Enzymatic methods for genome-wide profiling of protein binding sites

Robert A. Policastro and Gabriel E. Zentner

Abstract

Genome-wide mapping of protein–DNA interactions is a staple approach in many areas of modern molecular biology. Genome-wide profiles of protein-binding sites are most commonly generated by chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq). Although ChIP-seq has played a central role in studying genome-wide protein binding, recent work has highlighted systematic biases in the technique that warrant technical and interpretive caution and underscore the need for orthogonal techniques to both confirm the results of ChIP-seq studies and uncover new insights not accessible to ChIP. Several such techniques, based on genetic or immunological targeting of enzymatic activity to specific genomic loci, have been developed. Here, we review the development, applications and future prospects of these methods as complements to ChIP-based approaches and as powerful techniques in their own right.

Key words: ChIP-seq; DamID; ChEC-seq; ChIC; CUT&RUN; Calling Card-seq

Introduction

Faithful transcription, replication and repair of the genome rely on the association of transcription factors (TFs), chromatin remodelers and other chromatin-binding proteins (CBPs) with specific DNA sequences and/or chromatin features. The precise genome-wide mapping of CBP binding is thus of great interest in many areas of biological investigation. The most widely used method to determine binding sites for a CBP of interest is chromatin immunoprecipitation (ChIP). In a standard ChIP protocol, proteins are first cross-linked to DNA, most commonly with formaldehyde, though other agents such as ultraviolet light have been used. Chromatin is then fragmented mechanically by sonication or enzymatically using micrococcal nuclease (MNase). Immunoprecipitation of the CBP of interest then enriches for bound DNA fragments. Although this standard workflow has remained relatively unchanged since the first published descriptions of ChIP, its throughput and spatial resolution have improved by orders of magnitude over the last three decades. Initial ChIP studies published by Lis and colleagues [1–5] and Varshavsky and colleagues [6, 7] used Southern blotting as a readout for protein occupancy of a sequence of interest. In 1993, the first combination of ChIP with polymerase chain reaction (PCR) was published, in which Orlando and Paro [8] used adapter modification of immunoprecipitated DNA to generate probes for Southern blot analysis. In 1999, Blat and Kleckner [9] performed what may be considered a precursor to ChIP-chip when they used a membrane spotted with 133 DNA sequences spanning chromosome III of *Saccharomyces cerevisiae* as a readout for cohesin ChIP. ChIP truly entered the genomic era in 2000 with the first published ChIP-chip study [10], and in 2007, several groups used high-throughput sequencing as a readout for ChIP (ChIP-seq) [11–14]. The adaptation of ChIP to the sequencing readout has, in conjunction with the rapidly decreasing cost of high-throughput sequencing (https://www.genome.gov/sequencingcostsdata/), led to a quantum leap in our understanding of genomic regulation by CBPs.

While ChIP-seq has been a key technique in uncovering genome-scale regulatory programs, it is potentially susceptible to the introduction of biases during various steps of the procedure. Formaldehyde cross-linking is induced by the formation of covalent methylene bridges between certain macromolecules in the cell, including DNA bases and proteins [15]. Despite our
understanding of the chemical nature of these cross-links, the nature of these interactions in vivo and their effect on ChIP-seq results still remain a ‘black box’ to researchers [16, 17]. For example, formaldehyde may fail to cross-link highly dynamic protein–DNA interactions [18, 19], leading to false negatives. Furthermore, cross-linking duration is also an important consideration: long cross-linking times increase the possibility of capturing nonspecific interactions [20] and lessen the ability of ChIP to detect quantitative differences in binding site occupancy [21]. Long cross-linking times also confine sites that have low dissociation constants with sites that have fast on rates with variable dissociation constants [21]. Formaldehyde cross-linking may also lead to epitope masking, which could interfere with immunoprecipitation of the target protein [17, 22]. The efficiency of formaldehyde cross-linking also appears to vary from protein to protein [23, 24], with ChIP DNA recovery being markedly lower for proteins that associate with DNA via protein–protein interactions, though they can be more efficiently cross-linked with a second longer range cross-linking agent [25]. Beyond the drawbacks associated with formaldehyde cross-linking, sonication may also generate bias due to preferential shearing of open chromatin [26, 27]. This has been actively exploited by investigators to map chromatin accessibility genome-wide as in Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq) and Sono-seq, in which cross-linked and sonicated DNA is purified and sequenced [28, 29]. Sonication also generates a heterogeneous mixture of DNA fragment lengths, limiting binding-site resolution, though the addition of nuclease digestion steps has been shown to greatly improve ChIP resolution [30–32]. DNA fragmentation by MNase circumvents limitations of sonication, but its endonucleolytic cleavage activity is biased toward AT-rich regions of DNA [33]. Importantly, ChIP relies on the availability of a highly specific antibody, which may be expensive or unavailable, particularly for poorly studied proteins. Finally, there are biases within ChIP-seq whose causes are not yet understood. First, in a pair of articles published in 2013, two labs described what has been termed ‘hyper-ChIPable artifacts’, which are regions of false-positive signal that generally occur over highly transcribed genes [34, 35]. Such artifacts were revealed by ChIP-seq analysis of non-DNA-binding proteins, including nuclear-localized green fluorescent protein and the Golgi protein Mnn10, in budding yeast. Other false positives, called phantom peaks, have been observed in ChIP-seq experiments performed in Drosophila melanogaster embryos and mouse cells that were protein-null for the target of immunoprecipitation [36, 37]. Meta-analysis of D. melanogaster nonhistone protein ChIP-seq data sets generated by the modENCODE consortium also revealed that 31% of these data sets had 20% or more peaks overlapping with phantom peaks [36]. The question of how many peaks in a given ChIP-seq experiment are experimental artifacts is an important one and is beginning to be addressed in a systematic fashion. For example, Verfaillie et al. [38], using motif analysis and reporter assays, concluded that up to 90% of peaks in a given TP53 ChIP-seq experiment are artifacts. Artifactual peaks may also partially explain the finding reported by ENCODE that many ChIP-seq peaks for sequence-specific transcription factors lack consensus motifs [39]. Furthermore, a high fraction of ChIP-seq peaks being artifacts could help explain the finding that, in general, a relatively low fraction of factor-bound genes change expression upon knockdown of that factor [40]. Finally, we note that the above caveats tend to be applicable more to ChIP-seq analysis of transcription factors and other proteins that are not considered intrinsic components of chromatin; ChIP analysis of histones and their modifications is generally considered robust and reproducible.

The above concerns illustrate the need for complementary approaches to mapping the binding sites of CBPs on a global scale. Such methods often involve targeting of a specific enzymatic activity to regions bound by the CBP of interest. This targeting is achieved either through genetic fusion of the CBP and an enzyme or by immunological targeting of an enzymatic activity to a protein of interest. Here, we discuss the principles of these techniques and systems and questions to which they have been applied. We also consider recent developments of these powerful genome-wide mapping approaches.

**DNA adenine methyltransferase identification**

DNA adenine methyltransferase identification (DamID) is the most well-established enzymatic alternative to ChIP-seq. DamID works by fusing the *Escherichia coli* DNA adenine methyltransferase (Dam) to a CBP, which when expressed results in the methylation of adenines in the ‘GATC’ sequence context near where the protein is bound [41] (Figure 1A). Subsequent digestion of DNA by DpnI, which cuts within methylated GATC sites, then generates fragments that can be sequenced to resolve regions where the protein contacted DNA [42, 43]. DamID is advantageous in that it does not require formaldehyde cross-linking and is not based on immunoprecipitation, thus circumventing limitations associated with chromatin solubilization and antibody quality. It can also be performed in vivo. The major limitations of DamID are the requirement for a fusion protein, its spatial resolution (estimated to be ~1 kb), and the requirement for GATC sequences proximal to binding sites. When tested side by side, DamID performed favorably compared with ChIP. DamID and ChIP of GAGA factor in Drosophila cells yielded highly concordant binding profiles when analyzed by microarray hybridization, with DamID outperforming ChIP in terms of dynamic range and sensitivity [44].

DamID expanded on an earlier study that fused the cytosine CpG methyltransferase SssI to a zinc finger protein to allow for targeted gene repression [45]. The decision to switch to an adenine methyltransferase when designing DamID was made for two reasons. First, unlike CpG methylation, GATC adenine methylation is rare in eukaryotes, so its origin can be more confidently traced back to binding of the chimeric protein [46]. Second, Dam was previously used to probe in vivo chromatin state in *D. melanogaster* without any adverse effects on development and growth, unlike the lethality induced when a cytosine methyltransferase was expressed [47, 48], consistent with the apparent lack of effect of adenine methylation on DNA topology [49].

The throughput and flexibility of the DamID system have improved considerably since its introduction. Early single-locus studies relied on quantitative PCR primers that amplified across GATC sequences near putative binding sites of the tagged protein. If the GATC was methylated it would be protected from cutting by DpnII and the subsequent loss of PCR amplification, and thus the proportion of protected GATC sites at a location could serve as a readout for relative occupancy [41]. In 2001, the throughput of DamID was improved to hundreds of sites using DpnI to fragment DNA at methylated GATC sites and then hybridizing the resulting fragments to tiling microarrays [50]. Complete genome-wide mapping of DamID methylation sites was achieved in 2011 when DpnI-released fragments were subjected to high-throughput sequencing [51]. Beyond the increase of throughput, the flexibility of DamID has been demonstrated...
through its use in numerous experimental systems including flies [41], worms [52], fission yeast [53], budding yeast [54], plants [55], fish [56] and mammalian cells [57]. The utility of DamID has been extended in recent years for genomic co-occupancy, tissue-specific and single-cell studies. Many DNA-binding proteins bind DNA as dimers or as parts of large complexes, but interrogating co-localization of complex members on DNA has been limited by the low efficiency of sequential ChIP [58]. Split DamID (SpDamID) presents a promising solution by splitting the Dam protein in half and fusing each half to proteins believed to interact on DNA. If the two proteins come in close proximity on DNA, the two halves of DamID complement one another and methylate DNA near the site of co-occupancy [59]. A recent modification of the DamID protocol, targeted DamID (TaDa), allows for tissue-specific profiling without laborious isolation of complex cells via fluorescence-activated cell sorting (FACS) [60], laser capture microdissection [61] or isolation of nuclei tagged in specific cell types (INTACT) [62]. TaDa employs the well-characterized Gal4-UAS system to express a CBP-Dam fusion in a cell type of interest. To avoid the toxicity and nonspecific methylation that accompany strong Dam overexpression, including with the Gal4-UAS system [63–65], the Dam fusion protein is expressed as a bicistronic message, with the factor of interest as the primary ORF and Dam as the secondary ORF [63]. Expression of this transcript driven by a tissue-specific promoter results in leaky low-level expression of the chimeric protein in the target tissue due to occasional reinitiation of translation by ribosomes that have in close proximity on DNA, the two halves of DamID complement one another and methylate DNA near the site of co-occupancy [59]. A recent modification of the DamID protocol, targeted DamID (TaDa), allows for tissue-specific profiling without laborious isolation of complex cells via fluorescence-activated cell sorting (FACS) [60], laser capture microdissection [61] or isolation of nuclei tagged in specific cell types (INTACT) [62]. TaDa employs the well-characterized Gal4-UAS system to express a CBP-Dam fusion in a cell type of interest. To avoid the toxicity and nonspecific methylation that accompany strong Dam overexpression, including with the Gal4-UAS system [63–65], the Dam fusion protein is expressed as a bicistronic message, with the factor of interest as the primary ORF and Dam as the secondary ORF [63]. Expression of this transcript driven by a tissue-specific promoter results in leaky low-level expression of the chimeric protein in the target tissue due to occasional reinitiation of translation by ribosomes that have in close proximity on DNA, the two halves of DamID complement one another and methylate DNA near the site of co-occupancy [59]. A recent modification of the DamID protocol, targeted DamID (TaDa), allows for tissue-specific profiling without laborious isolation of complex cells via fluorescence-activated cell sorting (FACS) [60], laser capture microdissection [61] or isolation of nuclei tagged in specific cell types (INTACT) [62]. TaDa employs the well-characterized Gal4-UAS system to express a CBP-Dam fusion in a cell type of interest. To avoid the toxicity and nonspecific methylation that accompany strong Dam overexpression, including with the Gal4-UAS system [63–65], the Dam fusion protein is expressed as a bicistronic message, with the factor of interest as the primary ORF and Dam as the secondary ORF [63]. Expression of this transcript driven by a tissue-specific promoter results in leaky low-level expression of the chimeric protein in the target tissue due to occasional reinitiation of translation by ribosomes that have completed translation of the primary ORF. Notably, TaDa and ChiP-seq of RNA Polymerase II gave highly concordant results. More recently, the Flp/FRT system has been implemented to allow lineage-specific expression of a Dam fusion in D. melanogaster by selective recombinational removal of a transcriptional stop element upstream of the fusion protein-encoding transgene [64]. Finally, the power of DamID to interrogate chromatin-based phenomena at single-cell resolution was displayed in a recent study of nuclear lamin–DNA interactions in vivo in mammalian cells. Prior to this study, low-resolution single-cell measurements of DamID relied on fluorescently tagged catalytically dead DpnI to bind to the sites of methylation, which allowed general binding regions to be viewed by microscopy [66]. Single-cell measurements were then combined with high-throughput sequencing by fusing Dam to Lamin B1 and FACS single cells [67]. Cell lysis, DpnI digestion, adapter ligation and PCR amplification then occurred in the same well to avoid loss of the small quantities of DNA, which successfully resulted in a genome-wide readout of Dam methylation using only a single cell.

In addition to the development of new applications for DamID, investigators have also sought to increase its spatial resolution. Using structural and mutational studies of Dam as a guide, Moore and colleagues [68] developed a variant of DamID termed Dam immunoprecipitation (DamIP). DamIP employs a fusion between the CBP of interest and a Dam K9A point mutant. The K9A form of Dam displays moderately reduced DNA binding and methylation activity and has an alteration in sequence specificity between the CBP of interest and a Dam K9A point mutant. The K9A form of Dam displays moderately reduced DNA binding and methylation activity and has an alteration in sequence specificity from GATC to ATC [69], in principle providing increased mapping resolution due to a target sequence occurring every 64 bases rather than 256 bases by chance. This alteration in sequence specificity precludes the release of methylated DNA by DpnI, and so methylation is detected by immunoprecipitation using an antibody against N⁴-methyladenosine after sonication to fragment the genome. DamIP was combined with high-throughput sequencing (DamIP-seq) to map the genomic binding of estrogen receptor α [70]. DamIP displayed comparable resolution to ChIP-seq, suggesting that the use of the K9A Dam mutant indeed improves mapping resolution. The relaxed sequenced specificity of K9A Dam was also used to the advantage of researchers studying protein association with RNA Pol III-transcribed genes, 98% of which have an ATC sequence within 150 bp in the human genome [71]. Notably, the reduced activity of Dam K9A allowed the investigators to express Dam fusions from the strong constitutive CMV and Moloney murine leukemia virus promoters as well as a strong tet-inducible promoter without

Figure 1. Principles of enzymatic assays for mapping protein interactions within the genome. (A) DamID. A CBP is fused to Dam and expressed in the cell or tissue type of interest, resulting in adenine methylation within GATC sequences. (B) ChEC. A CBP is fused to MNase and expressed in cells. When calcium is added, the nucleolytic activity of MNase is activated, resulting in cleavage of DNA proximal to binding sites for the CBP. (C) CUT&RUN. Cells are incubated with an antibody directed against the CBP of interest. A fusion of MNase with protein A targets calcium-dependent cleavage to sites bound by the factor of interest via antibody recognition by protein A (D). Calling card. In yeast, a CBP is fused to Sir4 and expressed in cells along with the Ty5 integrase. Interaction between Sir4 and Ty5 integrase results in insertion of transposons proximal to binding sites for the Sir4-tagged CBP. In mammalian cells, the piggyBac transposase is instead directly fused to the CBP of interest. (A colour version of this figure is available online at: https://academic.oup.com/bfg)
saturating the genome with methylation [70, 71], in contrast to
the very low level induction required to prevent saturation when
using fusions with wild-type Dam.

Chromatin endogenous cleavage

One of the first attempts to supplant ChIP, the protein position
identification with nuclease tail (PIN’POINT) method, was
published in 1998 and involved fusion of the FokI endonuclease to
the Sp1 TF to generate cleavage around its binding sites [72].
Although PIN’POINT was not widely adopted, other techniques
based on nuclease fusions were subsequently developed. One
such method, chromatin endogenous cleavage (ChEC), is based
on fusion of a protein of interest to MNase. Cells expressing a
CBP-MNase fusion are permeabilized and calcium is added, induc-
ing the cleavage activity of MNase and generating double-
strand breaks around loci bound by the fusion (Figure 1B). ChEC
was published in 2004 by Laemmli and colleagues [73] who used
it to interrogate the binding of several regulatory factors to the
budding yeast genome in conjunction with Southern blotting.
The Laemmli group then used ChEC in conjunction with low-
resolution tailed microarray analysis to map the binding of nu-
clear pore components to the yeast genome [74]. ChEC has also
been extensively used to map protein binding to ribosomal DNA
in yeast [75–77]. The advantages of ChEC are similar to those of
DamID: it does not require cross-linking and does not depend
on immunoprecipitation, so issues with chromatin solubiliza-
tion and antibody quality are circumvented. In addition, ChEC-
seq enables targeted genomic cleavage due to the calcium
dependence of MNase as well as much higher spatial resolution
than DamID. From a practical standpoint, ChEC is a very simple
technique: once cells expressing a CBP-MNase fusion are con-
structed, cells are simply permeabilized and incubated with cal-
cium, and total DNA is then extracted [78]. As with DamID, a
drawback of ChEC is the requirement for a fusion protein.

In 2015, the Henikoff group published ChEC-seq, wherein
DNA obtained through the ChEC technique is subjected to high-
throughput sequencing for genome-wide identification of tran-
scription factor-binding sites [79]. ChEC-seq identified binding
sites for the yeast general regulatory factors (GRFs) Abf1, Rap1
and Reb1 with high spatial resolution and, of note, yielded the
majority of ChIP-determined binding sites for each factor.
Interestingly, sites displaying robust consensus motifs were rap-
 idly cleaved, whereas many sites with poor motif matches dis-
played gradual accumulation of cleavage. Such sites could rep-
cord loci transiently sampled by GRFs during target site
searching. The observation of temporal separation between
sites with high- and low-quality motif matches also suggests
that the kinetics of MNase digestion could be used to stringently
identify relevant binding sites. That is, sites displaying robust
cleavage at a short digestion time point (<1 min) are likely to be
robustly bound in a majority of cells during that time period.
Our lab has also successfully applied ChEC-seq to subunits of
the conserved and essential Mediator transcriptional coactivator
complex [80], indicating that ChEC-seq can be used to map the
binding of large complexes that do not directly contact DNA.
ChEC-seq was also able to capture the association of Mediator
with ribosomal protein genes [81], which has been difficult to as-
ssess with ChIP-based methods [82].

Chromatin immunocleavage

Chromatin immunocleavage (ChIC) was introduced in parallel
with ChEC [73]. In ChIC, cells are incubated with an antibody
directed against a protein of interest, followed by incubation with
a fusion of immunoglobulin-binding Protein A and MNase
(pA-MN) (Figure 1C). If necessary, a secondary antibody recog-
nizing the primary antibody can be used to augment antibody
recognition by pA-MN. The advantages of ChIC are thus similar
to ChEC, with the added benefit that a fusion protein is not
required. However, unlike ChEC, a high-quality antibody is ne-
necessary. Despite the potential power of ChIC, no studies employ-
ing it were published until early 2017, when it was adapted to a
sequencing readout in the form of cleavage under targets and
release using nuclease (CUT&RUN) [83]. CUT&RUN was success-
fully applied to the mapping of diverse CBPs in budding yeast
including the GRFs Abf1 and Reb1, the ATPases Mot1 and Sh1
and the histones H2A and Cse4. CUT&RUN also performed well
in human cells, generating low-background, high-resolution
binding maps of the TFs CTCF, Myc and Max as well as
H3K27me3. In each of the tested cases, CUT&RUN compared fa-
vorably with ChIP-seq in terms of specificity and sensitivity.
Furthermore, CUT&RUN was successfully applied to mapping of
CTCF binding sites in as few as 600 000 cells, suggesting that it
might be suitable for samples in which only a small amount of
material is available, such as clinical biopsies and specific neur-
onal or stem cell populations in conjunction with methods to
isolate particular cell populations such as INTACT [62, 84].
Lastly, CUT&RUN may be able to provide information on 3D gen-
ome architecture, thanks to MNase cleavage of distal regions
looped to binding sites for the MNase-targeted factor (CTCF in
the published case). We note that detection of protein-involved
genomic interactions is a potential application of all techniques
described here; indeed, DamID has been used in a locus-specific
manner to assay long-range genomic contacts [85].

Calling cards

While ChIP-seq and the other enzymatic alternatives discussed
in this review excel at mapping the binding sites of particular
proteins on DNA, they are limited to mapping one factor at a
time. To facilitate simultaneous interrogation of binding sites
for multiple proteins in a single population of cells, Mitra and
colleagues [86] introduced the Calling Card approach. The
Calling Card method takes advantage of the S. cerevisiae retro-
transposon Ty5, the insertion of which into DNA is directed by
the interaction of a fragment of the Sir4 protein with the Ty5
integrase [87]. By tethering the Sir4 protein to a CBP, the retro-
transposon will be inserted near where the chimeric protein
bound DNA [88] (Figure 1D). Generation of a library amenable to
next-generation sequencing (Calling Card-seq) can then be gen-
erated by fragmentation of the genome with restriction en-
zymes, circularization of the fragments and then inverse PCR
with primers complementary to the retrotransposon. Addition
of unique bar codes to the retrotransposons allows multiple
strains harboring different proteins tethered to Sir4 to be com-
bined and processed as one sample, thus allowing simultan-
eous generation of binding profiles for multiple CBPs [89].
Limitations of this method include sequence preferences for the
transposon insertion site, a preference for insertion at open
chromatin due to reduced steric hindrance and the reliance on
restriction sites near where the transposon inserted.

The power of Calling Card-seq was demonstrated in a study
of pseudohyphal growth, a form of cellular differentiation in re-
sponse to nutrient starvation in S. cerevisiae [90]. In this study,
the authors tagged 28 TFs believed to be involved in pseudohy-
phal growth gave each a uniquely barcoded retrotransposon
and then pooled the strains for analysis [91]. They found over
725 binding sites for the analyzed TFs and were able to build a TF network for pseudohyphal growth by considering the number of genes targeted by each protein that are thought to be important for that process.

In 2012, Calling Card-seq was adapted to a mammalian system by fusing the piggyBac (PB) transposase to a CBP, allowing the targeted integration of the PB transposon near where the fused protein bound [92]. The authors were motivated by the prospect of using Calling card-seq to uncover temporally restricted binding events, such as those that occur in certain developmental stages. With ChIP-seq and other comparable methods, the generated DNA binding profile of a protein is a snapshot of the protein–DNA interaction landscape when the cells were harvested. However, the transposon insertion events from Calling Card-seq are stable across cell divisions. This not only allows simultaneous cataloging of all binding events through development but could also be used to follow associations between cell fate and the activity of certain TFs. Calling Card-seq performed well for the Sp1 transcription factor, displaying high concordance with ChIP-seq-determined Sp1 binding sites [92].

### Data analysis

An important consideration when using the enzymatic mapping methods described above is analysis of the resulting data. As these techniques are used less widely than ChIP-seq, fewer method-specific analytical tools are available. For DamID-seq, a number of specifically tailored software packages are available to streamline analysis, including damidseq_pipeline [93] and iDEAR [56]. The analytical steps for ChEC-seq and CUT&RUN are similar to those for ChIP-seq, and so analysis of these data types is amenable to the use of established tools. Indeed, we have analyzed ChEC-seq data with both custom scripts (https://github.com/zentnerlab/chec-seq) and the HOMER software suite (http://homer.ucsd.edu) [94]. Custom software used for CUT&RUN data analysis is also available online (https://github.com/peteskene). To our knowledge, no tools for the analysis of Calling Card-seq data have been made available.

### Conclusions

Genome-wide mapping of protein-binding sites has become a staple of modern molecular biology. Although this mapping is most often performed with ChIP-based techniques, a number of enzymatic genome-wide mapping methods have been developed over the last few decades to avoid limitations associated with ChIP and discover new insights into the expression and regulation of the genome. Such methods offer notable advantages relative to ChIP-seq, mainly due to the fact that they do not depend on immunoprecipitation. DamID, ChEC-seq and Calling Card-seq do not require cross-linking or chromatin solubilization and circumvent issues arising from poor antibody quality. CUT&RUN offers the same advantages in terms of cross-linking and chromatin solubilization, provided that a suitable antibody against the factor of interest is available. Salient features of the methods described here are presented in Table 1.

A general limitation of the enzymatic methods described here, save for CUT&RUN, is the requirement for a fusion protein. Generation of enzymatic fusions controlled by endogenous regulatory elements is simple in yeast, where recombination-based tagging is simple and quick. Endogenous tagging is much more laborious in metazoan systems, but advances in CRISPR-based genome editing may facilitate generation of endogenous regulatory elements...
Enzymatically tagged loci. An alternative approach is expression of fusions from plasmid vectors, which is commonly used in DamID. For DamID, CBP-Dam fusions are generally placed under the control of a heat shock promoter in a plasmid that is maintained episomally or integrated into the genome following selection [57]. Following transfection, the leakage of the uninduced heat-shock promoter generates very low levels of CBP-Dam. In principle, this approach could be used for any enzymatic fusion method and in any system in which transgenesis is possible. The requirement for a fusion protein also precludes the use of enzymatic fusion methods in clinical samples, though CUT&RUN may be ideal in such situations if an appropriate antibody is available. It is also possible that the addition of a large enzymatic moiety to a protein could alter its function and/or its association with interacting partners, a concern that is particularly acute when the protein of interest is part of a large complex. To this end, structural consideration of the protein of interest may be useful. For instance, when choosing subunits of the yeast Mediator complex to tag with MNase, we used available structural data to select subunits with C-terminal ends that were predicted to be solvent exposed rather than buried within the complex [80].

Despite these limitations, each of the enzymatic-mapping strategies discussed here should be applicable to any system amenable to transgenesis. To determine how broadly applicable each technique is, it will be important to assess their performance in a variety of experimental systems. While DamID, the original publication of which occurred in 2000 [41], and its variant in a variety of experimental systems. While DamID, the each technique is, it will be important to assess their performance in a variety of experimental systems. While DamID, the original publication of which occurred in 2000 [41], and its various iterations have been applied to many systems from yeast to plants, ChEC-seq, CUT&RUN, Calling Card-seq and TRIBE have not been extensively tested in the literature, presumably due to their relative newness. We look forward to increased application of enzymatic methods for mapping genome-wide protein–chromatin interactions to further characterize their various advantages and limitations and discover new biological insights in many areas of inquiry.

Key Points

• Several enzymatic methods to profile the interactions of specific factors with the genome have been developed.

• Enzymatic methods for protein–DNA interaction mapping avoid limitations associated with ChIP-based approaches.

Acknowledgements

The authors thank Sebastian Gründer for helpful comments on the manuscript.

Funding

This work was supported by Indiana University Startup funds (to G.E.Z.).

References


81. Grünberg S, Zentner GE. Genome-wide characterization of Mediator recruitment, function, and regulation. Transcription 2017;e1291082.


