

# Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA

Neil P. Blackledge<sup>1</sup>, Hannah K. Long<sup>1,2</sup>, Jin C. Zhou<sup>1</sup>, Skirmantas Kriaucionis<sup>3</sup>, Roger Patient<sup>2</sup> and Robert J. Klose<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry, <sup>2</sup>Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital and <sup>3</sup>Ludwig Institute for Cancer Research Ltd, Oxford University, Oxford, UK

Received June 16, 2011; Revised November 16, 2011; Accepted November 18, 2011

## ABSTRACT

Across vertebrate genomes methylation of cytosine residues within the context of CpG dinucleotides is a pervasive epigenetic mark that can impact gene expression and has been implicated in various developmental and disease-associated processes. Several biochemical approaches exist to profile DNA methylation, but recently an alternative approach based on profiling non-methylated CpGs was developed. This technique, called CxxC affinity purification (CAP), uses a ZF-CxxC (CxxC) domain to specifically capture DNA containing clusters of non-methylated CpGs. Here we describe a new CAP approach, called biotinylated CAP (Bio-CAP), which eliminates the requirement for specialized equipment while dramatically improving and simplifying the CxxC-based DNA affinity purification. Importantly, this approach isolates non-methylated DNA in a manner that is directly proportional to the density of non-methylated CpGs, and discriminates non-methylated CpGs from both methylated and hydroxymethylated CpGs. Unlike conventional CAP, Bio-CAP can be applied to nanogram quantities of genomic DNA and in a magnetic format is amenable to efficient parallel processing of samples. Furthermore, Bio-CAP can be applied to genome-wide profiling of non-methylated DNA with relatively small amounts of input material. Therefore, Bio-CAP is a simple and streamlined approach for characterizing regions of the non-methylated DNA, whether at specific target regions or genome wide.

## INTRODUCTION

Cytosine methylation within the context of CpG dinucleotides is the most prevalent modification to vertebrate DNA and represents the best understood epigenetic modification [reviewed in ref. (1)]. Methylated CpGs dominate the vertebrate genomic landscape, occurring within both intragenic and intergenic regions (2). Despite pervasive CpG methylation, vertebrate genomes are punctuated by DNA elements called CpG islands (CGIs) that have a high concentration of CpGs that exist in a predominantly non-methylated state (3,4). Importantly, CGIs are associated with ~70% of annotated genes suggesting that they play an important functional role at gene regulatory elements (5,6).

The methylation status of CpG dinucleotides, especially within the context of CGI promoters, has implications for the gene expression. The most striking examples of this involve acquisition of DNA methylation at CGI promoters, which is usually coupled to silencing of the associated gene. This phenomenon has been documented during cellular differentiation and certain disease states, most notably cancer. For example, a broad number of cancers are associated with the acquisition of DNA methylation at CGI promoters of tumour suppressor genes (7–9). DNA methylation is also implicated in X chromosome inactivation, during which hundreds of CGIs on the X chromosome acquire DNA methylation, while genomic imprinting mechanisms often involve differential CGI methylation between maternal and paternal alleles (10,11).

The links between DNA methylation and gene expression have provoked tremendous interest in studying CpG methylation states and consequently a broad range of techniques have been developed for this purpose [reviewed in ref. (12)]. These include bisulfite conversion

\*To whom correspondence should be addressed. Tel: +01865613222; Email: rob.klose@bioch.ox.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

of cytosine (but not methylated cytosine) into uracil, followed by interrogation via sequencing-based approaches (bisulfite sequencing). Bisulfite sequencing reveals cytosine methylation states at base pair resolution and is unparalleled in this regard. Although bisulfite conversion has been coupled to massively parallel sequencing to yield genome-wide methylation profiles at a base pair resolution (2,13), this is an extremely costly endeavour due to the depth of sequencing required.

Techniques based on affinity capture of methylated DNA, notably methylated DNA immunoprecipitation (MeDIP) and methylated DNA capture by affinity purification (MethylCap), have proved of utility in DNA methylation studies [reviewed in refs (12,14)]. These methods have been combined with massively parallel sequencing to yield genome-wide methylation profiles, albeit at lower resolution than those obtained with bisulfite sequencing approaches (15). Given the abundance of DNA methylation, even methyl-CpG affinity approaches can require an economically prohibitive sequencing depth to obtain accurate locus-specific DNA methylation profiles (16). Using a zinc finger CxxC (CxxC) domain that specifically recognizes non-methylated CpGs (17), Bird and colleagues (18) recently developed a clever approach based on CxxC affinity purification (CAP) to alternatively profile non-methylated DNA. Given that only 1–2% of a typical vertebrate genome is non-methylated, the CAP assay specifically recovers a much smaller genomic fraction than methylated DNA affinity approaches yet retains the capacity to differentiate between methylated and non-methylated CpG dinucleotides. Using CAP followed by massively parallel sequencing, non-methylated regions of the genome can be profiled with comparatively low sequencing depth (18,19).

The conventional CAP assay works by manually fragmenting the genome into regions containing intact non-methylated DNA followed by application of the DNA material to an automated chromatography system encompassing a CxxC affinity resin. Non-methylated DNA binds to the resin and is eluted using either a linear salt gradient or step gradient. During the chromatographic run, fractions are automatically collected, analysed and combined to isolate purified non-methylated DNA. The existing CAP technique has several limitations that make it inaccessible to some research groups and impractical for certain experimental scenarios. For example, the incorporation of an automated chromatography step requires access to a high-resolution chromatographic system. Since this approach uses a 1 ml chromatography column packed with the affinity resin, it requires large amounts of the recombinant CxxC module (60 mg) (18,19). Processing samples via the 1 ml column configuration necessitates large elution volumes that subsequently require DNA precipitation prior to PCR or sequencing analysis. These large elution volumes and DNA handling steps (during which there is the potential for loss of DNA) in turn mean that the CAP assay requires large amounts of input DNA, restricting the utility of this approach in instances where the amount of DNA material available is limiting (for example rare cell types or valuable patient samples). From a pragmatic standpoint,

the DNA preparation, chromatography and processing time mean that CAP is labour-intensive, time-consuming and not amenable to parallel processing of multiple DNA samples.

To overcome many of the limitations inherent to conventional CAP, we have engineered a completely new CAP approach, called biotinylated CAP (Bio-CAP), which is fast, simple, requires no specialized equipment and efficiently isolates the non-methylated regions of genomic DNA. Furthermore, we demonstrate that Bio-CAP can be applied to very small quantities of genomic DNA, which is adaptable to parallel processing, and provides material suitable for massively parallel sequencing-based genome-wide analysis of non-methylated DNA.

## MATERIALS AND METHODS

### DNA constructs

A ZF-CxxC construct encoding amino acids 600–750 of human KDM2B was PCR amplified from a human KDM2B cDNA and engineered to encode a C-terminal avi-tag. The PCR product was inserted via ligation-independent cloning into a pNIC28 prokaryotic expression vector that has been modified to express an N-terminal 6-his tag followed by a tobacco etch virus (TEV) protease cleavage site as described before (20). The sequence integrity of the resulting construct was verified by sequencing.

### Protein expression and purification

The CxxC-avi protein was expressed by freshly transforming a BL21 expression strain carrying the pRAR2 plasmid. The 4 L cultures were grown in  $2 \times$  TB supplemented with 0.25 mM  $ZnCl_2$  and grown to an  $OD_{600}$  of 0.6 at 37°C. The culture was then cooled to 30°C and expression was induced with 1 mM IPTG for 3 h. After 3 h, the cultures were pelleted and the CxxC-avi protein was isolated in batch by Ni-NTA-mediated purification as described previously (Klose and Bird, 2004). The peak elution fractions from the Ni-NTA purifications were pooled and digested overnight at 4°C with His-tagged TEV protease. The following day the protein was desalted to remove the imidazole and reappplied to a Ni-NTA column to remove the cleaved tag and TEV protease. The cleaved CxxC-avi protein was then desalted into 10 mM Tris pH 8.0 containing 250 mM potassium glutamate in preparation for *in vitro* biotinylation. The yield of pure CxxC-avi protein after this step was 16 mg. *In vitro* biotinylation was carried out by adding His-tagged recombinant BirA to the protein and supplementing the reaction with 10 mM ATP, 10 mM  $Mg(OAc)_2$  and 50  $\mu$ M D-biotin. The reaction was allowed to proceed overnight. The efficiency of biotinylation was verified by mass spectrometry prior to and after *in vitro* biotinylation. The following day the reaction was supplemented with 20 mM imidazole and applied to Ni-NTA resin to remove the His-tagged BirA ligase. The CxxC-avi protein was then desalted into 20 mM HEPES pH 7.9, 150 mM KCl, 0.5 mM dithiothreitol (DTT) and 10% glycerol and stored aliquoted at  $-80^\circ C$ . The final yield of pure biotinylated CxxC-avi protein from 4 L was 6 mg.

The protein remains stable and functional when stored at  $-80^{\circ}\text{C}$  for  $>1$  year.

### Electrophoretic mobility shift assay (EMSA)

EMSA probes were generated and labelled as previously described (Blackledge *et al.*, 2010). EMSA was also performed as previously described with samples analysed on a 1.3% agarose gel.

### Cell culture

Murine V6.5 embryonic stem (ES) cells [C57BL/6 (F) x 129/sv (M)] were cultured on inactivated mouse embryonic fibroblasts (MEFs) in DMEM supplemented with 15% fetal calf serum (FCS), leukaemia-inhibiting factor, penicillin/streptomycin, L-glutamine and non-essential amino acids. Prior to genomic DNA extraction for Bio-CAP experiments, V6.5 cells were cultured for two passages under feeder-free conditions on 0.1% gelatin.

### Bio-CAP

For each individual Bio-CAP experiment, 25  $\mu\text{l}$  of NeutrAvidin Agarose Resin (Thermo Scientific, 29200) or NeutrAvidin-coated magnetic beads (Thermo Scientific, 7815-2104-011150) was washed with BC100 buffer and then incubated with 50  $\mu\text{l}$  of 0.5  $\mu\text{g}/\mu\text{l}$  biotinylated hKDM2b-CxxC protein diluted in BC100 buffer (or BC100 buffer alone for 'beads only' control) for 1 h at  $4^{\circ}\text{C}$ . The conjugated resin/CxxC protein was then washed with CAP100 buffer (12.5% glycerol, 0.1% Triton-x-100, 20 mM HEPES pH 7.9 and 100 mM NaCl).

Genomic DNA at a concentration of 0.35 mg/ml was sonicated to an average size of 150–250 bp using a Diagenode Bioruptor. Sonicated DNA was then diluted in CAP100 buffer to a concentration of 16  $\mu\text{g}/\text{ml}$  and a 100  $\mu\text{l}$  input sample was retained. For each Bio-CAP experiment, 500  $\mu\text{l}$  of diluted sonicated DNA, corresponding to  $\sim 8$   $\mu\text{g}$  of DNA (unless otherwise stated), was added to the conjugated CxxC resin. The DNA and resin were incubated at  $4^{\circ}\text{C}$  for 1 h with gentle mixing. The resin was then collected by centrifugation at 2000 rpm for 3 min at  $4^{\circ}\text{C}$  or by magnetization, and the unbound flowthrough (FT) material was removed. The resin and any associated DNA was washed twice with 1 ml of CAP100 buffer, before the first elution was performed by adding 50  $\mu\text{l}$  of CAP300 (12.5% glycerol, 0.1% Triton-x-100, 20 mM HEPES pH 7.9 and 300 mM NaCl) to the resin and incubating at room temperature for 10 min. Following centrifugation or magnetisation, a 50  $\mu\text{l}$  elution fraction was carefully collected. The elution process was repeated using another 50  $\mu\text{l}$  of CAP300 and the 300 mM elution fractions were pooled (giving a total volume of 100  $\mu\text{l}$ ). Subsequent elutions were performed in the same way using buffers with 500, 700 mM and 1 M NaCl sequentially. Each 100  $\mu\text{l}$  elution fraction, together with 100  $\mu\text{l}$  of both the input and FT samples, was purified using a PCR purification column (Qiagen) and DNA was eluted in a volume of 50  $\mu\text{l}$ .

For real-time quantitative PCR (qPCR) analysis, Bio-CAP samples were typically diluted 10-fold and 5  $\mu\text{l}$  of this was used per 15  $\mu\text{l}$  reaction. Analysis was

performed using Sybr Green (Quantace) on a Rotor-Gene 6000 (Corbett). Primer sets used for qPCR are available on request.

For the Bio-CAP spiking experiment, 200 bp regions containing set numbers of CpGs were amplified from human genomic DNA. The methylated human probe was generated by *in vitro* methylation of the 13 CpG probe with M.SssI (NEB) and the hydroxymethylated human probe was generated by amplifying the 13 CpG probe in a PCR reaction containing 5hmC. Approximately 10 pg of each 200 bp probe was added to 8  $\mu\text{g}$  of sonicated mouse genomic DNA and the Bio-CAP assay was then performed in the same way as above. The 200 bp human probes were interrogated by qPCR using primer sets nested within each region. Primer sets are available on request.

### MeDIP

Genomic DNA was sonicated using a Diagenode Bioruptor to an average size of 250 bp. Prior to immunoprecipitation, DNA blunting, dA overhang addition and adaptor ligation were completed according to Illumina library preparation recommendations. Immunoprecipitation was performed using 2  $\mu\text{g}$  of anti-mC antibody (Eurogentec) using a general MeDIP protocol (<http://www.epigenome-noe.net/WWW/researchtools/protocol.php?protid=33>) with some minor modifications and was sequenced as described below.

### Bisulfite sequencing

Bisulfite conversion of DNA was performed using the EZ DNA Methylation-Gold Kit (Zymo Research). PCR-amplified DNA was cloned into pGEM-T Easy (Promega) and sequenced. Sequenced clones were analysed using the web-based tool QUMA (<http://quma.cdb.riken.jp/>) (21). Primer sets used for bisulfite sequencing are available on request.

### High-throughput sequencing

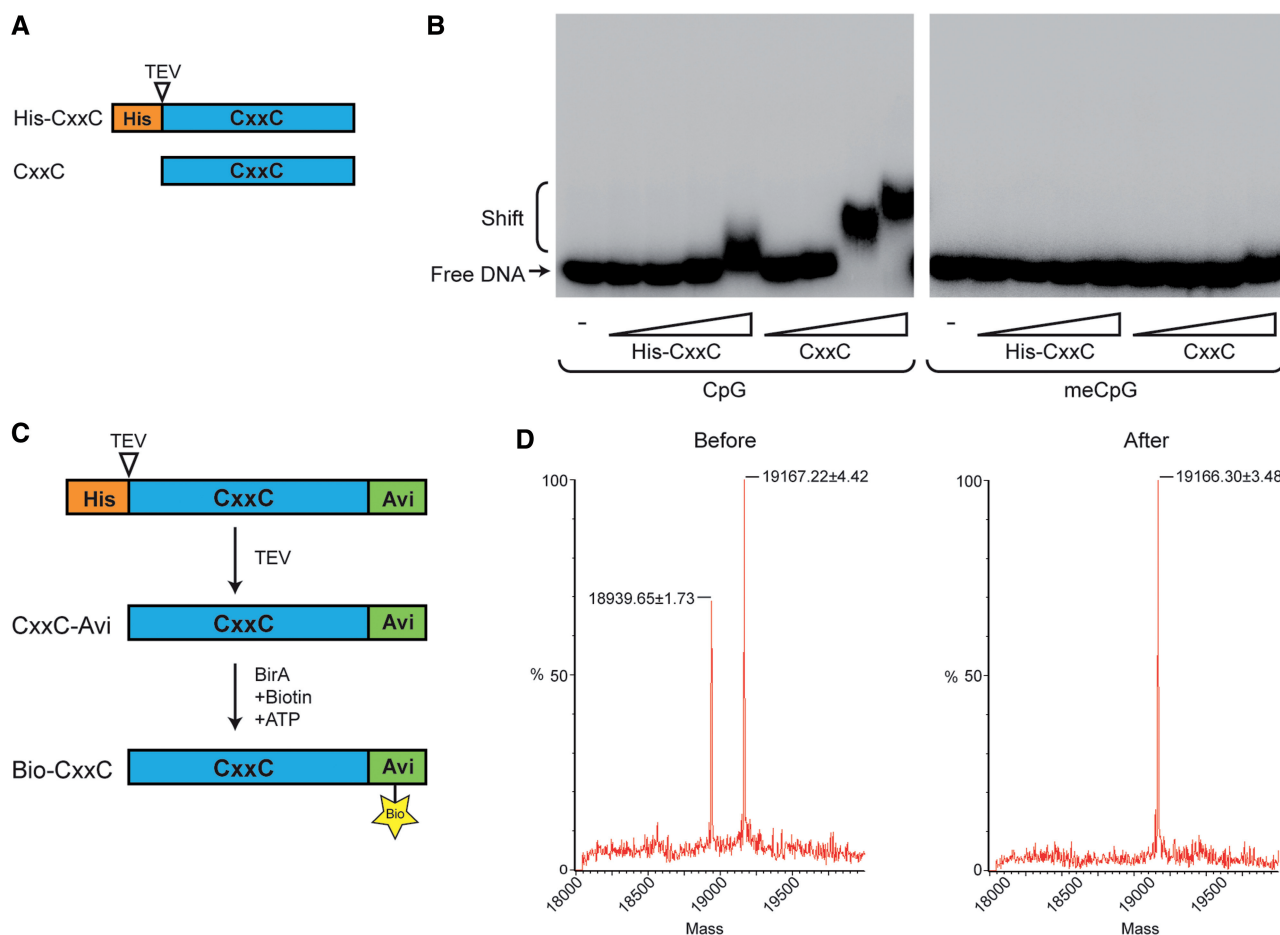
Bio-CAP DNA was prepared for Solexa 2G sequencing by blunting the DNA with a mixture of T4 DNA polymerase, Klenow DNA polymerase and T4 PNK (NEB) according to manufacturer's instruction. dA overhangs were then added and Illumina adapters ligated. Adapter-ligated DNA was subject to 18 cycles of PCR before size selection by agarose gel electrophoresis. Amplified DNA was purified using the Qiaquick gel extraction kit (Qiagen). The purified DNA was quantified both with an Agilent Bioanalyzer and Invitrogen Qubit and diluted to a working concentration of 10 nM prior to sequencing. Sequencing on a Solexa 2G instrument was carried out according to the manufacturer's instructions. Raw sequencing reads were aligned to the mouse mm9 genome using bowtie (22) retaining only reads that align to one position in the genome. BAM files corresponding to the aligned data were imported into Seqminer (23) to generate the heat map and scatter plot at CGI regions.

## RESULTS

### A biotinylated CxxC domain recognizes DNA containing non-methylated CpGs

Based on the limitations inherent to the conventional CAP technique, we set out to devise a simplified and more functional CAP-based approach to profile non-methylated DNA. First, we focussed our efforts on identifying a high-affinity recombinant CxxC domain fragment and a suitable solid support resin to affix the affinity module onto. We previously demonstrated that the histone H3 lysine 36 demethylase KDM2A binds specifically to non-methylated CpGs via its CxxC domain and *in vivo* this recruits the enzyme to non-methylated CGIs (20). KDM2B, a closely related H3K36 demethylase, also possesses a CxxC domain. Multiple sequence alignment shows conservation of the Zn-coordinating cysteines and DNA-binding residues in KDM2B suggesting that this protein will also bind non-methylated DNA. To test this possibility, recombinant His-tagged KDM2B CxxC domain (His-CxxC) was expressed in *Escherichia coli*

and purified on a nickel-NTA (Ni-NTA) column (Figure 1A). By EMSA, the His-CxxC protein bound to a DNA probe that contained non-methylated CpGs (Figure 1B, left panel), but binding was abrogated when the probe was *in vitro* methylated (Figure 1B, right panel). Interestingly, we noticed that binding of the recombinant KDM2B CxxC domain to non-methylated DNA appeared more efficient than a similar construct encompassing the KDM2A CxxC domain (unpublished observations) making this a good candidate fragment around which to develop a new CAP module. The recombinant KDM2B CxxC fragment was purified from *E. coli* using an N-terminal His tag. To examine whether the addition of this tag affects the functionality of the CxxC domain, we cleaved off the His tag taking advantage of a TEV protease cleavage site between the His tag and the CxxC domain (Figure 1A). Interestingly, cleavage of the tag resulted in increased binding to the non-methylated probe (Figure 1B, left panel) without compromising its specificity for non-methylated DNA. This suggests that in the context of the recombinant KDM2B CxxC



**Figure 1.** Generation and characterization of a Bio-CxxC protein. (A) Schematic illustrating recombinant CxxC protein with N-terminal His tag (His-CxxC). Cleavage with TEV protease removes His tag to give CxxC alone. (B) EMSA demonstrating that His-CxxC and CxxC bind to a DNA probe containing non-methylated CpGs in a concentration-dependent manner (left panel), and that DNA methylation blocks this binding (right panel). (C) Schematic illustrating the generation of a CxxC protein with a C-terminal Avi-Tag (CxxC-Avi) and the subsequent reaction catalysed by the *E. coli* biotin ligase BirA to give a Bio-CxxC protein. (D) Mass spectrometry analysis of CxxC-Avi protein before (left panel) and after *in vitro* BirA reaction (right panel). Expected mass of CxxC-Avi is 18939 Da and Bio-CxxC is 19166 Da.

fragment an N-terminal His tag interferes with DNA binding affinity. Based on these observations, the KDM2B CxxC domain lacking the N-terminal His tag appears to be an ideal affinity module for DNA-containing non-methylated CpG dinucleotides.

CAP requires that the CxxC module be affixed to a solid support to facilitate affinity capture of non-methylated DNA. Conventional CAP uses a His tag to bind the CxxC module to a Ni-NTA-based resin. This linkage is not ideal as Ni-NTA resin has intrinsic charge properties, is susceptible to leaching of bound protein in the presence of metal chelating or reducing agents, and is sensitive to pH and high salt. To overcome some of the disadvantages of a His tag linkage, we instead engineered a 15 amino acid AviTag<sup>TM</sup> (GLNDIFEAQKIEWHE) onto the C-terminus of the KDM2B CxxC fragment (to give CxxC-Avi, Figure 1C) (24,25). The CxxC-Avi protein was then *in vitro* biotinylated using recombinant biotin ligase BirA, an enzyme that specifically conjugates biotin to a lysine residue in the AviTag<sup>TM</sup> sequence (Figure 1C). The introduction of a site-specific biotin moiety into the affinity module permitted us to affix the CxxC domain to an avidin-based solid support. The avidin/biotin linkage is a significant improvement over the His tag linkage used previously as it is one of the strongest documented non-covalent interactions between a protein and its ligand, remaining intact even during stringent washing and manipulation conditions (26). Mass spectrometry was used to monitor biotinylation of the CxxC-Avi protein (Figure 1D). Prior to the *in vitro* BirA reaction, a proportion of the CxxC-Avi protein was already biotinylated exhibiting a mass of ~19167 Da (the expected mass for the biotinylated form of the protein) as opposed to 18393 Da (the expected mass of the unmodified protein) (Figure 1D, left panel). This is likely due to biotinylation of the protein by the endogenous *E. coli* BirA ligase system. Importantly, *in vitro* biotinylation by BirA yielded a homogeneously biotinylated species (Figure 1D, right panel). The biotinylated CxxC (Bio-CxxC) protein bound to a DNA probe containing non-methylated CpGs with similar affinity to an Avi-CxxC protein without the biotin moiety (data not shown). Hence neither the presence of the AviTag<sup>TM</sup> nor the biotin moiety interfered with the DNA-binding activity of the KDM2B CxxC domain.

### Bio-CAP specifically isolates CGI DNA

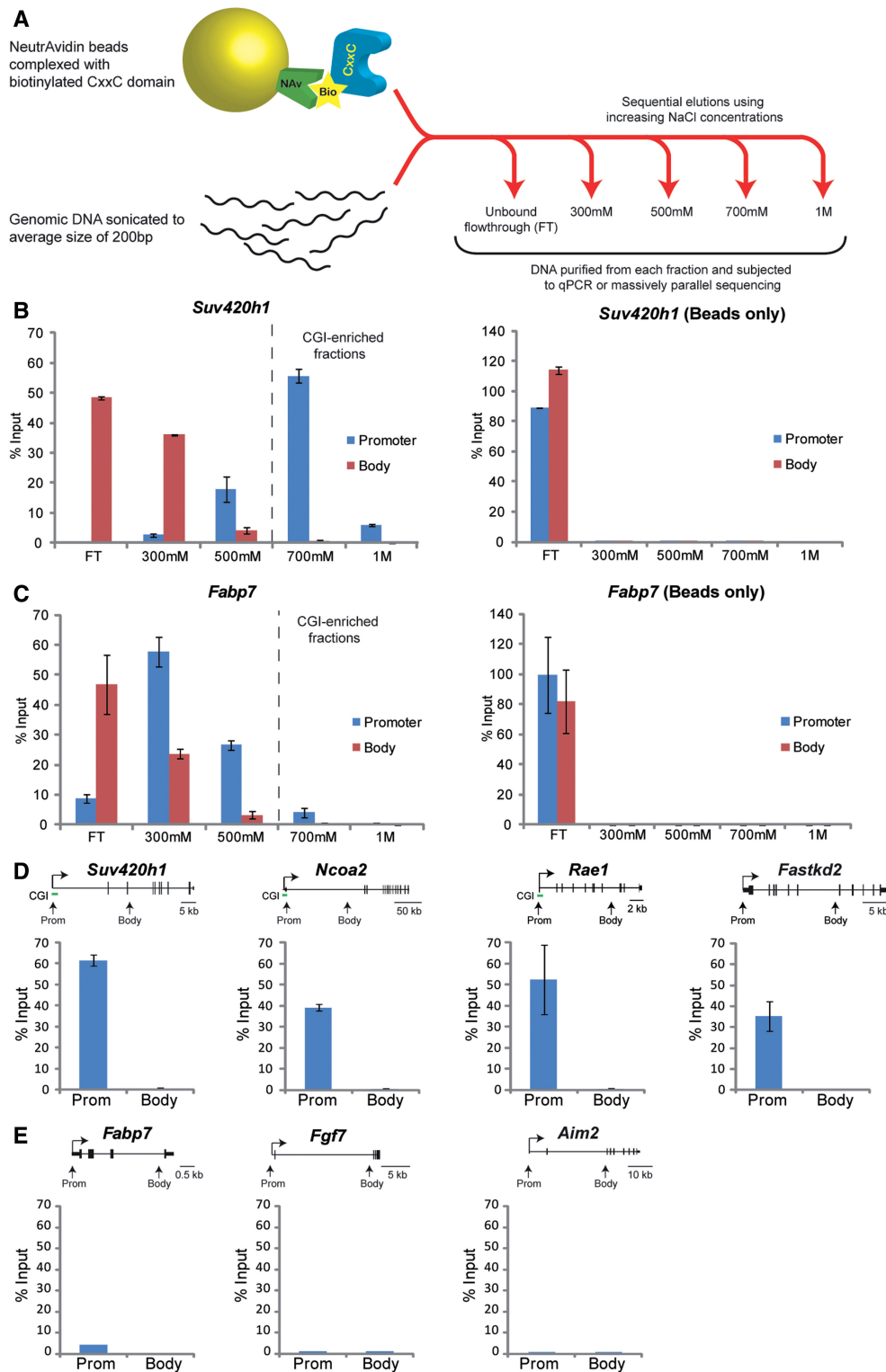
There are several avidin/streptavidin-based resins with different features that could function as a solid support to affix the Bio-CxxC protein. NeutrAvidin, a deglycosylated derivative of avidin, was chosen based on the fact that it features a near neutral isoelectric point and therefore exhibits exceptionally low-non-specific binding properties while retaining an extremely high affinity for biotin ( $K_d \sim 10^{-15} \text{M}^{-1}$ ) (27).

To determine whether the bead-immobilized CxxC protein could specifically isolate non-methylated DNA, we used genomic DNA from V6.5 mouse ES cells, a cell type used previously to study the CGI binding factor KDM2A (20). Mouse ES cell genomic DNA was firstly

sonicated to an average length of 200 bp. Approximately 8  $\mu\text{g}$  of sonicated DNA was added to 25  $\mu\text{g}$  of the CxxC binding module that had been immobilized on 25  $\mu\text{l}$  of NeutrAvidin beads. The DNA and CxxC resin was then incubated at 4°C for 1 h with gentle mixing to permit CxxC–DNA complexes to form (Figure 2A). After sedimenting the CxxC resin and associated DNA by centrifugation, the unbound FT DNA was collected. Material that bound to the CxxC domain was then subjected to a series of sequential elution steps using increasing salt concentrations and fractions were collected at each salt concentration (300, 500, 700 mM and 1 M NaCl) (Figure 2A).

DNA from the input, FT and each of the elution fractions was subjected to quantitative PCR (qPCR) using primer sets specific to the *Suv420h1* promoter CGI and gene body. This analysis revealed that DNA corresponding to the *Suv420h1* CGI was mostly present within the high-salt fractions (700 mM and 1 M), with little or none of the *Suv420h1* CGI present in the FT and low-salt (300 mM) fractions (Figure 2B). In contrast, DNA corresponding to the *Suv420h1* gene body was mostly present within the FT and low-salt fractions, with almost no DNA from this region present in the high-salt fractions (Figure 2B, left panel). Similar qPCR analysis was also performed using promoter and body primer sets from *Fabp7*, a gene that lacks a CGI at its promoter. While the *Fabp7* body region has an elution profile almost identical to that of the *Suv420h1* body, unlike *Suv420h1*, the *Fabp7* promoter region eluted mostly in the low-salt fractions indicating a lack of non-methylated CpG dinucleotides (Figure 2C, left panel). To verify that these elution profiles were dependent specifically upon the CxxC domain, a control experiment was performed using NeutrAvidin beads alone. In the beads only experiment, DNA corresponding to both the promoter and body regions of *Suv420h1* and *Fabp7* was exclusively present within the FT fraction demonstrating that the NeutrAvidin solid support exhibits no non-specific DNA binding (Figure 2B and C, right panels).

From the above analyses, it was evident that the high-salt fractions exhibited massive enrichment of the *Suv420h1* CGI compared to non-CGI regions. In order to understand whether this holds true for other regions of the genome, we designed promoter and body primer sets for a panel of genes with CGI promoters (Figure 2D), and genes with non-CGI promoters (Figure 2E). Each of these primer sets was used to assay the high-salt fractions from our CxxC-based purification by qPCR. These experiments demonstrate that in all cases the high-salt elution fractions (a sum of 700 mM and 1 M fractions) are massively enriched for CGIs (Figure 2D and E), with non-CGI regions exhibiting negligible enrichment. *In silico* algorithms have previously been used to predict the presence of CGIs, but these often overlook some regions that actually exhibit significant levels of non-methylated CpGs. One such example is the promoter of the *Fastkd2* gene, which was shown by a CAP-sequencing study to be non-methylated in mouse sperm yet is not bioinformatically defined as a CGI (19). Consistent with this previous observation, the *Fastkd2* promoter is strongly enriched in the high-salt fractions



**Figure 2.** Bio-CAP specifically purifies CGI DNA. (A) Schematic of the Bio-CAP procedure. The immobilized Bio-CxxC protein is first incubated with sheared genomic DNA, allowing DNA to bind to the CxxC domain. The unbound FT is then collected, followed by a series of elution fractions performed at increasing salt concentrations. DNA from each of these fractions is then interrogated by qPCR or massively parallel sequencing. (B) Analysis of *Suv420h1*, a gene with a CGI promoter, in a Bio-CAP experiment (left panel) or a ‘beads only’ control experiment (right panel). For both experiments, DNA from each fraction was subjected to qPCR using primer sets specific to the CGI promoter (blue) or body region (red) of *Suv420h1*. (C) As for (B) except that primer sets were specific to the promoter and body of *Fabp7*, a gene that has a non-CGI promoter. (D) Bio-CAP analyses for a panel of genes with CGI promoters. Promoter and body primer sets were used to assay the CGI-enriched fractions (700 mM + 1M) from a Bio-CAP assay. A schematic of each gene showing the position of primer sets and CGI from *in silico* predictions is shown above each data set. (E) As for (D) except primer sets were specific to a panel of genes with non-CGI promoters. All Bio-CAP data are from two biological replicates and error bars represent SEM.

of our assay, indicating that this region is also non-methylated in mouse ES cells (Figure 2D, far right panel). Collectively, these analyses demonstrate that within the context of our CxxC-based affinity purification, regions of the genome corresponding to CGIs are specifically retained with high affinity, eluting only under high-salt conditions. In contrast, non-CGI regions are not retained by the immobilized CxxC domain and are therefore present in the FT and low-salt elutions. Therefore, using our new affinity matrix and scaled-down CAP configuration we demonstrate simple, robust and specific isolation of non-methylated DNA. Based on the utilization of a biotinylated high-affinity CxxC domain, we call this new purification strategy Bio-CAP.

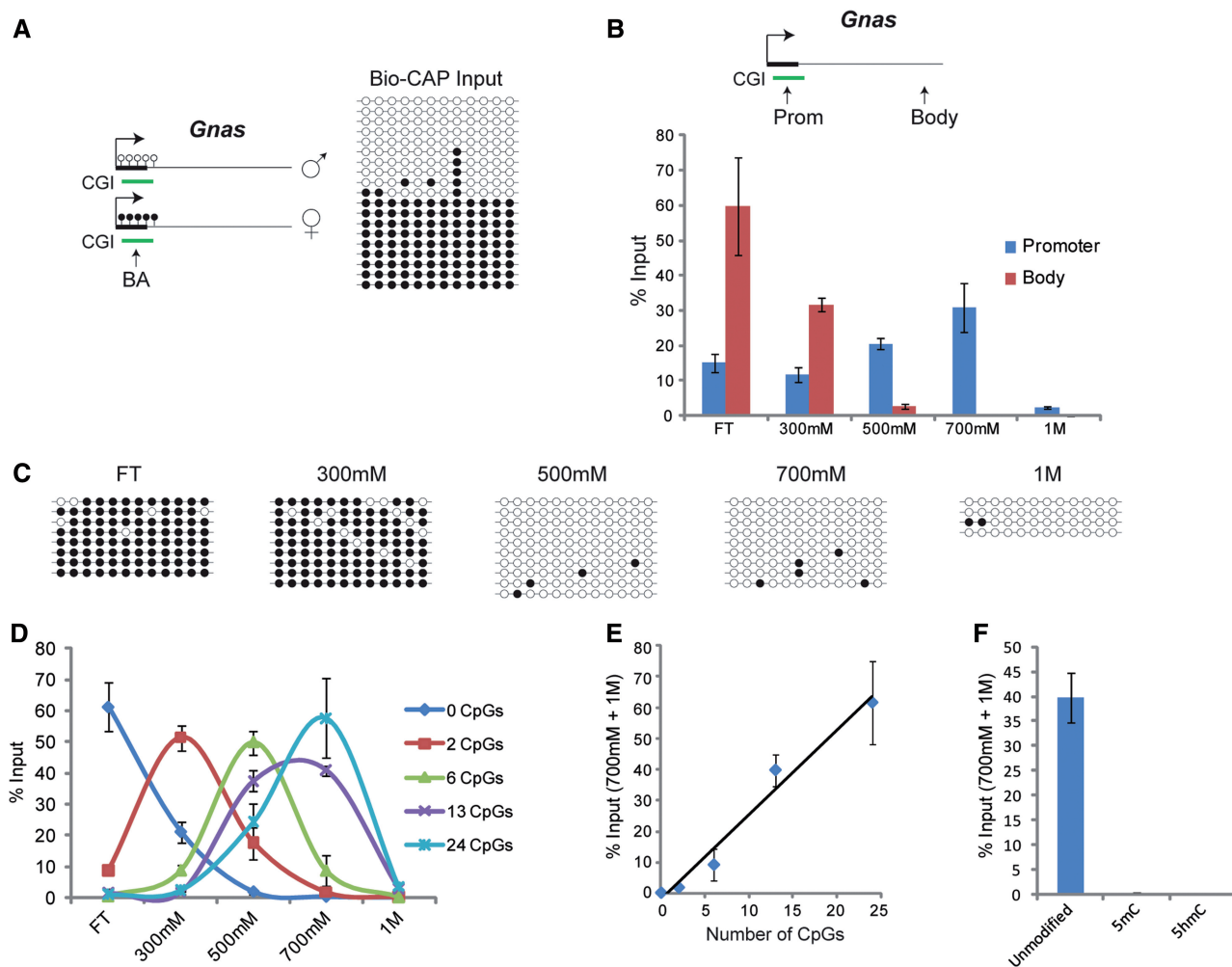
### Bio-CAP elution profiles are dictated by methylation status and density of CpGs

The above experiments demonstrate that the high-salt elutions of the Bio-CAP assay specifically represent the CGI fraction of the genome. Although the majority of CGIs remain free of DNA methylation, at certain developmental time points or in particular disease states-specific CGIs can acquire DNA methylation. To investigate the impact of DNA methylation on the CGI elution profiles within Bio-CAP, we took advantage of the *Gnas* CGI, which is subject to an allele-specific epigenetic imprinting process that results in dense methylation of the CGI on the maternal allele while the paternal allele remains non-methylated (28). Using bisulfite sequencing analysis, this imprinting was confirmed for the mouse ES cells used in our Bio-CAP experiments, with the expected 50:50 ratio of non-methylated to methylated alleles (Figure 3A). Using qPCR analysis, we then went on to quantify the presence of the *Gnas* CGI in Bio-CAP fractions. This revealed a relatively flat elution profile, with comparable amounts of the *Gnas* CGI in the FT, 300, 500 and 700mM fractions (Figure 3B). This is in contrast to typical CGI elution profiles that demonstrate negligible presence in the FT and low-salt (300mM) fractions, with massive enrichment in the high-salt fractions (especially 700mM) (compare Figure 2B with Figure 3B). To determine the methylation status of the *Gnas* CGI in each of the Bio-CAP fractions, material from each fraction was subjected to bisulfite sequencing. This revealed that the FT and 300mM Bio-CAP fractions contained only the methylated *Gnas* allele, while the 500, 700mM and 1M fractions contained only the non-methylated *Gnas* allele (Figure 3C). This striking result demonstrates that within the context of a Bio-CAP assay, the methylation status of a CGI dictates its binding affinity for the CxxC domain. Bio-CAP is therefore able to separate methylated CGIs from non-methylated CGIs.

Various computational algorithms based on CpG observed/expected ratio and GC percentage have been used to predict CGI genomic locations (5). However, it has since been experimentally shown that some regions of non-methylated DNA fall below the CpG and/or GC threshold required for *in silico* definition as a CGI (19).

To determine how CpG density influences Bio-CAP read-out, primers were designed to specifically amplify 200 bp regions of the human genome containing specific numbers of CpG dinucleotides (0, 2, 6, 13 and 24 CpGs). Mouse ES cell genomic DNA was spiked using each of these 200 bp human probes, such that each probe was at approximately the same molar concentration as the mouse genomic DNA, and this spiked DNA mixture was then used in a Bio-CAP experiment. The spiked Bio-CAP fractions were interrogated using primer sets nested within each individual 200 bp probe, therefore revealing the elution profile for each human fragment (Figure 3D). Unsurprisingly, the 200 bp probe containing 0 CpGs was exclusively present in the FT and 300mM fractions, indicating no specific binding of this probe to the CxxC domain. However, with increasing numbers of CpGs, the 200 bp probes exhibited elution profiles that shifted towards the higher salt elutions, indicating specific binding by the CxxC domain. Importantly, there was a near-linear relationship between the number of non-methylated CpGs and the percentage recovery in the high salt (700mM and 1M) fractions (Figure 3E). This would imply that by interrogating high salt Bio-CAP fractions, one gets a semiquantitative read-out of the absolute number of non-methylated CpGs present in a given DNA fragment. Importantly, this experiment also demonstrates that Bio-CAP is able to reveal the presence of non-methylated DNA, even when these regions fall below the threshold for traditional computational CGI definition. This is illustrated well by the 200 bp probe with six CpGs, which does not have the CpG density necessary for computational definition as a CGI, yet shows moderate recovery in the high salt Bio-CAP elutions (Figure 3D and E). To assay the impact of DNA methylation on recovery of 200 bp human probes, the 13 CpG probe was *in vitro* methylated. This resulted in complete abrogation of recovery in the high-salt fractions (Figure 3F), further demonstrating that methylation of CpGs inhibits CxxC domain binding. It has been recently shown that a small proportion of CpGs within the genome are hydroxymethylated (29,30). Importantly, a hydroxymethylated version of the 13 CpG probe also demonstrated negligible recovery in the Bio-CAP high-salt fractions (Figure 3F), indicating that within the context of a Bio-CAP experiment, the CxxC domain does not recognize hydroxymethylated CpGs. Collectively, these experiments using the imprinted *Gnas* CGI and spiking with human DNA probes demonstrate that specific retention of DNA in Bio-CAP assays is dictated by the density of CpGs within a given genomic region and is negatively impacted by both CpG methylation and hydroxymethylation.

As with other affinity-based techniques (such as MeDIP or MethylCap), Bio-CAP is dependent on the number of CpGs available for binding. Hence, it is difficult to quantitatively compare the absolute level of non-methylated DNA between two regions with differential CpG density. Nevertheless, a specific locus can be directly compared between samples for changes in methylation status.



**Figure 3.** Bio-CAP is sensitive to the methylation status and density of CpG dinucleotides. (A) A schematic of the imprinted *Gnas* gene, which has a CGI promoter that is methylated on the maternal allele but non-methylated on the paternal allele (left panel). Using the indicated bisulfite PCR amplicon (BA), this imprinting was confirmed in the mouse ES cells used for Bio-CAP (right panel). Empty and filled circles represent non-methylated and methylated CpG dinucleotides, respectively. (B) qPCR analysis of *Gnas* promoter and body regions in Bio-CAP. A schematic of *Gnas* indicating the position of primer sets and a CGI from *in silico* prediction is shown above the data set. (C) Using the same amplicon (BA) shown in (A), bisulfite sequencing was performed on each of the fractions from a Bio-CAP experiment. (D) A Bio-CAP experiment with mouse ES cell DNA was spiked with a panel of human-specific 200 bp probes each containing a known number of CpGs. Using qPCR analysis, Bio-CAP fractions were assayed for the presence of each probe. (E) Scatter graph showing relationship between number of CpGs and % input recovery in high-salt fractions (700 mM + 1 M) of Bio-CAP spiking experiment shown in (D). A line of best fit is shown. (F) A Bio-CAP experiment with mouse ES cell DNA was spiked with a 200 bp human probe containing 13 CpGs that were either unmodified, methylated (5mC) or hydroxymethylated (5hmC). Using qPCR analysis, Bio-CAP high-salt fractions (700 mM + 1 M) were assayed for the presence of each of probe variant. All Bio-CAP data are from at least two biological replicates and error bars represent SEM.

### Bio-CAP can specifically isolate non-methylated DNA from small quantities of DNA and can be adapted for high-throughput approaches

The conventional CAP technique uses a chromatography column configuration with large elution volumes (3 ml), meaning that prior to analysis by qPCR or massively parallel sequencing, purified CAP DNA must be concentrated by precipitation (18). Due to the large column volumes and precipitation step at which significant loss of DNA can occur, this CAP technique requires sizeable quantities of input DNA per purification (~100  $\mu$ g) (18). A requirement for such large quantities of input material limits the utility of conventional CAP in

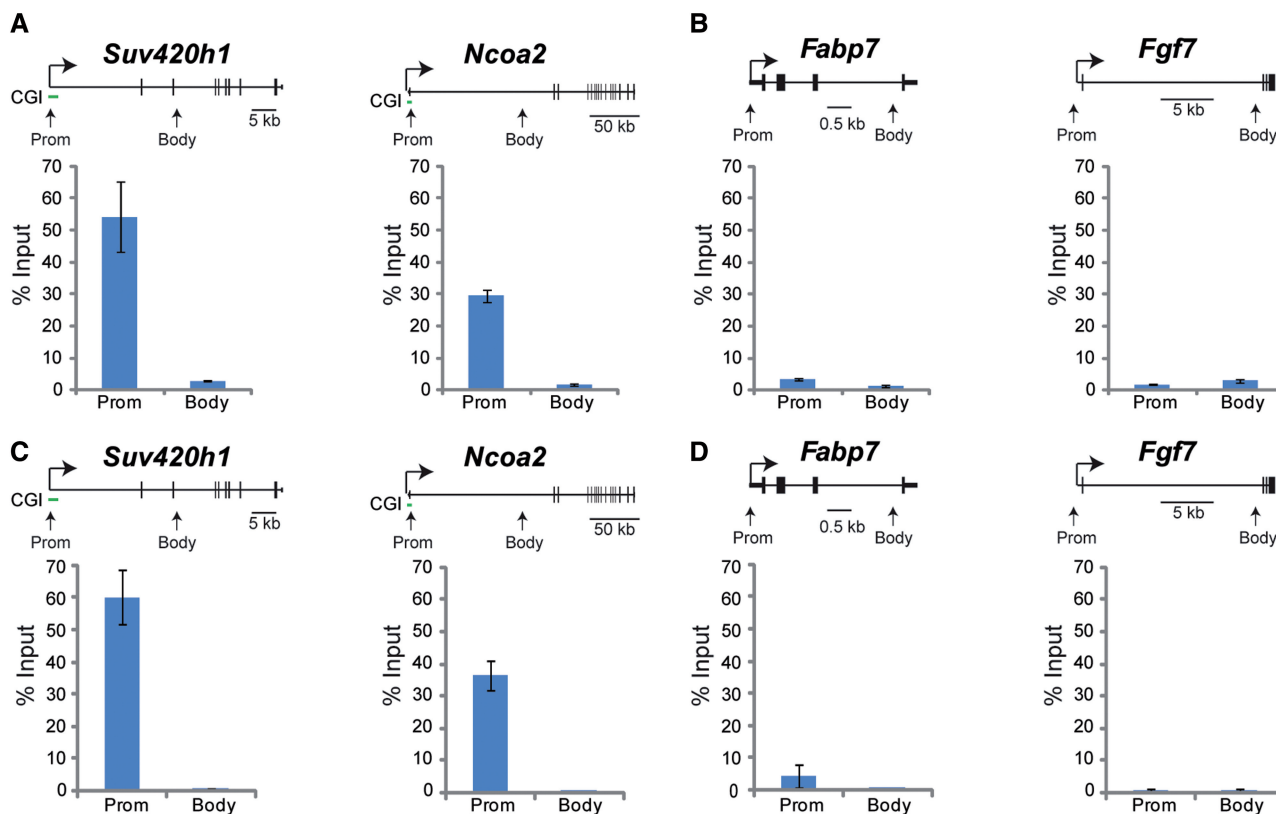
instances when DNA availability is limiting, such as very small or precious biological samples. For example, within a clinical setting, if one wanted to study the DNA methylation status of patient samples, obtaining large quantities of DNA for this purpose could be problematic. Unlike conventional CAP, the Bio-CAP assay uses small quantities of beads (25  $\mu$ l) that have very low-non-specific binding properties and small elution volumes (2  $\times$  50  $\mu$ l at each salt concentration). We therefore hypothesized that Bio-CAP may be amenable to much smaller amounts of input DNA than conventional CAP. To examine this possibility, Bio-CAP assays were performed using 100 ng of mouse ES cell DNA. The high-salt elutions were



interrogated by qPCR using promoter- and body-specific primer sets for *Suv420h1* and *Ncoa2*, two genes with non-methylated CGI promoters. These analyses revealed that the CGI promoter regions of both genes were massively enriched compared to the corresponding gene body regions (Figure 4A). Similar qPCR analysis was performed using promoter and body primer sets specific for *Fabp7* and *Fgf7*, genes that lack CGIs at their promoters. These analyses revealed little or no enrichment for either non-CGI promoter relative to the corresponding gene body (Figure 4B). Strikingly, the absolute percentage recovery values from the Bio-CAP assay with 100 ng input DNA are almost identical to the percentage recovery values for Bio-CAP using 8  $\mu$ g input DNA (compare Figure 4A and B with Figure 2D and E). The Bio-CAP assay is therefore extremely robust and yields consistent recovery, even when the amount of input DNA between samples differs by almost two orders of magnitude.

For reasons discussed above, including chromatography processing time and extensive DNA handling steps, the conventional CAP technique is somewhat laborious in nature. By comparison, Bio-CAP is simpler, scaled down and streamlined. These features mean that

Bio-CAP is amenable to parallel processing of multiple samples. Recently, systems such as the Diagenode SX-8G IP-Star Compact<sup>®</sup> that rely on magnetic separation of samples have been used for purification of methylated DNA (16). Therefore, to test whether Bio-CAP could be adapted for use in a magnetic configuration, we performed a modified Bio-CAP assay in which the Bio-CxxC domain was immobilized on NeutrAvidin-coated magnetic particles. Mouse genomic DNA sonicated to 150–250 bp was incubated with the magnetic CxxC resin for 1 h in the same way as a regular Bio-CAP experiment. The magnetic CxxC resin and associated DNA were then collected using a magnetic microcentrifuge tube rack, allowing unbound FT material to be removed. To further streamline the Bio-CAP procedure in this magnetic approach, two 10 min washes were performed in 500 mM NaCl to remove non-specifically bound material, followed by two 10 min elution steps in 1 M NaCl. Using this streamlined magnetic Bio-CAP approach, the entire purification procedure, including pull down, washes, elutions and DNA clean-up takes  $\sim$ 2 h. The 1 M elutions from the magnetic Bio-CAP assay were analysed by qPCR, using the same CGI and non-CGI primer sets as above.



**Figure 4.** Bio-CAP can be performed on low quantities of DNA and can be adapted for magnetic bead-based automation. (A) Bio-CAP was performed using 100 ng of mouse ES cell DNA and the high-salt fractions (700 mM + 1 M) were interrogated by qPCR using promoter and body primer sets specific to two genes with CGI promoters. A schematic of each gene showing the position of primer sets and CGI from *in silico* predictions is shown above each data set. (B) As for (A) except that primer sets were specific to two genes with non-CGI promoters. (C) A Bio-CAP experiment was performed using NeutrAvidin-coated magnetic beads and the high-salt fractions (700 mM + 1 M) were interrogated by qPCR using promoter and body primer sets specific to two genes with CGI promoters. (D) As for (C) except that primer sets were specific to two genes with non-CGI promoters. All Bio-CAP data are from two biological replicates and error bars represent SEM.

Strikingly, the magnetic Bio-CAP assay replicated almost identically the Bio-CAP assay with conventional NeutrAvidin beads (compare Figure 4C and D with Figure 4A and B), with CGI regions being massively enriched in the high-salt fraction compared to gene body and non-CGI promoter regions. This simplified magnetic recovery approach allows manual parallel processing of samples with a typical magnetic microcentrifuge rack capable of manipulating 16 samples per run. These experiments demonstrate that Bio-CAP can be used in conjunction with NeutrAvidin-coated magnetic particles to rapidly and specifically purify non-methylated DNA. Using this magnetic particle approach, it is likely that Bio-CAP would easily transit to an automated purification system such as the Diagenode SX-8G IP-Star Compact® further reducing hands on time and increasing throughput.

#### **Bio-CAP can be coupled to massively parallel sequencing to profile non-methylated DNA**

In all of the above analyses, we interrogated Bio-CAP material via qPCR with primer sets specific for individual genomic loci. We were also keen to understand whether Bio-CAP can be utilized for genome-wide profiling of non-methylated DNA. The conventional CAP technique has been previously coupled to massively parallel sequencing to map non-methylated CGIs in genomic DNA from mouse sperm (19), providing a comparative measure for application of Bio-CAP for the same purpose (Bio-CAP-seq). Starting with 8 µg of mouse sperm genomic DNA, the Bio-CAP high-salt fractions (700 mM and 1 M) yielded ~120 ng of purified DNA, of which 10 ng was used for library generation and Solexa 2G sequencing. When the Bio-CAP-seq reads were aligned to the mouse genome (complete data set to be published at a later date), huge enrichment was observed at computationally defined CGIs (Figure 5A). Importantly, non-CGI gene promoters showed little or no enrichment, confirming that Bio-CAP specifically captures regions containing non-methylated CpGs rather than promoter elements *per se*. Focussing on contiguous regions of the genome, the profile for non-methylated DNA obtained by Bio-CAP-seq shows a striking correlation with the existing conventional CAP-seq data set for mouse sperm (Figure 5). This striking correlation holds true when Bio-CAP-seq and CAP-seq are compared at CGIs genome wide (Figure 5B and C). We also observed that peaks in the Bio-CAP sequencing profile show extremely good overlap with sites of enrichment for the CGI binding factor KDM2A obtained from mouse ES cells (20) (Figure 5A). Importantly, an input DNA control sample, which gave a similar number of sequencing reads, resulted in a flat sequencing profile with no enrichment at CGI regions (Figure 5A). These analyses demonstrate that Bio-CAP can be coupled with massively parallel sequencing to profile non-methylated DNA genome wide.

#### **Bio-CAP-seq specifically and efficiently identifies non-methylated DNA**

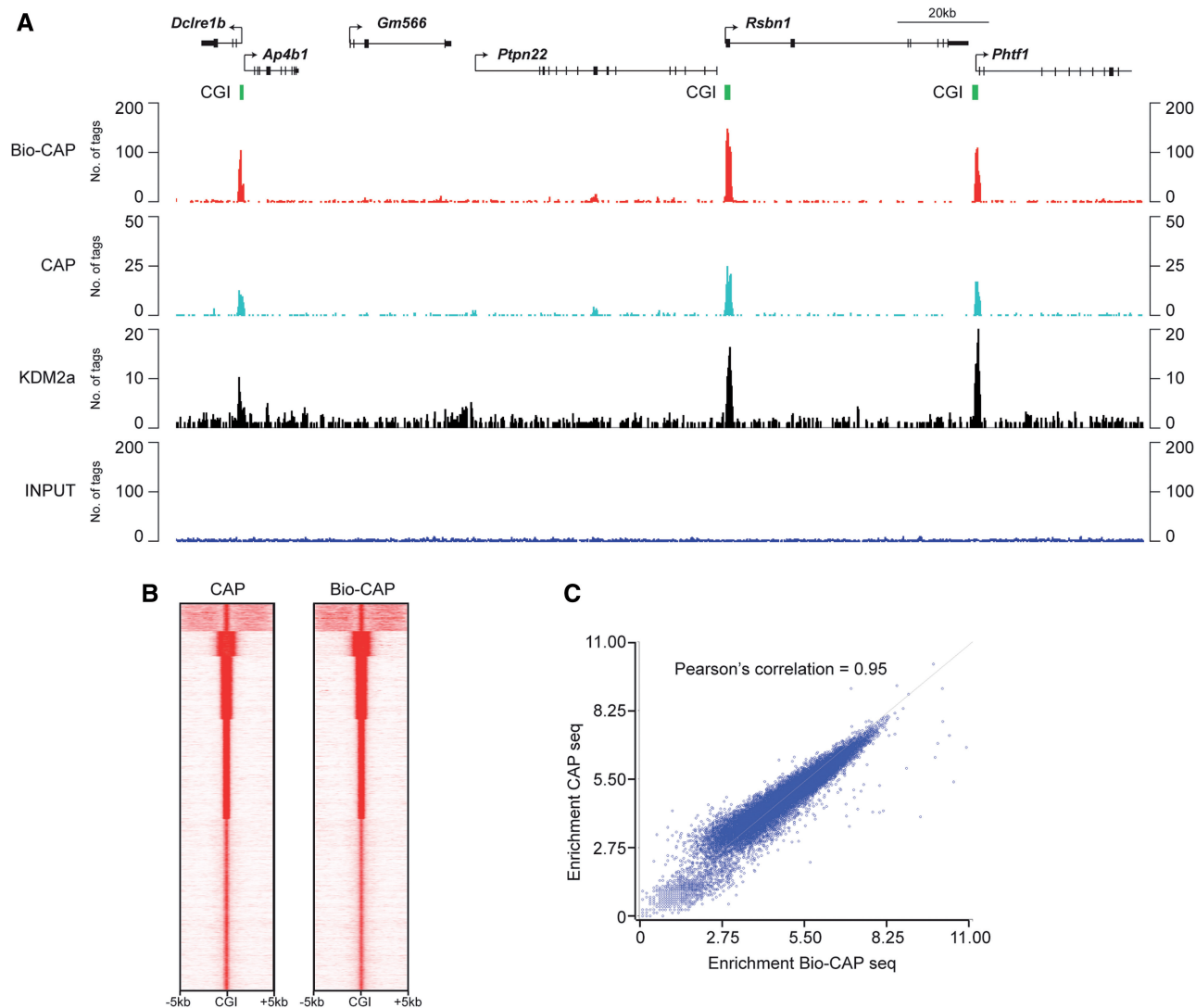
Techniques based on affinity capture of methylated DNA, notably MeDIP and MethylCap, have been previously

coupled to massively parallel sequencing to profile methylated CpGs genome wide (15,16,31,32). Using mouse ES cell genomic DNA, we performed Bio-CAP-seq and MeDIP-seq in order to highlight the utility of Bio-CAP for non-methylated CGI identification and to verify that these regions generally lack methylated DNA signal. When visualizing Bio-CAP-seq and MeDIP-seq signals at equivalent read depth, non-methylated CGIs are immediately evident in the Bio-CAP-seq profiles, but difficult to identify in the MeDIP-seq profiles (Figure 6A and B). Although non-methylated regions can be computationally inferred from MeDIP-seq data sets, this often requires increased sequencing depth and usually fails to identify non-methylated regions with a low-CpG density (33). Indeed, it was only with the advent of CAP that the large number of non-methylated regions falling below the CGI algorithm threshold became evident (18). Importantly, by comparing Bio-CAP-seq and MeDIP-seq signal at CGIs genome wide there was no correlation between Bio-CAP-seq signal and MeDIP-seq signal, verifying that Bio-CAP-seq effectively identifies non-methylated regions of the genome (Figure 6C). Together these observations highlight the utility of Bio-CAP-seq for identification of non-methylated DNA.

## **DISCUSSION**

Within mammalian genomes, the methylation status of CpG dinucleotides can impact gene expression and has been implicated in various developmental and disease processes. Techniques to study DNA methylation include the recently developed CAP approach to specifically capture regions of DNA containing non-methylated CpGs. The CAP assay takes advantage of a CxxC domain that specifically binds non-methylated (but not methylated) CpG dinucleotides. Although CAP has been successfully used to map non-methylated CGIs in a variety of cell types from different species (18,19), the technique has some limitations. Notably, the chromatography column-based approach requires specialist equipment that some laboratories may not have access to, while large elution volumes result in extensive, time-consuming DNA handling steps and demand sizable amounts of input DNA.

To overcome the limitations inherent to the conventional CAP assay, we have developed a new CxxC-based purification called Bio-CAP. Bio-CAP exploits a high-affinity CxxC domain from the histone H3 lysine 36 demethylase KDM2B, into which we have engineered a biotin moiety. This biotin tag enables the CxxC module to be immobilized with a high affinity onto an avidin-based solid support. The nature of this interaction, together with the choice of solid support resin, means that the entire Bio-CAP procedure, including pull-down, washes and elution steps, can be performed in standard microcentrifuge tubes. When compared to conventional CAP, the Bio-CAP procedure is therefore much simpler, requiring only standard laboratory equipment. Furthermore, unlike conventional CAP, the Bio-CAP technique is amenable to processing multiple samples in



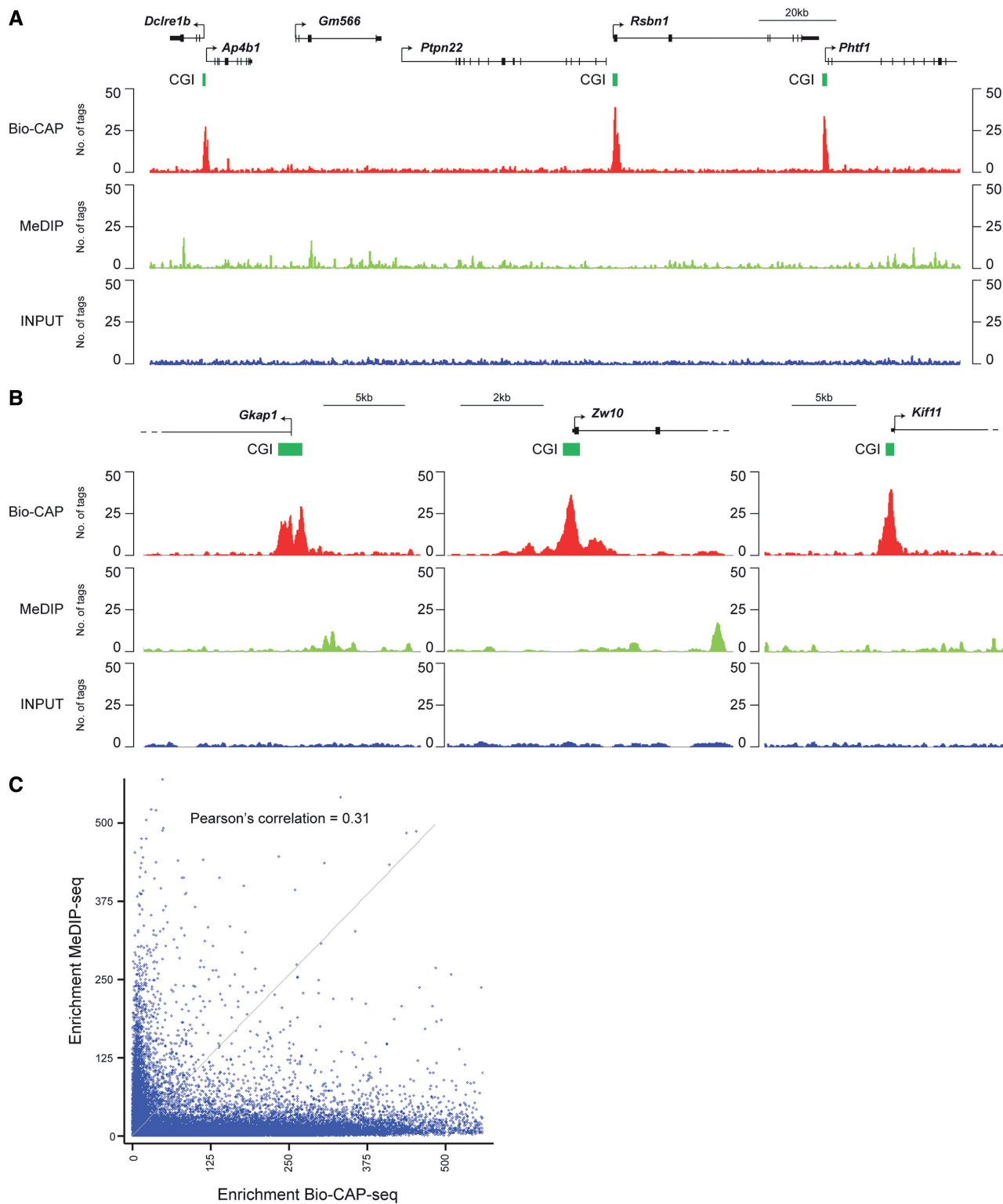
**Figure 5.** Bio-CAP followed by massively parallel sequencing reveals non-methylated CGIs genome wide. (A) The high-salt fraction (700 mM + 1 M) from a Bio-CAP experiment with mouse sperm genomic DNA was subjected to massively parallel sequencing and sequencing reads were aligned back to the mouse genome. Bio-CAP sequencing analysis is shown (top panel, red) for an ~180 kb region of chromosome 3 (chr3: 103 598 131–103 782 151), with comparative analysis from conventional CAP sequencing (second panel, light blue), KDM2A ChIP sequencing in mouse ES cells (third panel, black) and input material (bottom panel, dark blue). Annotated genes in this region are illustrated above the sequencing traces, with arrows indicating transcription start sites and CGIs from *in silico* predictions indicated in green. (B) Heat map comparing CAP-seq and Bio-CAP-seq data sets centered at CGIs including 5 kb windows upstream and downstream. (C) Bio-CAP-seq enrichment plotted against CAP-seq enrichment for all CGIs genome wide, illustrating the extremely high correlation between Bio-CAP-seq and CAP-seq.

parallel and even offers the capacity for robotic automation.

Importantly, we have demonstrated that Bio-CAP can be coupled to qPCR to assay non-methylated DNA at individual loci, or to massively parallel sequencing (Bio-CAP-seq) to profile non-methylated DNA genome wide. Techniques based on affinity capture of methylated CpGs, such as MeDIP, suffer from a reduced capacity to detect non-methylated genomic regions that have a low-CpG density (33). In contrast, traditional CAP and now Bio-CAP are able to enrich for non-methylated DNA, even at low-CpG density, enabling sensitive profiling of non-methylated CGIs (18). We anticipate that identification of non-methylated CGIs using

Bio-CAP-seq, as opposed to inferring their location from extensive sequencing of the large methylated genomic fraction (by MeDIP-seq, for example), will improve our capacity to accurately identify non-methylated regions of the genome, particularly in regions with low-CpG density.

Bio-CAP recovers DNA fragments with an efficiency that is directly proportional to the density of non-methylated CpGs and specifically discriminates between methylated and non-methylated CGIs. Furthermore, in a significant advance on the conventional CAP assay, Bio-CAP can be performed on nanogram quantities of DNA without any apparent loss of resolution. We therefore envisage that Bio-CAP could be



**Figure 6.** Comparison with MeDIP-seq reveals the utility of Bio-CAP-seq in identifying non-methylated CGIs. (A) Mouse ES cell genomic DNA was subjected to both Bio-CAP-seq and MeDIP-seq. Aligned sequencing data are shown for Bio-CAP-seq (top panel, red), MeDIP-seq (middle panel, green) and matched input material (bottom panel, blue) for an ~180 kb region of chromosome 3 (chr3: 103 598 131–103 782 151). Annotated genes in this region are illustrated above the sequencing traces, with arrows indicating transcription start sites and CGIs from *in silico* predictions indicated in green. (B) Bio-CAP-seq and MeDIP-seq data for another three genic regions, each including a CGI. The figure is annotated as for (A). (C) Bio-CAP-seq enrichment plotted against MeDIP-seq enrichment for all CGIs genome wide. The plot demonstrates that Bio-CAP-seq and MeDIP-seq exhibit no correlation, consistent with the fact that Bio-CAP-seq identifies non-methylated regions of the genome.

used to assay non-methylated DNA in an almost limitless number of biological scenarios, both at specific loci and genome wide. This could include situations in which only small amounts of DNA are available, potentially allowing CGI methylation status profiling in rare subpopulations of cells, at very early stages of development, or in valuable experimental samples. Due to the sensitivity of Bio-CAP, we also believe that this technique could reveal subtle methylation differences between samples, for example different developmental time points or different disease states. Finally, given its sensitivity and capacity for robotic automation, it is possible that Bio-CAP could be used in diagnostic applications.

## ACKNOWLEDGEMENTS

We would like to thank the Oxford Wellcome Trust Centre for Human Genetics Genomics Facility for Solexa Sequencing, David Staunton for help with mass spectrometry analysis, Mark Howarth for the BirA expression construct, Ryo Koyama-Nasu for the human KDM2B cDNA, Buyu Li, Christian Edlich and Ernest Laue for sharing unpublished observations regarding KDM2B prior to publication, Kim Nasmyth for the TEV protease and Neil Brockdorff and his laboratory for advice and fruitful discussion. We are also grateful to Anca Farcas and Xuan Shirley Li for critical reading of the manuscript.

## FUNDING

Wellcome Trust; Medical Research Council; Cancer Research UK; European Molecular Biology Organization; Lister Institute of Preventative Medicine; Ludwig Institute for Cancer Research Ltd; Nuffield Department of Clinical Medicine. Funding for open access charge: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Klose,R.J. and Bird,A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
- Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
- Cooper,D.N., Taggart,M.H. and Bird,A.P. (1983) Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.*, **11**, 647–658.
- Bird,A., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, **40**, 91–99.
- Gardiner-Garden,M. and Frommer,M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.
- Larsen,F., Gundersen,G., Lopez,R. and Prydz,H. (1992) CpG islands as gene markers in the human genome. *Genomics*, **13**, 1095–1107.
- Herman,J.G., Latif,F., Weng,Y., Lerman,M.I., Zbar,B., Liu,S., Samid,D., Duan,D.S., Gnarr,J.R., Linehan,W.M. *et al.* (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl Acad. Sci. USA*, **91**, 9700–9704.
- Merlo,A., Herman,J.G., Mao,L., Lee,D.J., Gabrielson,E., Burger,P.C., Baylin,S.B. and Sidransky,D. (1995) 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat. Med.*, **1**, 686–692.
- Herman,J.G. and Baylin,S.B. (2003) Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.*, **349**, 2042–2054.
- Edwards,C.A. and Ferguson-Smith,A.C. (2007) Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.*, **19**, 281–289.
- Reik,W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.
- Laird,P.W. (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.*, **11**, 191–203.
- Ji,H., Ehrlich,L.I., Seita,J., Murakami,P., Doi,A., Lindau,P., Lee,H., Aryee,M.J., Irizarry,R.A., Kim,K. *et al.* (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*, **467**, 338–342.
- Shiraishi,M., Sekiguchi,A., Oates,A.J., Terry,M.J., Miyamoto,Y. and Sekiya,T. (2004) Methyl-CpG binding domain column chromatography as a tool for the analysis of genomic DNA methylation. *Anal. Biochem.*, **329**, 1–10.
- Down,T.A., Rakyen,V.K., Turner,D.J., Flicek,P., Li,H., Kulesha,E., Graf,S., Johnson,N., Herrero,J., Tomazou,E.M. *et al.* (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.*, **26**, 779–785.
- Bock,C., Tomazou,E.M., Brinkman,A.B., Muller,F., Simmer,F., Gu,H., Jager,N., Gnirke,A., Stunnenberg,H.G. and Meissner,A. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.*, **28**, 1106–1114.
- Voo,K.S., Carlone,D.L., Jacobsen,B.M., Flodin,A. and Skalnik,D.G. (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell Biol.*, **20**, 2108–2121.
- Illingworth,R., Kerr,A., Desousa,D., Jorgensen,H., Ellis,P., Stalker,J., Jackson,D., Clee,C., Plumb,R., Rogers,J. *et al.* (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.*, **6**, e22.
- Illingworth,R.S., Gruenewald-Schneider,U., Webb,S., Kerr,A.R., James,K.D., Turner,D.J., Smith,C., Harrison,D.J., Andrews,R. and Bird,A.P. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.*, **6**, e1001134.
- Blackledge,N.P., Zhou,J.C., Tolstorukov,M.Y., Farcas,A.M., Park,P.J. and Klose,R.J. (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell*, **38**, 179–190.
- Kumaki,Y., Oda,M. and Okano,M. (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.*, **36**, W170–W175.
- Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Ye,T., Krebs,A.R., Choukrallah,M.A., Keime,C., Plewniak,F., Davidson,I. and Tora,L. (2011) seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.*, **39**, e35.
- Schatz,P.J. (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology (N Y)*, **11**, 1138–1143.
- Beckett,D., Kovaleva,E. and Schatz,P.J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.*, **8**, 921–929.
- Green,N.M. (1963) Avidin. 1. The use of (14-C)biotin for kinetic studies and for assay. *Biochem. J.*, **89**, 585–591.
- Hiller,Y., Gershoni,J.M., Bayer,E.A. and Wilchek,M. (1987) Biotin binding to avidin. Oligosaccharide side chain not required for ligand association. *Biochem. J.*, **248**, 167–171.

28. Liu, J., Yu, S., Litman, D., Chen, W. and Weinstein, L.S. (2000) Identification of a methylation imprint mark within the mouse Gnas locus. *Mol. Cell Biol.*, **20**, 5808–5817.
29. Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
30. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
31. Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. and Schubeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*, **37**, 853–862.
32. Brinkman, A.B., Simmer, F., Ma, K., Kaan, A., Zhu, J. and Stunnenberg, H.G. (2010) Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*, **52**, 232–236.
33. Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y. *et al.* (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.*, **28**, 1097–1105.