

Enzymatic methods for genome-wide profiling of protein binding sites

Robert A. Policastro and Gabriel E. Zentner

Corresponding author: Gabriel E. Zentner, Indiana University. Tel.: 812-856-7377; Fax: 812-855-6082; E-mail: gzentner@indiana.edu

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 with microarray analysis (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

Robert A. Policastro is a graduate student in the Zentner laboratory at Indiana University. He is interested in the interplay of protein factors and DNA sequence in promoter regulation.

Gabriel E. Zentner is an assistant professor of biology at Indiana University. His lab uses genome-wide methodologies to study basic mechanisms of transcriptional regulation.

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 conflate 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 have 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

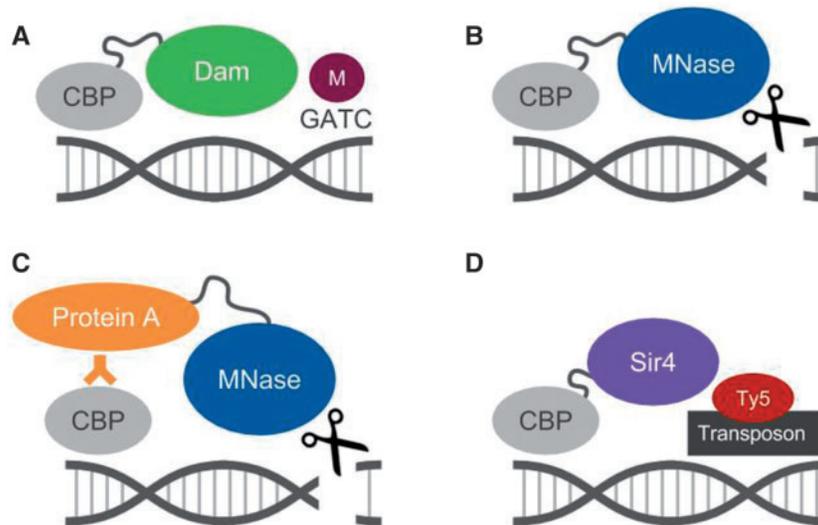


Figure 1. Principles of enzymatic assays for mapping protein interactions within the genome. (A) DamID. A CBP is fused to Dam and expressed in the cell or tissue type of interest, resulting in adenine methylation within GATC sequences. (B) ChEC. A CBP is fused to MNase and expressed in cells. When calcium 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.

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 specific cell types 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 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 N⁶-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 150bp 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

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 broadly 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, inducing 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 tiled microarray analysis to map the binding of nuclear 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 solubilization and antibody quality are circumvented. In addition, ChEC-seq enables controlled 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 constructed, cells are simply permeabilized and incubated with calcium, 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 transcription 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 rapidly cleaved, whereas many sites with poor motif matches displayed gradual accumulation of cleavage. Such sites could represent 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 assess 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 recognizing 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 necessary. Despite the potential power of ChIC, no studies employing 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 successfully applied to the mapping of diverse CBPs in budding yeast including the GRFs Abf1 and Reb1, the ATPases Mot1 and Sth1 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 favorably 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 neuronal 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 genome 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* retrotransposon 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 retrotransposon 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 generated by fragmentation of the genome with restriction enzymes, 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 combined and processed as one sample, thus allowing simultaneous 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 response to nutrient starvation in *S. cerevisiae* [90]. In this study, the authors tagged 28 TFs believed to be involved in pseudohyphal growth gave each a uniquely barcoded retrotransposon and then pooled the strains for analysis [91]. They found over

Table 1. Salient features of the enzymatic mapping methods discussed in this review

Enzyme	Method	Reported input (cells)	Spatial resolution	Tested organisms	Advantages	Disadvantages	Sequencing reference
DNA adenine methyltransferase	DamID-seq	1–10 ⁶	1 kb	Sc, Sp, Ce, Dm, Mm, Hs, At	–No antibody, sonication or cross-linking –Potential for single-cell resolution	–Genetic fusion –Constitutive methylation	[51]
	SpDamID	10 ² –10 ⁶	1 kb	Hs	–No antibody, sonication or cross-linking –Provides information on protein co-occupancy	–Genetic fusion –Constitutive methylation	[59]
	TaDa	10 ² –10 ⁴	1 kb	Dm	–No antibody, sonication or cross-linking –Cell type specific	–Genetic fusion –Constitutive methylation	[63]
	DamIP-seq	Not reported	200–300 bp	Hs	–No cross-linking –Improved resolution	–Genetic fusion –Constitutive methylation	[70]
Micrococcal nuclease	ChEC-seq	5 × 10 ⁷ –10 ⁸	10–200 bp	Sc	–No antibody, sonication or cross-linking –Inducible	–Genetic fusion	[79]
	CUT&RUN	6 × 10 ⁵ –10 ⁷	10–200 bp	Sc, Hs	–No cross-linking or sonication –Inducible	–Requires a specific antibody	[83]
Transposases	Calling Card-seq	5 × 10 ⁵ –10 ⁶	200–300 bp	Sc, Hs	–No antibody, sonication or cross-linking	–Genetic fusion	[92]

Sc: *Saccharomyces cerevisiae*; Sp: *Schizosaccharomyces pombe*; Ce: *Caenorhabditis elegans*; Dm: *Drosophila melanogaster*; Mm: *Mus musculus*; Hs: *Homo sapiens*; At: *Arabidopsis thaliana*.

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 cataloguing 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

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 leakiness 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 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 Grünberg for helpful comments on the manuscript.

Funding

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

References

- Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci USA* 1984;**81**:4275–9.
- Gilmour DS, Lis JT. *In vivo* interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol Cell Biol* 1985;**5**:2009–18.
- Gilmour DS, Pflugfelder G, Wang JC, et al. Topoisomerase I interacts with transcribed regions in *Drosophila* cells. *Cell* 1986;**44**:401–7.
- Gilmour DS, Lis JT. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol Cell Biol* 1986;**6**:3984–9.
- Gilmour DS, Lis JT. Protein-DNA cross-linking reveals dramatic variation in RNA polymerase II density on different histone repeats of *Drosophila melanogaster*. *Mol Cell Biol* 1987;**7**:3341–4.
- Solomon MJ, Varshavsky A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. *Proc Natl Acad Sci USA* 1985;**82**:6470–4.
- Solomon MJ, Larsen PL, Varshavsky A. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* 1988;**53**:937–47.
- Orlando V, Paro R. Mapping polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* 1993;**75**:1187–98.
- Blat Y, Kleckner N. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* 1999;**98**:249–59.
- Ren B, Robert F, Wyrick JJ, et al. Genome-wide location and function of DNA binding proteins. *Science* 2000;**290**:2306–9.
- Albert I, Mavrich TN, Tomsho LP, et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 2007;**446**:572–6.
- Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;**129**:823–37.
- Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* 2007;**316**:1497–502.
- Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;**448**:553–60.
- Feldman MY. Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog Nucleic Acid Res Mol Biol* 1973;**13**:1–49.
- Gavrilov A, Razin SV, Cavalli G. *In vivo* formaldehyde cross-linking: it is time for black box analysis. *Brief Funct Genomics* 2015;**14**:163–5.
- Hoffman EA, Frey BL, Smith LM, et al. Formaldehyde cross-linking: a tool for the study of chromatin complexes. *J Biol Chem* 2015;**290**:26404–11.
- Schmiedeberg L, Skene P, Deaton A, et al. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One* 2009;**4**:e4636.
- Tian B, Yang J, Brasier AR. Two-step cross-linking for analysis of protein-chromatin interactions. In: Vancura A (ed). *Transcriptional Regulation: Methods and Protocols*. New York, NY: Springer New York, 2012, 105–20.
- Baranello L, Kouzine F, Sanford S, et al. ChIP bias as a function of cross-linking time. *Chromosome Res* 2016;**24**:175–81.
- Poorey K, Viswanathan R, Carver MN, et al. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* 2013;**342**:369–72.
- Partridge EC, Watkins TA, Mendenhall EM. Every transcription factor deserves its map: scaling up epitope tagging of proteins to bypass antibody problems. *Bioessays* 2016;**38**:801–11.
- Toth J, Biggin MD. The specificity of protein-DNA crosslinking by formaldehyde: *in vitro* and in *Drosophila* embryos. *Nucleic Acids Res* 2000;**28**:e4.

24. Aoki T, Wolle D, Preger-Ben Noon E, et al. Bi-functional cross-linking reagents efficiently capture protein-DNA complexes in *Drosophila* embryos. *Fly* 2014;**8**:43–51.
25. Zeng P-Y, Vakoc CR, Chen Z-Z, et al. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques* 2006;**41**:694–8.
26. Teytelman L, Özaydın B, Zill O, et al. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One* 2009;**4**:e6700.
27. Vega VB, Cheung E, Palanisamy N, et al. Inherent signals in sequencing-based chromatin-immunoprecipitation control libraries. *PLoS One* 2009;**4**:e5241.
28. Gaulton KJ, Nammo T, Pasquali L, et al. A map of open chromatin in human pancreatic islets. *Nat Genet* 2010;**42**:255–9.
29. Auerbach RK, Eusikirchen G, Rozowsky J, et al. Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci USA* 2009;**106**:14926–31.
30. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 2011;**147**:1408–19.
31. Skene PJ, Hernandez AE, Groudine M, et al. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife* 2014;**3**:e02042.
32. He Q, Johnston J, Zeitlinger J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotech* 2015;**33**:395–401.
33. Allan J, Fraser RM, Owen-Hughes T, et al. Micrococcal nuclease does not substantially bias nucleosome mapping. *J Mol Biol* 2012;**417**:152–64.
34. Teytelman L, Thurtle DM, Rine J, et al. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci USA* 2013;**110**:18602–7.
35. Park D, Lee Y, Bhupindersingh G, et al. Widespread misinterpretable ChIP-seq bias in yeast. *PLoS One* 2013;**8**:e83506.
36. Jain D, Baldi S, Zabel A, et al. Active promoters give rise to false positive ‘Phantom Peaks’ in ChIP-seq experiments. *Nucleic Acids Res* 2015;**43**:6959–68.
37. Krebs W, Schmidt SV, Goren A, et al. Optimization of transcription factor binding map accuracy utilizing knockout-mouse models. *Nucleic Acids Res* 2014;**42**:13051–60.
38. Verfaillie A, Svetlichnyy D, Imrichova H, et al. Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res* 2016;**26**:882–95.
39. Neph S, Vierstra J, Stergachis AB, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 2012;**489**:83–90.
40. Cusanovich DA, Pavlovic B, Pritchard JK, et al. The functional consequences of variation in transcription factor binding. *PLoS Genet* 2014;**10**:e1004226.
41. van Steensel B, Henikoff S. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat Biotechnol* 2000;**18**:424–8.
42. Wu F, Yao J. Spatial compartmentalization at the nuclear periphery characterized by genome-wide mapping. *BMC Genomics* 2013;**14**:591.
43. Wu F, Olson BG, Yao J. DamID-seq: genome-wide mapping of protein-DNA interactions by high throughput sequencing of adenine-methylated DNA fragments. *J Vis Exp* 2016;**e53620**.
44. Moorman C, Sun LV, Wang J, et al. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2006;**103**:12027–32.
45. Xu G-L, Bestor TH. Cytosine methylation targeted to predetermined sequences. *Nat Genet* 1997;**17**:376–8.
46. Heyn H, Esteller M. An adenine code for DNA: a second life for N6-methyladenine. *Cell* 2015;**161**:710–3.
47. Wines DR, Talbert PB, Clark DV, et al. Introduction of a DNA methyltransferase into *Drosophila* to probe chromatin structure in vivo. *Chromosoma* 1996;**104**:332–40.
48. Lyko F, Ramsahoye BH, Kashevsky H, et al. Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in *Drosophila*. *Nat Genet* 1999;**23**:363–6.
49. Barras F, Marinus MG. The great GATC: DNA methylation in *E. coli*. *Trends Genet* 1989;**5**:139–43.
50. van Steensel B, Delrow J, Henikoff S. Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* 2001;**27**:304–8.
51. Luo SD, Shi GW, Baker BS. Direct targets of the *D. melanogaster* DSX^F protein and the evolution of sexual development. *Development* 2011;**138**:2761–71.
52. Schuster E, McElwee JJ, Tullet JMA, et al. DamID in *C. elegans* reveals longevity-associated targets of DAF-16/FoxO. *Mol Syst Biol* 2010;**6**:399.
53. Woolcock KJ, Gaidatzis D, Punga T, et al. Dicer associates with chromatin to repress genome activity in *Schizosaccharomyces pombe*. *Nat Struct Mol Biol* 2011;**18**:94–9.
54. Lebrun E, Fourel G, Defossez P-A, et al. A methyltransferase targeting assay reveals silencer-telomere interactions in budding yeast. *Mol Cell Biol* 2003;**23**:1498–508.
55. Germann S, Juul-Jensen T, Letarnc B, et al. DamID, a new tool for studying plant chromatin profiling in vivo, and its use to identify putative LHP1 target loci. *Plant J* 2006;**48**:153–63.
56. Gutierrez-Triana JA, Mateo JL, Ibberson D, et al. iDamIDseq and iDEAR: an improved method and computational pipeline to profile chromatin-binding proteins. *Development* 2016;**143**:4272–8.
57. Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protocols* 2007;**2**:1467–78.
58. Shankaranarayanan P, Mendoza-Parra M-A, Walia M, et al. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Methods* 2011;**8**:565–7.
59. Hass Matthew R, Liow H-h, Chen X, et al. SpDamID: marking DNA bound by protein complexes identifies notch-dimer responsive enhancers. *Mol Cell* 2015;**59**:685–97.
60. Bryant Z, Subrahmanyam L, Tworoger M, et al. Characterization of differentially expressed genes in purified *Drosophila* follicle cells: Toward a general strategy for cell type-specific developmental analysis. *Proc Natl Acad Sci USA* 1999;**96**:5559–64.
61. Bonner RF, Emmert-Buck M, Cole K, et al. Laser capture microdissection: molecular analysis of tissue. *Science* 1997;**278**:1481–3.
62. Deal RB, Henikoff S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev Cell* 2010;**18**:1030–40.
63. Southall Tony D, Gold Katrina S, Egger B, et al. Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Dev Cell* 2013;**26**:101–12.
64. Pindyurin AV, Pagie L, Kozhevnikova EN, et al. Inducible DamID systems for genomic mapping of chromatin proteins in *Drosophila*. *Nucleic Acids Res* 2016;**44**:5646–57.
65. Bianchi-Frias D, Orian A, Delrow JJ, et al. Hairy transcriptional repression targets and cofactor recruitment in *Drosophila*. *PLoS Biol* 2004;**2**:e178.
66. Kind J, Pagie L, Ortaobzkoyun H, et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 2013;**153**:178–92.

67. Kind J, Pagie L, de Vries Sandra S, et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 2015; **163**:134–47.
68. Xiao R, Roman-Sanchez R, Moore DD. DamIP: a novel method to identify DNA binding sites in vivo. *Nucl Recept Signal* 2010; **8**:e003.
69. Horton JR, Liebert K, Bekes M, et al. Structure and substrate recognition of the *Escherichia coli* DNA adenine methyltransferase. *J Mol Biol* 2006; **358**:559–70.
70. Xiao R, Sun D, Ayers S, et al. The estrogen receptor α cistrome defined by DamIP. *Mol Endocrinol* 2012; **26**:349–57.
71. Orioli A, Praz V, Lhôte P, et al. Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. *Genome Res* 2016; **26**:624–35.
72. Lee J-S, Lee C-H, Chung JH. Studying the recruitment of Sp1 to the β -globin promoter with an *in vivo* method: protein position identification with nuclease tail (PIN*POINT). *Proc Natl Acad Sci USA* 1998; **95**:969–74.
73. Schmid M, Durussel T, Laemmli UK. ChIC and ChEC: genomic mapping of chromatin proteins. *Mol Cell* 2004; **16**:147–57.
74. Schmid M, Arib G, Laemmli C, et al. Nup-PI: the nucleopore-promoter interaction of genes in yeast. *Mol Cell* 2006; **21**:379–91.
75. Merz K, Hondele M, Goetze H, et al. Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. *Genes Dev* 2008; **22**:1190–204.
76. Goetze H, Wittner M, Hamperl S, et al. Alternative chromatin structures of the 35S rRNA genes in *Saccharomyces cerevisiae* provide a molecular basis for the selective recruitment of RNA polymerases I and II. *Mol. Cell Biol* 2010; **30**:2028–45.
77. Reiter A, Hamperl S, Seitz H, et al. The Reb1 homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *EMBO J* 2012; **31**:3480–93.
78. Grünberg S, Zentner GE. Genome-wide mapping of protein-DNA interactions with ChEC-seq in *Saccharomyces cerevisiae*. *J Vis Exp* 2017; **124**:e55836.
79. Zentner GE, Kasinathan S, Xin B, et al. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat Commun* 2015; **6**:8733.
80. Grünberg S, Henikoff S, Hahn S, et al. Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. *EMBO J* 2016; **35**:2435–46.
81. Grünberg S, Zentner GE. Genome-wide characterization of Mediator recruitment, function, and regulation. *Transcription* 2017; e1291082.
82. Jeronimo C, Robert F. Kin28 regulates the transient association of Mediator with core promoters. *Nat Struct Mol Biol* 2014; **21**:449–55.
83. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 2017; **6**:e21856.
84. Mo A, Mukamel Eran A, Davis Fred P, et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 2015; **86**:1369–84.
85. Cléard F, Karch F, Maeda RK. DamID as an approach to studying long-distance chromatin interactions. In: Graba Y, Rezsöházy R (eds). *Hox Genes: Methods and Protocols*. New York, NY: Springer New York, 2014, 279–89.
86. Wang H, Johnston M, Mitra RD. Calling cards for DNA-binding proteins. *Genome Res* 2007; **17**:1202–9.
87. Zhu Y, Zou S, Wright DA, et al. Tagging chromatin with retrotransposons: target specificity of the *Saccharomyces Ty5* retrotransposon changes with the chromosomal localization of Sir3p and Sir4p. *Genes Dev* 1999; **13**:2738–49.
88. Zhu Y, Dai J, Fuerst PG, et al. Controlling integration specificity of a yeast retrotransposon. *Proc Natl Acad Sci USA* 2003; **100**:5891–5.
89. Wang H, Mayhew D, Chen X, et al. Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res* 2011; **21**:748–55.
90. Gimeno CJ, Ljungdahl PO, Styles CA, et al. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: Regulation by starvation and RAS. *Cell* 1992; **68**:1077–90.
91. Mayhew D, Mitra RD. Transcription factor regulation and chromosome dynamics during pseudohyphal growth. *Mol Biol Cell* 2014; **25**:2669–76.
92. Wang H, Mayhew D, Chen X, et al. “Calling Cards” for DNA-binding proteins in mammalian cells. *Genetics* 2012; **190**:941–9.
93. Marshall OJ, Brand AH. damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* 2015; **31**:3371–3.
94. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010; **38**:576–89.