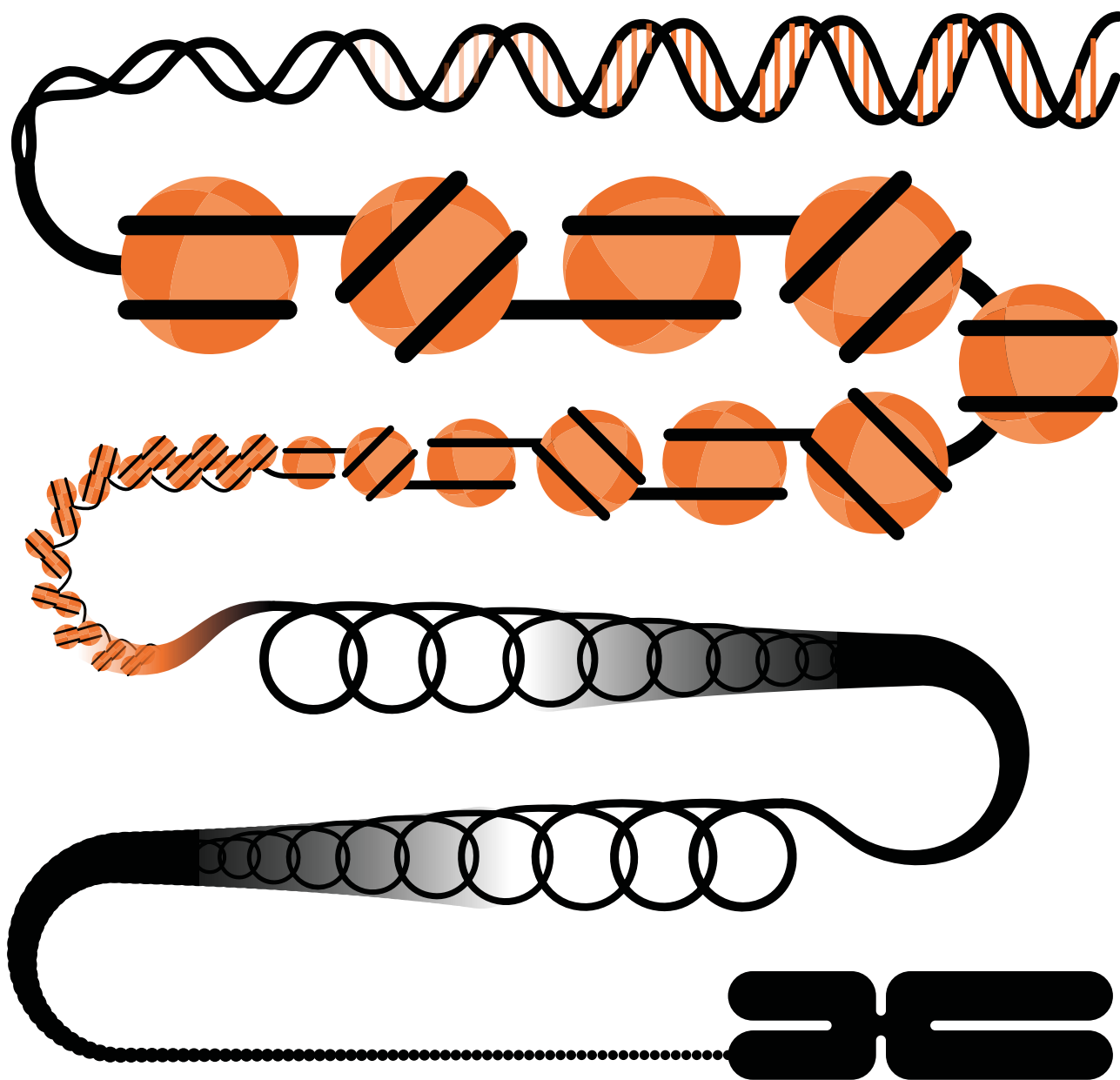


Epigenetics application guide



Contents

Introduction	5
- Why is epigenetics important?	5
- How to study epigenetics	5
Chromatin accessibility and architecture	6
- Methods to study DNA accessibility and nucleosome positioning	7
- DNase-seq	7
- MNase-seq	7
- ATAC-seq	8
- Chromosome conformation techniques	9
- Chromatin conformation capture (3C)	9
- Circularized chromosome conformation capture (4C)	10
- Carbon copy chromosome conformation capture (5C)	10
- Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)	10
- ChIP-loop	11
- Hi-C	11
- Capture-C	11
- References	12
Histone modifications	13
- Histone modifications in detail	14
- Acetylation	14
- Methylation	14
- Phosphorylation	15
- Ubiquitylation	15
- Studying histone modifications by ChIP	16
- Histone modifying enzymes: writers and erasers	17
- Characterizing histone methylation pathways	18
- Characterizing demethylase activity	18
- Characterizing histone acetylation pathways	18
- Characterizing histone deacetylase activity	19
- Inhibiting writers and erasers	19
- Histone modification readers/translators	19
- Effector domains recognize specific histone modifications	20
- Multivalency enables histone code complexity	20
- References	21
Studying epigenetics with ChIP	22
- Applications of ChIP	22
- Combined analysis	22
- ChIP expands the scope and precision of epigenetic research	23

- Protocol overview: how ChIP works.....	24
- Sample Preparation: X-ChIP vs N-ChIP	26
- Chromatin fragmentation differences.....	26
- Antibody selection	27
- Controls.....	28
- Sample controls.....	28
- Antibody controls	28
- Quantitative PCR controls.....	28
- Protocol optimization.....	28
- ChIP with low cell numbers	30
- ChIP from tissue	31
- Troubleshooting.....	31
- ChIP readout.....	33
- qPCR	33
- ChIP-on-Chip.....	33
- ChIP-seq.....	33
- Data analysis.....	34
- References.....	35
Chromatin profiling using ChIC / CUT&RUN.....	36
- Applications of ChIC/CUT&RUN.....	37
- Chromatin Modifications.....	37
- Transcription factors and chromatin-associated complexes.....	37
- Comparison of ChIC/CUT&RUN with ChIP-seq	38
- ChIC/CUT&RUN protocol overview.....	39
1. Nuclei extraction and binding to beads.....	40
2. Permeabilization and antibody binding	40
3. MNase binding and cleavage of target sequences	40
4. DNA recovery	40
5. Library preparation and sequencing.....	41
- Sample preparation	42
- Antibody selection	42
- Controls.....	42
- Optimization	43
- Permeabilizations.....	43
- Antibody concentration.....	43
- Duration of antibody incubation.....	43
- Use of a secondary antibody.....	43
- Digestion time.....	43
- Step-by-step protocol links	43
- FAQs.....	44

- References	46
DNA methylation and demethylation	47
- Bisulfite sequencing	48
- Bisulfite-based applications	49
- Bisulfite conversion: technical considerations	50
- DNA immunoprecipitation (DIP)	51
- DIP-based applications	52
- DIP: technical considerations	52
- Alternate methods to capture 5hmC, 5fC, and 5caC	53
- 5hmC mapping	53
- 5fC/5caC mapping	53
- Comparison of DNA modification sequencing methods	54
- Liquid chromatography tandem-mass spectrometry (LC/MS-MS)	56
- LC/MS-MS: technical considerations	56
- DNA modification IHC/ICC	56
- DNA modification IHC/ICC: technical considerations	57
- Methyl binding domain proteins (MBDs)	57
- MBDs: Technical considerations	58
- Novel DNA modifications	58
- Novel DNA modifications: technical considerations	58
- References	59
RNA modifications	61
- RNA immunoprecipitation (RIP)	62
- RIP: technical considerations	62
- CLIP	63
- CLIP: technical considerations	63
- miCLIP	64
- miCLIP: technical considerations	65
- Liquid chromatography tandem-mass spectrometry (LC/MS-MS)	67
- LC/MS-MS: technical considerations	67
- RNA modification control experiments	66
- RNase treatment	66
- DNase treatment	66
- Competition assays	66
- Dot blot	67
- RIP-MS	67
- References	68

Introduction

Why is epigenetics important?

Completion of the human genome project and advances in next-generation sequencing technologies have revealed that genomic DNA has much less control over biological processes and disease states than initially thought. Instead, epigenetic factors dictate how DNA is translated, tightly regulating DNA structure to control which genes to express at what times.

Many of these epigenetic factors work together to orchestrate essential cellular programs, from developmental processes to cell death pathways. Dysfunction of any of these factors can upset genomic regulation, causing cellular processes to go awry, resulting in disease from cancers and autoimmune disorders to neurological conditions, infertility, and everything in between.

To understand any aspect of biology or disease, it is essential to examine epigenetic factors that may contribute. This guide provides an overview of epigenetic regulation and how to study these critical players.

How to study epigenetics

Epigenetic regulation occurs on many interacting levels, and it is essential to examine all of these levels in parallel to understand epigenetic contributions to biological processes. Tackling epigenetic studies from multiple angles with redundancy is key to ensuring accurate results.

Here we focus on five essential aspects of epigenetic regulation.

1. Chromatin architecture and accessibility

Genomic DNA is packaged and organized into chromatin to fit into the nucleus of each cell. Some regions of the genome are tightly packaged and inaccessible for transcription, resulting in gene silencing. Other areas are in an open conformation, allowing transcription factors and machinery to bind for active gene transcription. Understanding which genomic regions are active vs inactive across the genome in different cellular or disease states can help to identify critical pathways and associated genes.

2. Histone modifications

Histones are proteins responsible for packaging DNA. A variety of mechanisms can modify histones, including acetylation, methylation, and phosphorylation, to control their interactions with DNA and therefore DNA structure and gene activation. Examining histone modifications, and the activity of enzymes that control these modifications can reveal mechanisms of epigenetic regulation and dysregulation at specific gene sites or across the genome at large.

3. DNA binding proteins

Many different types of proteins bind to DNA to either directly or indirectly to regulate chromatin conformation and gene transcription. Identifying the presence or absence of such proteins in specific regions or across the genome can provide help build a complete picture of epigenetic regulation and dysregulation, as well as point to particular players and pathways involved. We can study these aspects of epigenetic regulation with chromatin immunoprecipitation (ChIP).

4. DNA modifications

Throughout the DNA sequence, many chemical modifications exist. The most well-studied of these is 5-methylcytosine (5mC), a modification most commonly recognized as a stable, repressive regulator of gene expression. There is a large body of research that shows 5mC and other chemical modifications within DNA to have epigenetic roles in gene regulation. Identifying these marks and their function in biology is a fascinating area of epigenetics right now.

5. RNA modifications

Scientists are continually discovering new RNA modifications and new functions for existing modifications. Many RNA modifications thought only to exist in bacteria are being found in eukaryotic cells while others presumed only to exist on certain RNA, species such as tRNAs, are now being found to have crucial roles in mammalian mRNA translation. RNA modifications are very hot right now, and there is still a lot to explore in this field of research.

Chromatin accessibility and architecture

The genome is efficiently packaged into the nucleus. DNA is wrapped around histones to form a nucleosome, comprised of 147 base pairs of DNA and eight core histone proteins. Nucleosomes are strung together like beads on a string and packaged into higher order chromatin architecture (Figure 1).

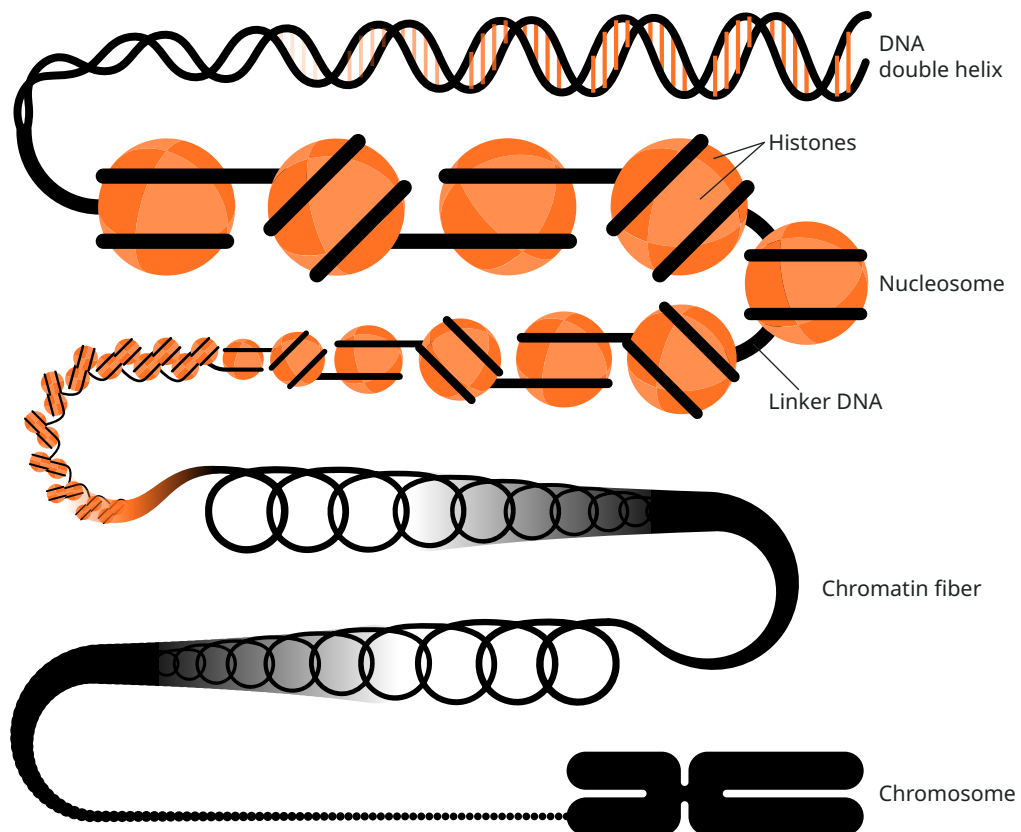


Figure 1. Chromatin structure. DNA winds around nucleosomes to form chromatin fiber and then chromosomes

DNA that is tightly bound in nucleosomes or compacted into higher order heterochromatin is inaccessible, preventing the binding of transcription factors, transcriptional machinery, and other DNA binding proteins, resulting in gene silencing. Meanwhile “linker” DNA and open euchromatin architecture are accessible to binding, allowing for active gene transcription.

Chromatin is actively and dynamically remodeled to alter gene expression and cellular programming, for example, during different developmental stages or in response to particular stimuli. Large genomic regions may be silenced or activated, or nucleosomes may be unraveled to access specific genes and DNA sequences.

Examining the chromatin structure and nucleosome positioning reveals epigenetic programs and mechanisms involved in specific cellular processes and disease states.

Methods to study DNA accessibility and nucleosome positioning

Surveying the genome for exposed regions accessible for active transcription vs those bound tightly into heterochromatin can be an essential first step to understanding the relationship between chromatin structure and function in different contexts. To take a snapshot of genomic architecture, researchers may use one of three methods: DNase-seq, MNase-seq, or ATAC-seq.

DNase-seq and ATAC-seq map exposed regions of DNA, whereas MNase-seq maps regions protected by nucleosomes. It is important to keep in mind that these methods provide snapshots of a dynamic process, often averaged across thousands of cells. If a particular region is dynamically changing, or different between cells within the population, the data may seem conflicting between methods. Some single cell analysis methods are evolving to resolve these challenges.

DNase-seq

DNase-seq uses DNase to digest exposed regions of the genome, whereas nucleosome-bound DNA is protected from DNase digestion. The small fragments generated by DNase digestion are then sequenced and mapped to the genome to identify regions of active transcription.

Advantages

- Most established and practiced method
- DNase cutting bias is well-understood
- Can be adapted to inversely examine protected genomic regions, called DNase footprinting, to identify transcription factor and nucleosome binding sites. However, it is important to use naked DNA as a control for such experiments as DNase I cutting bias can lead to false conclusions
- Possible to adapt for single cell analysis

Disadvantages

- Technically difficult to master, especially in optimizing digestion conditions for a given cell type/number
- Requires millions of cells, and may be challenging for analysis of rare patient samples

MNase-seq

In contrast to DNase-seq, MNase-seq uses micrococcal nuclease (MNase), from *Staphylococcus aureus*, to digest exposed genomic regions. Protected DNA bound to nucleosomes is then recovered and sequenced.

Advantages

- Common and well established in many cell types of many species, from yeast to humans, with some standardization of digestion and data analysis
- Can be used in combinations with chromatin immunoprecipitation (ChIP-seq), to study regulatory factors that bind to nucleosomes
- Can be adapted to generate base-pair resolution mapping
- Can be adapted to examine nucleosome positioning and DNA methylation state in nucleosome occupancy and methylome sequencing (NOMe-seq)

Disadvantages

- Requires large numbers of cells (10–20 million)
- Sequence-specific bias in the digestion of AT-rich regions (although most enzymes used in chromatin accessibility assays exhibit similar biases), but also unknown biases that may skew results
- Single-cell analysis not possible yet

ATAC-seq

Established in 2013, the assay for transposase-accessible chromatin (ATAC)-seq inserts sequencing adapters directly into accessible DNA using the enzyme Tn5 transposase. The DNA between the adapters is then amplified with qPCR and sequenced.

ATAC-seq uses a mutant hyperactive Tn5 transposase that is preloaded with DNA adaptors to simultaneously fragment and tag the genome with sequence adaptors (a process called transposition). PCR amplification and NGS follow this fragmentation and tagging. The frequency of sequences in a region correlates with open chromatin conformation.

Advantages

- Easiest method: no sonication, phenol-chloroform extraction, antibodies (ChIP-seq), or enzymatic digestion (DNase-seq, MNase-seq) are required
- Fastest method: <3 hours compared to up to a 4-day protocol
- Best signal-to-noise ratio
- Only 50,000 or fewer cells required (500–50,000 recommended)
- Single-nucleotide resolution possible
- Single-cell analysis is possible with adapted protocols utilizing flow cytometry/microfluidics

Disadvantages

- More expensive, requiring a kit from Illumina (Nextera DNA Library Preparation Kit)
- Least established method requires optimization of cell number and lysis conditions for specific cell types, tissues, and organisms to achieve ideal fragment distributions.
- Cell number defines the quality of the data, with too few cells or too many cells resulting in over or under-transposition that can skew results.

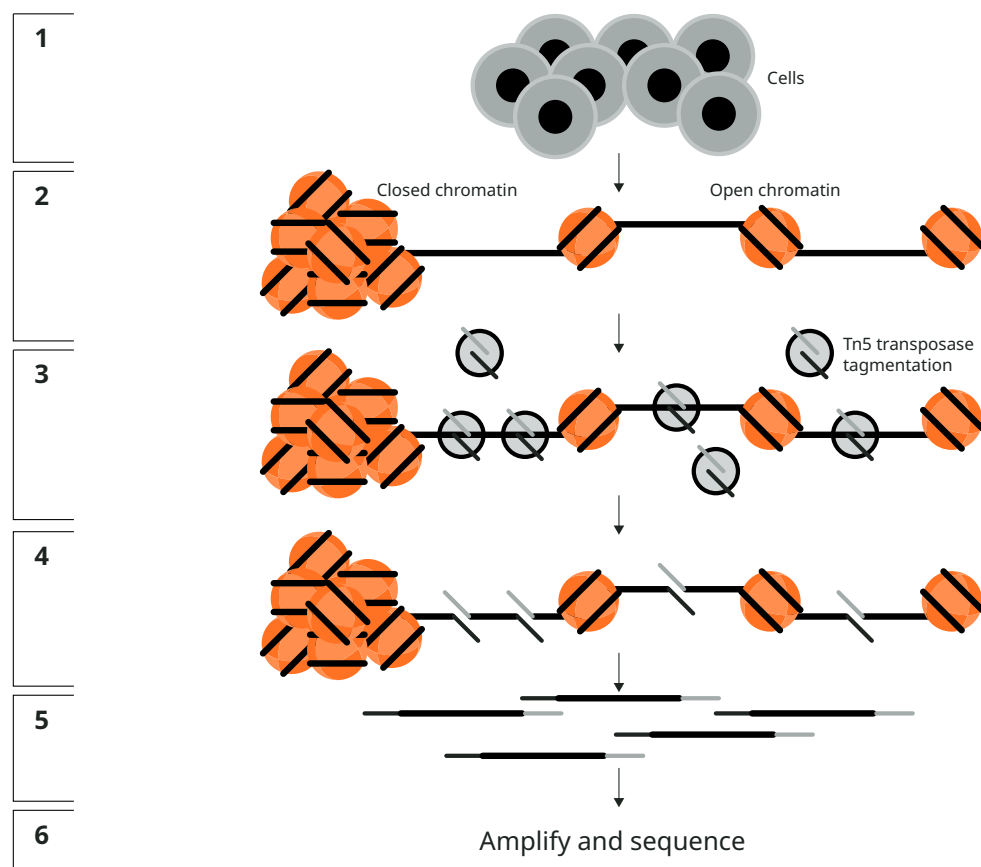


Figure 2. ATAC-seq protocol. Our step by step guide to ATAC seq can be found at www.abcam.com/atac

1. Harvest 50,000 cells. An accurate cell number is key to the success of the experiment.
2. Lyse cells to generate a crude nuclei preparation.
3. Tn5 tagmentation simultaneously fragments the genome and tags the resulting DNA with sequencing adapters.
4. Purify the fragmented and tagged DNA.
5. PCR amplify and purify the amplified DNA.
6. Sequence the library and correlate reads with open and closed chromatin.

Chromosome conformation techniques

We can assess the three-dimensional chromatin architecture with chromatin contact mapping to reveal physical interactions between distant genomic regions. This type of mapping is made possible by the advent of chromatin conformation capture (3C) and subsequent methods developed based on this approach. Each of these approaches has particular strengths for particular applications, but selecting a method for a specific purpose can be challenging due to the sheer variety of methodologies.

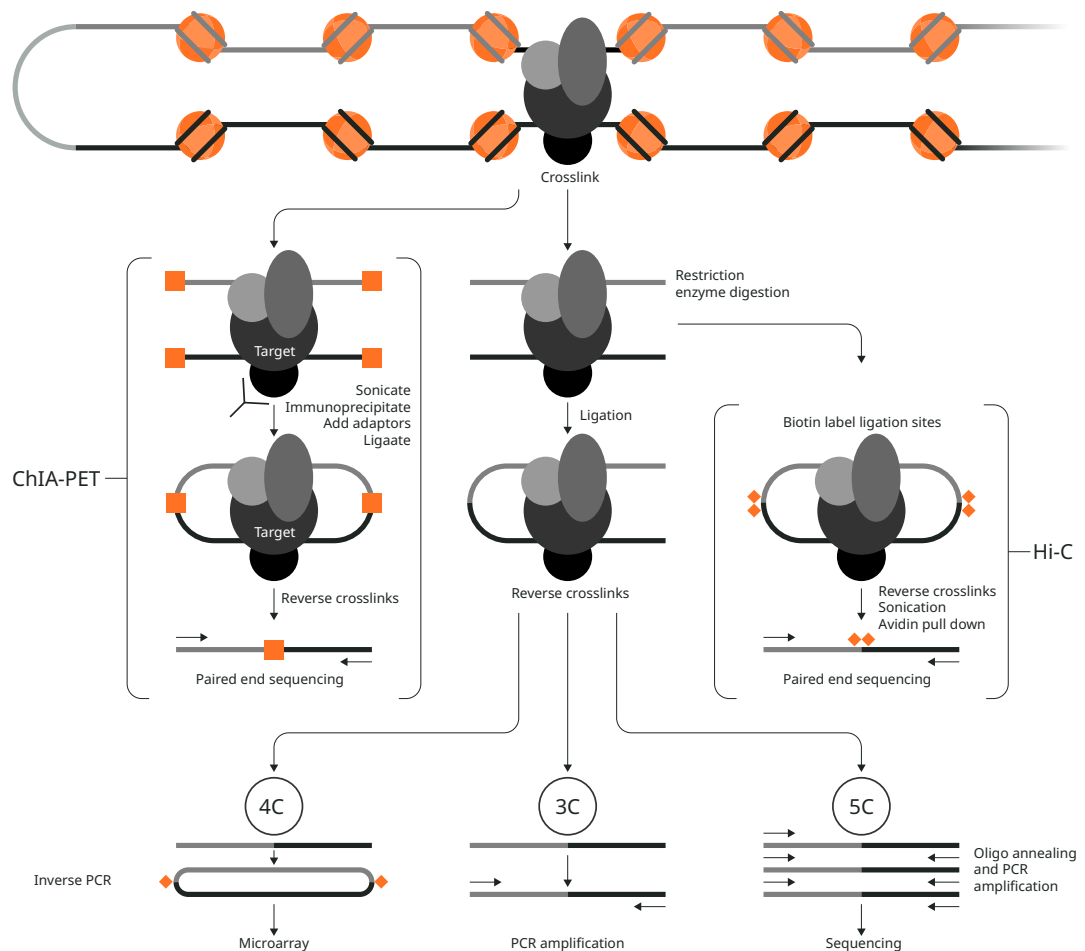


Figure 3. Chromosome conformation techniques. Various steps of 3C, 4C, 5C, ChIA-PET, and Hi-C.

Chromatin conformation capture (3C)

3C uses formaldehyde cross-linking to lock 3-dimensional chromatin structure in place, followed by restriction enzyme digestion. Excised DNA fragments are then analyzed by qPCR and sequencing to identify where distant DNA regions are connected. This approach for analyzing 3D chromatin structure and interactions in vivo was first developed in 2002 (Dekker et al., 2002), and has since become the foundation for a host of related techniques that have been developed to achieve greater scale, throughput, or specificity.

Circularized chromosome conformation capture (4C)

4C enables identification of previously unknown DNA regions that interact with a locus of interest, which makes 4C ideal for discovering novel interactions within a specific region (Dekker et al., 2006).

4C helpful hints:

- Choose the right restriction enzymes. More frequent cutters (ie four bp recognition sites) are better for local interactions between the region of interest and nearby sequences on the same chromosome (van der Werken et al., 2012).
- Optimize cross-linking. Lower formaldehyde concentrations promote undesirable region-of-interest self-ligations, but also prevent DNA “hairballs” that hinder restriction enzyme cutting. High formaldehyde concentrations lower self-ligation events but increase hairballs. An optimal formaldehyde concentration should be chosen for the specific experimental situation to balance these considerations. 1% formaldehyde treatment for 10 min is a good starting point for most experiments (van der Werken et al., 2012).

Carbon copy chromosome conformation capture (5C)

5C generates a library of any ligation products from DNA regions that associate with the target loci, which are then analyzed by NGS. 5C is ideal when great detail about all the interactions in a given region is needed, for example when diagramming a detailed interaction matrix of a particular chromosome. However, 5C is not truly genome-wide, since each 5C primer must be designed individually, so it is best suited to a specific regions (Dotsie and Dekker, 2007).

5C helpful hints:

- Select the right restriction enzyme. Choosing an enzyme that functions efficiently under your specific experimental conditions is essential. For example, BamHI is not recommended for most experiments due to inefficiency under 3C conditions (Dotsie et al., 2007).
- Optimize primer design. 5C uses two primers: a forward 5C primer that binds upstream of the ligation site, and a reverse primer that binds immediately downstream. Primer length should be adjusted so that the annealing temperature is about 65°C to allow primers to anneal exactly with their restriction fragments. Ensure that 5C primers are synthesized with a phosphate at the 5' end for ligation.
- Use a control template. This will control for differences in primer efficiency. A control library constructed from the entire genomic region under study is recommended. If this library is not constructed, then researchers should be aware that interaction frequencies would be less precise

Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)

ChIA-PET takes aspects of ChIP and 3C to analyze the interplay of distant DNA regions through a particular protein.

ChIA-PET is best used for discovery experiments involving a protein of interest and unknown DNA binding targets. Transcription factor binding sites, for example, are best studied with ChIA-PET since this technique requires the DNA to be bound by the transcription factor in vivo for the interaction to be called (Fullwood et al., 2009).

ChIA-PET helpful hints:

- Overlap PET tags to reduce background. Like most 3C technologies, background noise is a technical challenge. In ChIA-PET particularly, noise can make it difficult to find long-range interactions with the locus of interest. A useful tip to overcome this is to require PETs to overlap at both ends of the region to be a long-range interaction.

ChIP-loop

ChIP-loop is a mix of ChIP and 3C that employs antibodies targeted to proteins suspected to bind a DNA region of interest. ChIP-loop is ideal to find out if two known DNA regions interact via a protein of interest. ChIP-loop is also well suited to confirmation of suspected interactions, but not the discovery of novel ones (Horike et al., 2005).

ChIP-loop helpful hints:

- Avoid non-native loops. The biggest issue encountered with ChIP-loop is the formation of non-native loops forming during DNA concentration before ligation occurs. A simple way to avoid this is to choose a protocol that performs the precipitation after the ligation step (Simons et al., 2007).
- Validate ChIP-loop interactions. Another challenge in ChIP-loop can be accurate quantitation of ligation products. 3C technologies, especially ChIP-loop, often capture random interactions. To combat this, consider performing a ChIP experiment in parallel and using it to validate the ChIP-loop interactions. If a DNA-protein-DNA interaction identified by ChIP-loop is indeed real, then both DNA-protein interactions should also appear in the ChIP data (Simons et al., 2007).

Hi-C

Hi-C amplifies ligation products from the entire genome and assesses their frequencies by high-throughput sequencing. Hi-C is a great choice when broad coverage of the entire genome is required, and the resolution is not of great concern. For example, mapping the genome-wide changes in chromosome structure in tumor cells (Lieberman-Aiden et al., 2009).

Hi-C helpful hints:

- Optimize library amplification. Hi-C library amplification must generate enough product for analysis, while avoiding PCR artefacts. To do this, the PCR cycle number should be optimized (in the range of 9–15 cycles). If enough product cannot be produced (50 ng of DNA), multiple PCR reactions should be pooled rather than the cycle number increased, five reactions are usually sufficient (Belton et al., 2012).
- Balance read lengths. As with any sequence experiment, high-quality reads are paramount. The read length must be optimal to balance the need for long reads to map interactions, but not too long as to pass through the ligation junction into the partner fragment. Therefore, 50 bp reads are optimal in most cases (Belton et al., 2012).
- Choose an appropriate bin size. This is critical for data analysis. Bin size should be inversely proportional to the number of expected interactions in a region. Use smaller bins for more frequent intra-chromosomal interactions and larger bins for less frequent inter-chromosomal interactions (Belton et al., 2012).

Capture-C

Capture-C uses a combination of 3C and oligonucleotide capture technology (OCT), together with high-throughput sequencing to study hundreds of loci at once. Capture-C is ideal when both high resolution and genomic-wide scale are required. For example, analyzing the functional effect of every disease-associated SNP in the genome on local chromatin structure (Hughes et al., 2014).

Capture-C helpful hints:

- Carefully choose probe positions. It's best to position probes close to the restriction enzyme sites, even overlapping when possible (Hughes et al., 2014).
- Keep libraries complex. Maintaining library complexity is the top priority. A complex library means more high-quality interactions in the output. For this reason, anything that could decrease library complexity should be avoided, such as a Hi-C biotin capture (Hughes et al., 2014).
- Watch for false interaction in duplicated regions. The mapping process can stimulate strong interactions between these regions (such as pseudogenes) that are actually artefacts (Hughes et al., 2014).

References

- Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y and Dekker J (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*, 58, 268-76.
- Dekker J, Rippe K, Dekker M and Kleckner N (2002). Capturing chromosome conformation. *Science*, 295, 1306-1311.
- Dekker J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 3, 17-21.
- Dostie J and Dekker J (2007). Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc*, 2, 988-1002.
- Dostie J, Zhan Y and Dekker J (2007). Chromosome conformation capture carbon copy technology. *Curr Protoc Mol Biol*, Chapter 21, Unit 21.14.
- Horike S, Cai S, Miyano M, Cheng JF and Kohwi-Shigematsu T (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 37, 31-40.
- Fullwood MJ, et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 462, 58-64.
- Lieberman-Aiden E, et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289-293.
- Hughes JR (August 2014). Email interview.
- Hughes JR, et al. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*, 46, 205-212.
- Simonis M, Kooren J and de Laat W (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods*, 11, 895-901.
- Van de Werken H, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E and de Laat W (2012). 4C technology: protocols and data analysis. *Methods Enzymol*, 513, 89-112

Histone modifications

Introduction

Chromatin architecture, nucleosomal positioning, and ultimately access to DNA for gene transcription, is largely controlled by histone proteins. Each nucleosome is made of two identical subunits, each of which contains four histones: H2A, H2B, H3, and H4. Meanwhile, the H1 protein acts as the linker histone to stabilize internucleosomal DNA and does not form part of the nucleosome itself.

Histone proteins undergo post-translational modification (PTM) in different ways, which impacts their interactions with DNA. Some modifications disrupt histone-DNA interactions, causing nucleosomes to unwind. In this open chromatin conformation, called euchromatin, DNA is accessible to binding of transcriptional machinery and subsequent gene activation. In contrast, modifications that strengthen histone-DNA interactions create a tightly packed chromatin structure called heterochromatin. In this compact form, transcriptional machinery cannot access DNA, resulting in gene silencing. In this way, modification of histones by chromatin remodeling complexes changes chromatin architecture and gene activation.

At least nine different types of histone modifications have been discovered. Acetylation, methylation, phosphorylation, and ubiquitylation are the most well-understood, while GlcNAcylation, citrullination, crotonylation, and isomerization are more recent discoveries that have yet to be thoroughly investigated. Each of these modifications are added or removed from histone amino acid residues by a specific set of enzymes.

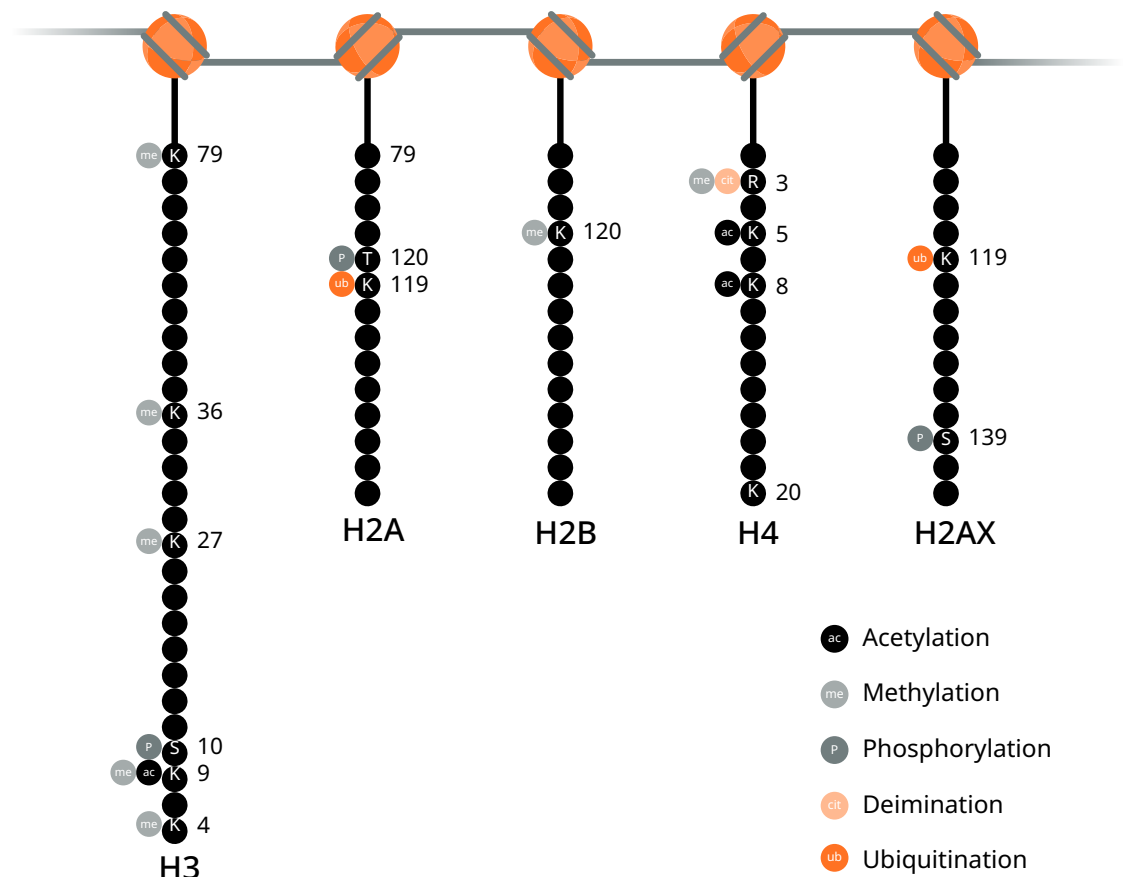


Figure 4. The most common histone modifications. To find out more see our full histone modifications poster at www.abcam.com/EpigeneticsPoster.

Together, these histone modifications make up what is known as the histone code, which dictates the transcriptional state of the local genomic region. Examining histone modifications at a particular region, or across the genome, can reveal gene activation states, locations of promoters, enhancers, and other gene regulatory elements.

Histone modifications in detail

Acetylation

Acetylation is one of the most widely studied histone modifications since it was one of the first discovered to influence transcriptional regulation. Acetylation adds a negative charge to lysine residues on the N-terminal histone tails that extend out from the nucleosome. These negative charges repel negatively charged DNA, which results in a relaxed chromatin structure. The open chromatin conformation allows transcription factor binding and significantly increases gene expression (Roth et al., 2001)

Histone acetylation is involved in cell cycle regulation, cell proliferation, and apoptosis and may play a vital role in regulating many other cellular processes, including cellular differentiation, DNA replication and repair, nuclear import and neuronal repression. An imbalance in the equilibrium of histone acetylation is associated with tumorigenesis and cancer progression.

Enzymatic regulation

Acetyl groups are added to lysine residues of histones H3 and H4 by histone acetyltransferases (HAT) and removed by deacetylases (HDAC). Histone acetylation is largely targeted to promoter regions, known as promoter-localized acetylation. For example, acetylation of K9 and K27 on histone H3 (H3K9ac and H3K27ac) is usually associated with enhancers and promoters of active genes. Low levels of global acetylation are also found throughout transcribed genes, whose function remains unclear.

Methylation

Methylation is added to the lysine or arginine residues of histones H3 and H4, with different impacts on transcription. Arginine methylation promotes transcriptional activation (Greer et al., 2012) while lysine methylation is implicated in both transcriptional activation and repression depending on the methylation site. This flexibility may be explained by the fact that that methylation does not alter histone charge or directly impact histone-DNA interactions, unlike acetylation.

Lysines can be mono-, di-, or tri-methylated, providing further functional diversity to each site of methylation. For example, both mono- and tri-methylation on K4 of histone H3 (H3K4me1 and H3K4me3) are activation markers, but with unique nuances: H3K4me1 typically marks transcriptional enhancers, while H3K4me3 marks gene promoters. Meanwhile, tri-methylation of K36 (H3K36me3) is an activation marker associated with transcribed regions in gene bodies.

In contrast, tri-methylation on K9 and K27 of histone H3 (H3K9me3 and H3K27me3) are repressive signals with unique functions: H3K27me3 is a temporary signal at promoter regions that controls development regulators in embryonic stem cells, including Hox and Sox genes. Meanwhile, H3K9me3 is a permanent signal for heterochromatin formation in gene-poor chromosomal regions with tandem repeat structures, such as satellite repeats, telomeres, and pericentromeres. It also marks retrotransposons and specific families of zinc finger genes (KRAB-ZFPs). Both marks are found on the inactive chromosome X, with H3K27me3 at intergenic and silenced coding regions and H3K9me3 predominantly in coding regions of active genes.

Enzymatic regulation

Histone methylation is a stable mark propagated through multiple cell divisions, and for many years was thought to be irreversible. However, it was recently discovered to be an actively regulated and reversible process.

Methylation: histone methyltransferases (HMTs)

- Lysine
 - SET domain containing (histone tails)
 - Non-SET domain containing (histone cores)
- Arginine
 - PRMT (protein arginine methyltransferases) family

Demethylation: histone demethylases

- Lysine
 - KDM1/LSD1 (lysine-specific demethylase 1)
 - JmjC (Jumonji domain-containing)
- Arginine
 - PAD4/PADI4

Phosphorylation

Histone phosphorylation is a critical intermediate step in chromosome condensation during cell division, transcriptional regulation, and DNA damage repair (Rossetto et al., 2012, Kschonsak et al., 2015). Unlike acetylation and methylation, histone phosphorylation establishes interactions between other histone modifications and serves as a platform for effector proteins, which leads to a downstream cascade of events.

Phosphorylation occurs on all core histones, with differential effects on each. Phosphorylation of histone H3 at serine 10 and 28, and histone H2A on T120, are involved in chromatin compaction and the regulation of chromatin structure and function during mitosis. These are important markers of cell cycle and cell growth that are conserved throughout eukaryotes. Phosphorylation of H2AX at S139 (resulting in γ H2AX) serves as a recruiting point for DNA damage repair proteins (Lowndes et al., 2005, Pinto et al., 2010) and is one of the earliest events to occur after DNA double-strand breaks. H2B phosphorylation is not as well studied but is found to facilitate apoptosis-related chromatin condensation, DNA fragmentation, and cell death (Füllgrabe et al., 2010).

Ubiquitylation

All histone core proteins can be ubiquitylated, but H2A and H2B are most commonly and are two of the most highly ubiquitylated proteins in the nucleus (Cao et al., 2012). Histone ubiquitylation plays a central role in the DNA damage response.

Monoubiquitylation of histones H2A, H2B, and H2AX is found at sites of DNA double-strand breaks. The most common forms are monoubiquitylated H2A on K119 and H2B on K123 (yeast)/K120 (vertebrates). Monoubiquitylated H2A is also associated with gene silencing, whereas H2B is also associated with transcription activation.

Poly-ubiquitylation is less common but is also important in DNA repair-- polyubiquitylation of H2A and H2AX on K63 provides a recognition site for DNA repair proteins, like RAP80.

Enzymatic regulation

Like other histone modifications, monoubiquitylation of H2A and H2B is reversible and is tightly regulated by histone ubiquitin ligases and deubiquitylating enzymes.

Monoubiquitylation

- H2A: polycomb group proteins
- H2B: Bre1 (yeast) and its homologs RNF20/RNF40 (mammals)
- Polyubiquitylation
 - H2A/H2AX K63: RNF8/RNF168

Quick reference guide to histone modifications

Most common histone modifications and where to find them:

Histone modification	Function	Location
H3K4me1	Activation	Enhancers
H3K4me3	Activation	Promoters
H3K36me3	Activation	Gene bodies
H3K79me2	Activation	Gene bodies
H3K9Ac	Activation	Enhancers, promoters
H3K27Ac	Activation	Enhancers, promoters
H4K16Ac	Activation	Repetitive sequences
H3K27me3	Repression	Promoters, gene-rich regions
H3K9me3	Repression	Satellite repeats, telomeres, pericentromeres
Gamma H2A.X	DNA replication	DNA double strand breaks
H3S10P	DNA replication	Mitotic chromosomes

Studying histone modifications by ChIP

ChIP uses antibodies to isolate a protein or modification of interest, along with the DNA to which it is bound (figure 5). The DNA is then sequenced and mapped to the genome to identify the protein or modification's location and abundance.

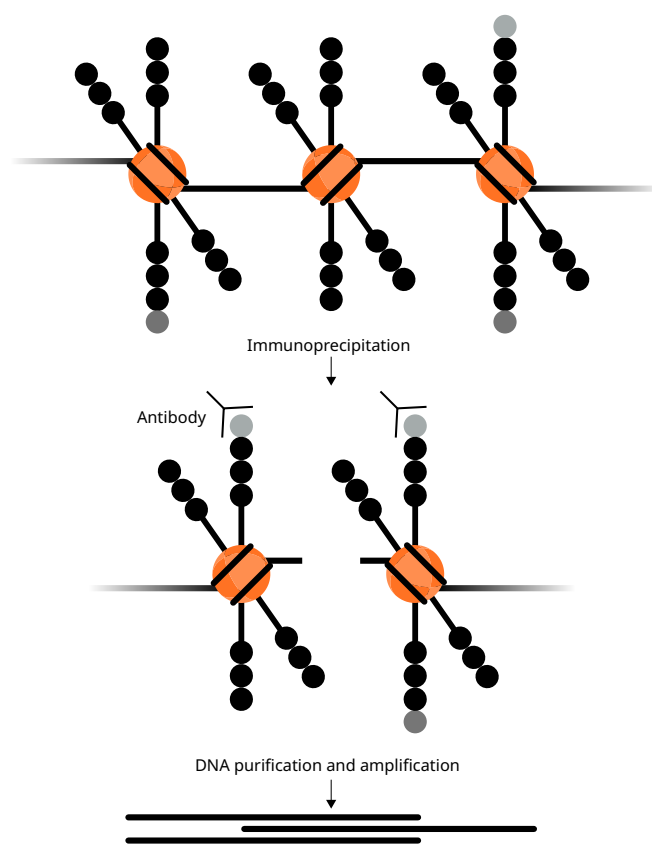


Figure 5. Histone modification ChIP. Antibodies bind directly to modified histone tails. Immunoprecipitation and DNA purification allow for the isolation and identification of the genomic regions that the modifications occupy.

Utilizing antibodies against specific histones and histone modifications in ChIP experiments can reveal the specific locations of

- Higher order chromatin structures, eg H3K9me3 marks heterochromatin and satellite repeats
- Active or silenced genes and genetic programs, eg H3K9ac marks gene activation
- Genetic elements like promoters and enhancers, eg H3K27me3 marks promoters in gene-rich regions, H3K4me1 marks active enhancers

If the function of a histone modification is known, ChIP can identify specific genes and regions with this histone modification signature and the corresponding function across the genome. These genes and regions can then be further examined for their role in the biological process of interest. Using ChIP against H3K4me1, for example, will reveal the locations and sequences of active enhancers throughout the genome, pointing to genes and genetic programs of interest.

Alternatively, if the function of the histone modification is not known, ChIP can identify sequences, genes, and locations with this signature, which can then be used to infer the function of the modification. This technique was pivotal in decoding much of the histone code and is still valuable in ascertaining the function of newly discovered modifications like ubiquitylation and other novel markings.

Histone modifying enzymes: writers and erasers

Histone modifications are dynamically added and removed from histone proteins by specific enzymes (table 1). The balance between these writers and erasers dictates which marks are present on histones, and at what levels, to ultimately control whether specific genetic programs and the cellular processes they orchestrate, are turned on or off.

Table 1. The major categories of histone writers and erasers.

Modification	Writers	Erasers
Acetylation	Histone acetyltransferases (HATs)	Histone deacetylases (HDACs)
Methylation	Histone methyltransferases (HMTs/KMTs) and protein arginine methyltransferases (PRMTs)	Lysine demethylases (KDMs)
Phosphorylation	Kinases	Phosphatases

For more details on the readers, writers, and erasers of histone modifications take a look at our epigenetic modification's poster at www.abcam.com/EpigeneticsPoster.

Identifying modification pathways and the specific writers and erasers at play can reveal

- Relevant cellular pathways, genetic programs and physiological effects for further investigation. For example, histone deacetylases (HDACs) activate immune developmental pathways, while histone acetyltransferases (HATs) play a crucial role in differentiation and proliferation.
- Imbalances between writers and erasers that alter genetic programming and underlie disease processes. Characterizing such imbalances, and the specific enzymes involved, can provide insights into disease pathology, from cancers to autoimmune disorders.
- New drug targets and therapeutic strategies. Once an imbalance is identified, drugs can be developed to impact the activity of these enzymes and correct the imbalance, offering new therapeutic strategies against diseases that have thus far evaded medical efforts. For example, many HDAC inhibitors are in development as novel drugs against cancers and inflammatory diseases like arthritis and type I diabetes.

For drug development efforts, compounds can easily be screened for their impact on writer and eraser activity.

Characterizing histone methylation pathways

In general, histone methyltransferase (HMT) assays are challenging to develop, and most have several drawbacks due to assay design. Typical HMT assays utilize 3H-SAM as a methyl donor and measure S-adenosylhomocysteine (SAH) as a general by-product of the methylation reaction. However, this requires

- Handling radioactive material
- High sensitivity to overcome low k_{cat} (turnover typically < 1 min⁻¹) and KM values for the methyl donor, SAM
- Prior purification of enzyme/protein complexes to assess activity of specific HMTs

Abcam HMT activity assays overcome these difficulties, assessing the activity of specific HMTs with antibodies that detect the specific methylated product, providing:

- Easy colorimetric or fluorometric detection, without radioactivity
- Compatibility with nuclear extracts, or purified proteins (assay is specific for the modification of interest)
- Data in 3 hours

For more information on our histone methylation assays visit www.abcam.com/HistoneMethylationAssays

Characterizing demethylase activity

Histone demethylase activity assays typically measure the formation of formaldehyde, a by-product of demethylation. They are therefore susceptible to interference from detergents, thiol groups and a range of ions. Similar to methylation assays, these assays are not specific for any demethylase and can only be performed with purified protein.

Abcam's histone demethylase assays circumvent these issues by directly measuring the formation of the demethylated product, providing:

- Increased sensitivity (20–1,000 fold) over formaldehyde-based assays
- More accurate data without interference from thiols, detergents or ions
- Compatibility with nuclear extracts or purified protein (due to the assay's specificity for the modification of interest)
- Measures demethylase activity from a broad range of species including mammalian cells/tissues, plants, and bacteria
- Fast microplate format with simple colorimetric or fluorometric readouts
- Data in 3 hours

For more information on our histone demethylase assays visit www.abcam.com/HistoneMethylationAssays

Characterizing histone acetylation pathways

Abcam offers kits to analyze overall, as well as H4-specific, HAT activity. These assays measure the HAT-catalyzed transfer of acetyl groups from the Acetyl-CoA donor to histone peptides, which generates the acetylated peptide and CoA-SH. The CoA-SH byproduct is then measured via colorimetric or fluorometric methods:

- Colorimetric assays- CoA-SH serves as an essential coenzyme for producing NADH, which reacts with soluble tetrazolium dye to generate a product that can be detected spectrophotometrically. This assay is ideal for kinetic studies, with continuous detection.
- Fluorometric assays- CoA-SH reacts with a developer and Probe to generate a product that is detected fluorometrically.

Characterizing histone deacetylase activity

HDAC proteins fall into four major groups (class I, class IIA, class IIB, class III, class IV) based on function and DNA sequence similarity. Classes I, IIA, and IIB are considered “classical” HDACs whose activities are inhibited by trichostatin A (TSA), whereas class III is a family of NAD⁺-dependent proteins (sirtuins (SIRT)) not affected by TSA. Class IV is considered an atypical class on its own, based solely on DNA sequence similarity to the others.

Each of these classes are associated with different cellular programs and may be assayed individually with various fluorometric assays. For example, SIRTs are typically associated with cancers and neurological diseases. Detecting SIRT activity, or identifying drugs that impact SIRT activity, may point to novel diagnostics or therapeutic strategies for these diseases.

Fluorometric assays utilize an acetylated peptide substrate with a fluorophore and quencher at its amino and carboxyl terminals. Once the substrate is deacetylated, it can be cleaved by a peptidase, releasing the fluorophore from the quencher. The subsequent increase in fluorescence intensity of the fluorophore is directly proportional to deacetylase activity.

Inhibiting writers and erasers

It can be useful to inhibit these modifying enzymes using small molecules and then assess downstream consequences to probe the involvement and biological functions of histone modifications. Thus, inhibitors of writers and erasers are vital tools for understanding the roles of epigenetic modification pathways. They are also essential for the validation of “druggable” targets in the context of pre-clinical studies both in academic and industry contexts.

To find out more about our range of histone methyltransferase and demethylase inhibitors visit www.abcam.com/HistoneWriterEraserInhibitors

Histone modification readers/translators

Histone modifications regulate the physical properties of chromatin, and its corresponding transcriptional state, either directly (eg acetyl groups that repel negatively charged DNA to create open chromatin conformation) or via protein adaptors termed effectors. Effector proteins recognize and bind to specific epigenetic marks, and subsequently, recruit molecular machinery to alter chromatin structure. These epigenetic readers determine the functional outcome of histone modifications by translating the histone code into action.

Effector domains recognize specific histone modifications

Effector proteins recognize and bind to histone modification marks through effector domains, known as modules (Table 2).

Table 2. Recognition of histone marks by modules or histone-binding proteins.

Histone-binding or effector module	Known histone marks
Chromodomain	H3K4me2/3, H3K9me2/3, H3K27me2/3
Tudor	H3K4me3, H4K20me2
MBT	H3K4me1, H4K20me1/2, H1K26me1
WD40 repeats	R2/H3K4me2
Bromodomain	Kac
PHD	H3K4, H3K4me3, H3K9me3, K36me3
14-3-3	H3S10ph
BRCT	H2A.XS139

These modules recognize specific histone modifications with amino acids that line the module's binding pocket. Meanwhile, residues outside of this binding pocket (particularly in the N⁺² and N⁻² positions) dictate specificity for the histone and amino acid residue being modified (eg H3K4 vs H4K20).

Slight variations in residues within or outside of the binding pocket allow for recognition of similar epigenetic marks. For example, effector proteins can distinguish between mono-, di-, or tri-methylation states with slight variations to the methyl-binding module's structure. For example, tudor domains may exclusively bind di- or tri-methylated lysines, while PHD finger modules may bind to both, or only to unmodified lysines (Ruthenburg et al., 2007).

Multivalency enables histone code complexity

Multiple histone-binding modules are often found in the same protein, and/or protein complex, that enable recognition of specific combinations of histone modifications. This allows for a more complex histone code, where histone modifications interact with each other rather than being interpreted in isolation.

Multivalent engagement of histone modifications is important for recognizing discrete marking patterns with composite specificity and enhanced affinity, while also enabling diverse and precise downstream actions. For example, a single epigenetic mark (like H3K4me3) may activate gene transcription in one context, but repress it in another, depending on the surrounding marks. Table 3 shows examples of some of the functional associations of different combinations of histone modifications (Ruthenburg et al., 2007).

Table 3. Functional associations of coexisting histone and DNA modifications

Histone marks	Chromatin state
H3K4me2/3 + H4K16ac	Transcriptionally active homeotic genes
H3K4me2/3 + H3K9/14/18/23ac	Transcriptionally active chromatin
H3S10ph + H3K14ac	Mitogen-stimulated transcription
H3K4me3 + H3K27me3	Bivalent domains
H3K9me3 + H3K27me3 + 5mC	Silent loci
H3K27me3 + H2AK119ub1	Silent homeotic genes
H3K9me3 + H4K20me3 + 5mC	Heterochromatin
H3K9me2/3 + H4K20me1 + H4K27me3 + 5mC	Inactive X-chromosome

Multiple effector modules in a protein or complex may interact with histone modifications on the same, or across, histones and/or nucleosomes. These interactions may be categorized as follows:

Intranucleosomal: binding to the same nucleosome

- Cis-histone: binding to the same histone tail
- Trans-histone: binding to different histone tails

Internucleosomal: binding to different nucleosomes

- Adjacent bridging: binding to adjacent nucleosomes
- Discontinuous bridging: binding to nonadjacent nucleosomes

References

Cao, J. & Yan, Q. Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front. Oncol.* 2, 26 (2012).

Füllgrabe, J., Hajji, N. & Joseph, B. Cracking the death code: apoptosis-related histone modifications. *Cell Death Differ.* 17, 1238–1243 (2010).

Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–57 (2012).

Kschonsak, M. & Haering, C. H. Shaping mitotic chromosomes: From classical concepts to molecular mechanisms. *BioEssays* 755–766 (2015)

Lowndes, N. F. & Toh, G. W.-L. DNA repair: the importance of phosphorylating histone H2AX. *Curr. Biol.* 15, R99–R102 (2005).

Pinto, D. M. S. & Flaus, A. Structure and function of histone H2AX. *Subcell. Biochem.* 50, 55–78 (2010).

Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* 7, 1098–1108 (2012)

Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annu. Rev. Biochem.* 70, 81–120 (2001)

Ruthenburg, A.J., Li, H., Taverna, S.D., Patel, D.J. & Allis, C.D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Rev. Mol. Cell Biol.* 8, (2007)

Studying epigenetics with ChIP

Introduction

Chromatin immunoprecipitation (ChIP) is a powerful technique that allows researchers to examine the interactions between epigenetic regulators and DNA in their natural context. With ChIP, researchers can identify specific genes and sequences where a protein of interest binds, across the entire genome, providing critical clues to their regulatory functions and mechanisms. By dissecting the temporal and spatial dynamics of protein-DNA interactions, ChIP provides insights into core biological processes and disease pathology.

ChIP is exceptionally versatile, with use in a broad scope of applications. From looking at sequence-specific protein binding to global regulatory processes, ChIP gives researchers the tools to integrate discoveries and paint a comprehensive picture of complex epigenetic regulatory systems.

Applications of ChIP

ChIP has played a central role in elucidating gene regulation, transcriptional machinery, and chromatin structure. Here are some of the key proteins you can detect using ChIP.

Transcription factors

By examining where, and when, specific transcription factors bind across the genome, researchers have identified specific binding sites and sequences, pinpointed downstream gene activation, and revealed genome-wide regulatory programs of transcription factors.

Further investigation by ChIP and other methods has revealed these transcription factors to be master regulators behind disease pathology, where they orchestrate epigenetic dysregulation that results in cancers, autoimmune diseases, allergy, and many others. By identifying these master regulators and their downstream genetic programs, ChIP has provided novel targets for diagnostic and therapeutic strategies against a wide variety of diseases.

Transcriptional machinery

ChIP studies examining the binding of RNA polymerase II, and other components of transcription, reveal promoter and enhancer sequences and novel mechanisms of transcriptional regulation.

Chromatin Structure

ChIP studies were pivotal in the discovery and characterization of the histone code. By mapping the locations of specific histone modifications and comparing to known gene activation states, researchers have documented how acetylation or methylation on particular histone residues influence gene activation or silencing and higher order chromatin structure. These histone modification signatures can now be used to predict those aspects of epigenetic regulation at specific regions of the genome via ChIP.

As new histone modifications and chromatin regulatory elements are discovered, ChIP continues to be an essential tool for revealing the functions of these elements, and complexities of their interactions, in genomic regulation.

Combined analysis

Combining ChIP analysis of multiple proteins is a powerful way to build a complete picture of genomic regulation. This approach enables researchers to study how different types of proteins, and protein complexes, interact spatially and temporally at specific sites along a particular gene, or across the entire genome, to regulate gene transcription (Barski et al., 2007).

ChIP expands the scope and precision of epigenetic research

ChIP facilitates the analysis of epigenetic mechanisms on a variety of scales, with unrivalled precision. These are some of the things you can achieve using ChIP.

Local epigenetic mechanism

- Map a protein of interest to a specific gene or genomic region of interest
- Identify specific binding site sequences of a protein of interest

Genome-wide epigenetic programming

- Localize a protein of interest, such as a transcription factor, at all of its binding sites across the genome
- Map proteins and chromatin characteristics across loci
- Compare enrichment of a protein-protein modification (eg histone acetylation) at different loci under different conditions

Dynamic epigenetic processes

- Quantify a protein/protein modification at an inducible gene over a time course
- Reveal essential mechanisms of epigenetic regulation and dysregulation involved in the biological process of interest by comparing ChIP results across different cellular states, conditions, and time points.
- Different tissues reveal epigenetic programs and genes responsible for differentiation and cell-type . specific functions and characteristics.
- Different cell cycle states reveal epigenetic programs and genes responsible for cell proliferation and cell cycle control, with implications for developmental processes and cancer pathology.
- Disease vs. healthy cells identify critical genes and programs that are dysregulated, to reveal underlying disease pathology and novel targets for diagnosis and treatment
- Treatment vs. no treatment reveals whether certain treatments or conditions may be effective at correcting epigenetic dysregulation that underlies disease pathology

Protocol overview: how ChIP works

The ChIP procedure utilizes an antibody to immunoprecipitate a protein of interest, such as a transcription factor, along with its associated DNA. The associated DNA is then recovered and analyzed by PCR, microarray or sequencing to determine the genomic sequence and location where the protein was bound.

The procedure can be broken down into five parts (Figure 6):

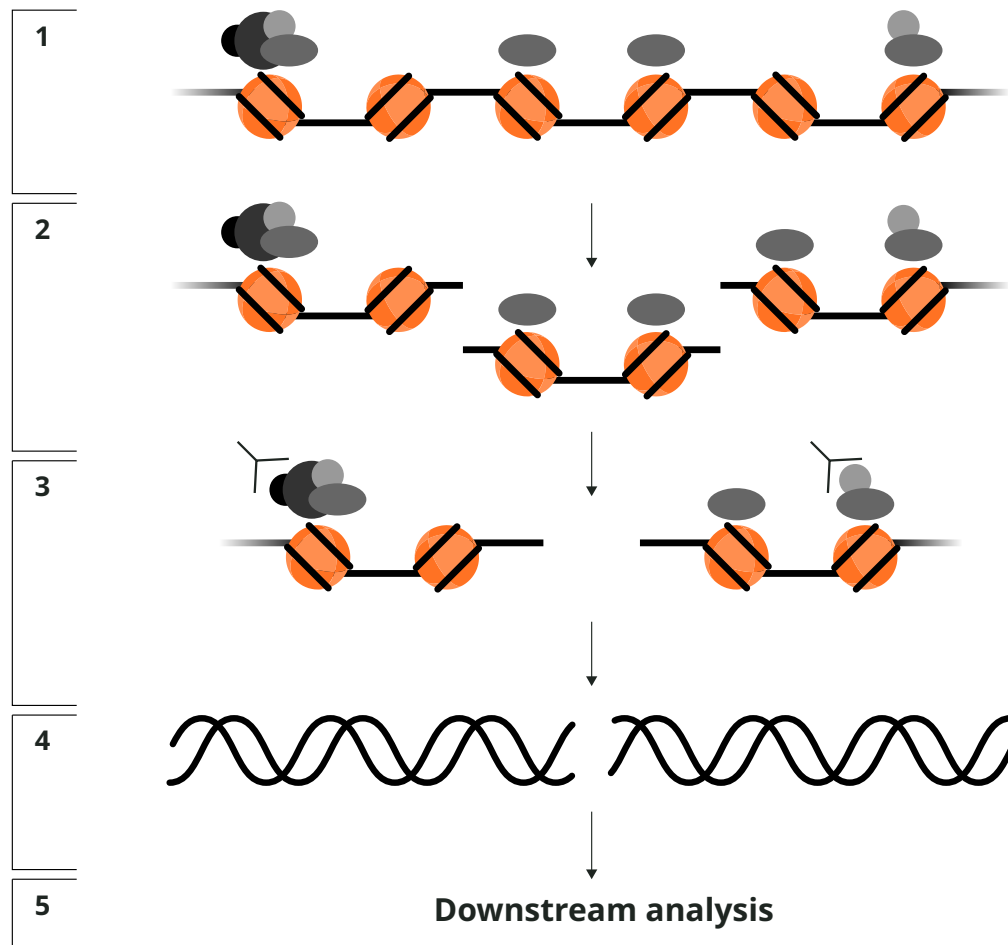


Figure 6. ChIP protocol workflow. Step-by-step approach to carrying out a ChIP experiment

1. Cross-link DNA and proteins (X-ChIP)
2. Chromatin fragmentation by sonication (X-ChIP) or by enzymatic methods (N-ChIP)
3. Immunoprecipitation of the chromatin fragments interacting with the target protein/modification
4. Reverse cross-links (X-ChIP) and DNA purification
5. Analysis of the material obtained to determine the abundance of the target sequences relative to the input

1. Cross-linking

In some cases, cross-linking of DNA and proteins may be required to stabilize their interactions, particularly for proteins that interact as part of large protein complexes and do not directly contact DNA. Crosslinking fixes these molecular interactions and freezes them at a particular point in time and is termed cross-linking ChIP (X-ChIP). In contrast, native ChIP (N-ChIP) which is performed without prior crosslinking.

Crosslinking is generally performed with formaldehyde, which reversibly crosslinks protein to DNA, RNA, and other proteins. Other chemicals like cisplatin can be used to selectively crosslink only between DNA and protein. Dual crosslinking with additional reagents may be required to study interactions between DNA and particularly large protein/RNA complexes. UV crosslinking is irreversible and therefore not compatible with ChIP.

2. Chromatin Fragmentation

Chromatin must be fragmented into small pieces for efficient immunoprecipitation and precise mapping of the target antigen to the genome. The size of the DNA fragments will ultimately determine the resolution of genomic mapping, so it is important to optimize the fragmentation protocol.

- N-ChIP provides the highest resolution mapping, with enzymatic digestion generating fragments the size of a single nucleosome at 175 base pairs
- X-ChIP relies on sonication to generate ideal fragment sizes of 200–1000 base pairs

Steps 1 and 2 are incredibly important to optimize for each experiment to get the highest quality DNA possible for subsequent ChIP analysis.

3. Immunoprecipitation

The protein of interest, and DNA fragment to which it is bound are then immunoprecipitated. The chromatin mixture is incubated with an antibody to the protein of interest, and either agarose or magnetic beads overnight at 4°C to form bead/antibody/protein/DNA complexes. The beads are then collected by either centrifugation, for Protein A, Protein G, or Protein A/G agarose beads, or by magnetic tube rack, for magnetic beads, to immunoprecipitate the antibody/protein/DNA complex. Non-specific binding is removed with subsequent washes.

4. DNA Recovery and Purification

The antibody/protein/DNA complex is eluted from the beads with SDS and heat. Crosslinking of protein and DNA must then be reversed with NaCl and heat for X-ChIP experiments. Any protein and RNA present are then degraded with proteinase K and RNase A, and the remaining DNA is purified with either phenolchloroform extraction or PCR purification kit.

You can find our complete X-ChIP protocol at www.abcam.com/X-ChIP

And our N-ChIP protocol at www.abcam.com/N-ChIP

5. Analyze the DNA

The purified DNA is then analyzed by qPCR, hybrid array (ChIP-on-chip) or next-generation sequencing (ChIP-seq) to identify and quantify the sequences that have been immunoprecipitated. These sequences are mapped to the genome to identify the genes and regions where the protein of interest was bound.

Sample Preparation: X-ChIP vs N-ChIP

The aim of cross-linking is to fix the antigen of interest to its chromatin binding site. Histones themselves generally do not require cross-linking, as they are already tightly associated with the DNA. Other DNA binding proteins that with weaker affinities for DNA or histones may require cross-linking to hold them in place.

- ChIP for histone modifications is unlikely to require cross-linking
- Non-histone proteins such as transcription factors and proteins contained in DNA binding complexes will most likely require cross-linking
- The further away from the DNA your interaction of interest lies, the less effective ChIP will be without cross-linking

Chromatin fragmentation differences

While N-ChIP and X-ChIP both require chromatin fragmentation to make interactions accessible to antibodies, they require different fragmentation procedures utilizing micrococcal nuclease or sonication, respectively (Neill et al., 2003).

N-ChIP

For N-ChIP experiments, enzymatic digestion with micrococcal nuclease should sufficiently fragment your sample into single nucleosomes (monosomes containing ~175 base pairs of DNA).

- Purified monosomes are not suitable for analyzing interactions with certain chromatin binders, like transcription factors, which often bind inter-nucleosomal DNA. Sonication is recommended for these instances.
- Nucleosomes are dynamic and may rearrange during the enzymatic digestion. This may be a problem for mapping certain areas of the genome, and any changes should be monitored with suitable controls (see detection controls for quantitative PCR). X-ChIP should be performed as a control to assess any dynamic and unwanted changes in the absence of cross-linking.
- Enzymatic cleavage will not produce entirely random chromatin fragments. Micrococcal nuclease favors certain areas of genome sequence over others and will not digest DNA evenly or equally. Certain loci could be overrepresented, while others may be absent, potentially impacting the accuracy of the data.
- To get consistency in digestions, aliquot stock enzyme after purchase and run a new time course with a fresh aliquot every time you set up an experiment. Although enzyme quality may vary over time in storage, the risk of variation within chromatin preparations (degree of compaction, etc) is far higher; one chromatin sample should not be treated as being the same as all others before it.

X-ChIP

Formaldehyde cross-linking restricts access of enzymes such as micrococcal nuclease to their targets, making enzymatic digestion ineffective in X-ChIP experiments. Instead, sonication is used to generate random DNA fragments of 500–700 base pairs (2–3 nucleosomes).

- Avoid foaming, which decreases energy transfer within the solution and decreases sonication efficiency. Sonication may also be affected by cross-linking time, cell density, or cell type.
- While sonicated chromatin can be snap frozen in liquid nitrogen and stored at -80°C for up to two months, avoid multiple cycles of freeze-thaw.
- Although sonication theoretically does not exhibit preferential cleavage of the genome, in practice this is rarely the case.
- DNA fragment sizes are typically larger, affecting the resolution of the assay. Regardless, fragments up to 1.5 kb resolve well for most purposes in ChIP. Micrococcal nuclease digestion can improve resolution in combination with sonication and may be useful with gentle or incomplete cross-linked samples.

Regardless of which fragmentation method is chosen, it is important to always run a fragmentation time course to optimize fragment size when setting up an experiment.

Table 4. Advantages and disadvantages of N-ChIP vs X-ChIP.

	N-ChIP	X-ChIP
Advantages	Efficient precipitation of DNA and histone proteins	For non-histone proteins. Can be performed on all cell types, tissues, and organisms.
	Specificity of the binding is more predictable	Enables DNA-protein, RNA-protein, and protein-protein cross-linking
	High resolution (~175bp/mononucleosome)	Reduced chances of chromatin rearrangements
Disadvantages	Only for histones	Over fixation can prevent effective sonication
	Selective nuclease digestion may bias input chromatin	Formaldehyde can alter the binding properties of the antigen
	High concentrations of nuclease may over-digest chromatin	Lower resolution compared to N-ChIP

Antibody selection

Not all antibodies are appropriate for ChIP experiments, and many antibodies are not of ChIP-quality or validated for ChIP applications. Choosing the right ChIP-grade antibody is essential for the success of your ChIP experiment.

If not commercially available, or if you would like to try an antibody that is not yet ChIP-tested:

- Antibodies approved for IP, IHC or ICC applications are good candidates. Similar to ChIP, these applications recognize the protein's native conformation, in contrast to western blot antibodies, which may only recognize the denatured peptide form.
- Antibody specificity is a major concern and should be fully-characterized before application in ChIP experiments. For N-ChIP applications, use peptide competition in western blot. However, for X-ChIP applications, this method will not guarantee antibody function as cross-linking can dramatically alter epitopes. Instead, compare ChIP and western blot results using that antibody to confirm equivalent performance.
- Ideally, antibodies for ChIP should be affinity-purified; however, many laboratories use sera as their antibody source and then overcome background problems that may arise with stringent buffers.

Controls

ChIP protocols and data analysis can be complex, so it is critical to include the right controls to ensure that the experiment worked as intended.

Sample controls

It is crucial to include an input sample control that has not been immunoprecipitated in all DNA recovery steps for comparison to pulldown sample results. It is ultimately this comparison that normalizes the data to provide interpretable results.

When immunoprecipitating for histone modifications, purified histone H3, and H1 can be used as positive controls for the quality of the histone preparation (histone H1 is commonly used for X-ChIP). Meanwhile, calf thymus histone preparation should be used as a positive control histone sample for checking antibody specificity in western blot.

Antibody controls

Various antibody controls are important to ensure that immunoprecipitation was successful and to rule out the possibility of contamination. Here are some examples of key controls:

- Positive controls for active gene loci: H3K4me3 and H3K9ac
- Negative controls for silent gene loci: H3K9me3, H3K9me, and H3K27me3
- Negative control for a non-chromatin epitope: anti-GFP antibody
- Negative IP control: isotype IgG antibody control or beads only IP

Also, chromatin remodeling may move or remove histones at a particular locus (eg an active promoter). To confirm the preservation of nucleosomes at particular genomic loci, use a control antibody against a non-modified histone such as histone H3. When analyzing histone modifications, normalize to histone content with an anti-H3 antibody.

Quantitative PCR controls

If analyzing data by qPCR, additional controls are necessary to ensure the quality of data analysis. Certain areas of the genome will purify better than others, and some nucleosomes may rearrange during enzymatic fragmentation. As a result, it is important to generate PCR primers to several regions in the starting material, as well as the purified/ChIP material, as controls for spurious results. Generate starting material by lysing the starting cells and take a sample for simple PCR of control regions in parallel with ChIP.

Also, during the qPCR stages, it is essential to perform positive and negative control qPCR for genomic loci where you know the protein of interest should or should not be bound. It is also necessary to perform a non-template control qPCR as a negative control to ensure there is no contamination in the PCR.

Protocol optimization

ChIP protocols must be optimized at multiple stages to achieve the best results. Here are a few things that may need a little extra optimization to give you the best ChIP results.

Cross-linking (X-ChIP only).

Formaldehyde is recommended for reversible cross-linking. Formaldehyde is an efficient DNA-protein crosslinker but not an effective protein-protein crosslinker making it difficult to ChIP proteins that do not bind directly to DNA. Alternative cross-linkers may be useful for cross-linking over various intermolecular distances.

Cross-linking is a time-critical procedure and should generally only last a few minutes. Excessive cross-linking can create several issues, including a reduction in antigen availability and sonication efficiency. For example, epitopes may be masked or altered, reducing antibody binding to the antigen and ensuing extraction of material from your sample.

- Always optimize cross-linking conditions with a time-course experiment (2–30 min crosslinking)
- Quench formaldehyde and terminate the cross-linking reaction with glycine
- Cross-links between proteins and DNA are disrupted by proteinase K, which cleaves peptide bonds adjacent to the carboxylic group of aliphatic and aromatic amino acids, to further aid DNA purification

Chromatin fragmentation

It is critical to optimize your chromatin input by fragmenting the chromatin to the appropriate size. Fragment sizes should be less than 1 kb, but ideally, 200–1000bp. The best resolution can be achieved with MNase digestion to single nucleosome level of 175 bp. Perform a time course of chromatin digestion over 2–30 minutes, purify DNA and run a gel alongside a DNA ladder to determine which conditions and timing achieve the optimal DNA fragment size (Figure 7). Different factors require optimization between N-ChIP and X-ChIP protocols.

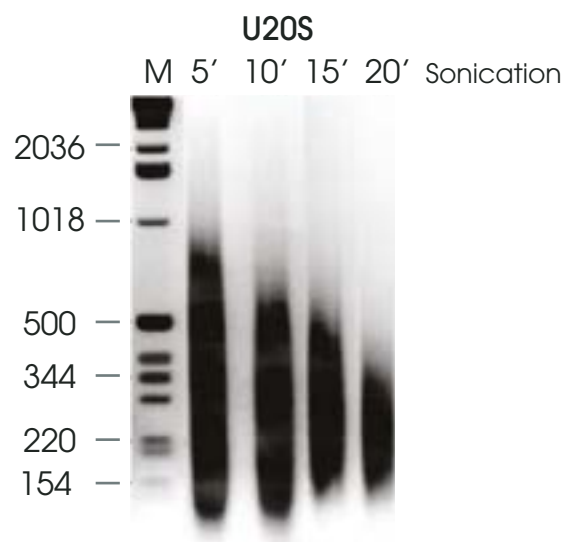


Figure 7. Example of sonication time course experiment. U2OS cells were sonicated for 5, 10, 15 and 20 min. The cross-links were reversed and the purified DNA was resolved on a 1.5% agarose gel. The fragment size decreases during the time course. The optimal fragment size is observed at 15 min.

Antibody concentration

It is important to titer the antibody to optimize the signal to noise ratio. Start with 3–5 µg of antibody for every 25–35 mg of pure monosomes. For quantitative ChIP, you may need to match the amount of chromatin with the same amount of antibody. ChIP typically requires a large amount of primary antibody (1–10 µg per µg of beads). As with many techniques, it is essential to optimize the amount of antibody at the beginning before you run your experiment.

Wash buffers

Determine the correct composition for appropriate stringency of wash steps, typically between 250–500 mM NaCl or LiCl. Higher concentrations of salt and detergent will give cleaner results. However, balance must be achieved between low background and detrimental effects on the target. If the buffer is too stringent, it will destroy specific antibody interactions, resulting in low signal. If the buffer is not stringent enough, non-specific interactions will remain, resulting in high background. NP-40 can be used as a detergent, while RIPA is commonly used for X-ChIP.

ChIP with low cell numbers

Standard ChIP workflows require a large number of cells. Approximately 10^6 to 10^7 cells as starting material, below which the assay is hindered by high background binding, poor enrichment efficiencies, and loss of enriched library complexity. However, these large sample sizes can be difficult to obtain, specifically when examining precious sample types like transgenic mouse tissues or clinical samples. To adjust for lower sample inputs, a number of strategies can be applied.

1. Improving enrichment efficiency and minimizing sample loss

Several adjustments to the ChIP workflow can increase enrichment efficiency and minimize sample loss for low input samples (Mao et al., 2013 and Dirks et al., 2016).

- The quality and properties of the sample itself are important considerations. Specifically, in formalin-fixed paraffin embedded (FFPE) samples, over-cross-linking can cause problems. Methods to extract soluble chromatin from FFPE samples may help (Cejas et al., 2016).
- The kinetics of the IP with low concentrations of antigen can be optimized by modifying variables like buffer pH, ionic strength, and time of incubation (Reverberi et al., 2007).
- Broad DNA fragment size distribution hinders analysis of low-input ChIP, which can be remedied by more limited sonication and/or MNase digestion for more uniform fragmentation (Gilfillian et al., 2012).
- While bacterial DNA is sometimes used as a blocking agent to reduce background for standard ChIP, it is not advised for low-input ChIP as it carries through the assay and confounds data analysis. Other blocking agents such as inert proteins or mRNA can reduce background binding in low-input ChIP, without contaminating the data.
- Miniaturization of the assay into microwell formats facilitates automation and increases the concentration of the antigen (target transcription factor) during the IP workflow – this avoids the “dilution effect” of low antigen concentrations that favor dissociation of the antibody-antigen complex and decreases the efficiency of ChIP.
- Maximize sample retention with single-tube assay formats and the use of magnetic bead purification rather than phenol-based extraction after each assay step.
- The immobilization of antibody and washes to remove non-antibody bound material is often overlooked. Standard protocols use Protein A/G, but alternatives like epitope-tagged proteins may run the risk of over-expression and introduction of artefacts (Xiong et al., 2017).

Abcam's high-sensitivity ChIP assay employs a unique chimeric protein to capture the antibody-bound protein-DNA complex, offering significant advantages.

- The capture protein is smaller than Protein A or G and is coated at high density on the surface of microtiter plate wells, providing a much higher number of IgG immobilization sites in a smaller area, which in turn ensures efficiency and concentration of eluted DNA.
- The chimeric capture protein shows superior stability across a wider range of pH and salt concentrations, which allows for higher stringency wash conditions.
- Successful ChIP starting with just 2×10^3 cells or 0.5 mg tissue
- Relative enrichment factors > 500x
- Fast and easy 5-hour protocol from cells/tissue to enriched DNA
- Microplate assay format for flexibility in sample throughput and automation (can be used in single-well, 8-well strip or 96-well plate format)

2. Readout and downstream data processing platforms

In addition to the assay itself, the choice and optimization of downstream processing (ie sequencing, array, or PCR) and bioinformatic analysis are also important.

- The detection platform impacts the assay's sensitivity. ChIP-sequencing (ChIP-seq) is the gold standard platform for high sensitivity, with consistently lower noise than ChIP-on-chip.
- The most common issues in low-input ChIP-seq are high numbers of unmappable reads, PCR duplicates, and poor library complexity. Therefore library preparation must be optimized for low input samples by optimization adapter ligation to avoid amplification-derived error and bias. Maximizing the efficiency of ChIP enrichment, as described above, can also help (Schmidl et al., 2015 and Bolduc et al., 2016).
- Bioinformatic workflows should be adapted to take into account likely process-derived biases in the data (Kiddler et al., 2011).

ChIP from tissue

Examining epigenetic mechanisms in specific tissues can reveal essential elements of tissue-specific genetic programming, development, and biological processes. ChIP can be a valuable tool for examining roles and mechanisms of tissue-specific transcription factors, gene activation and other aspects of epigenetic regulation. To perform ChIP from tissue samples requires specialized chromatin preparation protocols to ensure quality input material and reliable results.

The amount of tissue required will depend upon protein abundance, antibody affinity and the efficiency of cross-linking. The following protocol was optimized using 5–15 µg of chromatin for each ChIP assay, with 30 mg of liver tissue for each ChIP/antibody. Exact chromatin concentration should be determined for each tissue type before starting the X-ChIP assay.

For more information, you can find our ChIP from tissue samples protocol at www.abcam.com/TissueChIP

Troubleshooting

Even with optimization, your results may not be perfect on the first attempt. Here you can find some common issues and solutions you can use to fix them.

High background in non-specific antibody control

Potential problem	Solution
Non-specific binding to beads	Add additional washes or Add a pre-clearing step by incubating sonicated chromatin with Protein A/G beads for 1 hour prior to immunoprecipitation
Beads give high background	Try different brands of beads and different blocking strategies to see which provide the lowest background in your non-specific control
Contaminated wash buffers	Replace buffers

Low signal

Potential problem	Solution
Cells not efficiently lysed	RIPA buffer should work well
Not enough starting material	ChIP typically requires a large input with at least 25 µg chromatin (3–4 million cells) per IP condition
Chromatin fragment size may be too small	Run on a gel to ensure correct size, repeat fragmentation optimization if necessary
Not enough antibody	3–5 µg is usually sufficient, but up to 10 µg may be required if no signal is observed
Monoclonal antibodies may not be suitable, particularly for X-ChIP as crosslinking may mask the epitope	Try a polyclonal antibody or ChIP grade/approved monoclonal
Wash buffer is too stringent, eliminating specific antibody binding	NaCl in buffer should not exceed 500 mM. Wash buffer should be optimized as described above
Wrong affinity beads	Make sure antibody species and immunoglobulin bind to chosen beads or use a protein/AG mix
If using X-Chip, cells may have been cross-linked for too long, reducing availability of epitopes, or not long enough, reducing pull-down of DNA from the IP	X-ChIP may be required for analyzing proteins with weaker DNA affinity to keep proteins associated with DNA with crosslinking Further optimize your cross-linking time course

Note: Low signal may be real, with no antibody enrichment at the region of interest.

Include positive control antibody and locus to confirm ChIP is working.

The antigen may be present, but not at the expected genomic loci.

Low resolution with high background across large regions

Potential problem	Solution
DNA fragment size may be too large	Fragment sizes should be less than 1 kb, but ideally 200–1000bp. The best resolution can be achieved with MNase digestion to single nucleosome level of 175 bp. Run on a gel and further optimize chromatin fragmentation steps if necessary.

PCR amplification problems

High signal in all samples after PCR including non-template control

Potential problem	Solution
qPCR solution may be contaminated	Prepare new solutions from stock

No DNA amplification in samples

Potential problem	Solution
qPCR solution may be contaminated	Prepare new solutions from stock

What other treatments might affect my ChIP results?

Some antibodies are affected by relatively low concentrations of SDS. TSA, butyrate or colcemid addition do not generally affect ChIP.

Do not centrifuge sepharose beads at high speeds (we'd recommend not exceeding 3,000 x g, but you should refer to your manufacturer's protocol) as this will compact the beads and damage them.

ChIP readout

Once pulled down DNA fragments have been immunoprecipitated and purified, they can be analyzed by several different methods.

qPCR

Utilizes gene or target-specific primers to amplify known target loci among pulldown DNA

Limitations:

- Must know the genome sequence of target regions to design primers for readout

ChIP-on-Chip

Employs microarrays to examine the presence of many loci of interest, specific domains, etc. across the genome. Pulldown and control samples are amplified and labeled with complementary fluorescent probes. Samples are combined and hybridized to a microarray of interest. The ratio of fluorescent signals indicates enriched regions where proteins of interest have bound.

Limitations:

- Requires large cell numbers
- Not sensitive to repetitive elements
- Large number of arrays are necessary to cover the entire genome
- Susceptible to amplification bias after the ChIP procedure
- Lower resolution than ChIP-seq

ChIP-seq

Most commonly used method for genome-wide analysis with improved base pair resolution and none of the limitations of ChIP-on-Chip. Pulldown DNA and control samples are amplified, followed by high throughput sequencing of the fragments, which are then aligned to the genome. Overlapping fragments form a peak, indicating where the protein of interest was bound to the genome.

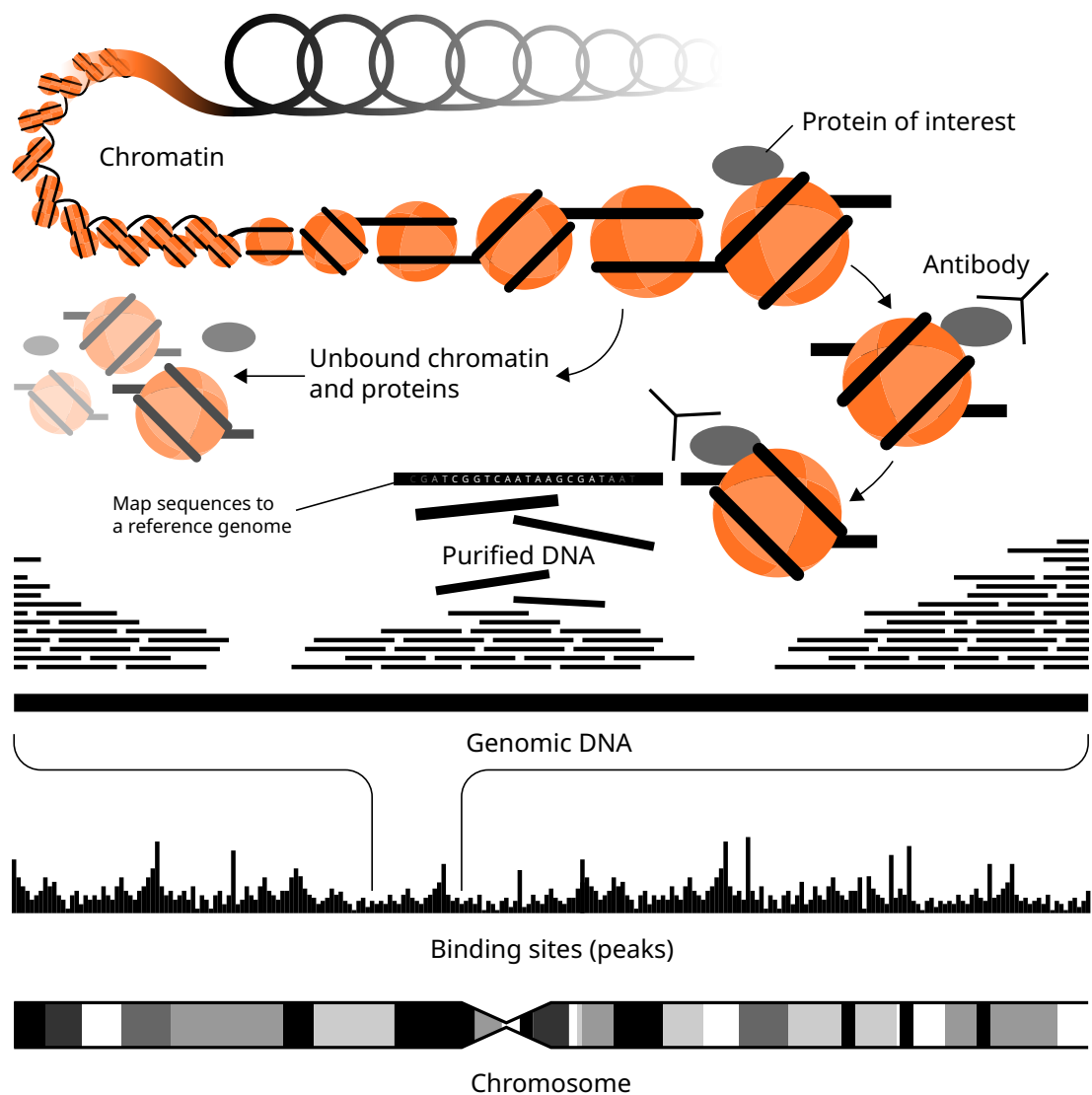


Figure 8. ChIP seq overview. After ChIP is carried out, the precipitated DNA can be used for library preparation, sequencing and then these sequences are mapped to a reference genome to then undergo peak calling to determine binding sites of your protein of interest.

Data analysis

Data should always be normalized for the amount of starting material to eliminate errors introduced by uneven sample quantities. To normalize data, take the final amplicon value and divide it by the amplicon value of input material. For histone modifications, the immunoprecipitated material is usually normalized to the input amount and the amount of the relevant immunoprecipitated histone. For example, ChIP with an H3K4me3 antibody will be expressed relative to the input amount and the amount of H3 immunoprecipitated.

Measuring the amounts (and quality) of starting material is the key to interpreting your results effectively.

Tutorial for ChIP-seq data analysis using online software

While ChIP-seq data analysis can be complex, it is arguably the most important part of the experiment. Robust data analysis is key to accurate and reliable results. A combination of online tools makes data interpretation accessible to bioinformatics specialists and wet lab biologists alike.

This step-by-step guide demonstrates how to extract reliable results from ChIP-seq data, and how to interpret data sets for successful ChIP-seq analysis (Hurtado et al., 2010 and Yan et al., 2013). For more information, see Abcam's data analysis webinar at www.abcam.com/ChIPanalysis

This webinar covers the following steps of ChIP data analysis:

1. QC of sequencing reads (FastQC)
2. Read alignment/mapping (Galaxy/bowtie)
3. Peak calling (Galaxy/macs)
4. Binding signal visualization (UCSC genome browser)
5. De novo motif discovery (MEME-ChIP)
6. Gene ontology of binding sites (GREAT)
7. Heatmap representation of binding signals (seqMINER)

References

- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 18;129(4):823-37(2007)
- Bolduc, N. Preparation of low-input and ligation-free libraries using template-switching technology. In *Current protocols in molecular biology* (Vol. Unit 7.26). Wiley & Sons. (2016).
- Cejas, P. Chromatin Immunoprecipitation from fixed clinical tissues reveals tumor-specific enhancer profiles. *Nature Medicine*. 22, 685. (2016).
- Dirks, R. Genome-wide epigenomic profiling for biomarker discovery. *Clinical Epigenetics*. 8, 122. (2016). ENCODE. (n.d.). ENCODE Platform Comparison. Retrieved from https://genome.ucsc.edu/ENCODE/platform_characterization.html
- Gilfillian, G. Limitations and possibilities of low cell number CHIP-SEQ. *BMC Genomics*. 13, 645. (2012).
- Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet*. 2010;43(1):27-33.
- Kidder, B. ChIP-Seq: Technical considerations for obtaining high quality data. *Nature Immunology*. 12, 918. (2011).
- Mao. Accounting for immunoprecipitation inefficiencies in the statistical analysis of ChIP-Seq data. *BMC Bioinformatics*. 14, 169 (2013).
- Neill O. P. L, Turner M. B. Immunoprecipitation of native chromatin: NChIP. *Methods*. Sep;31(1):76-82 (2003)
- Reverberi, R. Factors affecting the antigen-antibody reaction. *Blood transfusion*. 5, 227. (2007).
- Schmidl, C. ChIPmentation: fast, robust, low-input ChIP-Seq for histones and transcription factors. *Nature Methods*. 12, 963. (2015).
- Stelloo, S. Androgen receptor profiling predicts prostate cancer outcome. *EMBO Mol Med*. 7, 1450. (2015).
- Xiong, X. A scalable epitope tagging approach for high throughput ChIP-Seq analysis. *ACS Synth Biol*. (2017, Feb 19).
- Yan J, Enge M, Whittington T, Dave K, Liu J, Sur I, Schmierer B, Jolma A, Kivioja T, Taipale M, Taipale J. Transcription factor binding in human cells occurs in dense clusters formed around cohesion anchor sites. *Cell*. 2013; 154(4):801-13.

Chromatin profiling using ChIC/CUT&RUN

Introduction

The Henikoff lab has recently developed a new chromatin profiling methods: Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (Skene et al., 2018; Skene and Henikoff, 2017). This technique provides an exciting advance because it overcomes many of the drawbacks of conventional and widely used chromatin immunoprecipitation (ChIP) methods.

CUT&RUN is a genome-wide extension to chromatin immunocleavage (ChIC), which is a method developed by the Laemmli lab (Schmid et al., 2004). ChIC uses a Protein A-MNase fusion protein to cleave DNA regions associated with target proteins and recognized by specific antibodies. However, ChIC is limited to loci-specific analysis using Southern blotting.

A related method, chromatin endogenous cleavage (ChEC), uses a fusion between the protein of interest and MNase to analyze the protein's binding sites genome-wide (Schmid et al., 2004; Zentner et al., 2015). An obvious drawback of this method is the need to generate specific fusion proteins for each protein of interest.

ChIC/CUT&RUN is a major advancement over these two methods as it uses a recombinant Protein A/G-MNase tethered to the location of a protein of interest by an antibody (Meers et al., 2019; Skene and Henikoff, 2017). Importantly, this method recovers MNase-digested fragments and therefore is compatible with a sequencing-based genome-wide analysis of protein occupancy and histone modification positioning.

Advantages over ChIP-based methods include improved method simplicity achieved by the use of magnetic beads to immobilize nuclei, the compatibility with fresh and frozen tissue samples, and a shortened protocol (1–2 days) to generate material suitable for preparing DNA sequencing libraries. Furthermore, due to the *in situ* targeted cleavage of DNA on both sides of the protein of interest, only on-target DNA fragments are released from the nucleus and collected, leaving the off-target sequences behind. Thus, ChIC/CUT&RUN produces very little background signal compared to ChIP.

Since its development, ChIC/CUT&RUN has been adapted for a variety of experimental setups, including automation for high-throughput epigenetic profiling (Auto CUT&RUN) (Janssens et al., 2018), profiling of insoluble chromatin such as centromeric regions with ChIC/CUT&RUN. Salt (Thakur and Henikoff, 2018) and ChIC/CUT&RUN. ChIP for examining specific protein components within complexes released by CUT&RUN digestion (Brahma and Henikoff, 2019). Most certainly, the low input requirements and the high signal-to-noise mean that ChIC/CUT&RUN is compatible with single-cell analysis, for example, to investigate transcription factor occupancy in individual mouse embryo cells (Hainer and Fazio, 2019).

ChIC/CUT&RUN is a simple, versatile, and powerful method to profile DNA-protein interactions and should be in every molecular biologist's toolkit. This method can aid the genome-wide identification of specific gene and cis-regulatory elements marked by histone modifications or bound by a protein of interest, thereby providing insight into chromatin-based mechanisms of gene control. Due to the very low amount of starting material needed, ChIC/CUT&RUN is uniquely positioned to investigate rare cell types. On top of that, the easy-to-use and time-saving protocol allows for a quick turnaround, enabling the researcher to do multiple parallel analyses and, therefore, get a more comprehensive insight into the complexity of genome regulation.

Applications of ChIC/CUT&RUN

ChIC/CUT&RUN has been used to profile DNA-protein interactions in a variety of cell lines and tissue samples and a multitude of model organisms, including yeast, plants, and animal cells. So far, all these studies focused on mapping histone modifications and transcription factor binding sites.

Chromatin modifications

The first report of CUT&RUN profiled the histone modification H3K27me3 in a human cell line and H2A in yeast cells (Skene and Henikoff, 2017). Comparison of the equivalent data sets generated by ChIP-Seq and ChIC/CUT&RUN demonstrated close similarities between the two different chromatin-profiling techniques in discovering 'peak' regions.

Transcription factors and chromatin-associated complexes

Over the last decade, ChIP-seq has been the predominant method to identify the genome-wide location of transcription factor binding sites. Compared to ChIP, the new ChIC/CUT&RUN method is more cost-effective, quicker to use, require only a fraction of starting material including single cells, has a more favorable signal-to-noise ratio, detect more defined 'peaks', and is compatible with automation. This method is compatible with unfixed, native chromatin, even for transcription factor profiling, and therefore overcome some of the difficulties associated with cross-linked ChIP protocols. ChIC/CUT&RUN has been used to profile multiple transcription factors, such as CTCF and pluripotency factors, and large chromatin-associated complexes, such as Polycomb Repressive Complexes and chromatin remodellers.

Comparison of ChIC/CUT&RUN with ChIP-seq

Table 5. ChIC/CUT&RUN with ChIP seq

	ChIP-seq	ChIC/CUT&RUN
Starting material	Millions of cells in a standard protocol, with some low-input protocols reported	No more than 500,000 cells, with a standard of 50,000 cells. Availability of low input and single-cell protocols, compatible with automation.
Fixation	Standard for cross-linked protocols with formaldehyde, sometimes double fixation with DSG and formaldehyde	Not necessary or recommended
Nuclei isolation	Recommended in most protocols	Not necessary but compatible with methods
Sonication	Used in most cross-linked protocols to fragment and solubilize the chromatin	Not necessary
Lysis	Takes place before antibody-based immunoprecipitation	Takes place after antibody guided DNA fragmentation by MNase
Cleavage	No enzymatic cleavage in conventional cross-linked protocols, only in the native protocol using MNase to fragment DNA before immunoprecipitation	<i>In situ</i> DNA fragmentation using MNase or target regions identified with a specific antibody
Protocol length	About 1 week	Generally, 1 to 2 days
Secondary antibody	Not generally used	Not generally used, unless necessary due to low abundant target epitope
Sequencing library prep	End repair and adaptor ligation protocols necessary	End repair and adaptor ligation protocols for low-input DNA samples necessary
Sequencing depth	~ 20 million reads	~ 8 million reads
Cost	Relatively high cost per sample	Low cost per sample

ChIC/CUT&RUN protocol overview

During ChIC/CUT&RUN, an antibody against the protein of interest (such as a histone with a specific modification or a transcription factor) is used to guide Protein A/G-tagged MNase to the region in the genome where the protein is located. Activating the MNase then excises only those DNA sequences that are close to the protein of interest, and the small fragments are released from the nucleus. These DNA fragments are collected and used to generate NGS libraries with a low-input DNA library prep kit. The protocol can be separated into five parts.

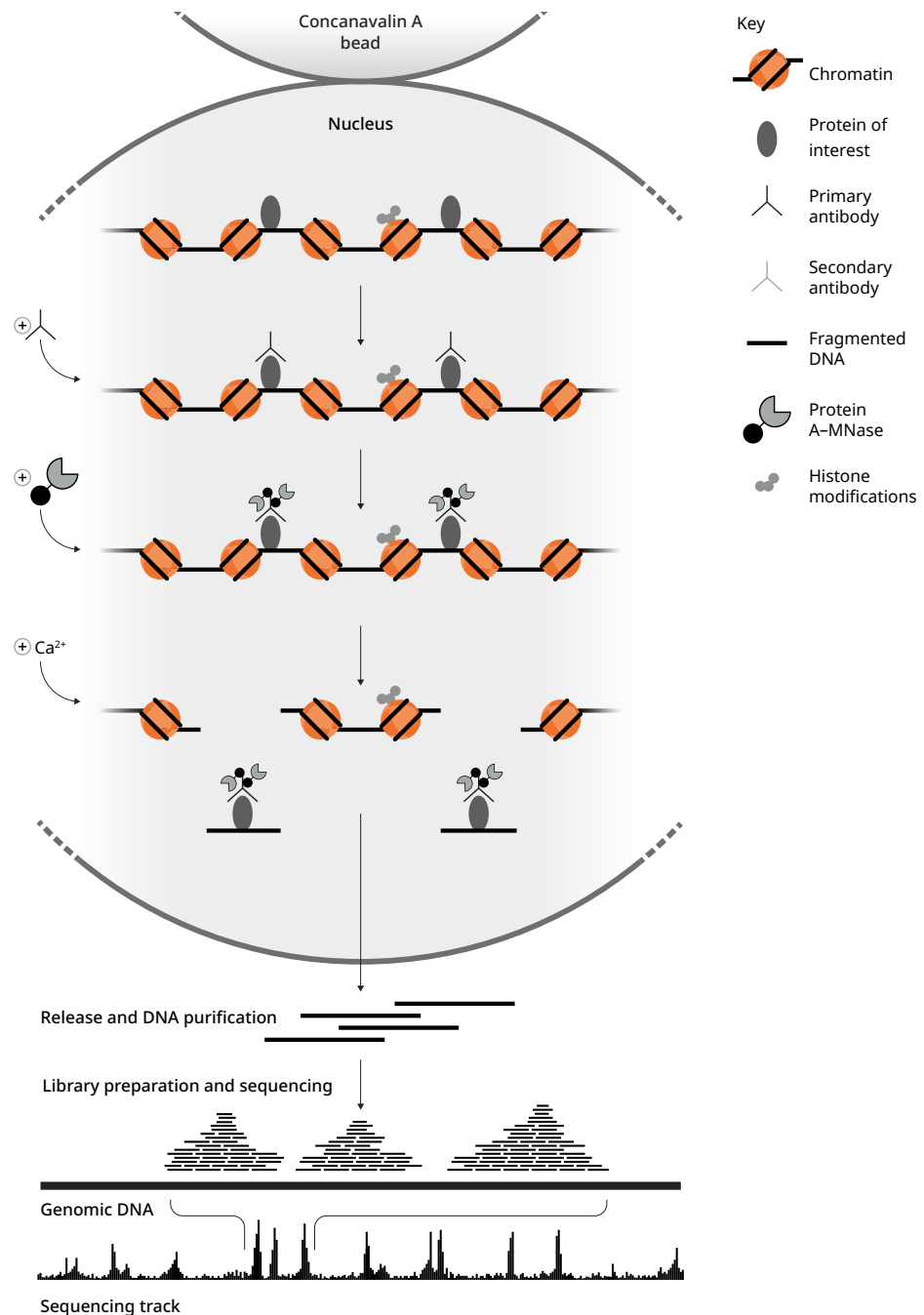


Figure 9. Schematic of the ChIC/CUT&RUN protocol. Nuclei are attached to magnetic Concanavalin A beads to allow ease of handling and safe liquid removal after each wash step. Nuclei are permeabilized and simultaneously incubated with an antibody against the protein of interest. Protein A/G-MNase fusion protein binds to the antibody against the protein of interest. When Ca^{2+} is added, the MNase cleaves the DNA on both sides of the formed complex and release DNA fragments that diffuse out of the nucleus. The DNA can be extracted and used in an end-repair and adapter ligation-based DNA library preparation. NGS informs the binding profile of the protein of interest via the frequency of sequences in a particular region.

1. Nuclei extraction and binding to beads

ChIC/CUT&RUN typically uses fresh, unfixed samples as starting material, but protocol adjustments can allow the use of samples cryopreserved in 10% DMSO (Janssens et al., 2018; Skene et al., 2018).

Very little material is necessary to perform the protocol, and the general recommendation is to start with less than 500,000 cells (mammalian).

Cells or nuclei released with the Henikoff recommended nuclear extraction buffer are bound to Concanavalin A magnetic beads, which have unique saccharide-binding properties. Other nuclear extraction protocols can be used to provide their compatibility (see suggestions below about the use of Triton X-100 containing buffers). Using magnetic stands allows for easy washing of the samples during the rest of the protocol.

2. Permeabilization and antibody binding

The nuclei bound to Concanavalin A beads are permeabilized and simultaneously incubated with an antibody against the protein of interest (primary antibody) in a buffer containing digitonin and EDTA. EDTA allows the rapid cessation of cell metabolism and thereby inhibits endogenous DNase activity, preserving the chromatin and reducing the overall background signal. The duration of this step can be adapted by the user; generally, ranging from 2 h to overnight. Antibody dilution of 1:100 or 0.5-1.0 µg is recommended as a starting point, but the amount of antibody used should be optimized. It is advised to always include positive (α-H3K27me3) and negative control (isotype control IgG) samples.

Of note, Protein A and Protein G have different binding efficiencies to different antibody species (see [here](#)). Even though antibody compatibility is increased by using Protein A/G over Protein A alone, it might be necessary to use a secondary antibody in some cases. A secondary antibody guides MNase to the target area by increasing the number of Protein A/G binding areas and thereby helps to recover low abundant target sequences.

3. MNase binding and cleavage of target sequences

To allow Protein A/G-MNase to be directed to the antibody-bound genomic target regions, the fusion protein is diluted in a wash buffer containing digitonin and incubated with the nuclei. Unbound enzyme fusion is then washed away, and MNase activity is activated at 0°C by adding Ca²⁺. Even though the cleavage itself is not particularly temperature-sensitive, the subsequent diffusion of the cleaved DNA fragments is sensitive, and a temperature rise would result in higher background. The digestion time can be adjusted if the final material contains a disproportional amount of high molecular weight fragments.

4. DNA recovery

The MNase activity is stopped by the addition of a STOP buffer containing EGTA. This buffer can optionally contain heterologous spike-in DNA to help calibrate the CUT&RUN profiles during data processing. The MNase generated fragments are released from the nuclei by increasing the incubation temperature and cleaned up using phenol/ chloroform/isoamyl alcohol extraction followed by ethanol precipitation. A variation of the protocol can further restrict the premature release of the MNase-fusion protein from its binding site, its diffusion, and potential for non-specific cleavage. This version of the protocol uses a combination of low salt buffers and a high Ca²⁺ concentration to activate MNase activity and is ideal for targets mainly found in active open chromatin, but it can also be used when an antibody shows high levels of background signal.

5. Library preparation and sequencing

Following clean-up of the cleaved DNA fragments, standard end repair and adapter ligation methods are used to generate low input DNA libraries according to the manufacturer's guidelines. The Henikoff lab originally recommended a TRUseq library preparation approach, and besides, there have now been multiple reports using the NEBNext® Ultra™ II DNA Library Kit. Alternatively, similar to the automated CUT&RUN protocol (Jenssens et al. 2018), the phenol/chloroform/isoamyl alcohol extraction steps can be omitted, and the released DNA fragments can be directly used in the end repair and adapter ligation-based protocol.

The size distribution and concentration of libraries can be determined by running capillary electrophoresis (eg Bioanalyzer or TapeStation). Multiple libraries can be pooled together to obtain about 8 million paired-end sequencing reads per library. Due to the low background of ChIC/CUT&RUN libraries, 8 million paired-end reads are enough to profile histone modifications and even transcription factors.

Most conventional ChIP-seq data analysis tools can analyze ChIC/CUT&RUN data. There are a few analysis tools, such as the SEACR peak caller (Meers et al., 2019b), which are specifically designed by the Henikoff lab for ChIC/CUT&RUN data. For calibrated ChIC/CUT&RUN, the heterologous spike-in DNA added with the STOP buffer can be used to normalize signals across samples. Alternatively, E. coli DNA carried over from the recombinantly-produced Protein A/G-MNase can be used as a spike-in.

Sample preparation

In general, sample preparation is straightforward for ChIC/CUT&RUN as this method can use fresh, unfixed samples. A single-cell suspension needs to be achieved in a way that is suitable for the cell type used. This can include using dissociation reagents like Accutase™, scraping cells off cell culture dishes, or mechanically dissociating tissues.

For ease of collection, samples can be cryopreserved in 10% DMSO and frozen down in a Mr. Frosty isopropyl alcohol chamber. Sample fixation is not required for ChIC/CUT&RUN. However, if the beads clump during washes, then a light fixation of the sample using 0.1% formaldehyde for only 2 min at room temperature before antibody incubation is beneficial.

Antibody selection

As for ChIP, not all antibodies will work for ChIC/CUT&RUN. Most antibodies are not yet tested for compatibility with this newer method, therefore testing and optimization will be required by the end user. ChIP-grade antibodies seem to largely work for ChIC/CUT&RUN, especially when shown to work for native ChIP.

When trialling/testing antibodies that are not labelled as ChIP grade, assuming specificity is fully characterized, then it's good to start with choosing antibodies that recognize the native form of the protein of interest; for example, antibodies that have been shown to work in immunoprecipitation or immunocytochemistry. Equally, when considering the concentration of antibody to use in a ChIC/CUT&RUN experiment, the concentration recommended for immunofluorescence-based methods seems to be a good starting point. Compatibility to Protein A and Protein G needs to be checked, and then appropriate secondary antibodies need to be used if necessary.

Controls

As with any other experiment type, including the correct controls is critical to ensure the experiment has worked as expected as well as to easily pinpoint areas to troubleshoot should the experiment fail.

Unlike ChIP, no input sample needs to be included in the protocols as a non-antibody guided MNase treatment would simply result in the identification of accessible chromatin. An IgG control should be included to get the background from the sample and set the baseline for the experiment. In terms of antibody controls, the Henikoff lab recommends using H3K27me3 as a positive control for ChIC/CUT&RUN experiments. The use of total unmodified histone controls, such as total H3, can be considered to allow the proportional representation of histone modifications.

Optimization

The optimization of ChIC/CUT&RUN protocol is required at several stages when using this technique.

Permeabilization

In the initial versions of the protocols, the digitonin concentration used in the wash buffers needed to be tested to ensure the efficient permeabilization of nuclei. This test might not be necessary if you follow improvements to the protocols, specifically by adding NP40 to the wash buffers. However, the efficiency of nuclei permeabilization should be kept in mind when using this technique on new material.

Antibody concentration

It is important to titrate the amount of antibody used per reaction, starting with recommended dilutions for ChIP or immunofluorescence assays.

Duration of antibody incubation

Primary antibody incubation for 1 h at room temperature is sufficient but this can be extended to 1 to 5 days in the cold room. This step can be optimized for the antibodies used in each experiment, keeping in mind that prolonged incubation might lead to an increase in background signal and, therefore, a less favorable signal-to-noise ratio overall.

Use of a secondary antibody

The use of a secondary antibody can be considered for ChIC/CUT&RUN if experiencing low recovery with primary antibodies alone. Another reason to include a secondary antibody is to avoid an unfavorable pairing of Protein A/G with the primary antibody, which can be achieved by using a more favorable secondary antibody class.

Digestion time

MNase digestion time can be adjusted depending on the protein of interest. Less abundant proteins may need more time to allow for the full recovery of all sites. It is important to keep in mind that prolonged digestion might result in untargeted cleavage and, therefore, higher background signals.

Step-by-step protocol links

- [Standard CUT&RUN protocol provided by the Henikoff lab \(version 3\)](#)
- [Specialized ChIC/CUT&RUN protocol to use with Drosophila tissue](#)
- [ChIC/CUT&RUN compatible library preparation protocol using NEBNext® Ultra™ II DNA library prep Kit](#)

FAQs

How many cells should I use as a starting material? The genome of my model organism is smaller than human/mouse, should I increase the cell number to start with?

Exceeding 500,000 cells per sample is not recommended. Generally, a good starting point is 50,000 cells. Unlike ChIP, ChIC/CUT&RUN is very sensitive and does not need millions of cells as a starting material. Using too many cells will reduce the yield and, more importantly, the complexity of the library.

If I use more cells, should I increase the amount of Concanavalin A beads I use?

The number of cells used per experiment should not exceed 500,000 cells. Since the number of beads in the protocol is optimized for the use of 50,000 to 500,000 cells, there should be no need to increase the number of beads recommended in the protocol.

How should I cryopreserve my cells to make them compatible with ChIC/CUT&RUN?

It is recommended to use 10% DMSO in an appropriate buffer or media and slowly freeze using a Mr. Frosty isopropyl alcohol chamber. Flash freezing is not recommended.

I have cells fixed and frozen for ChIP, can I use them for ChIC/CUT&RUN?

Depending on the fixation and cryopreservation conditions used for the ChIP sample, it might be possible. However, it is unlikely to work with the standard protocols because cell numbers for ChIP samples are likely to exceed recommendations for ChIC/CUT&RUN samples and the strong fixation conditions (eg double fixation and quenching) used will impair MNase.

The problem of cell numbers in ChIP samples exceeding the recommended cell starting material for ChIC/CUT&RUN can be easily circumvented by splitting the sample into multiples. However, the other problems of fixation and freezing of samples remain. Fixation used in ChIP protocols is likely too harsh to be compatible with either ChIC/CUT&RUN, as it will most likely impair the ability of the MNase to fragment the target sequences, and multiple reports have hinted at this incompatibility. Additionally, there is a possibility of epitope masking due to fixation. Sample fixation is, therefore, not recommended. When cryopreserving cells for ChIC/CUT&RUN, the use of 10% DMSO and a Mr. Frosty isopropyl alcohol chamber to gently freeze the samples is required. Flash freezing included in most ChIP protocols is not recommended.

My beads clump during the washes, what am I doing wrong?

The problem of clumping can be caused by using too high a ratio of cells to beads or the lysis of nuclei/cells in the digitonin containing buffer, which would release DNA and lead to clumping. The first thing would be to check that the recommended number of cells is not exceeded. Furthermore, a light fixation with formaldehyde (0.1% formaldehyde for 2 min at room temperature) before incubation with the antibody is recommended to reduce the clumping of beads in digitonin containing wash buffers. However, keep in mind that fixation could affect epitope availability and, therefore, will need to be tested for each antibody.

Why is there EDTA in the buffer for the primary antibody incubation?

The addition of EDTA in the permeabilization and antibody incubation buffer is recommended as it chelates Mg^{2+} and thereby stops any ATP-dependent cellular processes, including replication and chromatin remodelling, and it will also stop endogenous DNases.

Where can I get the Protein A/G-MNase from?

Both fusion proteins were initially provided by the Henikoff lab. There are protocols on how to make your own fusion proteins (Meers et al., 2019a), whereas the plasmids, Protein A-MNase and Protein A/G-MNase, are all commercially available from reagent suppliers.

Concerning E. coli DNA carryover, what is a normal percentage of E.coli reads to sample reads?

The carryover depends on the amount of starting cells as well as the abundance of the antibody epitope, with an inverse correlation between the number of starting cells and the number of sequencing reads that map to E. coli DNA. The carryover also depends on the amount of Concanavalin A beads used, whereby the more beads used, the more E. coli reads. Overall, calibration using E. coli carryover DNA should only be used for experiments that are comparable in the number of cells and beads used.

As a guide, the Henikoff lab has reported a range of E. coli read percentages from 0.01% to 11.5% of the total reads, with IgG samples, usually having a higher amount of E. coli reads (2% to 11.5%). They furthermore reported that when using only a few hundred or thousand cells in an experiment, the percentage of E. coli reads can reach about 30% to 70%. Additionally, when testing a commercially available fusion protein they only got 0.01% of the total reads back as E. coli reads, which might be too little to use for data normalization.

Is there a way of quality controlling the experiment before sequencing?

Similar to ChIP, qPCR could be used to examine carefully designed regions of interest. Keep in mind that the reactions need to be performed on the amplified library and not on the generated DNA fragment pool. In the case of ChIC/CUT&RUN, the released fragments often contain large genomic regions that have not been directly targeted by the antibody-directed MNase. These fragments will not contribute to the libraries prepared from these samples, but they would be amplified by qPCR. To avoid false positives, the template for qPCR should be an amplified library, and appropriate controls should be carried out as well.

The Bioanalyzer/TapeStation trace for my transcription factor experiment looks just like the IgG control, what did I do wrong?

Good news, most likely you didn't do anything wrong! Provided that your positive control sample (eg H3K27me3) returned a nice nucleosomal patterning, the only way to know whether your transcription factor sample has worked is to sequence the samples. Unfortunately, it is quite common for transcription factor libraries to look very similar to the IgG libraries, and yet after sequencing, they still produce good quality data sets. You could opt for a qPCR approach to compare signals in the IgG versus the transcription factor libraries for known transcription factor target sites.

References

- Brahma, S., Henikoff, S. RSC-associated subnucleosomes define MNase-sensitive promoters in yeast. *Mol Cell* 73, 238–249, e233 (2019).
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213–1218 (2013).
- Hainer, S.J., Fazzio, T.G. High-resolution chromatin profiling using CUT&RUN. *Curr Protoc Mol Biol* 126, e85 (2019).
- Janssens, D.H., et al. Automated in situ chromatin profiling efficiently resolves cell types and gene regulatory programs. *Epigenet Chromatin* 11, 74 (2018).
- Meers, M.P., Bryson, T.D., Henikoff, J.G., Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* 8: e46314 (2019).
- Meers, M.P., Tenenbaum, D., Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenet Chromatin* 12, 42 (2019).
- Schmid, M., Durussel, T., Laemmli, U.K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* 16, 147–157 (2004).
- Skene, P.J., Henikoff, J.G., Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 13, 1006–1019 (2018).
- Skene, P.J., Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6: e21856 (2017).
- Thakur, J., Henikoff, S. Unexpected conformational variations of the human centromeric chromatin complex. *Genes Dev* 32, 20–25 (2018).
- Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., Henikoff, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat Commun* 6, 8733 (2015).

DNA methylation and demethylation

Introduction

Throughout DNA, chemical modifications add a layer of regulation to the expression of genes encoded within the DNA sequence. The most well-studied of these chemical modifications is 5-methylcytosine (5mC), a modification most commonly recognized as a stable, repressive regulator of gene expression. The human genome consists of approximately 1% methylated cytosine making it the most abundant and widespread DNA modification (Moore et al 2012). There are several methods available to sequence 5mC throughout the genome, all of which have pros and cons, which we will discuss later in this guide. These methods include high-resolution approaches, such as whole-genome bisulfite sequencing, and antibody-dependent DNA immunoprecipitation (DIP) or MeDIP.

5mC was initially discovered to reside within CpG islands – stretches of DNA commonly found within promoter regions enriched in CpG dinucleotides. It is within these promoter regions that 5mC acts as a stable epigenetic mark repressing gene transcription. Within the mammalian genome, methylated cytosine is initially incorporated into the DNA during early development by the de novo methyltransferase enzymes DNMT3a and DNMT3b (Okano et al 1999). These methylation marks are maintained throughout the genome by an additional methyltransferase, DNMT1, which copies DNA methylation patterns to daughter strands during DNA replication (Vertino et al 1996).

Today the notion of 5mC being an entirely stable DNA modification is less concrete. Many methylated cytosines throughout the genome, particularly within gene bodies, undergo a process known as DNA demethylation – a process that ultimately results in the removal of 5mC back to an unmodified cytosine (C). DNA demethylation can occur in one of two ways: passive DNA demethylation, where methylated cytosine is diluted from the genome due to an absence of methylation maintenance enzymes. Or active DNA demethylation, which involves the oxidation of 5mC by ten-eleven translocation (TET) enzymes into oxidized derivatives of 5mC (reviewed in Wu et al 2017).

Active DNA demethylation occurs in a cycle, starting with 5mC and finishing with an unmodified C. 5mC is initially oxidized to 5-hydroxymethylcytosine (5hmC), which is further oxidized to 5-formylcytosine (5fC), and finally this is oxidized once more to 5-carboxylcytosine (5caC). 5fC and 5caC can be removed from DNA by thymine DNA glycosylase (TDG) in combination with base excision repair (BER) to result in an unmodified C (figure 10). 5hmC, 5fC, and 5caC have been the focus of many recent epigenetic studies. More and more are being found out about these epigenetic marks, including the potential for them to have stable epigenetic roles. Many sequencing methods have been developed to distinguish these marks throughout the genome including variations on MeDIP using 5hmC, 5fC, and 5caC antibodies, and variations on bisulfite sequencing such as TET assisted bisulfite sequencing (TAB-seq). The differences between these methods will be discussed later in this guide.

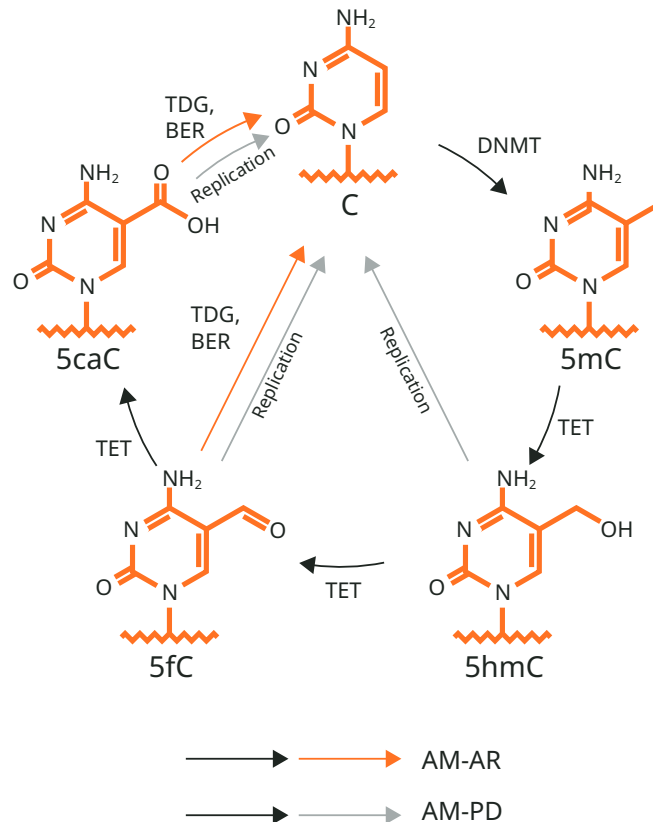


Figure 10. Cycle of DNA demethylation. Active DNA demethylation occurs by thymine DNA glycosylase (TDG) coupled with base excision repair (BER) or replication-dependent dilution of 5hmC, 5fC or 5caC. Active modification-passive dilution (AM-PD). active modification-active removal (AM-AR).

Bisulfite sequencing

It is not possible to detect 5mC using traditional DNA amplification approaches because the mark is not maintained during sample preparation and amplification. Bisulfite conversion is one of the most widely used approaches to convert DNA methylation marks into a suitable template for amplification and downstream analysis. Bisulfite conversion uses the treatment of DNA with NaOH and sodium bisulfite in a chemical reaction that converts cytosine bases into uracil (U), while methylated cytosines are protected from the conversion (figure 11).

During downstream analysis such as PCR or sequencing, unmethylated C bases that undergo deamination in the bisulfite reaction will be interpreted as thymine (T), whereas 5-mC bases will remain unchanged and still be detected as a C by the sequencing output. This allows you to determine the locations in the genome containing methylated cytosine (Frommer *et al.*, 1992)

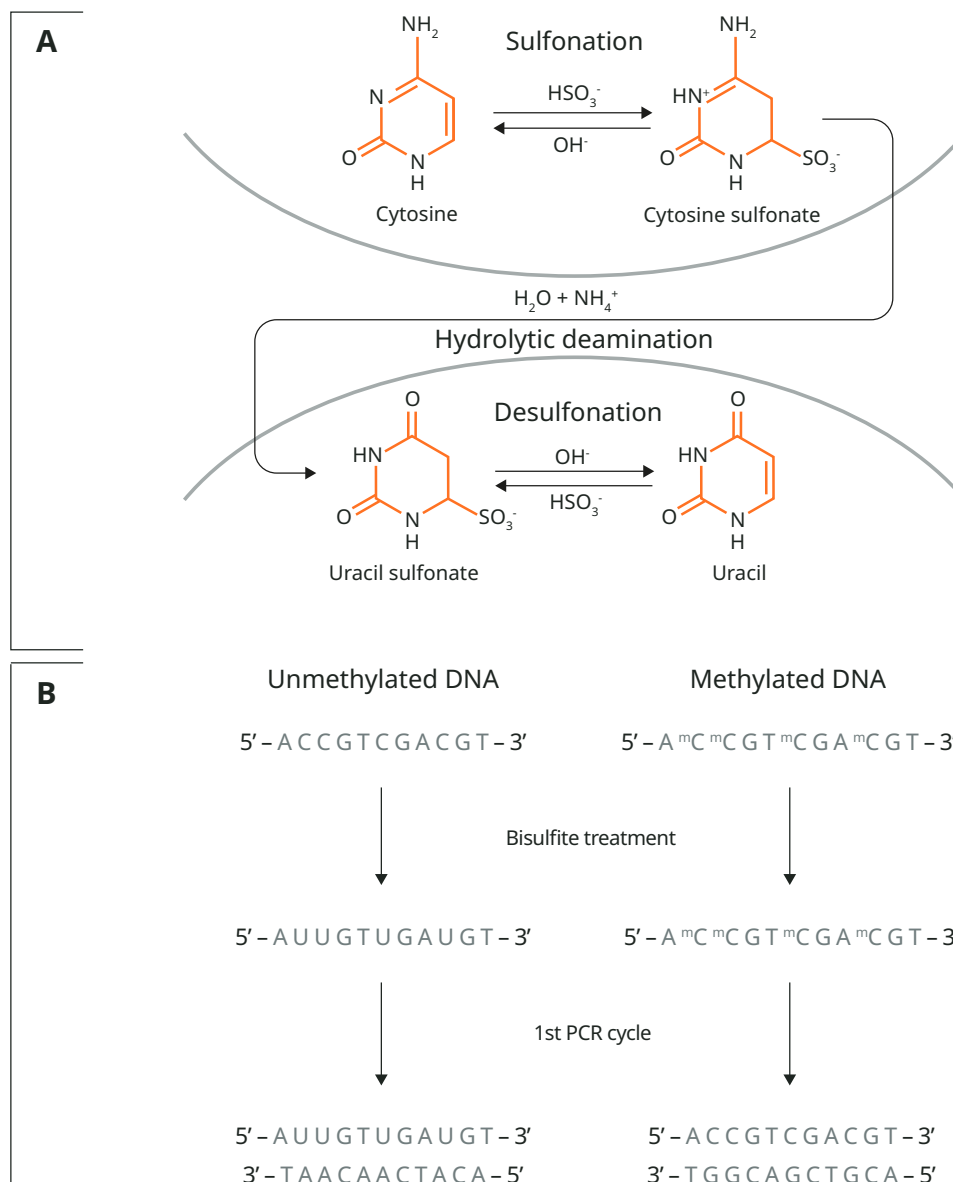


Figure 11. Bisulfite conversion. Treatment of DNA with bisulfite (sulphonation) leads to the deamination of cytosine residues and converts them to uracil, (A) while 5-methylcytosine residues remain the same (B)

Bisulfite-based applications

Bisulfite conversion has become the basis for several variations and applications designed for high throughput applications or the investigation of broader, whole genome-scale regions.

Here are some examples of bisulfite-based methods.

Genome-wide DNA methylation analysis

Whole genome bisulfite sequencing (WGBS; Lister et al 2009)

- Applies next-generation sequencing (NGS) techniques to bisulfite-converted input samples.
- WGBS produces single-base resolution DNA methylation maps that span the entire genome of an organism.

Reduced representation bisulfite sequencing (RRBS; Meissner et al., 2005)

- Combines the single-base resolution of bisulfite, and the genome-scale coverage of high throughput sequencing, with the use of methylation-sensitive restriction enzymes to enrich samples for high CpG content.
- Effectively limits sequencing to only the regions of high interest, where DNA methylation exists.

Targeted DNA methylation analysis

Methylation-specific PCR (MS-PCR; Herman et al., 1996)

- Applies PCR primers specific to bisulfite converted DNA templates that are either methylated or unmethylated. The differential PCR amplification indicates if DNA methylation modifications are present.

Pyrosequencing (Colella et al., 2003; Tost et al., 2007)

- Also known as sequencing by synthesis and can interrogate bisulfite-converted DNA at a specific region of interest. The level of 5mC is determined by comparing the ratio of C and T bases at an individual locus.

High resolution melting (HRM) analysis (Wojdacz and Dobrovic, 2007)

- Originally applied to SNP detection, but the process has also been adopted for DNA methylation. The real-time PCR-based protocol measures melting temperatures of PCR amplicons. The shift in melting temperatures, which vary on C-T content, corresponds to the level of DNA methylation in the sample.

Methylation-sensitive single-nucleotide primer extension (MS-SnuPE; Gonzalgo and Jones PA, 1997)

- Queries a CpG of interest by targeting bisulfite specific primers to the sequence immediately preceding a CpG. DNA polymerase terminating dideoxynucleotides allow the primer to extend a single base, which then can be quantitatively measured to determine C-T content, determining its DNA methylation status.

Bisulfite conversion: technical considerations

Incomplete conversion

Bisulfite conversion is a very powerful method because it is relatively simple to perform, and it can deliver single-base resolution of DNA methylation status. However, the method does have some drawbacks: incomplete conversion (or on occasion, over-conversion) can occur under sub-optimal reaction conditions leading to insufficient DNA denaturation, or when the DNA strands re-anneal before completion of the reaction.

Distinguishing 5hmC

DNA degradation is often a byproduct of the harsh bisulfite conversion reaction conditions, which can make working with smaller samples challenging. Insufficient desulfonation of the reaction will leave behind residues that can inhibit DNA polymerases used in PCR. Recent evidence indicates that bisulfite conversion does not distinguish between 5mC and 5hmC. Bisulfite conversion therefore lowers the overall complexity of the DNA sequence. This reduction sequence complexity can complicate primer design for downstream PCR-based interrogation or introduce challenges when attempting to uniquely map sequencing reads to a reference genome.

DNA immunoprecipitation (DIP)

Another method commonly used to map the location of DNA methylation marks is DIP. DIP relies heavily on having antibodies capable of recognizing the DNA modifications of interest. However, once you have this, DIP is a straightforward and effective method. It is also considerably cheaper and easier to analyze compared to WGBS sequencing, which requires the whole genome to be sequenced. DIP only requires sequencing of the small sheared DNA regions pulled down in your IP step.

DIP has been successfully carried out for the most well-characterized DNA modifications: 5mC, 5hmC, 5fC, and 5caC (Pastor et al., 2011, Shen et al., 2013). It has been used in a range of samples, including embryonic stem (ES) cells, brain tissue, and zebrafish fish embryos. The method is similar to ChIP, but your starting material is raw genomic DNA with no chromatin required. This genomic DNA will undergo shearing to approximately 150–300bp, and then this sheared DNA can undergo heat denaturation. This step is essential as the antibody will only be able to access the modifications within denatured (open) DNA.

After DNA denaturation the sheared DNA is incubated with the antibody recognizing your modification of interest, usually overnight, and then the samples undergo an IP step to pull down all the DNA bound to the antibody and washing away any unbound DNA. We recommend using magnetic beads for this type of IP step. When you carry out DIP, it is important to treat your initial genomic DNA with RNase to remove any RNA from the samples.

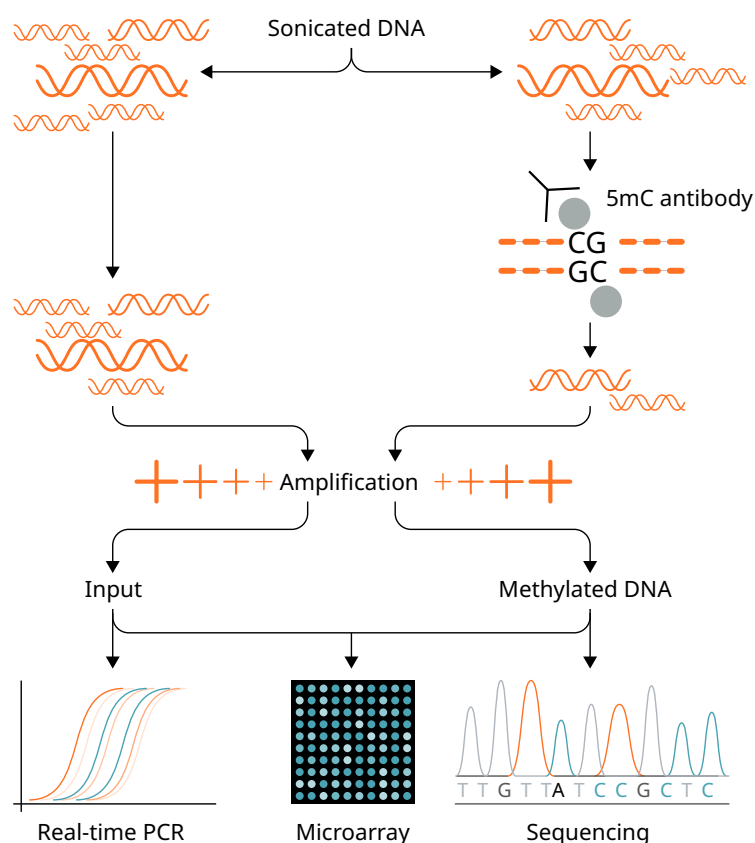


Figure 12. DIP methodology. Genomic DNA is sheared, and immunoprecipitation is carried out using antibodies against your DNA modification. Pulldown DNA and input samples can then be used for qPCR, microarray, or NGS.

DIP-based applications

Genome-wide DIP analysis

DIP-sequencing (Pastor et al., 2011)

- DIP is combined with NGS to map the location of DNA modifications across the whole genome.
- Due to the conservation of these chemical structures between species, it has been easy for researchers to sequence their DNA modification of choice in any organism.
- The library prep and analysis of DIP-sequencing is very similar to that of ChIP-sequencing.
- The small DNA fragments pulled down in your IP are used in library prep, and these can be sequenced at a relatively low read depth compared to WGBS as you are more selective about what you sequence, ie only regions bound to your antibody.

Targeted DIP analysis

DIP-PCR (Pastor et al., 2011)

- The pulldown DNA you obtain from your IP as described in the DIP-sequencing section above. However, this time instead of using the sheared DNA for library prep you can use it in a qPCR as template DNA.
- When you design primers for this type of DNA you have to consider that the template being genomic DNA will contain both exons and introns. This method can be very effective to determine levels of a modification across samples.
- It can be tricky to compare levels of different modifications as many factors including antibody affinity can affect this.

DIP: technical considerations

Shear your samples appropriately

Unlike WGBS, DIP is not single-base resolution. When you are shearing your DNA samples, it is important to get these DNA fragments to a good size of between 150–300bp, to try to improve the resolution of your DIP sequencing. Having larger fragments means you will inevitably pull down more DNA flanking your DNA modification of interest and not physically bound to it. This results in broad, unspecific peaks in your sequencing analysis.

Source high-performing antibodies

An antibody specific to your modification of interest is essential for DIP. Make sure there is minimal cross-reactivity with similar modifications, for example 5fC and 5hmC. The use of antibodies for this type of sequencing has the advantage that you are only limited by the antibodies available to you. So if you wanted to investigate a modification not previously characterized in DNA, eg m6A (more commonly associated with RNA), you could do so provided that you have a specific m6A antibody.

Alternate methods to capture 5hmC, 5fC, and 5caC

The biggest drawback of traditional bisulfite sequencing is that it is unable to distinguish the oxidized derivatives of 5mC and will profile only 5mC itself. Fortunately, there have been many variations on bisulfite sequencing and some entirely new approaches to tackling the problem of sequencing 5hmC, 5fC, and 5caC. Here we look at some of these new methods in more detail.

5hmC mapping

Tet-assisted bisulfite sequencing (TAB-seq; Yu et al., 2012)

- This method relies on the conversion of 5hmC into 5gmC. The addition of glucose in this glucosylation reaction acts to protect the 5hmC.
- TET enzymes are then added to the genomic DNA to convert all 5mC and 5fC present into 5caC. After Bisulphite conversion 5hmC is read as C.
- 5caC and unmethylated cytosines are all read as T. This method gives a clear differentiation between 5mC and 5hmC.
- The problems with this method are that all the conversions to T can make it difficult to map the end sequences produced. It also requires very deep sequencing to get a full coverage of the genome, so this can be more expensive than other methods.

Oxidative bisulfite sequencing (oxBS-seq; Booth et al., 2012)

- This is another method for detecting 5hmC at single-base resolution and uses potassium perruthenate (KRuO₄) to chemically convert 5hmC to 5fC.
- After this conversion, all 5mC remains unchanged. Subsequent bisulfite treatment and sequencing allows you to distinguish between 5mC and 5hmC sites by comparing the KRuO₄ treated and untreated samples.

hMe-Seal (Song et al., 2011)

- Similar to TAB-seq, hMe-Seal starts with the glucosylation of 5hmC to 5gmC, but the added glucose molecule is engineered to contain an azide group that can be chemically modified with biotin.
- 5hmC can then be enriched within the genome using the tight binding between biotin and streptavidin to carry out a pull-down for 5hmC.

Selective chemical labeling with exonuclease (SCL-exo; Sérandour et al., 2016)

- The initial steps for this are the same as hMe-Seal, azide-glucose glycosylation of 5hmC.
- Azide reaction with biotin allows for the 5hmC present to be linked to streptavidin however in this method the captured DNA undergoes exonuclease digestion which will stall at the biotin-5gmCs.

5fC/5caC mapping

M.SssI methylase-assisted bisulfite sequencing (MAB-seq; Wu et al., 2014)

- This method can quantitatively measure 5fC and 5caC at single-base resolution. This is achieved using M.SssI methyltransferase on your DNA to convert all unmodified cytosine to 5mC.
- After bisulfite-treatment all newly modified Cs, 5mC and 5hmC will be read in the sequencing as C, but all the 5fC and 5caC in the genome will be read as T.
- If you compare this to sequencing carried out without M.SssI treatment you can determine where the 5fC and 5caC modifications are within the genome.
- The biggest problem with this method is that it doesn't differentiate between 5fC and 5caC.

5fC chemically assisted bisulfite sequencing (fCAB-seq; Song *et al.*, 2013)

- This technique relies on the chemical protection of 5fC using O-ethylhydroxylamine (EtONH₂).
- This protection prevents bisulfite-mediated deamination of 5fC, and so this will appear as a C in the sequencing results (the same as 5mC and 5hmC).
- When this is compared to a sample not treated with EtONH₂ (where all 5fC modifications would be read as T), you can distinguish all the sites in the genome which have a 5fC.

5caC chemically assisted bisulfite sequencing (caCAB-seq; Lu *et al.*, 2013)

- caCAB-seq uses the modification of 5caC within the genome using 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC) to catalyze the formation of amide bonds to 5caC.
- This chemical modification prevents deamination of 5caC after bisulfite conversion allowing for it to be distinguished from 5fC in the sequencing.

Chemical-labeling-enabled C-to-T conversion sequencing (CLEVER-seq; Zhu *et al.*, 2017)

- CLEVER-seq is not only single-base resolution but can also be used on single cells. It is just for sequencing 5fC distribution and not 5caC.
- This method uses malononitrile to selectively label 5fC creating a 5fC-M adduct which is read as a T in the sequencing.

Comparison of DNA modification sequencing methods

It is important that you choose the best method for detecting DNA modifications that suit your needs. Consider things like whether you need single-base resolution, if you need to be able to quantify the absolute levels of the modification, and how feasible the method will be to use in your model system or sample type. Below you can find a table where we have summarized these key features for some of the available methods for sequencing 5hmC, 5fC, and 5caC.

Table 6. DNA modification sequencing methods**5hmC mapping only**

Name	Description	Single base resolution?	Allows absolute quantification of the modification?	Reference
5hmC-DIP	Using 5hmC specific antibodies to enrich for 5hmC.	No	No	Pastor, W. A. et al Nature 2011
TAB-seq	5hmC is converted to 5gmC to protect it. 5mC is converted to 5caC by TET enzymes. After Bisulphite conversion 5hmC is read as C. 5mC and 5caC are read as T.	Yes	Yes	Yu, M. et al Cell, 2012
oxBS-seq	Chemical conversion of 5hmC to 5fC using K ₂ Cr ₂ O ₇ allows the differentiation of 5mC and 5hmC at single base resolution.	Yes	Yes	Booth, M. J. et al Science, 2012
hMe-Seal	Glucosylation of 5hmC with an azide-containing glucose molecule and biotin allows for 5hmC enrichment using a biotin/streptavidin pulldown.	No	No	Song, C. X. et al. Nature Biotechnology 2011
SCL-exo	Azide-glucose glycosylation of 5hmC followed by a biotin reaction allows endonuclease activity to stall at biotin-5gmCs.	Yes	No	Sérandour, A. A. et al. Genome Biology 2016

5fC and 5caC mapping

Name	Description	Single base resolution?	Allows absolute quantification of the modification?	Reference
5fC/5caC DIP	Using 5fC and 5caC specific antibodies to enrich for these marks.	No	No	Shen, L. et al. Cell 2013
MAB-seq	M.SssI treatment of DNA converts all C into 5mC. Bisulfite conversion will then cause all C, 5mC, and 5hmC to read as C. All 5fC and 5caC will read as T's.	Yes	Yes	Wu, H., Nature Biotechnology 2014
fCAB-seq	EtONH ₂ protects all 5fC in the genome from oxidation after bisulfite treatment.	Yes	Yes	Song, X. et al. Cell 2013
caCAB-seq	EDC is used to catalyze the formation of amide bonds to 5caC preventing deamination of 5caC on bisulfite conversion.	Yes	Yes	Lu, X. et al. JACS 2013
CLEVER-seq	Malononitrile selectively labels 5fC creating a 5fC-M adduct which is read as a T in the sequencing.	Yes	Yes	Zhu, C. et al. Cell Stem Cell 2017

Liquid chromatography tandem-mass spectrometry (LC/MS-MS)

If you have access to LC/MS-MS, then this is the best way quantify the amount of a DNA modification within total genomic DNA (Le et al 2011 and Fernandez *et al.*, 2018). Using absolute quantification methods, LC/MS-MS gives you parallel quantification of all the DNA modifications found in total DNA from any organism and cell type (Zhang et al 2012). For absolute quantification, you are only limited which isotopic standards you have available to use as a standard to measure your sample against.

Using this technique combined with DIP (DIP-MS) allows you to determine if your DNA modification antibody is binding to your modification of interest and it will also allow you to see if it binds any other non-specific modifications. If you generate LC/MS-MS data of your DIP input and pull-down samples, you should see an enrichment of your modification of interest in the pulldown sample compared to the input. You can also then check other modifications with these same data to see if anything else came out as enriched in your samples to test for non-specific antibody binding. There is software being developed now that can even help you with this type of analysis.

LC/MS-MS: technical considerations

Technically challenging

MS equipment is costly and very specialized. The machine itself will require an enormous amount of maintenance and often requires its own technician to keep on top of things. Operating the machine is complicated and requires specialized training so it may be difficult to obtain this type of MS data on your own. Consider obtaining this data through collaborations or paid services if it is not feasible for you to purchase your own LC/MS-MS equipment.

DNA modification IHC/ICC

It is also possible to carry out IHC/ICC for DNA modifications. This can be done with a few simple additions to your standard IHC/ICC protocol. The most significant difference you will need to consider is that antibodies against DNA modifications cannot access and bind to the modification if it sits within double-stranded DNA. This means that you will need to denature the DNA making it single-stranded and accessible by the antibodies.

The most common form of DNA denaturation is to treat your samples with acid. This is usually 4N hydrochloric acid (HCL) applied directly to your IHC/ICC slides (Yamaguchi *et al.*, 2013 and Kaefer *et al.*, 2016). The best time to add this step to your protocol is before the addition of the primary antibody. Once you have permeabilized your cells or tissues with a detergent (eg PBS 0.1% Triton) you can wash and add 4N HCl to denature the DNA strands. It is important to thoroughly wash the acid off once the step is complete and neutralize the acid with an alkali (eg 100 μ M NaOH in PBS). After the acid is washed and neutralized you can proceed with your usual IHC/ICC steps and add the primary antibody.

When carrying out an IHC/ICC for DNA modifications you should also be wary that your antibody may recognize very similar modifications on RNA (eg 5mC on DNA and m5C on RNA). To avoid this problem, you can treat your samples with an RNase step to remove all RNA present. Again, this step should be optimized as leaving your sample in RNase for too long can also cause damage to the DNA present.

DNA modification IHC/ICC: technical considerations

Time your acid step

- It is crucial that you optimize the concentration and timings of the acid step before you start using your experimental samples.
- Too long in an acid treatment will ruin the samples, but the timing needs to be long enough to denature the DNA fully.
- Tissue samples will require the acid treatment for longer than cells used for ICC. You should try a range of timings from 10 minutes up to 40 minutes and see how your signal looks after this.
- ICC should only need 5–10 minutes maximum but again, it is important to test this first and optimize correctly.

Double IHC/ICC

- It can be difficult to carry out double IHC/ICC with a DNA modification given the effects of the acid treatment step. The acid treatment may denature proteins present in the sample or degrade epitopes required for recognition by your second primary antibody.
- If you want to carry out such a double immunoassay it will require careful optimization. Try to minimize the amount of time your sample spends in the acid treatment step to reduce damage to other proteins. You could also consider doing the primary antibody steps sequentially.
- For example, after you have applied the first primary antibody, fix this with a formaldehyde-based fixative before the acid treatment step and adding the second primary antibody (eg the DNA modification antibody).

Choose the right DNA stain

- You may find that because of the acid treatment step you cannot use your standard DNA stain. For example, DAPI may not bind so well as this recognizes the adenine-thymine bases present within double-stranded DNA.
- A good alternative commonly found in most labs is propidium iodide (PI). PI will recognize both double-stranded and single-stranded nucleotide chains.
- This means it will also pick up any RNA in your samples, so watch out for this. You can also find many commercially available DNA stains which will recognize single-stranded DNA.

Methyl binding domain proteins (MBDs)

5mC and its oxidized derivatives play an important role in gene silencing and promoting gene expression after DNA demethylation. It is now known that some of these DNA modifications can act as markers to recruit proteins to specific DNA sites, altering gene expression and acting as epigenetic marks. MBD3 and methyl CpG binding protein 2 (MECP2) have both been shown to bind 5hmC in addition to 5mC. Once bound to 5hmC they play a role in DNA accessibility and activation of transcription (Yildirim *et al.*, 2011 and Mellén *et al.*, 2012).

A common method to screen for binders of a DNA modification is to use a pull-down technique followed by MS to screen for any proteins pulled down. This method has been successfully used to find binders of 5mC, 5hmC, and 5fC (Iurlaro *et al.*, 2013 and Sprujit *et al.*, 2013). For this experiment, you need to create a synthetic DNA bait containing the modification you are interested in as well as baits containing other modifications and unmodified cytosine to act as controls. This DNA bait should be linked to a biotin molecule at one end that can be used to tether the bait to streptavidin-linked magnetic beads. Protein extract from your sample of interest can then be added to the tethered bait and flushed through with various wash steps to remove any non-specifically bound proteins. After this, you can elute the remaining proteins and carry out MS analysis to find out what your specific binders are.

MBDs: Technical considerations

DNA sequence

- When you design your synthetic DNA sequence, you may need to consider that the sequence itself may affect which binders you pull down.
- You may have a sequence in mind that you wish to use as bait, the promoter region of your gene of interest for example.
- Having a variety of sequences to use in your experiment will help to ensure that it is the modification which is the critical factor and not the DNA sequence.

The number of modifications

- The number of DNA modifications you have within your sequence may also influence the proteins binding to your bait.
- You should consider having just one modification or multiple modifications in a sequence to see how this is influencing your result.

Washing

- If you want to be sure that the proteins binding to your bait are true binders of your modification it is important to carry out very stringent washing.
- You can try high-salt washes to ensure you are removing everything non-specifically bound, but run the risk of removing everything, so you need to optimize this step to get the best results.

Novel DNA modifications

New DNA modifications could still be out there, just not discovered yet. It has been demonstrated that some modifications traditionally considered to be RNA modifications may also be present within DNA. One good example of this is N6-adenine methylation, known as m6A within RNA and 6mA within DNA. This modification is one of the most famous and abundant RNA modifications, but now it's known to also reside within DNA. One of the first studies to show this was from John Gurdon's lab in 2016 (Koziol *et al.*, 2016). They show that 6mA is within *Xenopus laevis*, mice, and the human genome using an antibody against 6mA to carry out DIP-seq.

Since this study, there have been several more claims that 6mA is present within DNA in zebrafish and pig genomes (Liu *et al.*, 2016), the mouse brain following environmental stress (Yao *et al.*, 2017), and within the *Arabidopsis thaliana* genome (Liang *et al.*, 2018). One study from 2018 took this one step further and uncovered the enzymes responsible for 6mA methylation and demethylation N6AMT1 and ALKBH1 respectively (Xiao *et al.*, 2018). The presence of enzymes actively adding and removing the DNA modification suggests that it has a real purpose to be there and potentially its own epigenetic function.

Novel DNA modifications: technical considerations

Antibody availability

- Until single-base resolution methods are available for individual modifications, many studies rely heavily on the use of antibody-based pull down (DIP-seq) to look for novel modifications within DNA.
- The biggest problem here is that you are then reliant on there being a specific, commercially available antibody for your modification which is quite often not the case. Many RNA modification antibodies will also recognize modifications within DNA, so this is one approach you could take.
- Treating your samples with RNase will help to ensure that you are targeting just DNA with your antibodies. You can look at the RNA modification antibodies available from Abcam at www.abcam.com/rnamods.

References

- Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., & Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, 336(6083), 934–937.
- Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., & Branco, M. R. (2014). Europe PMC Funders Group Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine, 8(10), 1841–1851.
- Colella, S., Shen, L., Baggerly, K. A., Issa, J. P. J., & Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing™ methylation analysis of CpG sites. *BioTechniques*, 35(1), 146–150.
- Fernandez, A. F., Valledor, L., Vallejo, F., Cañal, M. J., & Fraga, M. F. (2018). Quantification of Global DNA Methylation Levels by Mass Spectrometry. In J. Tost (Ed.), *DNA Methylation Protocols* (pp. 49–58). New York, NY: Springer New York.
- ML., G., & PA., J. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Research*, 25(12 PG-2529-31), 2529–2531.
- Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., & Baylin, S. B. (1996). Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands (DNA methylation/tumor suppressor genes/pl6/p15). *Proceedings of the National Academy of Sciences of the United States of America*, 93(September), 9821–9826.
- Kafer, G. R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I., & Carlton, P. M. (2016). 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *Cell Reports*, 14(6), 1283–1292.
- Koziol, M. J., Bradshaw, C. R., Allen, G. E., Costa, A. S. H., Frezza, C., & Gurdon, J. B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nature Structural and Molecular Biology*, 23(1), 24–30.
- Le, T., Kim, K. P., Fan, G., & Faull, K. F. (2011). A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Analytical Biochemistry*, 412(2), 203–209.
- Liang, Z., Shen, L., Cui, X., Bao, S., Geng, Y., Yu, G., ... Yu, H. (2018). DNA N6-Adenine Methylation in *Arabidopsis thaliana*. *Developmental Cell*, 45(3), 406–416.e3.
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-filippini, J., ... Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315–322.
- Liu, J., Zhu, Y., Luo, G. Z., Wang, X., Yue, Y., Wang, X., ... He, C. (2016). Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nature Communications*, 7(866), 1–7.
- Lu, X., Song, C., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., & He, C. (2013). Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J. Am. Chem. Soc.*, 135(25), 9315–9317.
- Lurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., ... Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, 14(10), 1–11.
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18), 5868–5877.

- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, 151(7), 1417–1430.
- Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1), 23–38.
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). Cell 1999 Okano, 99, 1–11.
- Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., ... Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347), 394–397.
- Sérandour, A. A., Avner, S., Mahé, E. A., Madigou, T., Guibert, S., Weber, M., & Salbert, G. (2016). Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets. *Genome Biology*, 17(1), 1–12.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H. L., ... Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, 153(3), 692–706.
- Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., ... He, C. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, 29(1), 68–75.
- Song, C. X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S. Q., Lin, L., ... He, C. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, 153(3), 678–691.
- Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., ... Vermeulen, M. (2013). Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*, 152(5), 1146–1159.
- Vertino, P. M., Yen, R. W., Gao, J., & Baylin, S. B. (1996). De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)methyltransferase. *Molecular and Cellular Biology*, 16(8), 4555–4565.
- Wu, H., Wu, X., Shen, L., & Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nature Biotechnology*, 32(12), 1231–1240.
- Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nature Reviews Genetics*, 18(9), 517–534.
- Xiao, C. Le, Zhu, S., He, M., De Chen, D., Zhang, Q., Chen, Y., ... Yan, G. R. (2018). N6-Methyladenine DNA Modification in the Human Genome. *Molecular Cell*, 1–13.
- Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., & Zhang, Y. (2013). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Research*, 23(3), 329–339.
- Yao, B., Cheng, Y., Wang, Z., Li, Y., Chen, L., Huang, L., ... Jin, P. (2017). DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nature Communications*, 8(1), 1–10.
- Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., ... Fazzio, T. G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, 147(7), 1498–1510.
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Zhang, L., Kim, A., ... He, C. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 149(6), 1368–1380.
- Zhang, L., Zhang, L., Zhou, K., Ye, X., Zhang, J., Xie, A., ... Cai, C. (2012). Simultaneous determination of global DNA methylation and hydroxymethylation levels by hydrophilic interaction liquid chromatography-tandem mass spectrometry. *Journal of Biomolecular Screening*, 17(7), 877–884.
- Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., ... Kee, K. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resource Resolution. *Stem Cell*, 1–12.

RNA modifications

The field of epigenetics is branching down many new and exciting avenues. One of these avenues is the area of RNA modification research. Recent advancements in the development of RNA modification detection and sequencing methods eg m6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP) has meant that it is becoming easier and faster to discover new modifications and map them to different species of RNA within any cell type or model organism. The advancements in this technology lead to a boom in the number of known RNA modifications. Currently, there are over 100 known RNA chemical modifications (Roundtree et al., 2017). You find these on mRNA, tRNAs, rRNAs and other non-coding RNAs including miRNAs. Each of these modifications also has its own function, including RNA structure, export, stability, and mRNA splicing. The future is bright for this field of research; there is still much to uncover regarding the function of some of these new modifications.

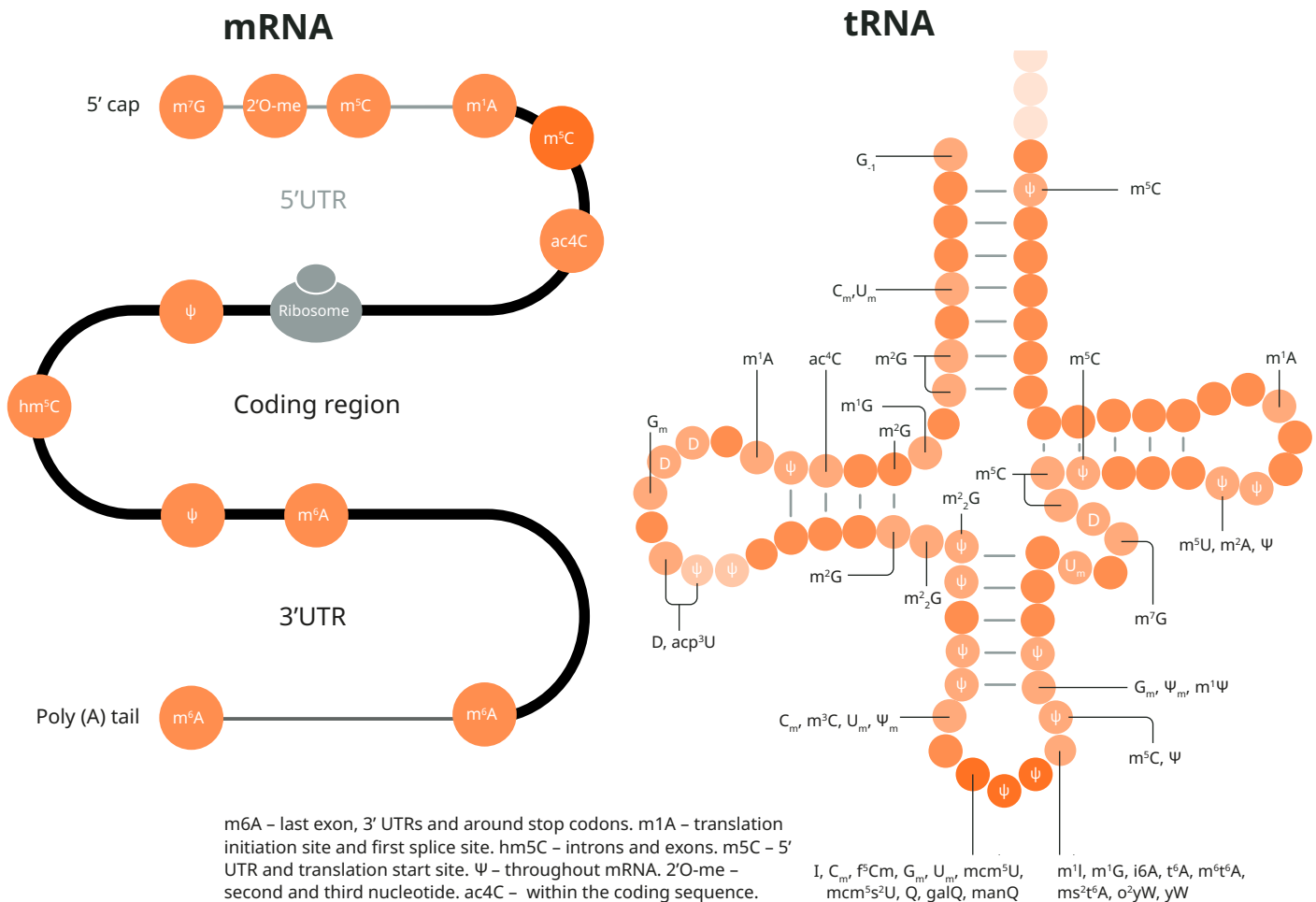


Figure 13. The distribution of RNA modifications on mRNA and tRNAs. To find out more take a look at our RNA modifications poster at www.abcam.com/RNAmodificationsposter

Of all the RNA species, tRNAs contain the most RNA modifications: almost one in five nucleotides within tRNAs are thought to contain RNA modifications (Kirchner et al., 2015). The modifications on tRNA are incredibly diverse and require step-by-step formation by multiple enzymes. You can commonly find modifications in the anticodon loop of the tRNA, which help promote translation efficiency by aiding codon-anticodon interactions and preventing frameshifting (Stuart et al., 2003).

The field of RNA modifications is relatively new but growing more and more every day. In this section, we will go through some of these protocols and offer a few tips and advice for working with RNA modifications.

Getting antibodies that are specific to your RNA modification of choice can be difficult. We have many well cited and validated RNA modification antibodies available including our m6A antibody, cited in several great publications including a Nature Methods paper which uses it for single-base resolution sequencing of m6A (Linder et al., 2015). This same m6A antibody also features in a Nature paper, looking at the role of m6A in mRNA stability (Mauer et al., 2017). For a full list of our RNA modification antibodies go to www.abcam.com/RNAmods

RNA immunoprecipitation (RIP)

RIP is an antibody-based technique used to map in vivo RNA-protein interactions. The RNA binding protein (RBP) of interest is immunoprecipitated together with its associated RNA for identification of bound transcripts (mRNAs, non-coding RNAs or viral RNAs). Transcripts are detected by real-time PCR, microarrays or sequencing.

Beyond transcription and subsequent translation, there is still much more to the function of RNA. For example, RNA-protein interactions can modulate mRNA and noncoding RNA function. This new appreciation for the potential of RNA has led to the development of novel methods allowing researchers to map RNA-protein interactions. RIP is one such protocol for the study of the physical association between individual proteins and RNA molecules.

Take a look at our full RIP protocol: www.abcam.com/RIP

Adapted from Khalila et al. (2009), Hendrickson et al. (2009), Hendrickson et al. (2008) and Rinn et al. (2007).

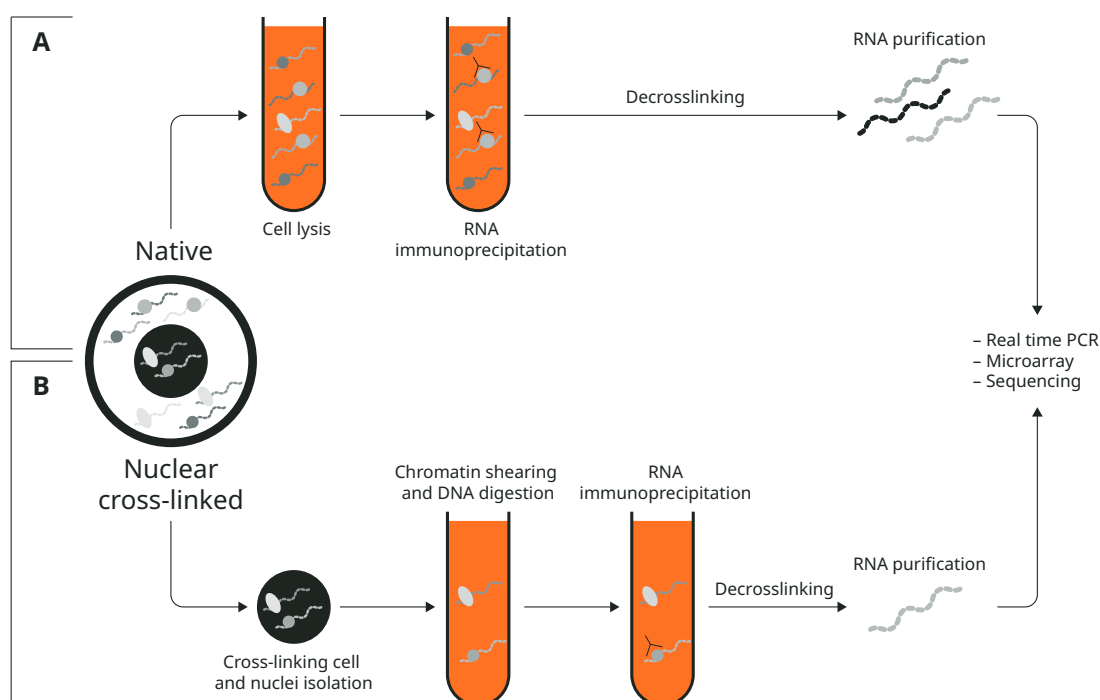


Figure 14. Schematic of RIP protocol workflow. (A) uses a native approach without cross linking. Method (B) uses formaldehyde cross linking.

RIP: technical considerations

RNase contamination

Avoid contamination using RNase-free reagents such as RNase-free tips, tubes, and reagent bottles. Use ultrapure distilled, DNase-free, RNase-free water to prepare buffers and solutions.

Plan your controls carefully

One or more negative controls should be maintained throughout the experiment, eg no antibody sample or immunoprecipitation from knockout cells or tissue. Knockdown cells are not recommended for negative control experiments.

Downstream analysis

The RNA isolated from your pull down can be analyzed using several techniques. Choose the best method based on the questions you want to answer and use multiple methods to confirm your results. For example, any novel results you obtain using RIP-seq should then be confirmed using RIP-qPCR.

CLIP

CLIP is an antibody-based technique used to study RNA-protein interactions related to RNA immunoprecipitation (RIP) but differs from RIP in that it uses UV radiation to cross-link RNA binding proteins to the RNA. This covalent bond is irreversible, allowing stringent purification conditions. Unlike RIP, CLIP provides information about the actual protein binding site on the RNA.

Different types of CLIP exist, including high-throughput sequencing-CLIP (HITS-CLIP), photoactivatable-ribonucleoside enhanced CLIP (PAR-CLIP), and individual CLIP (iCLIP).

You can find our full CLIP protocol adapted from König et al. J. Vis. Exp. 2011. "iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution." at www.abcam.com/CLIP

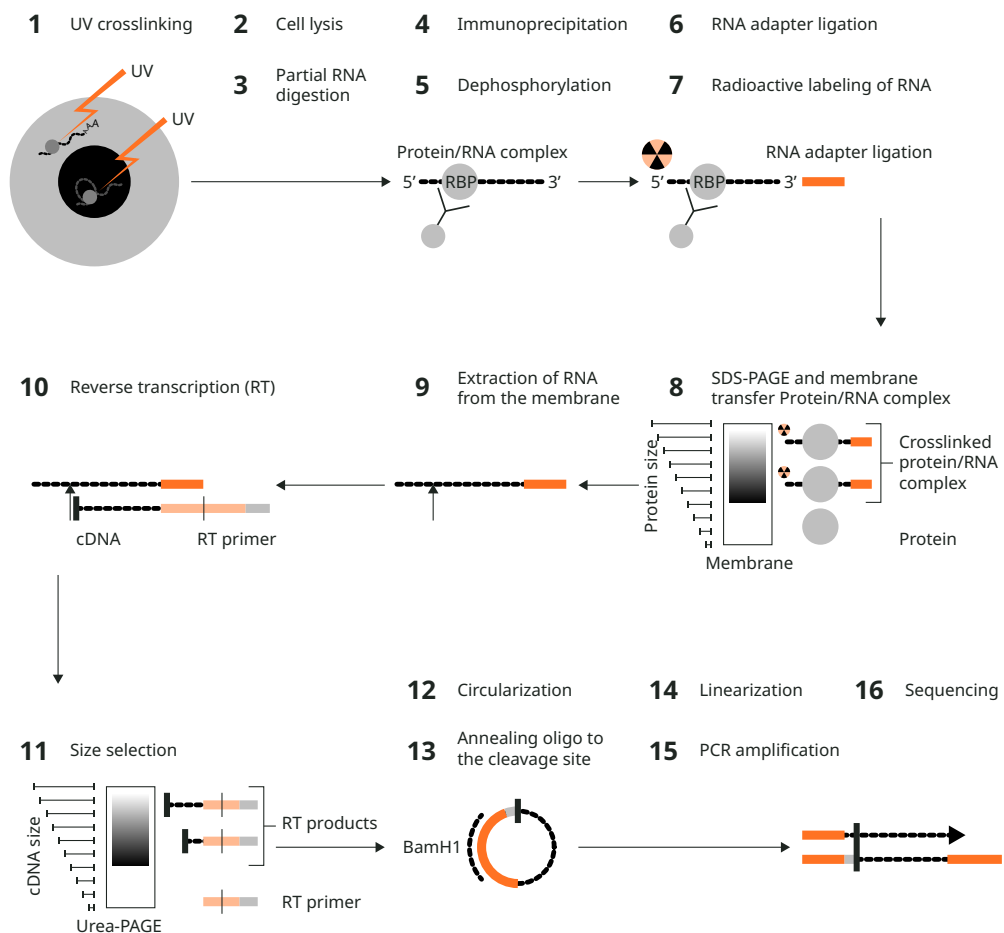


Figure 15. Schematic of CLIP protocol workflow.

CLIP: technical considerations

4-thiouridine pre-incubation

Optional 4-thiouridine pre-incubation and UV-A crosslinking may be necessary for certain proteins. 4-thiouridine enhances crosslinking of some proteins. Details for this can be found in the complete protocol.

Optimize antibody concentration

The amount of antibody required should be optimized before you start your experiment (Huppertz et al., 2014). A no-antibody sample is a good negative control. If your target of interest has not been studied using CLIP before, you could start by using an antibody that already works in IP, which is a good indication that it will work in CLIP.

miCLIP

Although m6A is the most abundant modified base in eukaryotic mRNA, current methods to accurately study it have limitations. New approaches to high-resolution mapping of m6A will be essential for understanding this epigenetic RNA modification.

miCLIP allows for high-resolution detection of single m6A residues and m6A clustering across the entire RNA (figure 16). Using miCLIP, it is possible to map m6A and the related dimethylated version m6Am (N6,2'-O-dimethyladenosine), at single-nucleotide resolution in human and mouse mRNA (Linder et al., 2015).

miCLIP is applicable to smaller RNAs. The authors of this protocol (Linder et al., 2015) discovered that m6A is present in small nucleolar RNAs (snoRNAs), a class of small non-coding RNAs. This was impossible to establish with previous applications due to lack of specificity and bioinformatic challenges.

You can find our full miCLIP protocol at www.abcam.com/miCLIP

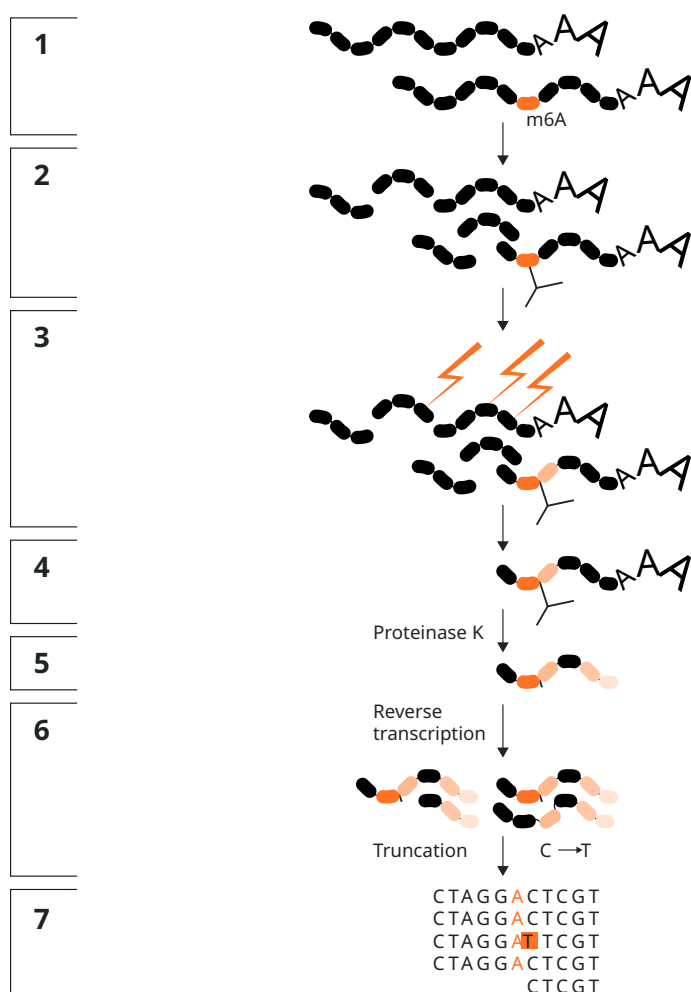


Figure 16. Schematic of miCLIP protocol workflow

1. RNA extraction.
2. Fragmentation of RNA to 30–130 nucleotide lengths, and incubation with anti-m6A antibody.
3. UV cross-linking of RNA to the bound antibody.
4. Recovery of antibody-RNA complexes with protein A/G affinity purification, SDS-PAGE and nitrocellulose membrane transfer.
5. Adapter ligation, and release of RNA with proteinase K.
6. Reverse transcription of RNA to cDNA, PCR amplification and sequencing.
7. Identification of C-T transitions or truncations and alignment against known genomic sequences. Mapping and annotation of these binding sites, identified as m6A/m6Am residues, to the transcriptome.

miCLIP: technical considerations

Not 100% accurate

The method is unable to identify the specific location of modified residues and only determines the general location of m6A sites.

Bias in bioinformatic m6A calling

Data analysis uses assumptions based upon known consensus sequences that harbor m6A residues, and so it misses modifications outside these motifs.

Liquid chromatography tandem-mass spectrometry (LC/MS-MS)

Similar to DNA modifications, if you have access to LC/MS-MS, then this is the best way quantify the amount of RNA modification within total RNA. Also similar to measuring DNA modifications you can use absolute quantification methods limited only by which isotopic standards you have available to use as a standard to measure your sample against.

Using this technique combined with RIP (RIP-MS) will allow you to determine if your RNA modification antibody is binding to your modification of interest and see if it binds any other non-specific modifications. If you generate LC/MS-MS data of your RIP input and pull-down samples, you should see an enrichment of your modification of interest in the pull-down sample compared to the input. You can also then check other modifications with these same data to see if anything else came out as enriched in your samples to test for non-specific antibody binding.

LC/MS-MS: technical considerations

Technically challenging

Mass spec equipment is costly and very specialized. The machine itself will require an enormous amount of maintenance and often requires its own technician to keep on top of things. Operating the machine is very complicated and requires specialized training so it may be difficult to obtain this type of mass spec data on your own. Consider obtaining this data through collaborations or paid services if it is not feasible for you to purchase your own LC/MS-MS equipment.

RNA modification control experiments

Due to the nature of RNA modifications their chemical structures are often very similar. To make sure you are getting the most accurate results from your antibodies, you need to test them in your model system thoroughly. Controls for RNA modification antibodies can be done using a range of applications. See below for some of our advanced controls and tips to make your RNA modification research easy.

RNase treatment

Whether you are carrying out ICC/IHC or RIP-qPCR, it is essential to have an RNase-treated control alongside your experimental samples (Delatte et al., 2016). For example, if you see a clear bright signal in your experimental IHC samples, but you get no signal in your RNase treated control samples, you can be confident that the signal you are getting is within the RNA and is not background signal from a non-specific source. This suggests that the antibody recognizes the modification within RNA and not the DNA.

It is crucial to ensure that you are not detecting high levels of non-specific background signal from DNA when using RNA modification antibodies. You can quickly add an RNase treatment step to your normal RNA modification IHC or RIP protocol. It is important not to leave the samples in the RNase solution for too long, this can lead to degradation of the DNA, and this then makes it difficult to carry out counterstains such as DAPI.

For each different sample type, you should test different concentrations of RNase and try leaving on your samples for varying lengths of time. For example, an IHC may need an RNase time of up to an hour depending on the tissue type whereas an ICC will require much less time – try 10–30 mins as a starting point.

DNase treatment

In addition to RNase-treated controls, you should carry out DNase-treated controls. If you are concerned that your RNA modification antibody is recognizing a similar modification within DNA, the best way to test for this is to treat your samples with DNase. Many modifications are within both RNA and DNA, so this is a common problem. For example, 5mC within DNA has the same chemical modification as m5C within RNA.

- If you carry out IHC using an RNA modification antibody, it is a good idea to have a DNase-treated control alongside your experimental samples. If you get a clear, strong signal from your experimental samples, but your DNase-treated control has no signal, it suggests that your antibody is binding to a modification within DNA.
- For this type of control, it is also important to optimize the conditions. Leaving your samples in DNase treatment for too long can lead to degradation of RNA, so be sure to test different DNase concentrations and the duration of the treatment.

Competition assays

Another way to ensure the specificity of your RNA modification antibody is to use a competition assay. This assay uses a synthetic modification-containing oligonucleotide which can be pre-incubated with your antibody (Meyer et al., 2012). When you then use this pre-incubated antibody for your applications, eg ICC/IHC or dot blot, you should see a reduction in the signal obtained when compared a sample stained with the antibody alone. You can try adding the competitor oligonucleotide to your antibody solution at increasing concentrations; a decreasing gradient of the signal reflects the amount of competitor you add to the antibody. For example, try a gradient of 0 ng, 10 ng, 100 ng, and 1 µg of the competitor oligonucleotide.

Dot blot

Carrying out a dot blot using RNA modification antibodies can be a quick and simple way to test for their specificity. A dot blot works like a simplified version of a western blot. For this technique, the sample is spotted directly on to the membrane, cross-linked, and then undergoes blotting. For more details take a look at our dot blot protocol. If you have access to synthetic RNA molecules containing your modification of interest, this can act as the perfect positive control. Similarly, loading an unmodified molecule or a molecule containing a different modification can serve as a negative control and help you to gauge any non-specific binding or cross-reactivity.

- For your experimental samples, it is possible to test whether your RNA modification antibody is specific by carrying out a dot blot with the right controls. For a negative control, use samples that contain a KO for the enzyme responsible for producing your specific RNA modification (Jia et al., 2011).
- If you load RNA from your wild-type and KO samples onto a membrane for dot blot, you should see a clear difference between the two samples. The wild-type sample will display a clear signal and the KO should appear blank when the membrane is stained using an antibody against your RNA modification of interest.

RIP-MS

If you have access to LC/MS-MS, then this is really the best way to test for RNA modification antibody specificity (Kellner et al., 2014). Using this technique combined with RIP (RIP-MS) allows you to determine if your antibody is binding to your modification of interest exclusively.

- Using either absolute or relative quantification methods, LC/MS-MS gives you parallel quantification of all the RNA modifications found in total RNA from any organism and cell type. If you generate LC/MS-MS data of your RIP input and pull-down samples you should see an enrichment of your modification of interest in the pulldown sample compared to the input.
- You can also then check other modifications with these same data to see if anything else came out as enriched in your samples to test for non-specific antibody binding. There is software being developed now that can even help you with this type of analysis (Yu et al., 2017).

References

- Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, Calonne E, Hassabi B, Putmans P, Awe S, Wetzel C, Kreher J, Soin R, Creppe C, Limbach PA, Gueydan C, Kruys V, Brehm A, Minakhina S, Defrance M, Steward R, Fuks F. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. (2016) *Science*. 2016 15:282-5
- Hendrickson DG, Hogan DJ, Herschlag D, Ferrell JE, and Brown PO (2008). Systematic Identification of mRNAs Recruited to Argonaute 2 by Specific microRNAs and Corresponding Changes in Transcript Abundance. *PLoS One* 3 (5), 2126.
- Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, and Brown PO (2009). Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. *PLoS Biology* 7 (11), 2643.
- Huppertz et al. iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods*. (2014).
- Jia G, Fu Y, Zhao X, Dai Q, Zheng G, et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7:885–87
- Kellner S, Ochel A, Thüring A, Spenkuch F, Neumann J, Sharma S, Entian KD, Schneider D, and Helm M. (2014) Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* 42(18): e142.
- Khalila AM, Guttman M, Huarte M, Garbera M, Rajd A, Morales DR, Thomas K, Pressera A, Bernstein BE, Oudenaarden AV, Regeva A, Lander ES, and Rinn JL (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *PNAS* 106, 11667–72.
- Kirchner, S., and Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* 16, 98–112.
- König et al. J. iCLIP -Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution. *Vis. Exp.* (2011).
- Linder B et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* 12:767-72 (2015)
- Mauer J et al. Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541:371-375 (2017).
- Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 1635-46
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, and Chang HY (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.
- Roundtree, I., Evans, M., Pan, T., & He, C. (2017) Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, 1187-1200
- Stuart, J.W., Koshlap, K.M., Guenther, R., and Agris, P.F. (2003). Naturally occurring modification restricts the anticodon domain conformational space of tRNA (Phe). *J. Mol. Biol.* 334, 901–918.
- Yu N, Lobue PA, Cao X, Limbach PA. (2017) RNAModMapper: RNA Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data. *Anal Chem* 10744-10752

