Published online 11 November 2012 Nucleic Acids Research, 2013, Vol. 41, No. 2 e38
doi:10.1093/nar/gks1034

A multi-parametric flow cytometric assay to analyze DNA–protein interactions

Mandana Arbab[1], Shaun Mahony[2], Hyunjii Cho[2], Joel M. Chick[3], P. Alexander Rolfe[2], John Peter van Hoff[1], Viveca W.S. Morris[1], Steven P. Gygi[3], Richard L. Maas[1], David K. Gifford[2] and Richard I. Sherwood[1,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 and [3]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

Received March 30, 2012; Revised September 25, 2012; Accepted October 5, 2012

ABSTRACT

Interactions between DNA and transcription factors (TFs) guide cellular function and development, yet the complexities of gene regulation are still far from being understood. Such understanding is limited by a paucity of techniques with which to probe DNA–protein interactions. We have devised magnetic protein immobilization on enhancer DNA (MagPIE), a simple, rapid, multi-parametric assay using flow cytometric immunofluorescence to reveal interactions among TFs, chromatin structure and DNA. In MagPIE, synthesized DNA is bound to magnetic beads, which are then incubated with nuclear lysate, permitting sequence-specific binding by TFs, histones and methylation by native lysate factors that can be optionally inhibited with small molecules. Lysate protein–DNA binding is monitored by flow cytometric immunofluorescence, which allows for accurate comparative measurement of TF-DNA affinity. Combinatorial fluorescent staining allows simultaneous analysis of sequence-specific TF-DNA interaction and chromatin modification. MagPIE provides a simple and robust method to analyze complex epigenetic interactions *in vitro*.

INTRODUCTION

The regulation of gene expression by transcription factors (TFs) and by epigenetic interactions is at the core of cellular function and development. DNA is bound by sequence-specific TFs, covalently modified by methylation and wrapped around histones with chemical alterations that vary its accessibility, all of which contribute to regulation of gene expression. Gaining a better understanding of these mechanisms will allow more accurate understanding and manipulation of tissue-specific gene expression, which carries significant implications for many areas of biology.

Traditionally, TFs have been considered the primary regulators of cell type-specific gene expression. TFs bind to 5–15 base pair sequences of DNA and can activate or repress gene expression by interaction with RNA polymerase II through coactivators (1,2) or by alteration of chromatin to facilitate or restrict gene expression indirectly (3,4). TFs bind *in vivo* to only a small fraction of the sequences they can bind to *in vitro* (5), and predictions of which sites a TF will occupy in a given cell type are imprecise. Our inability to accurately predict TF binding can be partly explained by differential chromatin structure and partly by the fact that TFs are known to operate in large multi-component complexes (6), and our knowledge of the logic of multi-TF DNA-binding interactions is rudimentary.

In addition to TFs, covalent modifications of DNA and histones have been found to affect gene expression. DNA methylation represses genes by recruitment of repressive DNA-binding proteins and interference with transcriptional activators (7). Furthermore, a wide array of histone modifications (e.g. methylation, acetylation, phosphorylation and ubiquitinylation) has been found to be positively and negatively correlated with gene expression at enhancer and promoter regions (8). However, how the DNA sequences and proteins contribute to sequence-specific chromatin alteration is largely unexplained.

Much research has focused on elucidating the mechanisms through which TFs recognize and bind specific DNA sequences and how they influence gene expression. *In vitro* techniques involve incubating nuclear lysate or purified

*To whom correspondence should be addressed. Tel: +1 617 525 4772; Fax: +1 617 525 4751; Email: rsherwood@partners.org
Present address:
Mandana Arbab, Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

TFs with DNA containing putative binding sites. In DNase I footprinting, the DNA is cleaved with DNase I after protein incubation, revealing binding sites by their ability to protect DNA from DNase I cleavage (9). In electrophoretic mobility shift assay (EMSA), DNA–protein complexes are run on a gel, revealing impeded mobility of DNA regions that are protein bound (10). More recently, high-throughput protein-binding microarrays (PBMs) have been developed in which TFs are incubated with an arrayed collection of small oligonucleotides and stained using immunofluorescence, leading to information on binding preference for all possible oligonucleotide sequences (11,12). These assays have proved powerful at elucidating the DNA-binding preferences of TFs, allowing for generation of position weight matrices (PWMs) that represent TF sequence specificity. However, the assay has been used primarily with purified TFs or defined combinations of TFs (13), thus overlooking complexities that exist within a cell. Furthermore, PBM assays do not easily allow for simultaneous multi-parametric analysis or simultaneous analysis of TF binding and chromatin modifications.

In vivo techniques are also useful at examining TF binding. In chromatin immunoprecipitation (ChIP), TFs are crosslinked to bound DNA in live cells, and by fragmenting DNA, immunoprecipitating a TF of interest and eluting DNA, *in vivo* bound DNA can be assayed for by PCR or microarray (14). Recently, genome-wide mapping of TF-binding sites *in vivo* has become feasible by combining ChIP with massively parallel DNA sequencing (ChIP-Seq) (5). ChIP has also been used to map genomic chromatin modifications (15). ChIP-Seq is a powerful technique for analyzing TF binding and chromatin modification, but it is not easily amenable to manipulation of binding sites nor does it allow for multi-parametric analysis. Mapping of open chromatin by DNase I hypersensitivity sequencing (DNase-Seq) (16) or formaldehyde-assisted isolation sequencing (FAIRE-Seq) (17) provides a broader view of active regulatory regions yet does not provide information about the underlying epigenetic events. Thus, *in vitro* techniques must be used to analyze DNA-binding sites identified in ChIP-Seq, DNase-Seq and FAIRE-Seq experiments.

We have devised a method, magnetic protein immobilization on enhancer DNA (MagPIE), to allow rapid, high-throughput, semi-quantitative, multi-factorial analysis of interactions among TFs, chromatin and DNA. In this assay, DNA is bound to magnetic beads using biotin–streptavidin (SA) interaction, and these DNA-coated magnetic beads are employed to pull down sequence-specific TFs and potentially their partners from crude nuclear lysate in a 10-min binding reaction (Figure 1). By immunostaining with fluorescently tagged TF-specific antibodies followed by flow cytometric analysis, TF occupancy on the DNA can be measured.

MagPIE allows analysis not only of TF binding but also of epigenetic modifications, as use of crude nuclear lysates allows DNA to be modified by methylation and to be packaged into chromatin, and we show using immunostaining for modified chromatin that both of these processes occur in a regulated fashion. By combinatorial fluorescent staining, both TF-DNA interaction and chromatin modification can be analyzed at once with future possibility to analyze a wide array of DNA-mediated interactions simultaneously. Thus, we have devised a method to study relative affinities of TF-DNA interactions while allowing simultaneous probing of multiple proteins and chromatin modifications.

MATERIALS AND METHODS

DNA amplification and bead binding

Oligonucleotides used in this study are listed in Supplementary Table S1 and were ordered from Integrated DNA Technologies. Significantly, all DNA regions were amplified and attached to beads using primers shown previously to have minimal affinity to TFs (11). PCR was performed using Onetaq 2X Master Mix or Phusion 2X Master Mix (New England Biolabs) using recommended PCR conditions.

SA-coated magnetic beads (0.5 µl M-280 Streptavidin Dynabeads per experiment, Life Technologies) are washed twice with 1× DNA-binding buffer (1× DBB: 10 mM Tris–HCl, pH 7.5, 1 M NaCl, 500 µM EDTA in H_2O) and precipitated magnetically using a $1'' \times 1'' \times 1''$ NIB magnet (McMaster Carr). Beads are resuspended in 2× DNA-binding buffer (2× DBB: 20 mM Tris–HCl, pH 7.5, 2 M NaCl, 1 mM EDTA in H_2O), and an equal volume of biotinylated DNA is added. DNA is incubated with beads for 20–60 min at RT with rotation. Beads are washed twice with 1× DBB, blocked with 4% milk (Sigma) and stored in 1× protein wash buffer (1× PWB: 10 mM Tris–HCl pH 8.0, 5 mM $MgCl_2$, 50 mM KCl, 50 mM NaCl, 4% glycerol, 0.05% Triton X-100 [Sigma] with 1:25 Protease Inhibitor Cocktail [PIC, Roche] and 10 mM DTT added fresh).

Cell culture

Undifferentiated 129P2/OlaHsd mouse ES cells and doxycycline-inducible V5-tagged Cdx2 or Onecut1 mouse ES cells (18) were maintained on gelatin-coated plates with mouse embryonic fibroblast feeders in Knockout DMEM (Invitrogen) supplemented with 15% ES tested fetal bovine serum (FBS) (HyClone), 0.1 mM nonessential amino acids (Invitrogen), Glutamax (Invitrogen), 0.55 mM 2-mercaptoethanol (Sigma) and 1× ESGRO LIF (Chemicon). For expansion, ES cells were passaged onto gelatin-coated 15 cm dishes and cultured in Knockout DMEM (Invitrogen) supplemented with 10% ES tested FBS (HyClone), 0.1 mM nonessential amino acids (Invitrogen), Glutamax (Invitrogen) and 0.55 mM 2-mercaptoethanol (Sigma). When necessary, cells were treated with 2 µg/ml doxycycline for 24–48 h before harvest to induce ectopic TF expression.

Cell lysate preparation

Cell lysate preparation was performed according to the CellLytic protocol (Sigma). Briefly, adherent 15 cm plates of mouse embryonic stem cells are washed twice with ice cold PBS, 2 ml PBS is added and cells are

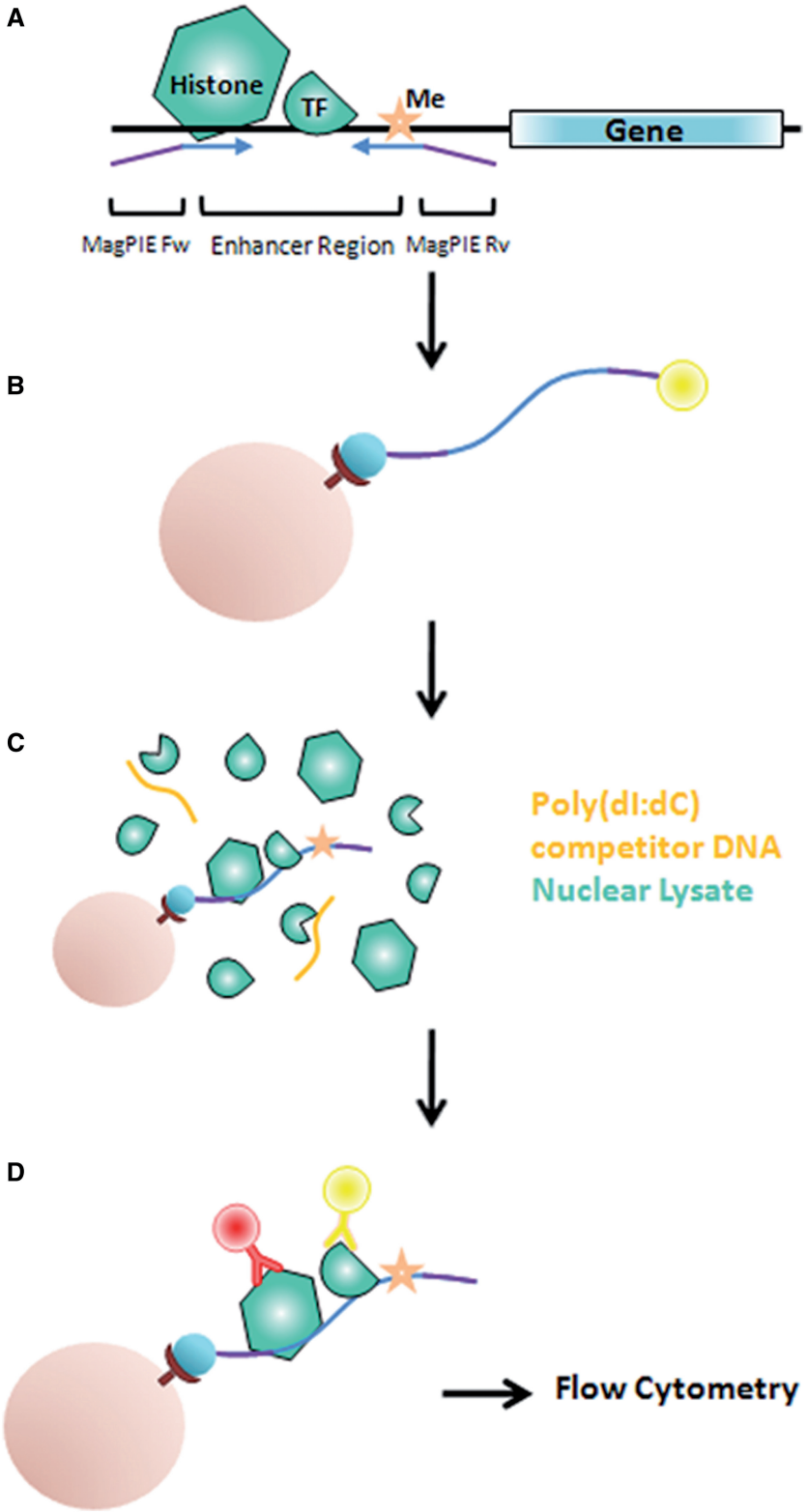

Figure 1. MagPIE experimental flowchart. (**A**) Genomic regions are amplified by PCR using sequence-specific primers linked to MagPIE primer sequences predicted to have low TF-binding affinity. The DNA is further amplified using a biotinylated MagPIE reverse primer and a MagPIE forward primer that may be tagged with a fluorophore. (**B**) Biotinylated DNA is captured on SA-coated magnetic beads. (**C**) Bead-immobilized DNA is incubated with crude nuclear protein lysate in the presence of competitor poly(dI:dC) DNA to pull down sequence-specific nuclear factors in a 10-min binding reaction at 37°C. (**D**) TFs and other DNA-bound proteins are immunostained with fluorescently tagged antibodies, and fluorescence intensity is analyzed by flow cytometry.

scraped and pelleted at 13 500 × g for 5 min at 4°C. Cells are resuspended in 1 ml/plate lysis buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$ and 10 mM KCl with 1:25 PIC and 10 mM DTT added fresh) and incubated for 15 min at 4°C with rotation; 60 µl/plate of 10% Igepal-630 (Sigma) is added, and the mixture is vortexed for 15 s and pelleted at 13 500g for 5 min at 4°C. Nuclei are resuspended in 1 ml/plate equilibration buffer (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 µM EDTA, 30 mM KCl and 25% glycerol with 1:25 PIC and 10 mM DTT added fresh) and pelleted at 12 000g for 4 min at 4°C. Nuclei are resuspended in 50–100 µl/plate nuclear lysis buffer (20 mM HEPES pH7.9, 1.5 mM $MgCl_2$, 420 mM NaCl, 200 µM EDTA and 25% glycerol with 1:25 PIC and 10 mM DTT added fresh), vortexed for 1 min and incubated for 15–30 min at 4°C with rotation. Lysate is centrifuged at 12 000g for 4 min at 4°C, and supernatant is passed through Ultrafree-MC columns (0.65 µm pore size, Millipore). Protein concentrations of 5–10 mg/ml are routinely obtained. Nuclear lysate is aliquotted, frozen in liquid nitrogen and stored at −80°C.

Protein–DNA binding and downstream analysis

Nuclear lysate is thawed on ice and diluted in protein-binding buffer (PBB), such that lysate is at a final concentration of ~1–3 mg/ml, and reaction has a final concentration of 10 mM Tris–HCl pH8.0, 5 mM $MgCl_2$, 50 mM KCl, 50 mM NaCl, 4% Glycerol, 0.05% Triton X-100 (Sigma) with 1:25 Protease Inhibitor Cocktail (PIC, Roche), 10 mM DTT, 0.01% BSA, 500 µM ATP and 7 µg/ml poly(dI:dC) (Sigma) added fresh. DNA-coated beads are precipitated magnetically, and 5 µl of lysate:PBB mix is added to 0.5 µl of DNA-coated beads, mixed thoroughly and incubated for 10 min at 37°C. Binding is terminated by placing the reaction on ice and adding 50 µl PWB. Beads are washed twice in PWB.

After magnetic precipitation, beads are resuspended in 10 µl antibody solution containing all primary and fluorescent secondary antibodies. Staining is performed for 20–60 min at 4°C with rotation. Beads are washed twice in PWB and resuspended in 50 µl PWB for flow cytometric analysis using an Accuri C6 flow cytometer (BD Biosciences).

For mass spectrometric preparation, reactions were scaled up 100-fold keeping all components proportional. Beads were washed eight times with PWB. Protein precipitation was performed on each sample using a final concentration of tricloroacetic acid of 20% and incubated on ice for 1 h. Samples were centrifuged at 10 000g for 5 min after which the supernatant was discarded and the pellet washed twice with 100% ice cold acetone. Each sample was then resuspended in 50 ml of 8 M urea and then subsequently diluted to 1.5 M urea using 50 mM Tris (pH 8.2). Samples were digested overnight at 37°C using trypsin. Each sample was desalted using self-packed C18 STAGE-tips (19), dried down using a vacuum centrifuge and resuspended in 1% formic acid. All LC−MS/MS data were obtained using an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher). Each sample was loaded onto a reverse phase column and separated using 120 min LC gradient of 5–27% buffer B at a flow of 0.5–1 µl/min. MS analysis was performed using top10 methods where the MS1 scan was acquired in the Orbitrap followed by 10 data-depended MS/MS scans on the 10 most intense ions in the LTQ with CID for fragmentation. MS/MS spectra assignments were made with the Sequest algorithm (20) using the entire mouse IPI database (version 3.6). Sequest searches were performed using a target-decoy strategy (21) with the mouse IPI database in correct orientation (forward database) and the same database but with all sequences in reverse orientation (reverse database). Sequest searching was performed with a precursor ion tolerance of 20 ppm with trypsin specificity, cysteine carboxyamidomethylation was considered a static modification and methionine oxidation was considered a dynamic modification. A protein level false discovery rate of less than 1% was used as a threshold for protein identifications using the target decoy strategy.

DNase I digestion of beads was performed by incubating DNA-coated beads either alone or after protein lysate binding with 1:100 final concentration of DNase I (Promega) in DNase I digestion buffer (Promega). Reactions were stopped after 3 min by addition of DNase I stop solution (Promega), and supernatant was run on a 2.5% agarose gel and imaged using a Typhoon Imager (GE Healthcare).

Antibodies used include mouse anti-V5 (R960-25, Invitrogen), mouse anti-Cdx2 (Biogenex), rabbit anti-Histone 3 Lysine 4 Monomethyl (Cell Signaling), sheep anti-Cytosine, 5-methyl (US Biological), sheep anti-Onecut1 (R&D Systems) and mouse anti-Sox2 (R&D Systems). Secondary antibodies include DyLight488 and DyLight649 conjugates (Jackson Immunoresearch) and FITC (Fisher Scientific).

Statistical analyses were performed using a Student's *t*-test unless otherwise stated.

Comparing Cdx2 MagPIE flow cytometric intensities to binding affinity metrics

Scores from three binding affinity metrics were compared with Cdx2 MagPIE intensities for the collection of 16 tested sequences normalized, such that the flow cytometric percentage of $Cdx2^+$ beads in the sequence predicted by Uniprobe to be the strongest binding sequence was set to 1. *Z* scores for each of the variable 8-mers were sourced from Cdx2 PBM contiguous 8-mer analysis files downloaded from the Uniprobe database, accession number UP00133 (11) and normalized, such that the highest score was set to 1. The bead-attached sequences were also scored against a Cdx2 consensus motif using two related metrics: (i) the standard log-likelihood ratio score from the PWM and (ii) the probability of binding occupancy defined by Granek and Clarke (22) The Cdx2 consensus motif was defined from Cdx2 ChIP-seq experiments. Specifically, binding events were predicted in the Cdx2 ChIP-seq data using GPS (23), and the GimmeMotifs (24) software package was used to perform *de novo* motif finding on 200-bp windows centered on the 2000 most enriched ChIP-seq-binding events.

RESULTS

MagPIE allows flow cytometric detection of TF-DNA binding

As a first step to using flow cytometry to study TF-DNA interactions, we monitored binding of PCR-amplified 5′-biotinylated DNA to SA-coated magnetic beads. By attaching a fluorophore, TYE665, to the non-biotinylated primer, we were able to monitor bead-bound DNA concentration by flow cytometry. We found DNA concentration to steadily increase with increasing amounts of DNA up to a saturation point of ~300 ng of 130 bp biotinylated DNA per µl of beads (Supplementary Figure S1). To assess the efficiency of DNA binding to beads, we performed restriction digest of the bead-bound DNA and DNA quantification, which confirmed that ~150 ng DNA or ~1 million DNA molecules were bound per bead (data not shown). For all future experiments, concentrations >500 ng of DNA per µl beads were used, such that beads were saturated with DNA.

To determine whether DNA–protein interactions can be analyzed using flow cytometry, DNA-bound SA beads were incubated with crude nuclear lysate, and DNA–protein interactions were monitored using indirect immunofluorescence staining. For these experiments, we focused on binding of Cdx2, a TF which is required for proper intestinal development (18,25), and whose *in vitro* and *in vivo* binding preferences have been determined extensively by PBM analysis and ChIP-Seq (26–28) (Sherwood *et al.*, manuscript in preparation). DNA amplified using MagPIE primers from either a library of randomized 130 bp sequences or from a 150 bp enhancer region displaying strong Cdx2 ChIP-seq binding (Sherwood *et al.*, manuscript in preparation) were incubated with nuclear lysate obtained from a mouse embryonic stem cell line that inducibly expresses V5 epitope-tagged Cdx2 protein (18) to allow monitoring of Cdx2 binding by means of an anti-V5 antibody (protocol detailed in Figure 1). A comparison of incubation temperature and timing identified 10-min incubation at 37°C as optimal for rapid analysis of DNA–protein interaction (data not shown), and incubation was performed in the presence of poly(dI:dC) competitor DNA (10). Flow cytometric analysis revealed a strong enrichment of Cdx2 binding to binding site-containing DNA, when compared with randomized DNA (Figure 2A and B).

To optimize binding conditions, concentration of lysate and competitor DNA were tested. Concentration of Cdx2-containing lysate appears to have a linear binding relationship with binding site-containing DNA up to a saturation point (Figure 2C) and usage of lysate from $>6 \times 10^3$ cells per µl of final reaction volume ($>3 \times 10^4$ cells per reaction) corresponding to ~1–3 µg/µl protein (5–15 µg per reaction), a relatively minute quantity of protein compared with *in vivo* binding techniques, was found to allow for optimal enrichment and was used for the rest of the experiments. Concentration of poly(dI:dC) competitor DNA profoundly affected the comparative binding of Cdx2 to binding site-containing DNA versus randomized DNA (Supplementary Figure S2) with specific interactions optimally enriched at 7–14 µg/ml poly(DI:DC). The amount of DNA-bound SA beads added to a defined concentration of lysate was also found to affect Cdx2 binding (data not shown), so bead amount was always kept constant at 0.5 µl (3×10^5 beads) per reaction. Although distinct TFs, especially if not ectopically overexpressed as Cdx2 was in these experiments, may require individualized optimization of lysate and competitor DNA concentration, these values provide rough estimates to test TF-DNA binding.

To confirm that Cdx2-DNA binding was dependent on the presence of a Cdx2 binding site, we analyzed the effect of binding site mutations. To facilitate this analysis, a 40 bp ChIP-seq Cdx2-binding site with a strong [predicted by Uniprobe to be functional (26)] and a weak (predicted to be non-functional) Cdx2 motif were chosen for mutation analysis. The motifs were amplified using the MagPIE forward and MagPIE biotinylated reverse primers making the total sequence 88 bp long. Although the strong 40 bp region displayed similar Cdx2-binding affinity to the 150-bp Cdx2-binding region (Figure 2D, sequence shown in Figure 2F as WT), missense mutation of the strong binding site resulted in a dramatic loss of Cdx2 binding to levels seen in DNA containing MagPIE primers alone, which are predicted to have a low TF-binding affinity (Figure 2E, missense mutation sequence shown in Figure 2F as no. 1 and MagPIE primer sequence shown in Supplementary Table S1). Mutation of the weak binding site did not affect Cdx2 binding (Figure 2F and G), confirming the prediction that Cdx2-DNA binding depends exclusively on the strong binding site. Additional analysis of 40 bp Cdx2-binding site regions containing 3–5 bp mutations at various locations revealed that mutations in three segments of the strong Cdx2-binding site each completely ablate Cdx2 binding, whereas mutations in multiple regions outside of the strong Cdx2-binding site do not affect Cdx2 binding (Figure 2F and G).

To ensure that the optimized conditions developed extend to factors other than Cdx2 and to lysates without ectopically expressed TFs, further MagPIE analysis was performed. Beads bound with a 138 bp region from the Onecut1 promoter containing a strong Onecut1 binding site and incubated with lysate ectopically expressing Onecut1 display stronger flow cytometric immunofluorescence for Onecut1 than MagPIE primer DNA or similar 188-bp Onecut1-binding site DNA with the Onecut1-binding site mutated (Figure 2H). Similarly, beads bound with an 88-bp DNA sequence containing a strong Sox2-binding site and incubated with lysate from mES cells which endogenously express Sox2 display stronger flow cytometric immunofluorescence for Sox2 than beads bound with a similar 88 bp DNA sequence in which the Sox2-binding site is mutated (Figure 2I). Thus, MagPIE allows flow cytometric analysis of TF-DNA interaction in the context of crude nuclear lysate, and binding depends strictly on the presence of high affinity binding sites.

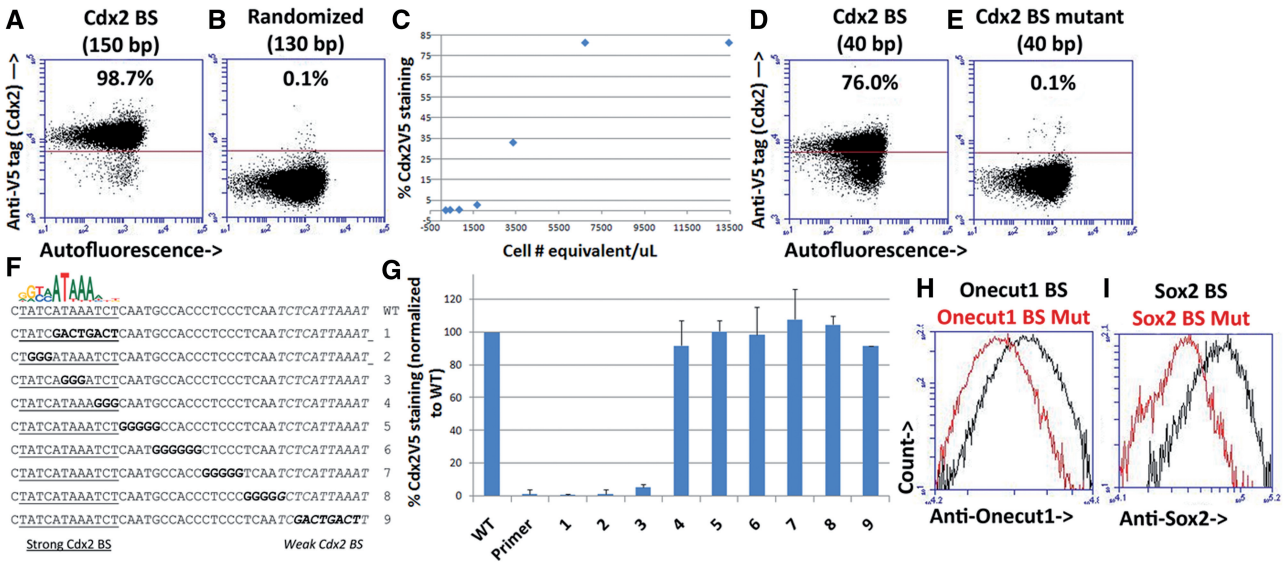

Figure 2. MagPIE allows flow cytometric detection of TF-DNA binding. Flow cytometric plots showing bead immunofluorescence intensity for V5 tagged Cdx2 (*Y* axis) vs. autofluorescence (*X* axis) for beads coated with 150 bp Cdx2-binding site containing DNA (**A**) or 130 bp randomized DNA (**B**) and incubated with lysate from mES cells ectopically expressing V5-tagged Cdx2. Intensity is provided as percentage of events above a threshold, which is set at background level to simplify the distinction of signal to noise. (**C**) Graph showing percentage of beads above a threshold level of flow cytometric V5-tagged Cdx2 immunofluorescence (*Y* axis) vs. concentration of lysate from mES cells ectopically expressing Cdx2 in cell number equivalent per microliter (*X* axis). Flow cytometric plots showing bead immunofluorescence intensity for V5-tagged Cdx2 (*Y* axis) vs. autofluorescence (*X* axis) for beads coated with 88 bp DNA sequences containing a 40 bp Cdx2 binding site (**D**, sequence shown as WT in panel **F**) or 88 bp DNA sequences containing a 40 bp Cdx2 mutant binding site (**E**, sequence shown as no. 1 in **F**) and incubated with lysate from mES cells ectopically expressing V5-tagged Cdx2. (F) Mutation series of a 40 bp enhancer fragment with Cdx2 ChIP-seq binding containing a strong Cdx2 binding site (underlined, with Uniprobe motif overlaid above sequence) and a weak Cdx2 binding site (italicized) with mutations bolded. Amplified enhancer fragments have a total DNA length of 88 bp. (**G**) Bar plot showing percentage of beads above a threshold level of flow cytometric V5 tagged Cdx2 immunofluorescence (*Y* axis, normalized to WT as 100%) for beads coated with DNA containing the 10 40 bp sequences shown in panel F and a negative control with 48 bp MagPIE primer sequence alone and incubated with lysate from mES cells ectopically expressing Cdx2. (**H**) Comparative flow cytometric plots showing bead immunofluorescence intensity for Onecut1 (*X* axis) for beads coated with 138 bp Onecut1 binding site-containing enhancer DNA (black) or 188 bp Onecut1 binding site mutant DNA (red) and incubated with lysate from mES cells ectopically expressing Onecut1. (**I**) Comparative flow cytometric plots showing bead immunofluorescence intensity for Sox2 (*X* axis) for beads coated with 88 bp DNA sequences containing a 40 bp Sox2 binding site (black) or 88 bp DNA sequences containing a 40 bp Sox2 mutant binding site (red) and incubated with lysate from wild-type mES cells.

MagPIE allows for comparative analysis of TF-DNA-binding affinity from crude lysates

To explore whether MagPIE allows accurate comparative analysis of TF-DNA-binding affinity, we tested a set of 16 8-mers with a range of experimentally determined Cdx2 affinities from the Uniprobe database for binding of Cdx2 in the context of crude nuclear lysate. Flow cytometric Cdx2 immunofluorescence intensity of the 8-mers is consistent with the rank of their similarity to the consensus Cdx2 motif (11) (Figure 3A), suggesting that MagPIE intensity correlates with binding affinity.

We performed correlation analysis of the average flow cytometric Cdx2 intensities for the 16 Cdx2 affinity 8-mers, averaged over three biological replicates, with several metrics of predicted binding affinity. MagPIE flow cytometric immunofluorescence displays highly significant correlation with the Cdx2 PBM *Z* score (Figure 3B; $r^2 = 0.87$, $P < 0.0001$), log-likelihood ratio scoring against the ChIP-Seq Cdx2 motif PWM (Figure 3C; $r^2 = 0.84$, $P < 0.0001$) and a statistical representation of probability of TF binding site occupancy (22) (Supplementary Figure S3; $r^2 = 0.89$, $P < 0.0001$). Using this TF-binding site occupancy prediction as a rough metric, above-background Cdx2-DNA binding in MagPIE can be detected when occupancy is predicted at 0.1% of DNA strands and UniPROBE PBM *Z* score is above 5 or PBM Enrichment Score is above 0.46 (29), displaying the sensitivity of MagPIE. This sensitivity depends on antigen–antibody strength and thus may vary for different TFs. As PBM scores have been shown to correlate well with the inverse of Kd values (11), we used the PBM Enrichment Score as an independent measure of TF binding strength. These data indicate that MagPIE also correlates inversely with Kd, and thus, in spite of using crude nuclear lysates, not purified TFs, MagPIE allows accurate comparative measurement of TF-DNA-binding affinity.

MagPIE allows simultaneous measurement of multiple distinct epigenetic events

Epigenetic modification of DNA encompasses methylation of DNA and the modification of its component histones. Because our experiments are performed using crude nuclear lysates that contain a wide range of nuclear proteins including DNA modifying enzymes, histones and histone modifying enzymes, we explored whether these epigenetic events can be studied and manipulated using MagPIE. To analyze DNA

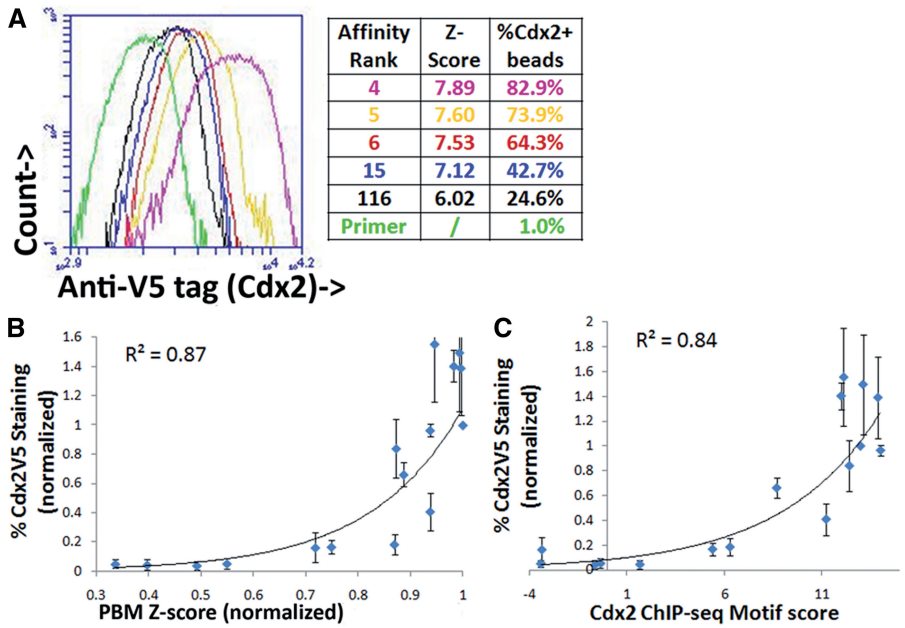

Figure 3. MagPIE allows accurate comparative analysis of TF-DNA binding affinities. (**A**) Comparative flow cytometric plots showing bead immunofluorescence intensity for V5 tagged Cdx2 (*X* axis) for beads coated with 48 bp MagPIE primer DNA (green) or beads coated with 62 bp DNA sequences containing 8-mers of differing predicted Cdx2 affinity (purple, yellow, red, blue and black) and incubated with lysate from mES cells ectopically expressing Cdx2. Graphs showing percentage of beads above a threshold level of flow cytometric V5 tagged Cdx2 immunofluorescence normalized, such that the flow cytometric percentage of $Cdx2^{+}$ beads in the strongest predicted binding sequence was set to 1 (*Y* axis) vs. Cdx2 PBM *Z* scores normalized, such that the highest score was set to 1 (**B**, *X* axis) or log-likelihood ratio scoring against the ChIP-Seq Cdx2 motif PWM (**C**, *X* axis). R^2 values for the best fit exponential curves are shown on the graphs.

methylation, 40-bp DNA-containing randomized sequences flanking two CG dinucleotide sequences to ensure the presence of a CG sequence was amplified with MagPIE primers and stained with an antibody that specifically recognizes methylated cytosine before and after incubation with nuclear lysate. We found that nuclear lysate incubation induces methyl cytosine reactivity; however pre-incubation of nuclear lysate with the non-nucleoside DNA methyltransferase inhibitor RG108 eliminates methyl cytosine reactivity ($t = 8.8$, $P < 0.001$ comparing methyl cytosine with and without RG108, Figure 4A), suggesting that DNA methylation occurs and can be selectively inhibited on the magnetic beads.

To further explore the potential of MagPIE as a tool for understanding epigenetic regulation, we examined whether bead-bound DNA is bound by histones upon protein lysate incubation. As an initial confirmation of histone binding during MagPIE, we performed mass spectrometry (LC-MS) on beads bound with Cdx2-binding enhancer DNA and length-matched control DNA sequences of 88 bp, 150 bp and 250 bp. In all experiments using these sequences, several histones H1, H2 and H3 are robustly detected (Table 1), indicating that histones bind to DNA during MagPIE, even when DNA is shorter than 147 bp, the minimum length shown to be required to assemble nucleosomes (30).

To determine whether histone binding and modification in MagPIE are regulated by sequence-dependent processes as they are *in vivo*, we turned to flow cytometric analysis of histone H3 monomethyl K4 (H3K4Me1), a marker of enhancer regions (15). We tested whether sequences that display H3K4Me1 marking in mES cells from published ChIP-Seq data (31) are preferentially marked by H3K4Me1 when incubated with mES lysate. Twelve sequences of similar length were tested (Supplementary Table S1), four of which display H3K4Me1 marking in mES cells and eight of which do not, and the sequences with *in vivo* H3K4Me1 marking were found to display significantly enriched H3K4Me1 reactivity by MagPIE ($P < 0.0001$, Figure 4B). The H3K4Me1 reactive sequences included a 200 bp strong consensus binding site for a Sox-Oct motif (Figure 4C), which is known bind the complex of Sox2 and Oct4, two TFs known to co-bind and co-regulate many stem cell-related genes (32) and the non-reactive sequences included a 226 bp sequence that does not display Sox2 binding by flow cytometry (Figure 4D). Enriched H3K4Me1 reactivity was not a result of increased histone binding, as flow cytometric analysis of total histone 3 revealed similar amounts of reactivity in H3K4Me1+ and H3K4Me1- regions (Figure 4E). To address whether histones are bound to DNA as a nucleosome, we used DNaseI digestion to analyze the effect of lysate incubation on DNaseI accessibility, which is known to be impeded by nucleosome presence (9). We synthesized a fluorescently labeled version of a 250 bp, H3K4Me1-region, bound it to beads and subjected it to DNaseI digestion before and after lysate incubation. We found that lysate incubation significantly impeded DNaseI digestion and resulted in cutting only of full-length DNA (Supplementary Figure S4), suggesting nucleosome presence. Additionally, to address whether histone methylation occurs *de novo* during

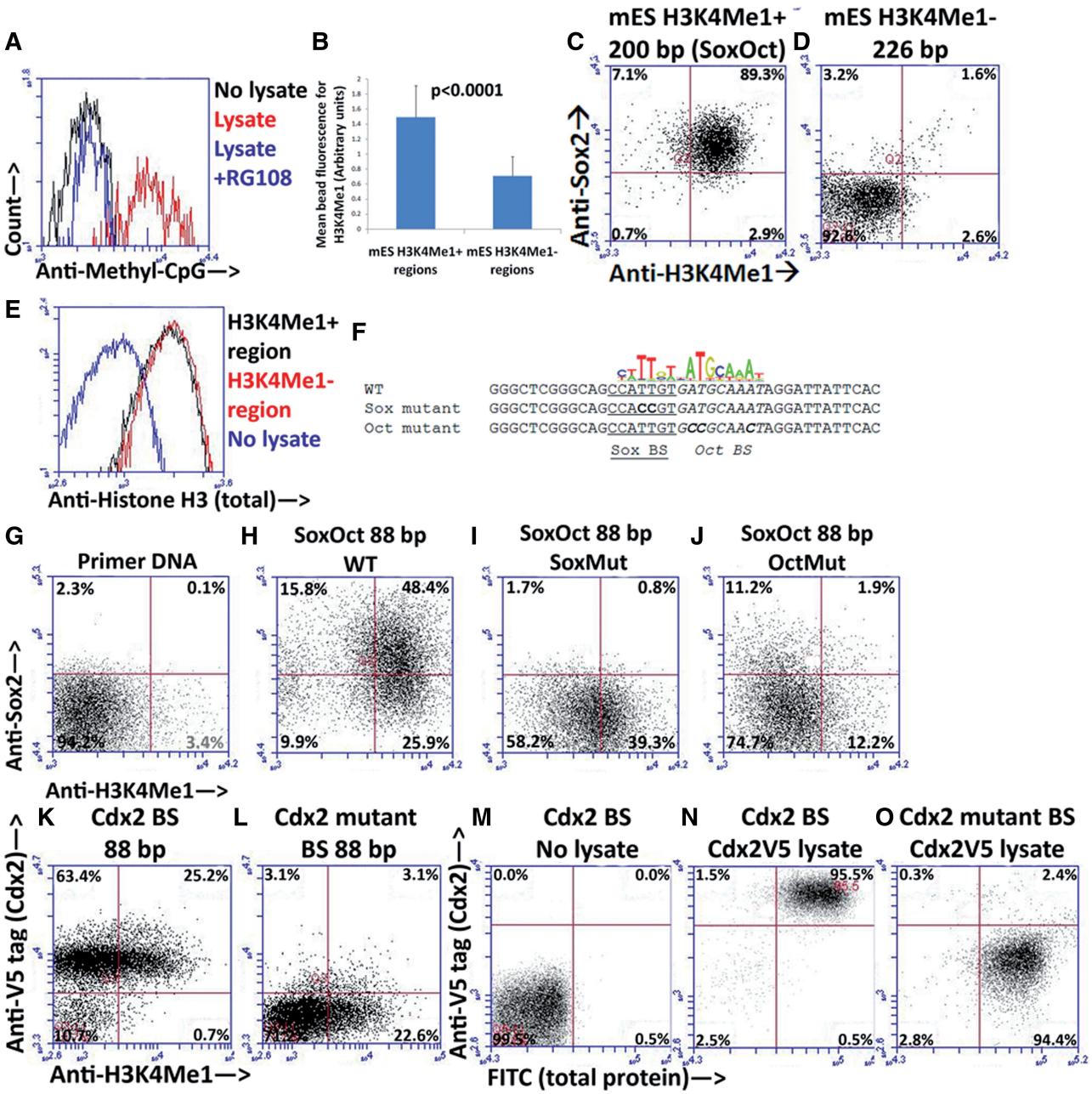

Figure 4. MagPIE allows simultaneous monitoring of protein–DNA binding and epigenetic DNA and histone methylation. (**A**) Comparative flow cytometric plots showing bead immunofluorescence intensity for methyl-CpG (*X* axis) for beads coated with 88 bp DNA containing 40 bp randomized sequences flanking two CG dinucleotides and incubated without lysate (black), with lysate from mES cells (red) or with lysate from mES cells and RG108 (blue). (**B**) Graph showing mean bead fluorescence for H3K4Me1 after incubation with mES lysate for a set of DNA regions with (left) or without (right) ChIP-Seq H3K4Me1 marking (sequences in Supplementary Table S1). Flow cytometric plots showing bead immunofluorescence intensity for Sox2 (*Y* axis) vs. bead immunofluorescence intensity for histone 3 lysine 4 monomethyl (*X* axis) for beads coated with a 200 bp DNA sequence that does (**C**) or a 226 bp sequence that does not (**D**) show Sox2 binding and H3K4Me1 marking in mES cells *in vivo*. (**E**) Comparative flow cytometric plots showing bead immunofluorescence intensity for Histone H3 total protein (*X* axis) for beads coated with DNA from a $H3K4Me1^{+}$ mES region (black) or from a $H3K4Me1^{-}$ mES region (red) and incubated with mES lysate and compared with DNA-coated beads without lysate (blue). (**F**) Mutation series of a 40 bp enhancer fragment with a strong SoxOct binding site (Sox binding site underlined, Oct binding site italicized with JASPAR motif overlaid above sequence) with mutations bolded. Amplified enhancer regions have a total DNA length of 88 bp. Flow cytometric plots showing bead immunofluorescence intensity for Sox2 (*Y* axis) vs. bead immunofluorescence intensity for histone 3 lysine 4 monomethyl (*X* axis) for beads coated with 48 bp primer DNA (**G**), 88 bp DNA sequences containing a 40 bp SoxOct wild-type binding site (**H**), 88 bp DNA sequences containing a 40 bp SoxOct Sox mutant binding site (**I**) or 88 bp DNA sequences containing a 40 bp SoxOct Oct mutant binding site (**J**) and incubated with lysate from mES cells. Flow cytometric plots showing bead immunofluorescence intensity for V5 tagged Cdx2 (*Y* axis) vs. bead immunofluorescence intensity for histone 3 lysine 4 monomethyl (*X* axis, **K** and **L**) or FITC total protein (*X* axis, **M** and **O**) for beads coated with 88 bp DNA sequences containing a 40 bp Cdx2 binding site (K, M and **N**) or 88 bp DNA sequences containing a 40 bp Cdx2 mutant binding site (L and O) and incubated with lysate from mES cells ectopically expressing Cdx2 (K, L, N and O) or without lysate (M).

Table 1. Mass spectrometric identification of histones in MagPIE samples

Histone subunit	Proteins with >5 peptides bound to 88 bp, 150 bp and 250 bp DNA
1	Hist1h1a, Hist1h1b, Hist1h1c and Hist1h1e
2	Hist1h2ba and Hist1h2bj
3	Histone 3 (variants too similar to distinguish)
4	None

MagPIE, we performed MagPIE in the presence of a non-hydrolyzable ATP analog, ATPγS. Addition of ATPγS significantly decreases H3K4Me1 reactivity (Supplementary Figure S4), suggesting that histone methylation is actively recruited during lysate incubation.

The ability to detect sequence-specific H3K4Me1 reactivity in mES lysate opens up the possibility of dissecting causal relationships between TF binding and histone methylation through simultaneous flow cytometric examination. To begin to address whether TF binding to a DNA sequence is correlated with H3K4Me1 marking of histones that have been wrapped by the bead-bound DNA, we synthesized a 40 bp region containing the strong consensus Sox-Oct binding site in the H3K4Me1+ enhancer analyzed in Figure 4C (Figure 4F) and amplified this region using the MagPIE primers. When incubated with mES lysate and stained simultaneously for Sox2 and H3K4Me1, this region displays strong Sox2 and H3K4Me1 reactivity, when compared with primer DNA (Figure 4G and H). H3K4Me1 reactivity and affinity to Sox2 of the 40 bp region is comparable to that of the full length 248 bp Sox-Oct sequence (Figure 4C). To test the relationship between TF binding and H3K4Me1 recruitment, we synthesized 88 bp sequences with point mutations in either the Sox2 or Oct4 binding site (Figure 4F). When incubated with mES lysate, both sites display decreased Sox2 binding and H3K4Me1 reactivity (Figure 4I and J); however, these processes are differently affected. Mutation of the Sox2 binding site ablates Sox2 binding but has only a mild effect on H3K4Me1 marking (Figure 4I), whereas mutation of the Oct4 binding site reduces the level of Sox2 binding and entirely extinguishes H3K4Me1 reactivity (Figure 4J) to the levels comparable to an H3K4Me1 non-reactive DNA sequence (Figure 4D). These initial data suggest that Oct4 binding more significantly affects H3K4Me1 recruitment by a Sox-Oct motif than Sox2 does. Furthermore, the diminished Sox2 binding after Oct4 site mutation (Figure 4J) indicates a cofactor relationship between Sox2 and Oct4 on the Sox-Oct site.

To determine whether known TF binding induces histone methylation in a distinct setting, 88 bp DNA sequences containing a strong Cdx2-binding site or a similar sequence with a mutated Cdx2-binding site were incubated with lysate from mES cells overexpressing Cdx2 and probed for both Cdx2 and H3K4Me1. Although Cdx2-binding site presence affects Cdx2 binding, H3K4Me1 reactivity is present with Cdx2-binding site containing DNA and Cdx2-binding site mutant DNA (Figure 4K and L). To investigate possible reasons for the lack of effect of Cdx2 mutation on histone methylation, we used the fluorescent amine dye FITC in addition to the other two probes to indicate levels of total bound protein. We find that total protein levels are unaffected by Cdx2-binding site mutation (Figure 4M–O), suggesting that histone methylation at this locus may be due to binding of a distinct DNA-binding protein. Significantly, this analysis required the simultaneous probing of three distinct fluorophores, possible due to the multi-parametric nature of MagPIE. These initial analyses indicate that MagPIE is capable of dissecting the roles of co-factor interactions on DNA binding and how TF binding regulates H3K4Me1 recruitment.

DISCUSSION

Interactions between DNA and its sequence-specific binding factors comprise a code that instructs cellular function and development, yet the complexities of this code are still far from being understood. Such understanding is limited by a dearth of techniques available for modeling of complex DNA–protein interactions. By adapting the commonly used method of *in vitro* protein–DNA binding to flow cytometric analysis, we have devised a multi-parametric assay to study interactions among TFs, chromatin and DNA.

We show that by binding DNA to magnetic beads using biotin–SA chemistry, incubating DNA-bound beads with crude nuclear lysate, performing immunofluorescent staining and analyzing beads by flow cytometry (Figure 1), we can monitor TF-DNA-binding interactions. As in the electrophoretic TF-DNA binding assay EMSA (10), sequence-specific binding of a TF and DNA depends on lysate concentration, concentration of competitor DNA and TF-DNA affinity (Figure 2). In MagPIE, each individual bead essentially represents an independent binding experiment with 1 million DNA molecules. The sensitivity provided by flow cytometry allows for rapid data generation (under 2 h total per experiment) using small quantities of protein (~5–15 μg/sample) that is easily scalable to high-throughput screening approaches and is highly reproducible across biological replicates. TF-DNA complexes formed on beads under these conditions are stable for several hours and can be stored at 4°C overnight with minor loss to fluorescence intensity by flow cytometric analysis (data not shown).

MagPIE provides a simple method to compare sequence-specific DNA-binding affinities of a TF in a complex, biologically relevant lysate (Figure 3). This comparative quantitation of TF-DNA affinity is not adequately provided by EMSA; in this regard, MagPIE provides similar information to PBM technology (11). PBM assays can explore more sequence space than MagPIE, and in direct comparison using Cdx2 as a test case, MagPIE produces above-background signal for a smaller subset of k-mers than PBM technology (Figure 3). However, PBM assays are typically performed

using purified TFs, limiting analysis of cofactor binding effects. Cofactors can affect TF-binding specificity, as is illustrated in our experiments by diminished Sox2 binding on Oct binding site mutant DNA (Figure 4) and more broadly in recent work on Hox family members, whose DNA-binding preferences depend on combinatorial interactions (33). Thus, MagPIE holds promise to dissect affinity differences in complex TF-DNA interactions. Although MagPIE could be adapted to use with purified TFs instead of crude lysates, allowing its utilization in the calculation of absolute TF-DNA affinity and specificity (34), one of the major benefits to flow cytometry is dissecting multiple simultaneous processes.

In this vein, when performed with crude lysates, MagPIE allows analysis of chromatin modification in concert with TF-DNA interaction. Within 10 min of incubation at 37°C, an array of epigenetic modifications including TF binding, DNA methylation and histone methylation can be detected and their relationship to one another can be examined (Figure 4 and Supplementary Figure S4). Small molecule inhibition indicates that these processes are occurring *de novo* during this incubation. We have collected multiple lines of evidence, including mass spectrometry, flow cytometric immunofluorescence and DNaseI digestion that together indicate that histones do bind to DNA during MagPIE. We cannot conclude from our data that intact nucleosomes are being formed, especially on fragments of DNA that are less than 147 bp, the standard size required to assemble an intact nucleosome (30); however, our consistent detection of histones on small DNA fragments could be a result of histone stabilization by cooperative interactions among the large number of DNA strands on each bead.

Nonetheless, it is clear that histone binding and methylation do occur in MagPIE (Figure 4 and Supplementary Figure S4), and histone methylation follows similar sequence-dependent patterns as *in vivo*. We have demonstrated this sequence-dependent histone methylation on >200 bp enhancer sites marked in mES cells by H3K4Me1, which show enriched H3K4Me1 in MagPIE experiments using mES cell lysate when compared with control H3K4Me1$^-$ regions. Additionally, we have begun to establish causal relationships between sequence-specific enhancer binding and histone methylation by performing mutation analysis on 40 bp enhancer fragments, which show comparable yet less dramatic differences in H3K4Me1 marking than the longer enhancers. These data, although currently limited to a small sample size of enhancer regions, already suggest that TFs can influence histone methylation to varying degrees, as loss of Oct4 results in a stronger loss of H3K4Me1 marking than loss of Sox2, and Cdx2 binding does not affect H3K4Me1 on one tested enhancer. Patterns of histone methylation on DNA fragments less than the 147 bp required to fully wrap one nucleosome correlate well with larger enhancers, yet the use of sub-nucleosomal DNA fragments raises concerns as to the physiological relevance of these data. It will be important to keep this point in mind in future mechanistic enhancer dissection experiments, as the ability to synthesize libraries of shorter DNA regions in a high-throughput manner (35) provides a major advantage for MagPIE, yet site-directed mutagenesis of specific binding sites within longer fragments of DNA will nonetheless be important to address the effects of intact nucleosomal architecture on histone methylation.

In combination, the experiments in this study hint that histone methylation machinery may be recruited to DNA in cell lysates by similar mechanisms as those that occur in vivo. These initial data point to protein–DNA interactions and modifications occurring in concert on specific DNA sequences; however flow cytometry only measures fluorescence at the level of individual beads, which contain 10^6 DNA strands each, and therefore, we are unable to detect whether such events are occurring on the same strands of DNA. The recent association of particular patterns of histone modification with states of activity in enhancers and promoters (36,37) has emphasized the importance of understanding the mechanism by which these marks are induced, yet the histone code is extremely complex (38,39), and the number of distinct types of histone modifications has now reached almost one hundred (40). Thus, methods such as MagPIE that permit the rapid analysis of the effect of DNA and protein composition on combinations of histone modifications will be powerful in deciphering how histone modifications are placed.

The ability to perform multi-parametric analysis using MagPIE holds great promise to uncovering mechanisms governing epigenetic interaction. It is common for flow cytometers to detect >10 distinct wavelengths of fluorescence, so the complexity of simultaneous protein–DNA interactions probed can be increased an order of magnitude using this assay when compared with similar assays. Such combinatorial analysis has been used to great effect in uncovering rare cell types with unique biological function (41), where the number of cell types identified has been limited solely by the number of antibody species used or the extent of directly fluorescently conjugated fluorescent antibodies available. Such an armory of directly fluorescently conjugated antibodies does not yet exist for TFs and histone modifications, but as these reagents accumulate, MagPIE should be able to help decipher the complexities of the histone code (42) by simultaneously monitoring the presence of multiple distinct TFs, multiple histone modifications, and DNA methylation on a single DNA sequence. Sequence-specific epigenetic states can then be correlated with activity using reporter assays to elucidate the epigenetic code.

MagPIE also holds promise in the analysis of distinct DNA-mediated events. Use of fluorescently labeled DNA (Supplementary Figure S1) could be useful in allowing analysis of protein–DNA binding using mixed populations of DNA or in modified *in vitro* footprinting analysis in which nuclease activity could be read out by the absence of a fluorescent tag. Use of advanced fluorimetric techniques such as FRET (43) could also allow precise localization of protein–protein and protein–DNA interactions to analyze events such as enhancer–promoter communication (44). MagPIE holds

promise as a tool for a wide array of DNA-based analytical techniques such as optimization of assays to detect novel protein–DNA interactions by mass spectrometry (45) and analysis of DNA damage using fluorescence. Thus, the combinatorial power of analysis of single events, in this case DNA-coated magnetic beads, with multi-parametric analysis broadens the scope of *in vitro* epigenetic analysis substantially.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

The authors thank Grigoriy Losyev and Jennifer Huynh for technical assistance and Trevor Siggers for advice.

FUNDING

The National Institutes of Health Common Fund [5UL1DE019581, DE019021 and 5TL1EB008540; 5R01HG002668 and P01-NS055923-01 to P.A.R.]; the Harvard Stem Cell Institute s Sternlicht Director's Fund award (to R.I.S.). Funding for open access charge: Internal funding.

Conflict of interest statement. None declared.

REFERENCES

1. Koch,F., Jourquin,F., Ferrier,P. and Andrau,J.C. (2008) Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.*, **33**, 265–273.
2. Kee,B.L., Arias,J. and Montminy,M.R. (1996) Adaptor-mediated recruitment of RNA polymerase II to a signal-dependent activator. *J. Biol. Chem.*, **271**, 2373–2375.
3. Rundlett,S.E., Carmen,A.A., Suka,N., Turner,B.M. and Grunstein,M. (1998) Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature*, **392**, 831–835.
4. Workman,J.L. and Kingston,R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.*, **67**, 545–579.
5. Johnson,D.S., Mortazavi,A., Myers,R.M. and Wold,B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497–1502.
6. Carey,M. (1998) The enhanceosome and transcriptional synergy. *Cell*, **92**, 5–8.
7. Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33(Suppl.)**, 245–254.
8. Young,R.A. (2011) Control of the embryonic stem cell state. *Cell*, **144**, 940–954.
9. Galas,D.J. and Schmitz,A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.*, **5**, 3157–3170.
10. Lane,D., Prentki,P. and Chandler,M. (1992) Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol. Rev.*, **56**, 509–528.
11. Berger,M.F., Philippakis,A.A., Qureshi,A.M., He,F.S., Estep,P.W. 3rd and Bulyk,M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.
12. Mukherjee,S., Berger,M.F., Jona,G., Wang,X.S., Muzzey,D., Snyder,M., Young,R.A. and Bulyk,M.L. (2004) Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.*, **36**, 1331–1339.
13. Siggers,T., Duyzend,M.H., Reddy,J., Khan,S. and Bulyk,M.L. (2011) Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol. Syst. Biol.*, **7**, 555.
14. Kuo,M.H. and Allis,C.D. (1999) In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, **19**, 425–433.
15. Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.
16. Boyle,A.P., Davis,S., Shulha,H.P., Meltzer,P., Margulies,E.H., Weng,Z., Furey,T.S. and Crawford,G.E. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**, 311–322.
17. Gaulton,K.J., Nammo,T., Pasquali,L., Simon,J.M., Giresi,P.G., Fogarty,M.P., Panhuis,T.M., Mieczkowski,P., Secchi,A., Bosco,D. *et al.* (2010) A map of open chromatin in human pancreatic islets. *Nat. Genet.*, **42**, 255–259.
18. Sherwood,R.I., Maehr,R., Mazzoni,E.O. and Melton,D.A. (2011) Wnt signaling specifies and patterns intestinal endoderm. *Mech. Dev.*, **128**, 387–400.
19. Rappsilber,J., Ishihama,Y. and Mann,M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.*, **75**, 663–670.
20. Eng,J.K., McCormack,A.L. and Yates,J.R. 3rd (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5**, 976–989.
21. Elias,J.E. and Gygi,S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods*, **4**, 207–214.
22. Granek,J.A. and Clarke,N.D. (2005) Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biol.*, **6**, R87.
23. Guo,Y., Papachristoudis,G., Altshuler,R.C., Gerber,G.K., Jaakkola,T.S., Gifford,D.K. and Mahony,S. (2010) Discovering homotypic binding events at high spatial resolution. *Bioinformatics*, **26**, 3028–3034.
24. van Heeringen,S.J. and Veenstra,G.J. (2011) GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics*, **27**, 270–271.
25. Chawengsaksophak,K., de Graaff,W., Rossant,J., Deschamps,J. and Beck,F. (2004) Cdx2 is essential for axial elongation in mouse development. *Proc. Natl Acad. Sci. USA*, **101**, 7641–7645.
26. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
27. Verzi,M.P., Shin,H., He,H.H., Sulahian,R., Meyer,C.A., Montgomery,R.K., Fleet,J.C., Brown,M., Liu,X.S. and Shivdasani,R.A. (2010) Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev. Cell*, **19**, 713–726.
28. Verzi,M.P., Hatzis,P., Sulahian,R., Philips,J., Schuijers,J., Shin,H., Freed,E., Lynch,J.P., Dang,D.T., Brown,M. *et al.* (2010) TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc. Natl Acad. Sci. USA*, **107**, 15157–15162.
29. Newburger,D.E. and Bulyk,M.L. (2009) UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.*, **37**, D77–D82.
30. Widom,J. (1998) Structure, dynamics, and function of chromatin in vitro. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 285–327.
31. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M.,

Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA*, **107**, 21931–21936.
32. Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
33. Slattery,M., Riley,T., Liu,P., Abe,N., Gomez-Alcala,P., Dror,I., Zhou,T., Rohs,R., Honig,B., Bussemaker,H.J. *et al.* (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*, **147**, 1270–1282.
34. Stormo,G.D. and Zhao,Y. (2010) Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.*, **11**, 751–760.
35. Melnikov,A., Murugan,A., Zhang,X., Tesileanu,T., Wang,L., Rogov,P., Feizi,S., Gnirke,A., Callan,C.G. Jr, Kinney,J.B. *et al.* (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277.
36. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
37. Rada-Iglesias,A., Bajpai,R., Swigut,T., Brugmann,S.A., Flynn,R.A. and Wysocka,J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279–283.
38. Ruthenburg,A.J., Li,H., Patel,D.J. and Allis,C.D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.*, **8**, 983–994.
39. Ruthenburg,A.J., Allis,C.D. and Wysocka,J. (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell.*, **25**, 15–30.
40. Tan,M., Luo,H., Lee,S., Jin,F., Yang,J.S., Montellier,E., Buchou,T., Cheng,Z., Rousseaux,S., Rajagopal,N. *et al.* (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.
41. Bryder,D., Rossi,D.J. and Weissman,I.L. (2006) Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am. J. Pathol.*, **169**, 338–346.
42. Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
43. Haugland,R.P., Yguerabide,J. and Stryer,L. (1969) Dependence of the kinetics of singlet-singlet energy transfer on spectral overlap. *Proc. Natl Acad. Sci. USA*, **63**, 23–30.
44. Fullwood,M.J., Liu,M.H., Pan,Y.F., Liu,J., Xu,H., Mohamed,Y.B., Orlov,Y.L., Velkov,S., Ho,A., Mei,P.H. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58–64.
45. Yaneva,M. and Tempst,P. (2006) Isolation and mass spectrometry of specific DNA binding proteins. *Methods Mol. Biol.*, **338**, 291–303.