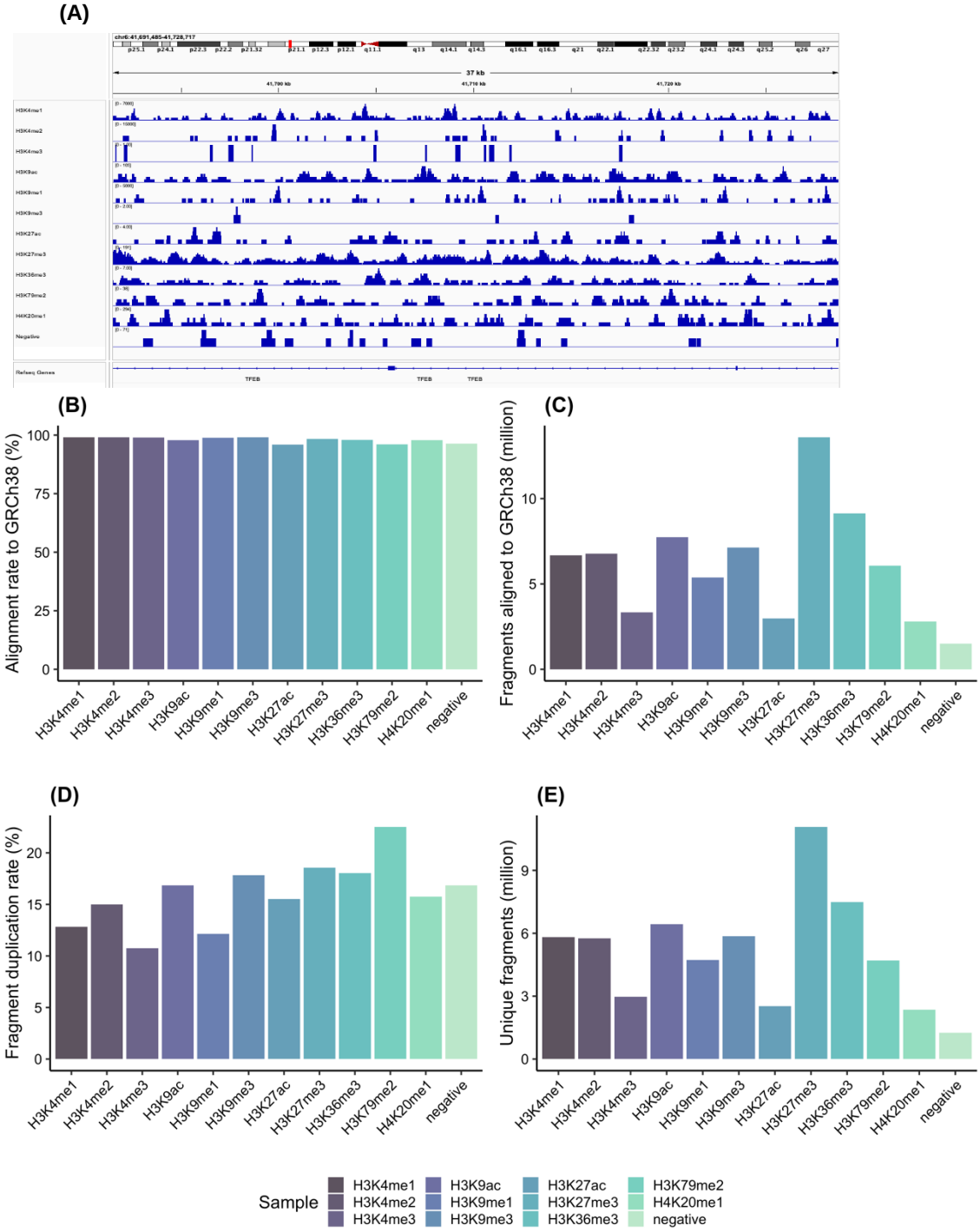


Fig.2: Quality control metrics of CUT&RUN sequencing data Eleven different histone PTMs were profiled in human K562 cell line using the modified CUT&RUN protocol (n=1). Negative IgG control targets an absent epitope in the cell line. The resulting fragments were extracted, amplified, and sequenced by pair-end next-generation sequencing. The quality of the sequencing data was assessed before generating peak profiles. **(A)** A representative display of chromatin landscapes of CUT&RUN samples using Integrative Genomic Viewer. **(B)** Alignment rates of CUT&RUN reads to human reference genome GRCh38; **(C)** Number of CUT&RUN reads aligned to GRCh38; **(D)** Duplication rates of aligned fragments; **(E)** Number of aligned fragments after removing duplicates, calculated by number of aligned fragments*(1-duplication rate/100).

RESULTS

Quality control of CUT&RUN data

To benchmark the performance of CUT&RUN against ENCODE ChIP, 11 histone PTMs were profiled in human K562 cell line using a modified CUT&RUN protocol. Quality control metrics of the sequencing data were summarised in **Fig.2**. The alignment rates to reference genome GRCh38 were over 95% across samples **(Fig.2B)**. The number of aligned fragments partially reflected the relative abundance of each mark **(Fig.2C)**. All samples had enough fragments since low-background experiments require just one million aligned reads. (34) H3K27me3, a wide-spread mark across human genome (i.e., heterochromatin areas), (53) had the most aligned fragments; the negative IgG control targeting an absent epitope had the least, which represented the background cleavage of MNase during experiments. Less than 25% reads were duplicates **(Fig.2D)**. CUT&RUN data continued to be processed with duplicates removed as suggested by the lab's previous work. (49) Number of fragments after duplicates being removed was expressed in **Fig.2E** as

Unique fragments = Number of fragments aligned to GRCh38 * (1-duplication rate/100).

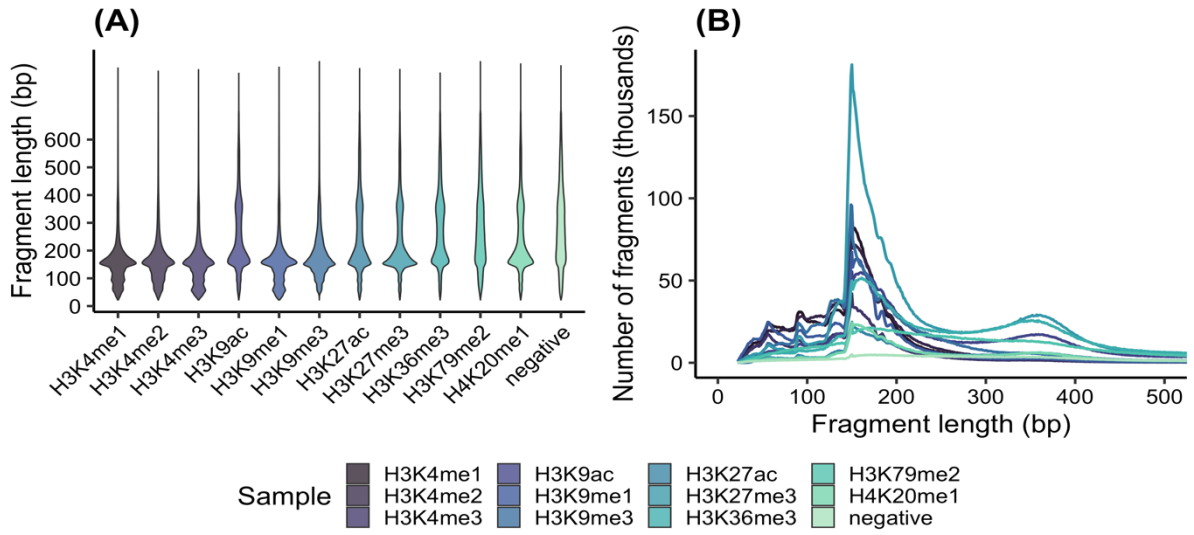


Fig.3: Size distribution of CUT&RUN fragments After alignment to reference genome, size distribution of aligned fragments was assessed. Negative IgG control represents fragments cleaved by MNase background action. **(A)** A violin graph showing size distribution of CUT&RUN fragments of each single histone PTM; **(B)** Fragment length was plotted against fragment number, showing relative abundance and distribution between different PTM samples. (n=1)

Size distribution of the fragments was characteristic of successful CUT&RUN reactions **(Fig.3)**. MNase, oYo-linked directly to antibodies, cleaved DNA on either side of chromatin particles, generating fragments of nucleosomal length or multiples of the length. Two distinct peaks were observed, one at around 150-180 bp (mono-nucleosome) and the other at 360 bp (di-nucleosome). The only exception was H3K79me2, the distribution of which resembled the IgG control, though it had more fragments than the control. Smaller fragments between 50 to 100 bp were widely present across all PTMs. These metrics showed that the CUT&RUN data generated using the modified protocol were qualified to generate robust PTM profiles.

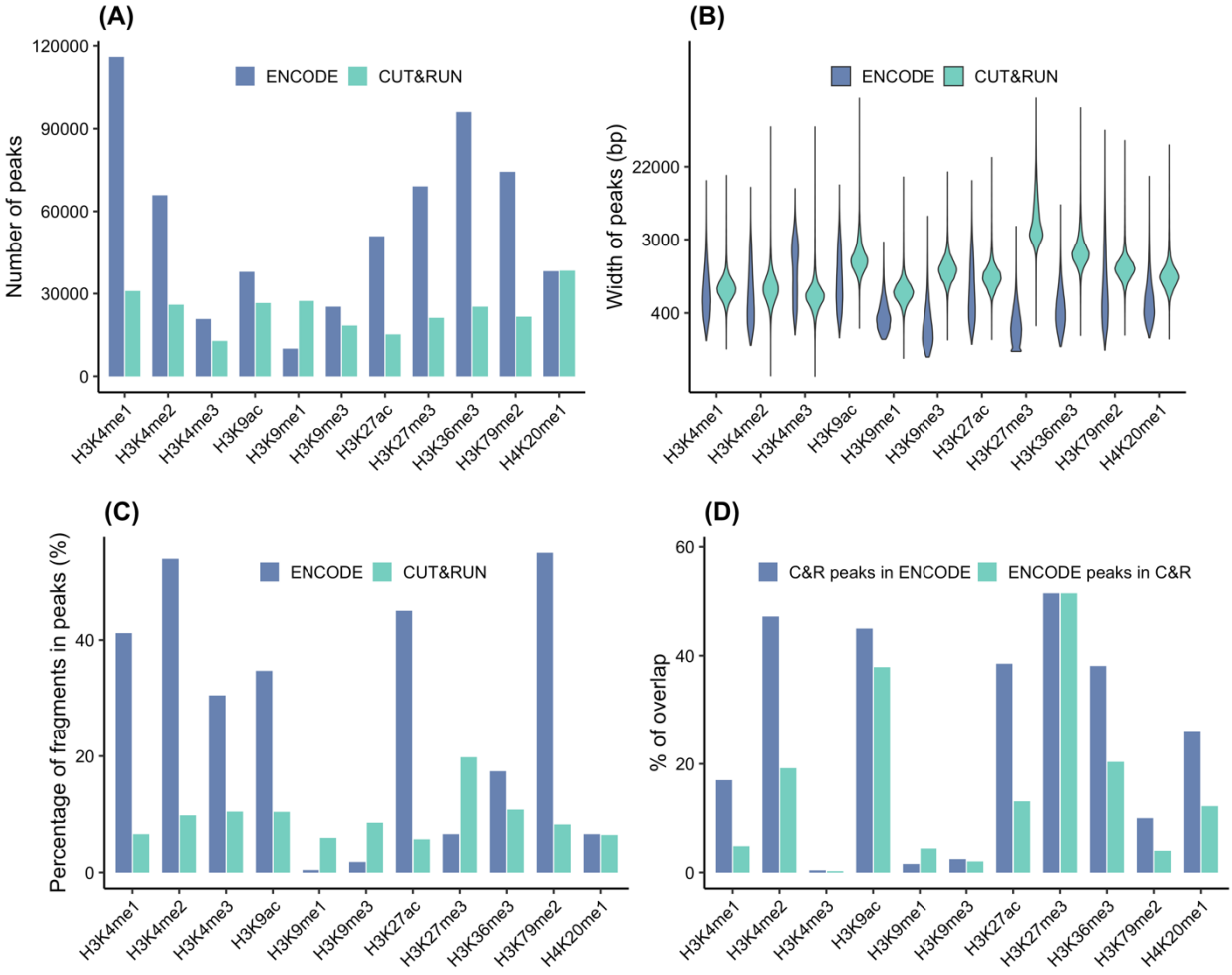


Fig.4: Comparison of peak profiles derived from CUT&RUN and ENCODE ChIP After quality control, CUT&RUN peaks of 11 histone PTMs in human K562 cells were called at a false discovery rate of 0.01. The peaks were compared to the corresponding ENCODE ChIP results of the same cell line. Overlaps with hg38 blacklisted regions were removed before the comparison. **(A)** Number of peaks called in CUT&RUN and ENCODE ChIP; **(B)** Distribution of widths of CUT&RUN and ENCODE ChIP peaks; **(C)** Background noisiness of the two techniques measured by fraction of aligned reads falling into peak regions (FRiP); **(D)** Overlap of CUT&RUN and ENCODE ChIP profiles. CUT&RUN peaks overlapped with ENCODE ChIP and ENCODE ChIP peaks overlapped with CUT&RUN were assessed respectively. (n=1 for CUT&RUN samples; n=2 for ENCODE ChIP samples)

Comparison of CUT&RUN peaks to ENCODE ChIP

Results of CUT&RUN and ChIP-seq are presented as peaks (genomic regions enriched with aligned fragments). After quality control, peaks were called from CUT&RUN data and compared to ENCODE ChIP peaks of the same cell line for benchmarking. (see **Supplementary Table 3** for data accession) (44,45) Peaks overlapped with hg38 blacklisted regions were filtered out to ensure unbiased PTM profiles. In general, CUT&RUN identified fewer peaks, except for H3K9me1 and H4K20me1, but the variation of peak number between different PTMs was greater in ENCODE ChIP **(Fig.4A)**. In terms of peak width, CUT&RUN-derived peaks were wider than ENCODE ChIP peaks and had a more compact distribution of width **(Fig.4B)**. Background noise of the two techniques was assessed by fraction of aligned reads falling into peak regions (FRiP) **(Fig.4C)**. Surprisingly, CUT&RUN had very low FRiP values (below 20%) across samples. The highest FRiP occurred in the positive control H3K27me3, the CUT&RUN conditions of which had been fully optimised in K562 cells by the creators of the method, though it did not reach the expectation either. (34) Most CUT&RUN data showed worse signal-to-noise performance compared to ENCODE ChIP, except for H3K9me1, H3K9me3, and H3K27me3. The overlap of CUT&RUN and ENCODE ChIP peaks was shown in **Fig.4D**. Since one CUT&RUN peak could correspond to multiple ENCODE peaks, and vice versa, percentage of CUT&RUN peaks found in ENCODE ChIP and percentage of ENCODE ChIP peaks recovered by CUT&RUN were calculated respectively. The overall overlap was less than 60%, which suggested that CUT&RUN generated very different epigenetic profiles to ENCODE ChIP. The positive control H3K27me3 had the best overlap rates (exceeding 50%) while both types of overlap were limited in H3K4me3, H3K9me1, and H3K9me3, implying suspicious reliability of the profiles derived. It was concluded that, compared to ENCODE ChIP, CUT&RUN peaks were fewer in number but wider in width, the signal-to-noise level of which generally fell short of the one of ChIP-seq; the peak overlap between the two techniques was limited.

Functional annotations of CUT&RUN and ENCODE ChIP peaks

To functionally analyse the difference between PTM profiles generated by CUT&RUN and ChIP-seq, peaks were categorised into different chromatin states of K562 cells **(Fig.5A)**. The major discrepancy occurred in peak profiles of H3K4me1, H3K4me3, H3K9me1, and H3K79me2. In particular, for H3K4me1 and H3K4me3, the CUT&RUN and ENCODE ChIP profiles were almost opposing, with CUT&RUN peaks mostly falling into heterochromatin while ChIP-seq peaks being enriched in enhancers of different kinds. In addition, the proportion of peaks being assigned to repressed and heterochromatic regions was always higher in

CUT&RUN samples than in ChIP-seq. The peaks were further annotated with the nearest genes and the genomic regions where they were located. **Fig.5B** showed that peaks derived

Fig.5: Functional annotations of CUT&RUN and ENCODE ChIP peaks (A) CUT&RUN and ENCODE ChIP peaks of 11 histone PTMs were annotated with chromatin states in human K562 cell line. Percentages of peaks falling into each state were plotted in the heatmap. The horizontal axis lists names of chromatin states; the vertical axis is the PTM profiles, derived from CUT&RUN data and ENCODE ChIP respectively. Txn, transcribed regions; Heterochrom.lo, heterochromatin or low signal; CNV, copy number variation. **(B)** Peaks were annotated with the nearest genes. The horizontal axis shows percentages of peaks falling into different regulatory regions; the vertical axis stands for PTM profiles. (n=1 for CUT&RUN samples; n=2 for ENCODE ChIP samples)

from CUT&RUN and ChIP-seq had distinct distributions of regulatory elements. Except for H3K27me3, H3K36me3, and H4K20me1, fewer promoters, especially promoters lying less than 1 kb away, were characterised by CUT&RUN. The extreme example was H3K4me3, the CUT&RUN profile of which mostly consisted of introns and distal intergenic areas, in contrast to over 80% being promoters in the corresponding ChIP-seq profile. Overall, the functional annotations suggested that CUT&RUN-derived peaks were functionally different from ENCODE ChIP peaks in chromatin states and regulatory elements.

To compare genes annotated by peaks of CUT&RUN and ENCODE ChIP and the roles in the biological context of K562 cells, KEGG pathway analysis and GO enrichment analysis were conducted for each histone PTM. Regarding KEGG analysis, many pathways were enriched in ENCODE ChIP of H3K4me1, H3K4me3, H3K9me1, H3K9me3, and H3K79me2 but seemed missing in the corresponding CUT&RUN profiles **(Fig.6)**. The pathway profiles of the remaining PTM samples looked alike between CUT&RUN and ENCODE ChIP; the pathways of high significance were mostly identified by both techniques. Similar results were observed in GO analysis, with only H3K4me1, H3K4me3, H3K9me1, H3K9me3, and H3K79me2 having distinct profiles of molecular functions between CUT&RUN and ENCODE ChIP **(Fig.7)**. Both KEGG and GO analysis led to the same conclusion that genes identified by CUT&RUN and ENCODE ChIP peaks were involved in similar biological processes for some PTMs but not others, indicating that the performance of CUT&RUN was not consistent across different PTMs.

DISCUSSION

This study introduced a modified version of CUT&RUN for increased compatibility to biological samples for the first time. The replacement of protein A/G-MNase fusion protein by oYo-Link MNase did not change the nature of the technique, as proved by the quality control metrics which resembled the features presented in the original CUT&RUN publication. (34) However, the overall performance of CUT&RUN was not comparable to ENCODE ChIP. Despite the low cell input, fewer peaks were called at wider widths and low signal-to-noise ratios for most PTMs. The hypothesis had to be partially rejected because very different PTM profiles were derived from CUT&RUN, as demonstrated by peak overlap lower than 60% overall and a few marks showing almost opposing functional annotations of CUT&RUN and ChIP-seq peaks. Similar results were observed in other CUT&RUN experiments (42,43) and other enzyme-tethering approaches (e.g., CUT&Tag recovered up to 50% of ENCODE ChIP

Fig.6: KEGG pathway analysis of CUT&RUN and ENCODE ChIP peaks Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed to compare CUT&RUN and ENCODE ChIP of 11 histone PTMs in human K562 cells. The horizontal axis specified PTM profiles generated from CUT&RUN and ENCODE ChIP data; the vertical axis specified the names of the enriched pathways. Gene ratio was calculated as the number of genes characterised by CUT&RUN or ENCODE ChIP peaks over the number of expected genes from each KEGG category. Adjusted p-values were shown by colour gradients; ≤ 0.01 was considered statistically significant (pure red). (n=1 for CUT&RUN samples; n=2 for ENCODE ChIP samples)

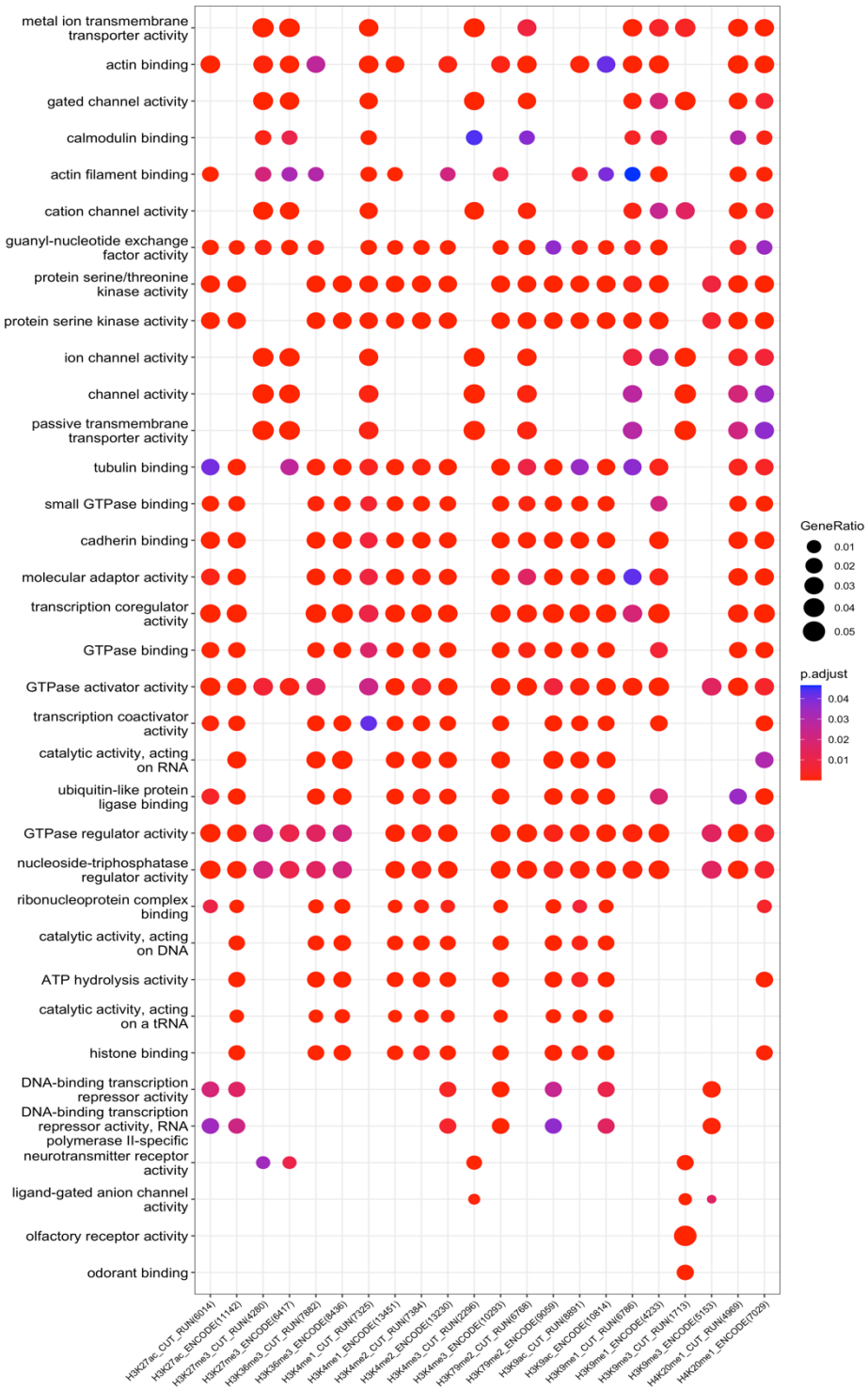


Fig.7: GO enrichment analysis of CUT&RUN and ENCODE ChIP peaks Gene ontology (GO) enrichment analysis compared the molecular functions of genes annotated by CUT&RUN and ENCODE ChIP peaks in human K562 cells. The horizontal axis specified PTM profiles generated from CUT&RUN and ENCODE ChIP data; the vertical axis specified the names of enriched GO terms. Gene ratio was calculated as the number of genes characterised by CUT&RUN or ENCODE ChIP peaks over the number of expected genes from each GO category. Adjusted p-values were shown by colour gradients; ≤ 0.01 was considered statistically significant (pure red). (n=1 for CUT&RUN samples; n=2 for ENCODE ChIP samples)

peaks), (49) suggesting an area to be investigated.

Regarding the signal-to-noise performance, although a high-calcium/low-salt condition had been followed to minimise small fragments resulted from MNase digestion of premature release of DNA fragments, (46) high background was detected in most CUT&RUN samples, represented by <20% FRiP scores, as opposed to around 75% claimed by CUT&RUN's inventors. (34) However, CUT&RUN seemed to outperform ChIP-seq when profiling condensed genomic regions. It showed better noise performance than ENCODE ChIP in H3K9me1, H3K9me3, and H3K27me3, all of which are marks of heterochromatin. Indeed, heterochromatin was found to be resistant to sonication in ChIP-seq but did not affect MNase accessibility to linker DNA. (54,55)

Small fragments of 50-100 bp observed in the fragment size distribution graphs might not be pure background but rather associated with chromatin structures of higher orders because MNase could cleave DNA on nucleosome surfaces depending on the proximal three-dimensional structure of chromatins, as addition to cleavage at linker regions which produced mononucleosomes and oligonucleosomes. (56,57) To reduce noise, digestion time of MNase could be optimised since it was proved that longer digestion would result in higher yields of CUT&RUN fragments but also higher background. (46)

The benchmarking result showed that CUT&RUN generated distinct peak profiles of PTMs to ENCODE ChIP, which was consistent to a few CUT&RUN publications (38,57) but not others. (42,43) Fewer peaks were called in CUT&RUN as expected but they were not due to CUT&RUN's superior resolution (demonstrated by the low FRiP scores). Different peak numbers were not the only reason for low overlap rates of CUT&RUN and ENCODE ChIP. H4K20me1, which had comparable peaks from the two techniques, also showed limited extent of overlap. Moreover, CUT&RUN peaks were mostly wider, which means one peak tended to cover multiple ENCODE peaks, but the overlap remained low. It was therefore concluded that there were fundamental differences between peaks called in CUT&RUN and ENCODE ChIP.

The statement above was further confirmed by functional characterisation of the peaks. It was suspected that CUT&RUN peaks were biased towards transcriptionally inaccessible regions, causing lower proportions of promoters being captured. H3K4me1 and H3K4me3 were well-established marks for active transcription and were mostly detected at enhancers and transcription start sites while CUT&RUN detected enrichments in heterochromatin and repressed areas. (58,59) Profiles of another active mark H3K27ac were not identical to ENCODE either according to a CUT&Tag study. (49) This could be a problem common to enzyme-tethering methods, or it reflected that ENCODE data missed some relevant peaks in transcriptionally inaccessible areas due to ChIP-seq limitations. In contrast, CUT&RUN

profiles might be preferred over ENCODE ChIP in heterochromatic marks H3K9me1, H3K9me3, and H3K27me3. The ENCODE data of these three marks suffered from insufficient reading depth, rendering the results less reliable. (44,45) Given that CUT&RUN was invented using H3K27me3, the performance in active marks should be tested further. Comparing CUT&RUN data to other enzyme-tethering methods and carrying out functional studies on ENCODE-specific genes could verify whether ENCODE is indeed the gold standard in epigenetics.

One of the limitations of CUT&RUN is that as ChIP-seq, it heavily depends on antibodies of high-quality. Functional profiles of CUT&RUN and ENCODE ChIP were distinct for H3K79me2 because the antibody used was not compatible to CUT&RUN, as shown by the abnormal fragment size distribution. In contrast, H3K27me3, the positive control, retained the best performance of every metric as it was conducted in the optimised conditions (i.e., antibody products; dilutions) by CUT&RUN's inventors, though preference to heterochromatin needs to be accounted. PTMs with large discrepancy between CUT&RUN and ChIP-seq results are planned to be profiled again using optimised antibody conditions in the future.

The major limitation of this study was that the benchmarking was conducted using histone PTMs only; the performance of CUT&RUN in profiling transcription factors and chromatin-associated complexes remained unknown. Since these elements are involved in many biological processes, (22,60,61) benchmarking should be the immediate next step. Another limitation was the lack of biological replicates. With replicates, consensus peaks could be selected to generate more reliable profiles. Ultimately, based on the results presented, it would be useful to perform CUT&RUN on various human post-mortem samples but only for marks of heterochromatin. The combinational effect of multiple histone PTMs (histone code hypothesis) could be studied in human tissues by using antibodies that bind to several marks simultaneously. (62) Integration of genome sequencing data and epigenomic annotations could also provides insights into disease mechanisms.

In conclusion, this study successfully set up a CUT&RUN protocol compatible to more biological samples. The benchmarking against ENCODE ChIP showed that CUT&RUN generated distinct profiles to ENCODE ChIP, in terms of peaks called and genes annotated, though the extent varied by PTMs. CUT&RUN could be preferred to ChIP-seq when profiling transcriptionally inaccessible heterochromatin while further experiments and optimisations are required to verify its performance in transcriptionally active regions. It was hoped that this work could guide future studies that use CUT&RUN to assess cell type-specific epigenomic profiles for understanding mechanisms of pathogenic genetic variants contributing to disease and develop effective treatment.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Nathan Skene for providing useful suggestions throughout the project and my daily supervisor Dr Jose Torres-Pérez for guiding my literature research and teaching me the technique. I would also like to acknowledge Brian Schilder, Dr Eugene Duff, Sera Choi, Leyla Abbasova, and other members from Neurogenomics Lab for sharing their experience on computational analysis. Lastly, I would like to thank my dear family and friends for being with me during the stressful time. Without any one of you, I would not be able to complete this work and enjoy the process of science.

REFERENCES

1. Shi X, Zhai Z, Chen Y, Li J, Nordenskiöld L. Recent Advances in Investigating Functional Dynamics of Chromatin. Front Genet. 2022 Apr 5;13:870640.

2. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011 Mar;21(3):381–95.

3. Marzi SJ, Leung SK, Ribarska T, Hannon E, Smith AR, Pishva E, et al. A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. Nat Neurosci. 2018 Nov;21(11):1618–27.

4. Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, et al. H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nat Commun. 2021 Dec;12(1):719.

5. Rabaneda-Bueno R, Mena-Montes B, Torres-Castro S, Torres-Carrillo N, Torres-Carrillo NM. Advances in Genetics and Epigenetic Alterations in Alzheimer's Disease: A Notion for Therapeutic Treatment. Genes. 2021 Dec 8;12(12):1959.

6. Lin Y, Qiu T, Wei G, Que Y, Wang W, Kong Y, et al. Role of Histone Post-Translational Modifications in Inflammatory Diseases. Front Immunol. 2022 Feb 24;13:852272.

7. Kaur G, Rathod SSS, Ghoneim MM, Alshehri S, Ahmad J, Mishra A, et al. DNA Methylation: A Promising Approach in Management of Alzheimer's Disease and Other Neurodegenerative Disorders. Biology. 2022 Jan 7;11(1):90.

8. Varela RB, Cararo JH, Tye SJ, Carvalho AF, Valvassori SS, Fries GR, et al. Contributions of epigenetic inheritance to the predisposition of major psychiatric disorders: Theoretical framework, evidence, and implications. Neurosci Biobehav Rev. 2022 Apr;135:104579.

9. Dachet F, Brown JB, Valyi-Nagy T, Narayan KD, Serafini A, Boley N, et al. Selective time-dependent changes in activity and cell-specific gene expression in human postmortem brain. Sci Rep. 2021 Dec;11(1):6078.

10. Jarmasz JS, Stirton H, Davie JR, Del Bigio MR. DNA methylation and histone post-translational modification stability in post-mortem brain tissue. Clin Epigenetics. 2019 Dec;11(1):5.

11. Nott A, Schlachetzki JCM, Fixsen BR, Glass CK. Nuclei isolation of multiple brain cell types for omics interrogation. Nat Protoc. 2021 Mar;16(3):1629–46.

12. Zhang F, Lupski JR. Non-coding genetic variants in human disease: Figure 1. Hum Mol Genet. 2015 Oct 15;24(R1):R102–10.

13. Eating Disorders Working Group of the Psychiatric Genomics Consortium, International Headache Genetics Consortium, 23andMe Research Team, Bryois J, Skene NG, Hansen TF, et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. Nat Genet. 2020 May;52(5):482–93.

14. Andrews SJ, Fulton-Howard B, Goate A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. Lancet Neurol. 2020 Apr;19(4):326–35.

15. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019 Sep;51(9):1339–48.

16. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019 Mar;51(3):404–13.

17. Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet. 2020 Oct;52(10):1024–35.

18. Schueller E, Paiva I, Blanc F, Wang XL, Cassel JC, Boutillier AL, et al. Dysregulation of histone acetylation pathways in hippocampus and frontal cortex of Alzheimer's disease patients. Eur Neuropsychopharmacol. 2020 Apr;33:101–16.

19. Lee MY, Lee J, Hyeon SJ, Cho H, Hwang YJ, Shin J, et al. Epigenome signatures landscaped by histone H3K9me3 are associated with the synaptic dysfunction in Alzheimer's disease. Aging Cell [Internet]. 2020 Jun [cited 2021 Sep 5];19(6). Available from: https://onlinelibrary.wiley.com/doi/10.1111/acel.13153

20. Pan RY, He L, Zhang J, Liu X, Liao Y, Gao J, et al. Positive feedback regulation of microglial glucose metabolism by histone H4 lysine 12 lactylation in Alzheimer's disease. Cell Metab. 2022 Apr;34(4):634-648.e6.

21. Tan T, Shi P, Abbas M, Wang Y, Xu J, Chen Y, et al. Epigenetic modification regulates tumor progression and metastasis through EMT (Review). Int J Oncol. 2022 Apr 21;60(6):70.

22. Ilina A, Khavinson V, Linkova N, Petukhov M. Neuroepigenetic Mechanisms of Action of Ultrashort Peptides in Alzheimer's Disease. Int J Mol Sci. 2022 Apr 12;23(8):4259.

23. Balsalobre A, Drouin J. Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol [Internet]. 2022 Mar 9 [cited 2022 Apr 30]; Available from: https://www.nature.com/articles/s41580-022-00464-z

24. Lu Y, Qu W, Min B, Liu Z, Chen C, Zhang C. Modelling epigenetic regulation of gene expression in 12 human cell types reveals combinatorial patterns of cell-type-specific genes. IET Syst Biol. 2014 Jun;8(3):104–15.

25. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. 2012;13(9):R53.

26. Singh R, Lanchantin J, Robins G, Qi Y. DeepChrome: deep-learning for predicting gene expression from histone modifications. Bioinformatics. 2016 Sep 1;32(17):i639–48.

27. Alzheimer Disease Genetics Consortium (ADGC), The European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES), Kunkle BW, Grenier-Boley B, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar;51(3):414–30.

28. Skene NG, Grant SGN. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Front Neurosci [Internet]. 2016 Jan 27 [cited 2021 Sep 13];10. Available from: http://journal.frontiersin.org/Article/10.3389/fnins.2016.00016/abstract

29. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet. 2014 Nov;15(11):709–21.

30. Montanera KN, Anwar Z, Shibin SM, Rhee HS. Chapter Sixteen - ChIP-exo: A method to study chromatin structure and organization at near-nucleotide resolution. In: Tollefsbol T, editor. Epigenetics Methods [Internet]. Academic Press; 2020 [cited 2022 Apr 18]. p. 323–52. (Translational Epigenetics; vol. 18). Available from: https://www.sciencedirect.com/science/article/pii/B9780128194140000161

31. He Q, Johnston J, Zeitlinger J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nat Biotechnol. 2015 Apr;33(4):395–401.

32. Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. High-resolution mapping of transcription factor binding sites on native chromatin. Nat Methods. 2014 Feb;11(2):203–9.

33. Bachiller S, Jiménez-Ferrer I, Paulus A, Yang Y, Swanberg M, Deierborg T, et al. Microglia in Neurological Diseases: A Road Map to Brain-Disease Dependent-Inflammatory Response. Front Cell Neurosci. 2018 Dec 18;12:488.

34. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife. 2017 Jan 16;6:e21856.

35. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Dec;10(1):1930.

36. S Policicchio S, Jonathan P Davies not provided, Chioza B, Joe Burrage not provided, Jonathan Mill not provided, L Dempster E. Fluorescence-activated nuclei sorting (FANS) on human post-mortem cortex tissue enabling the isolation of distinct neural cell populations for multiple omic profiling v1 [Internet]. 2020 [cited 2022 Apr 21]. Available from: https://www.protocols.io/view/fluorescence-activated-nuclei-sorting-fans-on-huma-bmh2k38e

37. oYo-Link® MNase [Internet]. AlphaThera. [cited 2022 Apr 29]. Available from: https://alphathera.com/mnase-product/p/mnase

38. Liu N, Hargreaves VV, Zhu Q, Kurland JV, Hong J, Kim W, et al. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell. 2018 Apr;173(2):430-442.e17.

39. Trizzino M, Zucco A, Deliard S, Wang F, Barbieri E, Veglia F, et al. EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. Sci Adv. 2021 Jan 13;7(3):eaaz8836.

40. Liu Z, Liu Y, Dang L, Geng M, Sun Y, Lu Y, et al. Integrative Cistromic and Transcriptomic Analyses Identify CREB Target Genes in Cystic Renal Epithelial Cells. J Am Soc Nephrol. 2021 Jun 23;ASN.2021010101.

41. Arabzade A, Zhao Y, Varadharajan S, Chen HC, Jessa S, Rivas B, et al. ZFTA–RELA Dictates Oncogenic Transcriptional Programs to Drive Aggressive Supratentorial Ependymoma. Cancer Discov. 2021 Sep;11(9):2200–15.

42. Zheng X yu, Gehring M. Low-input chromatin profiling in Arabidopsis endosperm using CUT&RUN. Plant Reprod. 2019 Mar;32(1):63–75.

43. Wu SY, Lee CF, Lai HT, Yu CT, Lee JE, Zuo H, et al. Opposing Functions of BRD4 Isoforms in Breast Cancer. Mol Cell. 2020 Jun;78(6):1114-1132.e10.

44. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep;489(7414):57–74.

45. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018 Jan 4;46(D1):D794–801.

46. Meers MP, Bryson TD, Henikoff JG, Henikoff S. Improved CUT&RUN chromatin profiling tools. eLife. 2019 Jun 24;8:e46314.

47. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc. 2018 May;13(5):1006–19.

48. Zheng Y, Ahmad K, Henikoff S. CUT&Tag Data Processing and Analysis Tutorial [Internet]. [cited 2021 Jul 14]. Available from: https://yezhengstat.github.io/CUTTag_tutorial/#I_Introduction

49. Hu D, Abbasova L, Schilder BM, Nott A, Skene NG, Marzi SJ. CUT&Tag recovers up to half of ENCODE ChIP-seq peaks [Internet]. Genomics; 2022 Mar [cited 2022 Apr 6]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.03.30.486382

50. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Mol Cell Biol. 2016 Mar;36(5):662–7.

51. EpiCompare: QC and Benchmarking of Epigenetic Datasets [Internet]. neurogenomics; 2022 [cited 2022 Apr 6]. Available from: https://github.com/neurogenomics/EpiCompare

52. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019 Dec;9(1):9354.

53. Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, et al. H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nat Commun. 2021 Dec;12(1):719.

54. Becker JS, McCarthy RL, Sidoli S, Donahue G, Kaeding KE, He Z, et al. Genomic and Proteomic Resolution of Heterochromatin and its Restriction of Alternate Fate Genes. Mol Cell. 2017 Dec 21;68(6):1023-1037.e15.

55. Chereji RV, Bryson TD, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. Genome Biol. 2019 Dec;20(1):198.

56. Davis IJ, Pattenden SG. Chapter 1-3 - Chromatin Accessibility as a Strategy to Detect Changes Associated With Development, Disease, and Exposure and Susceptibility to Chemical Toxins.

In: McCullough SD, Dolinoy DC, editors. Toxicoepigenetics [Internet]. Academic Press; 2019 [cited 2022 Apr 26]. p. 85–103. Available from: https://www.sciencedirect.com/science/article/pii/B9780128124338000034

57. Liu N, Xu S, Yao Q, Zhu Q, Kai Y, Hsu JY, et al. Transcription factor competition at the γ-globin promoters controls hemoglobin switching. Nat Genet. 2021 Apr;53(4):511–20.

58. Rada-Iglesias A. Is H3K4me1 at enhancers correlative or causative? Nat Genet. 2018 Jan;50(1):4–5.

59. Howe FS, Fischl H, Murray SC, Mellor J. Is H3K4me3 instructive for transcription activation? BioEssays. 2017 Jan;39(1):e201600095.

60. Rai SN, Tiwari N, Singh P, Mishra D, Singh AK, Hooshmandi E, et al. Therapeutic Potential of Vital Transcription Factors in Alzheimer's and Parkinson's Disease With Particular Emphasis on Transcription Factor EB Mediated Autophagy. Front Neurosci. 2021 Dec 14;15:777347.

61. Phalke S, Rivera-Correa J, Jenkins D, Flores Castro D, Giannopoulou E, Pernis AB. Molecular mechanisms controlling age-associated B cells in autoimmunity*. Immunol Rev. 2022 May;307(1):79–100.

62. Prakash K, Fournier D. Evidence for the implication of the histone code in building the genome structure. Biosystems. 2018 Feb;164:49–59.

63. Wang S, Meyer DH, Schumacher B. H3K4me2 regulates the recovery of protein biosynthesis and homeostasis following DNA damage. Nat Struct Mol Biol. 2020 Dec;27(12):1165–77.

64. Gates LA, Shi J, Rohira AD, Feng Q, Zhu B, Bedford MT, et al. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J Biol Chem. 2017 Sep 1;292(35):14456–72.

65. Rivera C, Saavedra F, Alvarez F, Díaz-Celis C, Ugalde V, Li J, et al. Methylation of histone H3 lysine 9 occurs during translation. Nucleic Acids Res. 2015 Oct 30;43(19):9097–106.

66. Nicetto D, Zaret KS. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Current Opinion in Genetics & Development. 2019 Apr;55:1–10.

67. DiFiore JV, Ptacek TS, Wang Y, Li B, Simon JM, Strahl BD. Unique and Shared Roles for Histone H3K36 Methylation States in Transcription Regulation Functions. Cell Reports. 2020 Jun;31(10):107751.

68. Li T, Liu Q, Garza N, Kornblau S, Jin VX. Integrative analysis reveals functional and regulatory roles of H3K79me2 in mediating alternative splicing. Genome Med. 2018 Dec;10(1):30.

69. Shoaib M, Chen Q, Shi X, Nair N, Prasanna C, Yang R, et al. Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. Nat Commun. 2021 Dec;12(1):4800.

SUPPLEMENTARY MATERIALS

Supplementary Table 1: Description and potential roles in gene expression regulation of 11 histone PTMs profiled in CUT&RUN experiments.

Histone PTM	Description	Epigenetic implications in mammalian cells
H3K4me1	mono-methylation of the 4th lysine on histone protein H3	Marks for active and primed enhancers (58)
H3K4me2	di-methylation of the 4th lysine on histone protein H3	Shows enrichment in promoters of active or primed genes (63)
H3K4me3	tri-methylation of the 4th lysine on histone protein H3	Shows enrichment at transcription start sites, activating gene expression (59)
H3K9ac	acetylation of the 9th lysine on histone protein H3	Marks for active promoters (64)
H3K9me1	mono-methylation of the 9th lysine on histone protein H3	Marks for functional heterochromatin (65)
H3K9me3	tri-methylation of the 9th lysine on histone protein H3	Marks for repeat-rich heterochromatin at telomeres and centromeres (66)
H3K27ac	acetylation of the 27th lysine on histone protein H3	Marks for active promoters and enhancers (3)
H3K27me3	tri-methylation of the 27th lysine on histone protein H3	Marks for heterochromatin (4)
H3K36me3	tri-methylation of the 36th lysine on histone protein H3	Shows enrichment in gene bodies, i.e., exons, implicated in transcription activation (67)
H3K79me2	di-methylation of the 79th lysine on histone protein H3	Shows enrichment in intragenic regions (68)
H4K20me1	mono-methylation of the 20th lysine on histone protein H4	Marks for active transcription (69)

Supplementary Table 2: The table lists the recipes of CUT&RUN buffers. HEPES (Sigma, #H3375); KCl (Sigma, #P3911); CaCl2 (Fisher Scientific, #BP510); MnCl2 (Sigma, #203734); NaCl (Sigma, #S5150-1L); spermidine (Sigma, S0266); Roche Complete Protease Inhibitor EDTA-free tablets (Sigma; #5056489001); EDTA (Research Organics, #3002E); digitonin (Sigma, 300410); EGTA (Sigma, #E3889); RNAse A (Thermo Fisher, #EN0531); glycogen (Roche, 10901393001); Spike-in *E.coli* DNA (EpiCypher, #18-1401). Abbreviations: HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; EGTA, ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid.

Name	Components
Binding buffer	20 mM HEPES, pH 7.5; 10 mM KCl; 1 mM $CaCl_2$; 1 mM $MnCl_2$
Wash buffer	20 mM HEPES, pH 7.5; 150 mM NaCl; 0.5 mM spermidine; 1 Roche Complete Protease Inhibitor EDTA-free tablet
Antibody buffer	2 mM EDTA in wash buffer
High-calcium/low-salt rinse buffer	20 mM HEPES, pH 7.5; 0.5 mM spermidine; 0.05% digitonin
Incubation buffer	3.5 mM HEPES, pH 7.5; 10 mM $CaCl_2$; 0.05% digitonin
Stop buffer	170 mM NaCl; 20 mM EGTA; 0.05% digitonin; 50 µg/mL RNAse A; 25 µg/mL glycogen; 0.02 ng/mL spike-in *E.coli* DNA

Supplementary Table 3: ENCODE accessions for ChIP-seq datasets used in benchmarking Peak files of 11 histone PTMs were obtained from ENCODE4 (Bernstein Laboratory) in bed narrowPeak format. (44,45) The genome assembly is GRCh38. Biological replicates n=2.

Histone PTM	Method	Cell line	Data accession
H3K4me1	ChIP-seq	human K562 cell line	ENCFF183UQD
H3K4me2	ChIP-seq	human K562 cell line	ENCFF749KLQ
H3K4me3	ChIP-seq	human K562 cell line	ENCFF885FQN
H3K9ac	ChIP-seq	human K562 cell line	ENCFF891CHI
H3K9me1	ChIP-seq	human K562 cell line	ENCFF462AVD
H3K9me3	ChIP-seq	human K562 cell line	ENCFF963GZJ
H3K27ac	ChIP-seq	human K562 cell line	ENCFF864OSZ
H3K27me3	ChIP-seq	human K562 cell line	ENCFF881ONN
H3K36me3	ChIP-seq	human K562 cell line	ENCFF520LHY
H3K79me2	ChIP-seq	human K562 cell line	ENCFF945LQZ
H4K20me1	ChIP-seq	human K562 cell line	ENCFF909RKY