Genome-wide mapping of protein-DNA interaction dynamics


Steve Henikoff, HHMI and Fred Hutchinson Cancer Research Center
ENCODE 2019, Seattle, July 9, 2019
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
served miRNAs (miRNAs) that have been implicated in development. Although miRNA genes are not likely to account for more than a miniscule amount of the transcriptome, it is much greater problem. These trans not be discounted as spurious proc far as they were repeatedly detect different methods, including tilin; rays, ditag sequencing and 5′-RACE. Still, the lowest-level transcripts might represent transcriptional noise resulting from l meraase transitting through them. However, we know that at least some of the transcripts found between and within protein-coding genes function in gene regulation, such as the conerved microRNAs (miRNAs) that have been implicated in development. Although miRNA genes are not likely to account for more than a miniscule amount of the transcriptome, it is served miRNAs (miRNAs) that have been implicated in development. Although miRNA genes are not likely to account for more than a miniscule amount of the transcriptome, it is
Chromatin Immunoprecipitation (ChIP)

ENCODE datasets

ChIP-seq: 58% of the total

# Google Scholar publications

"ChIP-seq"
Chromatin Immunoprecipitation (ChIP)

- DNA with interacting proteins
- DNA-proteins complexes
- DNA sequencing

Limitations

- Low efficiency (millions of cells)
- High background (ten millions of reads)
- Low resolution (hundreds of bp)
- Most antibodies fail

As part of the US National Institutes of Health (NIH) Protein Capture Reagents Program (PCRP), we have generated a collection of 1,406 highly validated immunoprecipitation- and/or immunoblotting-grade mouse monoclonal antibodies (mAbs) to 737 human transcription factors, using an integrated production and validation pipeline. We used HuProt human protein microarrays as a primary validation tool to identify mAbs with high specificity for their cognate targets. We further validated PCRP mAbs by means of multiple experimental applications, including immunoprecipitation, immunoblotting, chromatin immunoprecipitation followed by sequencing (ChIP-seq), and immunohistochemistry.

305 were tested by ChIP-seq:
50 passed by ENCODE standards
Chromatin Immunoprecipitation (ChIP)

Limitations

Low efficiency (millions of cells)
High background (ten millions of reads)
Low resolution (hundreds of bp)
Most antibodies fail

Artifacts

31% of modENCODE profiles >20% overlap with “Phantom Peaks”
Chromatin profiling by enzyme tethering

**ChEC**

---

**DamID** – van Steensel & Henikoff (2000) *Nat Biotech*

**ChIC and ChEC** – Schmid *et al.* (2004) *Mol Cell*

**ChEC-Seq** – Zentner *et al.* (2015) *Nat Commun*
CUT&RUN
(Cleavage Under Targets & Release Using Nuclease)

ConA Bead

TF complex diffuses out

TF-specific antibody

Protein A-MNase

Ca++ @ 0°C

Stop + spike-in

Supernatant

Paired-end sequencing

Skene & Henikoff (2017) eLife
Human K562 cells

H3K27me3 (CST#9733 mRabAb)

chr7:125,975,280-135,942,588

ENCODE
56 million reads

ENCODE
7.5 million reads

CUT&RUN (# of cells)

CUT&RUN provides base-pair resolution maps

CTCF
CUT&RUN ≤120 bp fragment ends

Human K562 cells

44 bp

5 sec 15 sec 45 sec 3 min 9 min 27 min

Normalized Counts (% of max signal)

Distance from site (bp)

CTCF

ENCODE

10 million reads each

200 bp

CUT&RUN (≤120 bp)

ChIP-exo (forward)

ChIP-exo (reverse)

Skene & Henikoff (2017) eLife

Pete Skene
High-throughput CUT&RUN of patient-derived xenografts

Thaw flash-frozen mouse brain tumors

Array

CUT&RUN + barcoded adapters

PCR, clean-up on robot, QC, mix up to 48 samples

Paired-end 25x25 sequencing

Mix with ConA magnetic beads

~$50 for ~6 million mapped human fragments per sample

Patient A (45 mg)  Patient B (15 mg)

Disaggregate

Mix with ConA magnetic beads

~$50 for ~6 million mapped human fragments per sample

Derek Janssens & Jay Sarthy
CUT&RUN profiling of pediatric glioma-derived PDXs

Average ~6 million mapped fragments per sample

H3K27me3
H3K4me2

CDKN2A (p16) CDKN2B (p15)

Patient A
Cells 1
Cells 2
PDX 1
PDX 2

Patient B
Cells 1
Cells 2
PDX 1
PDX 2

Janssens et al. (2018) Epigenetics Chromatin
AutoCUT&RUN datasets on the 4DN Data Portal

4D Nucleome Data Portal
A platform to search, visualize, and download nucleomics data.

https://data.4dnucleome.org
CUT&RUN with salt fractionation

Specific antibody

Protein A-MNase

Ca^{++} @ 0°C

Stop + spike-in

Supernatant

Beads/Nuclei

Pellet

High NaCl

Low NaCl

Prepare sequencing libraries

Thakur & Henikoff (2018) *Genes Dev*
CUT&RUN with ChIP

Specific antibody

Protein A-MNase

Ca++ @ 0°C

Stop + spike-in

Supernatant (= ChIP Input)

ChIP

Sequencing libraries + Protein A beads

+ FLAG peptide + 2nd antibody

Brahma & Henikoff (2019) Molecular Cell

Sandipan Brahma
**H2A.Z CUT&RUN ChIP**

Budding yeast
chrI:83,173-130,676

Nucleosomes*  
H2A.Z ChIP-seq**  
H2A.Z CUT&RUN

5 kb

Brahma & Henikoff (2019) *Molecular Cell*  
*Chereji et al. 2018; **Weiner et al. 2015*
How do TFs find their sites in chromatin?

- ‘Site-exposure’
- Nucleosome occludes TF binding site
- ‘Pioneering’

Transient unwrapping

Mike Meers
A CUT&RUN assay for pioneering activity
Sparse Enrichment Analysis for CUT&RUN (SEACR)

<table>
<thead>
<tr>
<th>Target IgG</th>
<th>Meets threshold</th>
<th>Fails threshold</th>
<th>IgG overlap (filtered out)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total signal threshold</th>
<th>Fraction above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Threshold = Highest Target/IgG

SEACR (AUC only)

Sox2

<table>
<thead>
<tr>
<th>hESC</th>
<th>17050</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoderm</td>
<td>1</td>
<td>6215</td>
</tr>
</tbody>
</table>

SEACR (Union)

Sox2

<table>
<thead>
<tr>
<th>hESC</th>
<th>19690</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoderm</td>
<td>4</td>
<td>7327</td>
</tr>
</tbody>
</table>

MACS2

<table>
<thead>
<tr>
<th>Sox2</th>
<th>FoxA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>hESC</td>
<td>10728</td>
</tr>
<tr>
<td>Endoderm</td>
<td>695</td>
</tr>
</tbody>
</table>

HOMER

<table>
<thead>
<tr>
<th>Sox2</th>
<th>FoxA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>hESC</td>
<td>7095</td>
</tr>
<tr>
<td>Endoderm</td>
<td>87</td>
</tr>
</tbody>
</table>

SEACR also calls “broad domains”

Chr2:176,830,000-177,180,000

H3K27me3

Meers, et al. bioRxiv, Epigenetics & Chromatin, in press
SEACR web server (coming soon)

Input fields for bedgraph files

Normalization option

Relaxed or stringent options

We gratefully acknowledge our funders:
Finding TF binding sites in CUT&RUN data

Target

IgG

Target

IgG

Target

IgG

IgG overlap (filtered out)

Meets threshold

Fails threshold

Target

IgG

= Contiguous signal blocks

= Contiguous signal blocks

meeting threshold

≤120 bp

>150 bp

Enriched Chromatin Occupancy (EChO)

Fragment size (bp)

CTCF

Small fragment foci

Large fragment foci

Meers et al. Molecular Cell (2019)
FoxA2 pioneers 29% of binding sites from Day 3-5

CTCF sites

FoxA2 sites

Meers et al. Molecular Cell (2019)
CUT&RUN used in publications from others
(pA-MNase sent on request to ~600 labs)


+ 10 preprints using CUT&RUN on bioRxiv
CUT&Tag
(Cleavage Under Targets & Tagmentation)

1 day from live cells to sequencing-ready libraries
CUT&Tag: High signal-to-noise

Human K562 cells

Genes

CUT&Tag
H3K4me2

CUT&RUN
H3K4me2

ChIP-seq
H3K4me2

ATAC-seq

Kaya-Okur et al. (2019) Nature Commun
**CUT&Tag: Highest signal-to-noise**

![Graph showing signal-to-noise comparison between different techniques.](attachment:image.png)

- **H3K4me2 or ATAC-seq**
  - CUT&Tag: Highest signal-to-noise

**K562 cells**

- # Fragments sampled
- Fraction of reads in peaks (FRiP)

*GSM733651, **GSM2695562

Kaya-Okur et al. (2019) *Nature Commun*
CUT&Tag: Efficient for low cell numbers

Human K562 cells  H3K27me3 (CST#9733 mRabAb)  chr7:132,600,000-133,100,000

Genes

ENCODE (56M)

CUT&Tag (4.6M) 6,000 cells

CUT&Tag (4.0M) 600 cells

CUT&Tag (2.8M) 60 cells

500 kb

Human/E.coli fragments vs. # Cells

$R^2 = 0.98$

=> Use carry-over E. coli DNA for calibration (replaces ‘spike-in’)

Meers et al. (2019) eLife
Kaya-Okur et al. (2019) Nature Commun
CUT&Tag: Efficient for low cell numbers

Human K562 cells  H3K27me3 (CST#9733 mRabAb)  chr7:132,600,000-133,100,000

Genes

ENCODE (56M)

CUT&Tag (4.6M) 6,000 cells

CUT&Tag (4.0M) 600 cells

CUT&Tag (2.8M) 60 cells

Aggregate scCUT&Tag (908 single cells)

CUT&Tag: Efficient for low cell numbers

CUT&Tag (Takara iCELL8 platform)

Kaya-Okur et al. (2019) Nature Commun
Single-cell CUT&Tag maps silent domains

Human K562 cells  H3K27me3 (CST#9733 mRabAb)  chr7:132,600,000-133,100,000

Genes

ENCODE (56M)

CUT&Tag (4.6M) 6,000 cells

CUT&Tag (4.0M) 600 cells

CUT&Tag (2.8M) 60 cells

Aggregate scCUT&Tag (908 single cells)

CUT&Tag 908 single cells

Kaya-Okur et al. (2019) Nature Commun
Single-cell CUT&Tag maps ‘active’ nucleosomes

Human K562 cells

ATAC-seq
15.5M fragments

Bulk CUT&Tag
~2000 cells

Aggregate scCUT&Tag
808 single cells

CUT&Tag 808 single cells

Kaya-Okur et al. (2019) Nature Commun
Single-cell CUT&Tag distinguishes cell types

H3K27me3 profiles of: 479 H1 ES cells 479 K562 cells

Cluster 1
100%

Cluster 2
84%

$k$-means clustering 479 cells ea.

Kaya-Okur et al. (2019) Nature Commun
number of cells with a single hydrogel bead, that is, typically 5,000 cells out of 15,000 starting cells (Supplementary Fig. 5c). To confirm that barcodes were unique to a single cell, we performed an experiment with a mixture of mouse and human cell lines, which showed that 97% of the barcodes were unambiguously assigned to a single species (Supplementary Fig. 6), which is consistent with the percentage of occupied droplets containing single cells (95%; see Methods).

**Detection of single-cell chromatin landscapes in vitro.** Next, we validated the efficiency and accuracy of the scChIP-seq procedure pooled and immunoprecipitated. For H3K4me3, we achieved an average coverage of 1,630 and 1,823 reads per million mapped reads (RPM) in the scChIP-seq replicates (see Methods; Supplementary Fig. 7a–c, respectively), and a high correlation across replicates (see Methods; Supplementary Fig. 7d,e). For both single-cell chromatin profiles identified by consensus clustering two standard deviations from each cell line (Fig. 1b and Supplementary Videos 1 and 2). The microfluidics workflow includes live mor
Single-cell CUT&Tag versus Single-cell ChIP-seq

H3K27me3 (CST mAB #9733)

Number of unique fragments per cell vs. Number of single human cells

CUT&Tag

K562 6858

ChIP-seq

Jurkat 863
Ramos 1380
### Single-cell CUT&Tag versus Single-cell ChIP-seq

<table>
<thead>
<tr>
<th>Technique</th>
<th>Replicates</th>
<th>Cell Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENCODE/input</td>
<td>1x</td>
<td>K562</td>
</tr>
<tr>
<td>ENCODE/ChIP</td>
<td>15x</td>
<td>K562</td>
</tr>
<tr>
<td>CUT&amp;Tag bulk</td>
<td>65x</td>
<td>K562</td>
</tr>
<tr>
<td>scCUT&amp;Tag</td>
<td>23x</td>
<td>K562</td>
</tr>
<tr>
<td>scChIP-seq</td>
<td>2.2x</td>
<td>Jurkat</td>
</tr>
<tr>
<td>scChIP-seq</td>
<td>3.1x</td>
<td>Ramos</td>
</tr>
</tbody>
</table>

**H3K27me3**

<table>
<thead>
<tr>
<th>Region</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr7:132,467,734-133,789,152</td>
<td>3.8x</td>
</tr>
</tbody>
</table>

**Genes:**
- CHCHD3
- EXOC4
Antibody variability remains a challenge

H3K27me3

An assessment of histone-modification antibody quality

Thea A Egelhofer1,16, Aki Minoda2,3,16, Sarit Klugman4,5,16, Kyungjoon Lee6, Paulina Kolasinska-Zwierz7,8, Artyom A Alekseyenko9,10, Ming-Sin Cheung9,8, Daniel S Day9, Sarah Gadel11, Andrey A Gorchakov9,10, Tingting Gu11, Peter V Kharchenko12, Samantha Kuan4,5, Israelabetorre13, Daniela Linder-Basso13, Ying Liu14, Queminh Ngo4,5, Marc Perry11, Andreas Rechtsteiner1, Nicole C Riddle11, Yuri B Schwartz12, Gregory A Shanower12, Anne Vielle7,8, Julie Ahringer7,8, Sarah C R Elgin11, Mitzi I Kuroda9,10, Vincenzo Pirrotta12, Bing Rent3, Susan Strome1, Peter J Park4, Gary H Karpen2,3, R David Hawkins5,6 & Jason D Lieb7,15

We have tested the specificity and utility of more than 200 antibodies raised against 57 different histone modifications in Drosophila melanogaster, Caenorhabditis elegans and human cells. Although most antibodies performed well, more than 25% failed specificity tests by dot blot or western blot. Among specific antibodies, more than 20% failed in chromatin immunoprecipitation experiments. We advise rigorous testing of histone-modification antibodies before use, and we provide a website for posting new test results (http://combio.med.harvard.edu/antibodies/).

http://combio.med.harvard.edu/antibodies/
Antibody variability remains a challenge

**Figure 3. Bar graphs illustrating the frequency of datasets related to a defined antibody vendor reads (TMRs) interval.** Datasets for each target were categorized on the basis of their sequencing mapped million reads; intervals of 10 million and a last category from 100–500 million), as well as on their quality grade (Y-axis: from “A” to “D”, defined on the basis of a read count intensity dispersion (dRCI) threshold criterion of 2.5%). The fraction of datasets related to a given vendor per TMRs interval. A to D. Frequency bar graphs correspond to H3K27ac and H3K4me3, respectively.
Antibody specificity may be uncertain

seminal insights into histone post-translational modification (PTM) regulation and distribution (Barski et al., 2007; Guenther et al., 2007; Heintzman et al., 2007, 2009; Mikkelsen et al., 2007; Rada-Iglesias et al., 2011; Santos-Rosa et al., 2002; Vermeulen et al., 2007; Wysocka et al., 2006; Leroy et al., 2013; Bieberstein et al., 2012; Sims et al., 2007).

To this end, we assessed the specificity of 52 commercial “ChIP grade” antibodies using histone peptide microarrays and ICeChIP (Figures S2 and S3). In this incubated with slide-immobilized pe

HMD ≤ 0 (7,666 Refseq genes). Vertical axis represents position in sorted H3K4me3 signal. Green oval shows ICeChIP methylform specificity for each antibody, orange circle with E indicates antibody validated to ENCODE standards, purple circle with M indicates monoclonality, and red oval shows percentage of peaks in each ENCODE dataset found in all three of the other ENCODE datasets.

Average SD of on-target signal. (B) Abbreviation codes, specificities in ICeChIP and peptide arrays, and target IP enrichments for antibodies referred to in the main text. Values.
CUT&RUN: Targeted in situ genome-wide profiling with high efficiency for low cell numbers

Derek Janssens¹, Steven Henikoff²
¹Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, Washington, USA 98109...

CUT&Tag

Bench top CUT&Tag

Hatice Kaya-Okur¹, Steven Henikoff¹
¹Fred Hutchinson Cancer Research Center

Human Cell Atlas Method Development Community

CONTACT

Derek Janssens

Steven Henikoff

**Tagmentation (1 hr)**

DNA extraction (1 hr)

30 To stop tagmentation and solubilize DNA fragments, add 10 μL 0.5M EDTA, 3 μL 10% SDS and 2.5 μL 20 mg/mL Proteinase K to each sample.

31 Mix by full-speed vortexing ~2 s, and incubate 1 hr 50 °C or 37 °C overnight to digest. 01:00:00

It is typical for the beads to form a large clump during incubation owing to the viscoelasticity of DNA. However, for abundant genome-wide epitopes, large-scale fragmentation of the genome will normally result in reduced clumping and release of beads into suspension, turning the liquid brownish relative to negative controls.

Step 30) Left: H3K27me3; Right: IgG.

Total views across all versions: 26,345

Total views across all versions: 15,438
Acknowledgments (posters)

Mike Meers #11 (Pioneer factors)
Hatice Kaya-Okur #28 (CUT&Tag)
Jay Sarthy #29 (H3 variants in cancer)
Sandipan Brahma #26 (CUT&RUN.ChIP)
Derek Janssens #27 (AutoCUT&RUN)

Email for pA-Tn5: cutnrun@fredhutch.org (sent to ~400 labs)