Biochemical identification of non-methylated DNA by BioCAP-seq

Hannah K. Long[A, B], Nathan R. Rose[A], Neil P. Blackledge[A] and Robert J. Klose[A]*

*Corresponding author: Robert J Klose, rob.klose@bioch.ox.ac.uk

Affiliations:

(A) Department of Biochemistry and (B) Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford University, Oxford, UK

Present Address: Hannah K. Long, Department of Chemical and Systems Biology and Institute for Stem Cell Biology & Regenerative Medicine, Stanford University School of Medicine, Stanford University, Stanford, CA 94305, USA.

Keywords

DNA methylation, non-methylated DNA, CpG island, non-methylated island, biotinylated CxxC affinity purification

Summary

CpG islands are regions of vertebrate genomes that often function as gene regulatory elements and are associated with most gene promoters. CpG island elements usually contain non-methylated CpG dinucleotides, while the remainder of the genome is pervasively methylated. We developed a biochemical approach called biotinylated CxxC affinity purification (BioCAP) to unbiasedly isolate regions of the genome that contain non-methylated CpG dinucleotides. The resulting highly pure non-methylated DNA is easily analysed by quantitative PCR to interrogate specific loci or via massively parallel sequencing to yield genome-wide profiles.

1. Introduction

The methylation of cytosine at the C-5 position is one of the most widely studied and best understood of the epigenetic modifications [1–6]. DNA methylation occurs mostly within the context of CpG dinucleotides in vertebrates and is found pervasively throughout the genome [7–10]. In contrast, short contiguous regions of DNA are found interspersed across vertebrate genomes that lack methylation on CpG dinucleotides [11]. These non-methylated regions, called CpG islands (CGIs), exhibit an increased density of CpG dinucleotides compared to the remainder of the genome [11–13] and are associated with up to two thirds of vertebrate gene promoters [14,15]. CGIs specifically recruit a group of proteins that contain a ZF-CxxC domain that binds to non-methylated DNA (Figure 1A) [16–19]. ZF-CxxC proteins tend to associate with chromatin modifying activities suggesting non-methylated DNA may function to alter chromatin structure at gene regulatory elements [20–27]. CGIs generally remain free of DNA methylation in most tissues and are thought to be protected from DNA methylation by DNA demethylases [28,29], nucleotide composition and transcription

factor binding [8,30–33]. However acquisition of DNA methylation at distal CGIs, and in rare cases at gene promoters during development, can lead to gene silencing. Therefore CGI methylation can play a regulatory function in specific developmental contexts, and in human diseases such as cancer [3,34].

A number of techniques are available for the analysis of methylated DNA at the genome scale. While many of these approaches also provide information about where non-methylated DNA is located, this often requires extensive sequencing depth and is costly to implement. Given that only a very small fraction of most vertebrate genomes contains non-methylated DNA, techniques have been developed to directly detect non-methylated DNA. Importantly these approaches require significantly less sequencing depth. These methods generally utilise a selective purification approach to specifically isolate the non-methylated fraction of the genome prior to massively parallel sequencing. For example, the MRE-seq approach utilises methylation-sensitive restriction enzymes (MSREs) to digest DNA only when their recognition site is non-methylated [35]. Small, double-cut DNA fragments released from genomic DNA can then be analysed by massively parallel sequencing to generate non-methylated DNA profiles [36,37]. However this technique is inherently biased due to the use of restriction sites which are not uniformly distributed through the genome. Another method utilises a chemical-labelling approach called methyltransferase-directed transfer of activated groups (mTAG) to covalently label unmodified CpG dinucleotides, which can then be used to isolate non-methhyated regions of the genome [38]. While this method uses unbiased shearing prior to isolation, it is dependent on an enzymatic reaction which must proceed to completion to ensure sensitivity and accuracy. A third approach known as CxxC Affinity Purification (CAP) was developed to affinity purify non-methylated DNA via the DNA binding selectivity of the ZF-CxxC domain for non-methylated CpG dinucleotides [11,39]. Briefly, genomic DNA is fragmented by sonication and then applied to an affinity matrix containing a recombinant ZF-CxxC protein. The non-methylated DNA binds under low salt conditions and methylated DNA is removed by extensive washing. The non-methylated DNA is then eluted with high salt to disrupt the ZF-CxxC domain

interaction with non-methylated DNA. The purified non-methylated DNA is then amenable to downstream analysis including massively parallel sequencing. Importantly, this approach is not subject to the same biases as the other methods described above, because the genome is not fractionated with restriction enzymes prior to affinity purification and is not dependent on an enzymatic reaction to modify non-methylated CpGs prior to affinity purification.

We modified the CAP technique by engineering a high affinity biotinylated ZF-CxxC domain which is immobilized to a neutravidin matrix, chosen for its low non-specific binding to DNA (Figure 1). Known as BioCAP, this approach has several advantages over traditional CAP in that it is carried out in batch in micro centrifuge tubes, is amenable to the use of small amounts of genomic input DNA, and is streamlined for automation allowing parallel sample processing [40]. Extensive BioCAP analysis by massively parallel sequencing (BioCAP-seq) has demonstrated that that this approach is highly effective at generating genome-wide profiles of non-methylated DNA in diverse vertebrate genomes and can function to compare the location of non-methylated DNA between different tissues in the same organism [41].

Here we provide a detailed protocol for the analysis of non-methylated DNA using BioCAP.

2. Materials

Prepare all solutions described below using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at 4 °C unless indicated otherwise.

2.1 *Expression and purification of BioCAP ZF-CxxC affinity capture protein*

1) 1000x antibiotic stock solutions: Dissolve 50 mg kanamycin sulfate in 1 mL water, and filter using a 0.22 µm syringe filter to make a 50 mg/mL stock. Dissolve 34 mg chloramphenicol in 1 mL ethanol to make a 34 mg/mL stock. Store both at -20 °C.

2) 2xTY medium: dissolve 16 g tryptone, 10 g yeast extract, 5 g NaCl in 1L distilled water. Adjust pH to 7.0 and autoclave to sterilise. Store at room temperature.
3) Lysis buffer: 20 mM Tris pH 8.0, 500 mM NaCl, 0.1% NP40, 1x Complete EDTA-free Protease Inhibitor Cocktail added fresh prior to lysis.
4) 1 M $ZnCl_2$ solution: dissolve 272.6 mg $ZnCl_2$ in 2 mL distilled water and filter using 0.22 µm syringe filter. Store at room temperature.
5) IPTG stock: dissolve 476.6 mg IPTG in 2 mL distilled water and filter with 0.22 µm syringe filter. Store at -20 °C.
6) PBS: dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 in 1 L distilled water. Store at room temperature.
7) IMAC sepharose beads: 6 Fast Flow, GE Healthcare, 17-0921-07.
8) BioRad column: Polyprep chromatography column, 731-1550.
9) Wash buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole; adjust pH to 8.0, and add 1x Complete EDTA-free Protease Inhibitor Cocktail fresh prior to use.
10) Elution buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole; adjust pH to 8.0, and add 1x Complete EDTA-free Protease Inhibitor Cocktail fresh prior to use.
11) TEV protease: Sigma, T4455-1MG.
12) HiPrep 26/10 desalting column: GE Healthcare, 17-5087-01.
13) Biotinylation buffer: 20 mM Tris, pH 8.0, 250 mM potassium glutamate. Store at room temperature.
14) BioMix: 10 mM ATP, 10 mM $Mg(OAc)_2$, 50 µM D-biotin. Store at -80 °C.
15) 1 M imidazole: dissolve 68 mg imidazole in 1 mL distilled water. Store at room temperature.
16) BC150 buffer: 20 mM HEPES pH 7.9, 150 mM KCl, 0.5 mM dithiothreitol (DTT), 10% v/v glycerol.

2.2 *Preparation of genomic DNA, performing BioCAP and picogreen quantification of DNA*

1) Extraction buffer: 20 mM Tris HCl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 0.5% SDS (add after homogenisation). Add fresh: 200 µg/mL Proteinase K. If purifying genomic DNA from sperm, add 40 mM DTT fresh. Store extraction buffer without Proteinase K and DTT at room temperature.
2) RNaseA/T1: Thermo Scientific, EN0551.
3) 70% ethanol: 70% EtOH and 30% distilled water by volume. Store at room temperature.
4) Genomic-tip 100/G: QIAGEN, 10243.
5) NeutrAvidin Agarose Resin: Thermo Scientific, 29200.
6) NeutrAvidin-coated magnetic beads: Thermo Scientific, 7815-2104-011150.
7) Protein LoBind tube: Eppendorf, 022431081.
8) CAP buffers: X mM NaCl, 0.1% Triton-x-100, 20 mM HEPES pH 7.9, 12.5% v/v Glycerol. CAP100, X = 100 mM NaCl. CAP300, X = 300 mM NaCl. CAP500, X = 500 mM NaCl. CAP700, X = 700 mM NaCl. CAP1000, X = 1000 mM NaCl. Check conductivity of buffers using a conductivity meter.
9) 1.5 mL TPX microtubes: Diagenode, M-50050.
10) PicoGreen: Invitrogen, P11496.
11) 1xTE: 10 mM Tris, pH 8.0, 1 mM EDTA.

3. Methods

3.1 *Expression and purification of BioCAP ZF-CxxC affinity capture protein*

Recombinant KDM2B ZF-CxxC protein was expressed and purified from *E. coli*. The protocol is illustrated with appropriate stopping points indicated in Figure 2.

1) Transform pNIC28-hKdm2b-CxxC-PHD (See Note 1) into 100 µL Rosetta2 *E. coli* (or any strain containing pRARE2 plasmid) following a standard heat shock transformation protocol. Plate on agar plate containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL) and incubate overnight at 37 °C. This plasmid encodes a bacterial expression cassette to produce a His-tagged human KDM2B ZF-CxxC domain with a C-terminal avi-tag and a TEV protease cleavage site immediately after the N-terminal His_6 tag.
2) Inoculate 100 mL 2xTY medium (+ 100 µL kanamycin and chloramphenicol 1000x stock solutions) with a single colony of Rosetta2 *E. coli* from the overnight transformation (Step 1). Grow overnight at 37 °C.
3) Add 15 mL overnight culture to 650 mL 2xTY in a 2 L baffled Erlenmeyer flask (add 650 µL kanamycin and chloramphenicol 1000x stock solutions, and supplement with 162.5 µL of 1 M $ZnCl_2$ solution. Prepare 6 flasks in total, resulting in approximately 4 L of culture. Grow at 37 °C until the OD_{600} = 0.6, then cool to 30 °C and induce expression with the addition of 325 µL of 1 M IPTG (isopropyl β-D-1-thiogalactopyranoside) stock solution to each flask.
4) Grow at 30 °C for 5 h before harvesting by centrifugation at 10,000 rcf for 20 min at 4 °C.
5) Wash pellet once by resuspending in 20 mL PBS, transferring to a 50 mL Falcon tube and centrifuging for 10 min at 3,600 rcf. Remove supernatant. At this point the pellet can be stored at -80 °C until convenient to continue.
6) Add lysis buffer corresponding to 4x the volume of the cell pellet, and resuspend pellet gently without introducing air bubbles. Sonicate on ice in 20 mL aliquots at 60% amplitude (using Sonics Vibra-cell sonicator), alternating 30 s sonication with 30 s rest on ice.
7) Centrifuge the lysate at 48,000 rcf for 20 min. Filter the supernatant through a 0.22 µm syringe filter and set aside. This is the soluble protein fraction from *E. coli* that contains the recombinant His_6 tagged ZF-CxxC domain.
8) Charge IMAC sepharose beads with 0.2 M $NiSO_4$ following the manufacturer's instructions to generate Ni-NTA beads.

9) Prepare the Ni-NTA beads by transferring 4 mL of 50 % bead slurry to a 15 mL Falcon tube and centrifuging at 850 rcf for 3 min. Remove the supernatant and resuspend the beads in 10 mL lysis buffer. Centrifuge again at 850 rcf for 3 min, remove supernatant, and repeat this wash once. Remove the supernatant.
10) Add the filtered lysate supernatant (from Step 7) to the washed beads and rotate the suspension for 1 h at 4 °C to allow the His_6-tagged protein to bind to the beads.
11) Transfer the suspension to an empty 25 mL BioRad column and collect the flowthrough. Wash the beads twice with 10 mL wash buffer. The protein is now bound to the Ni-NTA beads.
12) Elute protein from the column in batch by adding 1mL of elution buffer gently to the bead bed in the BioRad column. Collect the eluate in a 1.5 mL Eppendorf tube. Repeat the elution 10 times.
13) Run 7.5 µL of each of the lysate supernatant, flowthrough, washes and eluate fractions on a 12 % SDS-PAGE gel. Pool the eluate fractions which contain significant amounts of CxxC protein, which is approximately 19 kDa in size. Check protein concentration by Bradford assay or Nanodrop.
14) Add the appropriate amount of TEV protease to pooled fractions and rotate overnight at 4 °C to cleave off the His_6-tag as per the manufacturer's recommedations.
15) Desalt the CxxC/TEV mixture on a HiPrep 26/10 desalting column (GE Healthcare) into wash buffer following the manufacturer's instructions to remove excess imidazole, and add the eluate to 300 µL pre-washed Ni-NTA beads (prepare according to Step 8) by washing 600 µL slurry in wash buffer). Rotate for 1 h at 4 °C to bind the His_6-tagged TEV protease and cleaved His_6-tag from the recombinant protein, then transfer to an empty 25 mL BioRad column and collect flowthrough. Retain flowthrough as this now contains the CxxC protein without the His_6-tagged TEV protease or cleaved His_6 tag. Also elute bound TEV with 2 mL

elution buffer and retain sample for SDS-PAGE analysis to verify efficient cleavage of the His_6-tag.

16) Desalt the CxxC protein (from the flowthrough in Step 15) on a HiPrep 26/10 desalting column, into biotinylation buffer. (See Note 2). Check the concentration of the protein by Bradford assay or Nanodrop spectrophotometry.
17) Add His_6-tagged BirA biotin ligase (add 2.5 µg for every 10 nmol CxxC substrate, see reference [42] for expression and purification details; also see Note 3). Also add 500 µL BioMix. Rotate for 2 h at 4 °C, then add the same quantity of His_6-tagged BirA and 500 µL BioMix again. Allow the reaction to go to completion by rotating overnight at 4 °C.
18) Adjust the imidazole concentration of the reaction mixture to 20 mM imidazole by adding the appropriate amount of 1 M imidazole. Add 300 µL prewashed Ni-NTA beads to the reaction mixture (prepare according to Step 9 by washing 600 µL slurry in wash buffer), and rotate for 1 h at 4 °C to remove the His_6-tagged BirA. Transfer to an empty BioRad column. Collect flowthrough (retain flowthrough as this contains the biotinylated CxxC protein). Elute bound BirA with 2 mL elution buffer and retain sample for SDS-PAGE analysis.
19) Desalt the CxxC protein into BC150 buffer on a HiPrep 26/10 desalting column, and check concentration by Bradford assay. Aliquot and store the biotinylated CxxC protein in volumes of 50 µL to avoid repeat freeze thaw cycles. To verify complete biotinylation, check the mass of the purified protein by mass spectrometry. The unmodified protein is 18,939 Da and the biotinylated protein is 19,165 Da.

3.2 *Prepare genomic DNA*

3.2.1 Genomic DNA extraction protocol from tissue samples

1) To prepare genomic DNA, add 950 µL extraction buffer to 25 mg tissue (without SDS or Proteinase K).

2) Homogenise tissue in a 1 mL dounce homogeniser until tissue is dissociated into a cell suspension (some connective tissue may remain).
3) Add SDS and Proteinase K to homogenised tissue (and DTT if preparing genomic DNA from sperm, see Note 4).
4) Incubate at 50 °C for 3 h (or until the sample is homogeneous).
5) Add 10 µL of RNaseA/T1 (0.01 volumes, 2 mg/mL, 10,000 U/mL) and incubate at 37 °C for 1 h.
6) Split the extract into two 2 mL eppendorfs and add a further 400 µL extraction buffer to each (without SDS, Proteinase K, and DTT) to obtain a volume that is easy to extract by phenol chloroform extraction.
7) To carry out a phenol chloroform extraction, add an equal volume (900 µL) of phenol:chloroform:isoamylalcohol (25:24:1 saturated with 10 mM Tris pH8.0 and 1mM EDTA) and vortex briefly. Separate phases by centrifugation for 5 min at 850 rcf in a benchtop centrifuge at room temperature. Transfer the upper aqueous phase into a fresh tube being careful not to disturb the interface between the upper and lower phases.
8) Repeat the phenol chloroform extraction twice more.
9) Perform a final extraction with an equal volume of chloroform (900 µL), centrifuge as above and transfer the upper aqueous phase to a fresh tube. Each tube will have less than 900 µL due to loss of the sample during each transfer step.
10) To perform ethanol precipitation, split the upper phase from the final extraction into four 2 mL eppendorf tubes for ease of ethanol precipitation (around 400 µL extract per tube). Add 2.5 volumes ethanol (1 mL) and 0.1 volumes sodium acetate (3M NaOAc, pH 5.2, 40 µL).
11) Place the samples at -80 °C for 30 min – 1 h.
12) Centrifuge sample at 21,100 rcf in a benchtop centrifuge for 20 min at 4 °C.
13) Carefully decant the supernatant without disturbing the pellet and wash the pellet by adding 1 mL 70% ethanol per eppendorf tube.
14) Centrifuge sample at 21,100 rcf in a benchtop centrifuge for 10 min at 4 °C.

15) Carefully decant the supernatant without disturbing the pellet and pulse spin samples. Carefully remove the remaining liquid with a pipette taking care not to disturb the pellet.
16) Air dry the pellet for 15 min at room temperature.
17) Resuspend each pellet in 50 µL MilliQ water. Warm the samples to 37°C and gently flick the tubes if necessary to aid resuspension.
18) Combine the samples and determine the concentration using a Nanodrop 1000 spectrophotometer at 260 nm.
19) Store genomic DNA at -20 °C.

3.2.2 Alternative purification of genomic DNA

Alternatively, we have had success purifying genomic DNA using the 100/G Genomic Tip kit and routinely use this method for purification of genomic DNA from 80-100 mg of tissue (depending on the tissue-type). The manufacturer's instructions were followed for extraction of genomic DNA for tissue samples. A brief outline of the procedure is given below.

1) Homogenise 80-100 mg tissue in a 1 mL dounce using Buffer G2 with RNaseA, final concentration 200 µg/mL.
2) Incubate the sample at 50 °C for 2 h with Proteinase K (1 mg/mL final concentration).
3) If particulate matter remains in the sample, centrifuge for 5000 rcf for 10 min at 4 °C. Discard the pellet.
4) Vortex for 10 s and apply the sample to a 100/G Genomic-tip column pre-equilibrated with 4 mL Buffer QBT.
5) Wash the column twice with 7.5 mL Buffer QC.
6) Elute genomic DNA with 5 mL of Buffer QF (warmed to 50 °C).
7) Precipitate genomic DNA by addition of 0.7 volumes of room temperature isopropanol. This is most easily done by aliquoting 1 mL of the sample into ten 2 mL eppendorf tubes and adding

700 µL isopropanol. Mix the samples by inversion and immediately centrifuge samples for 15 min at 21,100 rcf in a benchtop centrifuge at 4 °C.

8) Wash the DNA pellet with 1 mL 4 °C 70 % ethanol.

9) Air dry the pellet and resuspend in 50 µL MilliQ water.

10) Combine the samples and determine the concentration using a Nanodrop 1000 spectrophotometer at 260 nm.

11) Store genomic DNA at -20 °C.

3.3 *Performing BioCAP*

3.3.1 Preparation of beads

1) For each BioCAP experiment, use 25 µL of packed NeutrAvidin Agarose Resin or NeutrAvidin-coated magnetic beads.

2) Pipette 50 µL beads (50 % slurry) into a Protein LoBind tube and either pellet by centrifugation for 3 min at 850 rcf in a swing-bucket centrifuge by placing the centrifuge tubes in the top of a 15 mL falcon tube with the cap removed (for agarose beads), or immobilise by magnetisation for magnetic beads.

3) Remove the supernatant by pipetting with a gel-loading tip and resuspend the beads in 1 mL of BC150 buffer by inversion.

4) Pellet the beads again, or immobilise using the magnet, and remove the supernatant.

5) Incubate the beads with 50 µL of 0.5 µg/µL biotinylated hKDM2B ZF-CxxC protein diluted in 425 µL BC150 buffer for 1 h at 4 °C, rotating the beads on a flywheel at 15 rpm.

6) Pellet the conjugated resin-CxxC protein, or immobilise on the magnet, and remove the supernatant.

7) Wash the beads three times with 1 mL CAP1000 buffer and once with 1 mL CAP100 buffer by inverting the tubes and rotating for 5 min at 4 °C on a flywheel (see Note 5). Between washes, remove the supernatant from the beads by centrifuging or immobilising on a magnet.

3.3.2 Preparing genomic DNA for use in BioCAP

1) Typically 100 µL of approximately 350 ng/µL genomic DNA is used per BioCAP experiment.
2) Sonicate the genomic DNA to an average size of 150-250 bp. Sonication is performed in 1.5 mL TPX microtubes using a Diagenode Bioruptor. Sonication is performed on the high setting with 30 s on and 30 s off for 2 h. During sonication, spin the samples down at 10 min intervals by pulse centrifugation in a Mini Centrifuge.
3) Separate by gel electrophoresis 3 µL sonicated sample and 3 µL pre-sonication sample on a 1% agarose gel to ensure the sample is sonicated to 150-350 bp in length (see Figure 1B).
4) Dilute the sonicated DNA to approximately 17.5 µg/mL in CAP100 buffer: e.g. 50 µL sample is diluted in 950 µL CAP100 buffer.
5) 500 µL of this material is used for the BioCAP procedure and the remaining volume is saved as an input control.

3.3.3 BioCAP

The BioCAP protocol is shown schematically in Figure 1C.

1) For each BioCAP experiment, add 500 µL of diluted sonicated DNA, corresponding to approximately 8 µg of DNA, to the conjugated CxxC resin.
2) Incubate beads plus genomic DNA at 4 °C for 1 h, rotating on a flywheel at 15 rpm.
3) Collect the resin by centrifugation at 850 rcf for 3 min at 4 °C or by magnetisation, and remove the unbound flow through (FT) material.

4) Wash the ZF-CxxC resin with bound DNA twice with 1 mL of CAP100 buffer by inverting the tube several times, and removing the supernatant following centrifugation at 850 rcf for 3 min at 4°C or by magnetisation.
5) The first BioCAP elution is performed by addition of 50 µL of CAP300 buffer to the ZF-CxxC resin with incubation at room temperature for 10 min with gentle agitiation (the tube should not be inverted due to the small volume of elution buffer). Following centrifugation or magnetisation, the 50 µL elution fraction is carefully collected using a gel-loading tip.
6) The elution process is repeated using another 50 µL of CAP300 for 10 min and the 300 mM elution fractions are pooled (giving a total volume of 100 µL).
7) Perform subsequent elutions in the same manner using buffers CAP500, CAP700 and CAP1000 sequentially, with two elutions per buffer.
8) Each 100 µL elution fraction, and 100 µL of both the input and FT samples, should then be purified using a PCR purification column, eluting DNA in 50 µL distilled water.
9) For real-time qPCR analysis, BioCAP samples are typically diluted 10-fold and 5 µL used in a 15 µL quantitative PCR reaction volume.
10) For BioCAP sequencing, the BioCAP recovery should be verified for all elution fractions by qPCR at several loci and the amount of DNA in the CAP700 and CAP1000 elutions quantified using PicoGreen reagent (see 'PicoGreen quantification', 3.3.5). The combined CAP700 and CAP1000 elutions can then be used for library preparation for massively parallel sequencing.

3.3.4 Alternative rapid magnetic BioCAP

Mouse genomic DNA, sonicated to 150-350 bp, is incubated with the prepared magnetic CxxC resin (see 3.3.1 'Preparation of beads') for 1 h in the same manner as a conventional BioCAP experiment (see **Error! Reference source not found.** 'BioCAP'). The magnetic CxxC resin and associated DNA are then collected using a magnetic microcentrifuge tube rack, allowing unbound FT material to be removed. Two 10 min washes are performed with 50 µL CAP500 (500 mM NaCl), followed by two 10

min elution steps with 50 µL CAP1000 (1M NaCl). The CAP1000 elutions are combined for downstream analysis. The alternative magnetic BioCAP protocol may be preferred when rapid isolation of hypomethylated DNA is desired without interrogating the stepwise elution of increasingly hypomethylated DNA fractions. Equally, this method is convenient for rapid analysis of large numbers of samples, and is amenable to automated processing.

3.3.5 PicoGreen quantification

To quantify DNA at low concentrations, PicoGreen quantification is used.

1) Determine the concentration of input DNA (i.e. sonicated DNA diluted to around 17.5 µg/mL in CAP100 buffer for BioCAP experiments, see 3.3.2) using a Nanodrop 1000 spectrophotometer at 260 nm.
2) Dilute the input DNA to 1 µg/mL in 1xTE and make a dilution series of 200 ng/mL, 100 ng/mL, 40 ng/mL and 8 ng/mL of the 1 µg/mL input DNA using 1xTE. Mix 2 µL of each dilution with 2 µL of PicoGreen (diluted 1:200 in 1xTE). Create a calibration curve by measuring the fluorescence of the dilution series using a Nanodrop 3300. The concentration of the samples can then be inferred from this curve.
3) To calculate the concentration of the purified CAP elutions, dilute 1 µL of each samples 10-fold in 1xTE. Then mix 2 µL of the 10-fold diluted or undiluted samples with 2 µL diluted PicoGreen. Mix by flicking the tubes and pulse spin.
4) Detect fluorescence of the samples using a Nanodrop 3300 and infer the concentration of the samples from the calibration curve (from Step 2). Average the two inferred DNA concentration values from the undiluted and 10-fold diluted DNA samples to calculate an estimate of the DNA concentration for each sample.

4. Notes

1) AddGene, plasmid 49216: pNIC28-hKdm2b-CxxC-PHD.

2) Note that BirA activity is inhibited to some degree by NaCl, so this desalting step is crucial to ensure efficient biotinylation.
3) AddGene, plasmid 20857: pET21a-BirA.
4) Sperm heads are resilient to extraction by proteinase K and SDS alone as their membrane is enriched for disulfide bonds [43]. Therefore for genomic DNA extraction of sperm or testes tissue, the strong reducing agent dithiothreitol (DTT) was added to the gDNA extraction buffer to reduce these disulfide bonds.
5) This wash step is important, as it removes any bacterial CpG-containing DNA that might be bound to the CxxC domain from the protein expression and purification steps.

5. References

[1] R.J. Klose, A.P. Bird, Genomic DNA methylation: The mark and its mediators, Trends Biochem. Sci. 31 (2006) 89–97. doi:10.1016/j.tibs.2005.12.008.

[2] H. Cedar, Y. Bergman, Programming of DNA methylation patterns, Annu Rev Biochem. 81 (2012) 97–117. doi:10.1146/annurev-biochem-052610.

[3] P.A. Jones, Functions of DNA methylation: islands, start sites, gene bodies and beyond., Nat. Rev. Genet. 13 (2012) 484–92. doi:10.1038/nrg3230.

[4] Y. Bergman, H. Cedar, DNA methylation dynamics in health and disease, Nat Struct Mol Biol. 20 (2013) 274–281. doi:nsmb.2518 [pii]\r10.1038/nsmb.2518.

[5] S. Seisenberger, J.R. Peat, T. a Hore, F. Santos, W. Dean, W. Reik, Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers., Philos. Trans. R. Soc. London. 368 (2013) 20110330. doi:10.1098/rstb.2011.0330.

[6] D. Schübeler, Function and information content of DNA methylation, Nature. 517 (2015) 321–326. doi:10.1038/nature14192.

[7] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, a H. Millar, J. a

Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences., Nature. 462 (2009) 315–22. doi:10.1038/nature08514.

[8] M.B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E.J. Oakeley, D. Gaidatzis, V.K. Tiwari, D. Schübeler, DNA-binding factors shape the mouse methylome at distal regulatory regions., Nature. 480 (2011) 490–5. doi:10.1038/nature10716.

[9] L. Jiang, J. Zhang, J.J. Wang, L. Wang, L. Zhang, G. Li, X. Yang, X. Ma, X. Sun, J. Cai, J. Zhang, X. Huang, M. Yu, X. Wang, F. Liu, C.I. Wu, C. He, B. Zhang, W. Ci, J. Liu, Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos, Cell. 153 (2013) 773–784. doi:10.1016/j.cell.2013.04.041.

[10] M.E. Potok, D.A. Nix, T.J. Parnell, B.R. Cairns, Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern, Cell. 153 (2013) 759–772. doi:10.1016/j.cell.2013.04.030.

[11] R. Illingworth, A. Kerr, D. DeSousa, H. Jørgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, A. Bird, A novel CpG island set identifies tissue-specific methylation at developmental gene loci, PLoS Biol. 6 (2008) 0037–0051. doi:10.1371/journal.pbio.0060022.

[12] K.E. Conway, B.B. McConnell, C.E. Bowring, C.D. Donald, S.T. Warren, P.M. Vertino, TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers, Cancer Res. 60 (2000) 6236–6242. doi:10.1038/321209a0.

[13] M. Gardiner-Garden, M. Frommer, CpG Islands in vertebrate genomes, J. Mol. Biol. 196 (1987) 261–282. doi:10.1016/0022-2836(87)90689-9.

[14] F. Larsen, G. Gundersen, R. Lopez, H. Prydz, CpG islands as gene markers in the human genome, Genomics. 13 (1992) 1095–1107. doi:10.1016/0888-7543(92)90024-M.

[15] A. Bird, M. Taggart, M. Frommer, O.J. Miller, D. Macleod, A fraction of the mouse genome

that is derived from islands of nonmethylated, CpG-rich DNA, Cell. 40 (1985) 91–99. doi:10.1016/0092-8674(85)90312-5.

[16] T. Cierpicki, L.E. Risner, J. Grembecka, S.M. Lukasik, R. Popovic, M. Omonkowska, D.D. Shultis, N.J. Zeleznik-Le, J.H. Bushweller, Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia, Nat. Struct. & Mol. Biol. 17 (2009) 62–68. doi:10.1038/nsmb.1714.

[17] J. Song, O. Rechkoblit, T.H. Bestor, D.J. Patel, Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation., Science. 331 (2011) 1036–40. doi:10.1126/science.1195380.

[18] C. Xu, C. Bian, R. Lam, A. Dong, J. Min, The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain., Nat. Commun. 2 (2011) 227. doi:10.1038/ncomms1237.

[19] J. Song, M. Teplova, S. Ishibe-Murakami, D.J. Patel, Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation, Science. 335 (2012) 709–712. doi:10.1126/science.1214453.

[20] N.P. Blackledge, J.C. Zhou, M.Y. Tolstorukov, A.M. Farcas, P.J. Park, R.J. Klose, CpG Islands Recruit a Histone H3 Lysine 36 Demethylase, Mol. Cell. 38 (2010) 179–190. doi:10.1016/j.molcel.2010.04.009.

[21] J.P. Thomson, P.J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A.R.W. Kerr, A. Deaton, R. Andrews, K.D. James, D.J. Turner, R. Illingworth, A. Bird, CpG islands influence chromatin structure via the CpG-binding protein Cfp1., Nature. 464 (2010) 1082–6. doi:10.1038/nature08924.

[22] A.M. Farcas, N.P. Blackledge, I. Sudbery, H.K. Long, J.F. McGouran, N.R. Rose, S. Lee, D. Sims, A. Cerase, T.W. Sheahan, H. Koseki, N. Brockdorff, C.P. Ponting, B.M. Kessler, R.J. Klose, KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands, Elife. 2012 (2012) 1–26. doi:10.7554/eLife.00205.

[23] M. Boulard, J.R. Edwards, T.H. Bestor, FBXL10 protects Polycomb-bound genes from hypermethylation, Nat. Genet. 47 (2015) 1–9. doi:10.1038/ng.3272.

[24] X. Wu, J.V. Johansen, K. Helin, Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation, Mol. Cell. 49 (2013) 1134–1146. doi:10.1016/j.molcel.2013.01.016.

[25] J. He, L. Shen, M. Wan, O. Taranova, H. Wu, Y. Zhang, Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes., Nat. Cell Biol. 15 (2013) 373–84. doi:10.1038/ncb2702.

[26] H.K. Long, N.P. Blackledge, R.J. Klose, ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection., Biochem. Soc. Trans. 41 (2013) 727–40. doi:10.1042/BST20130028.

[27] N.P. Blackledge, A.M. Farcas, T. Kondo, H.W. King, J.F. McGouran, L.L.P. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H.K. Long, T.W. Sheahan, N. Brockdorff, B.M. Kessler, H. Koseki, R.J. Klose, Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation, Cell. 157 (2014) 1445–1459. doi:10.1016/j.cell.2014.05.004.

[28] K. Williams, J. Christensen, K. Helin, DNA methylation: TET proteins—guardians of CpG islands?, EMBO Rep. 13 (2011) 28–35. doi:10.1038/embor.2011.233.

[29] K.D. Rasmussen, K. Helin, Role of TET enzymes in DNA methylation, development, and cancer, Genes Dev. 30 (2016) 733–750. doi:10.1101/gad.276568.115.

[30] F. Lienert, C. Wirbelauer, I. Som, A. Dean, F. Mohn, D. Schübeler, Identification of genetic elements that autonomously determine DNA methylation states., Nat. Genet. 43 (2011) 1091–1097. doi:10.1038/ng.946.

[31] A.R. Krebs, S. Dessus-Babus, L. Burger, D. Sch??beler, High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions, Elife. 3 (2014) e04094. doi:10.7554/eLife.04094.

[32] E. Wachter, T. Quante, C. Merusi, A. Arczewska, F. Stewart, S. Webb, A. Bird, Synthetic CpG islands reveal DNA sequence determinants of chromatin structure, Elife. 3 (2014) e03397. doi:10.7554/eLife.03397.

[33] H.K. Long, H.W. King, R.K. Patient, D.T. Odom, R.J. Klose, Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved., Nucleic Acids Res. 44 (2016) gkw258. doi:10.1093/nar/gkw258.

[34] I.J.H. Van Vlodrop, H.E.C. Niessen, S. Derks, M.M.L.L. Baldewijns, W. Van Criekinge, J.G. Herman, M. Van Engeland, Analysis of promoter CpG island hypermethylation in cancer: Location, location, location!, Clin. Cancer Res. 17 (2011) 4225–4231. doi:10.1158/1078-0432.CCR-10-3394.

[35] P.W. Laird, Principles and challenges of genomewide DNA methylation analysis, Nat. Rev. Genet. 11 (2010) 191–203. doi:10.1038/nrg2732.

[36] R.A. Harris, T. Wang, C. Coarfa, R.P. Nagarajan, C. Hong, S.L. Downey, B.E. Johnson, S.D. Fouse, A. Delaney, Y. Zhao, A. Olshen, T. Ballinger, X. Zhou, K.J. Forsberg, J. Gu, L. Echipare, H. O'Geen, R. Lister, M. Pelizzola, Y. Xi, C.B. Epstein, B.E. Bernstein, R.D. Hawkins, B. Ren, W.-Y. Chung, H. Gu, C. Bock, A. Gnirke, M.Q. Zhang, D. Haussler, J.R. Ecker, W. Li, P.J. Farnham, R. a Waterland, A. Meissner, M. a Marra, M. Hirst, A. Milosavljevic, J.F. Costello, Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications., Nat. Biotechnol. 28 (2010) 1097–1105. doi:10.1038/nbt.1682.

[37] A.K. Maunakea, R.P. Nagarajan, M. Bilenky, T.J. Ballinger, C. D'Souza, S.D. Fouse, B.E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V.M. Heine, D.H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S.J.M. Jones, D. Haussler, M.A. Marra, M. Hirst, T. Wang, J.F. Costello, Conserved role of intragenic DNA methylation in regulating alternative promoters., Nature. 466 (2010) 253–7. doi:10.1038/nature09165.

[38] E. Kriukienė, V. Labrie, T. Khare, G. Urbanavičiūtė, A. Lapinaitė, K. Koncevičius, D. Li, T. Wang, S. Pai, C. Ptak, J. Gordevičius, S.-C. Wang, A. Petronis, S. Klimašauskas, DNA unmethylome

profiling by covalent capture of CpG sites., Nat. Commun. 4 (2013) 2190. doi:10.1038/ncomms3190.

[39] R.S. Illingworth, U. Gruenewald-Schneider, S. Webb, A.R.W. Kerr, K.D. James, D.J. Turner, C. Smith, D.J. Harrison, R. Andrews, A.P. Bird, Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome, PLoS Genet. 6 (2010) e1001134. doi:10.1371/journal.pgen.1001134.

[40] N.P. Blackledge, H.K. Long, J.C. Zhou, S. Kriaucionis, R. Patient, R.J. Klose, Bio-CAP: A versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA, Nucleic Acids Res. 40 (2012) e32. doi:10.1093/nar/gkr1207.

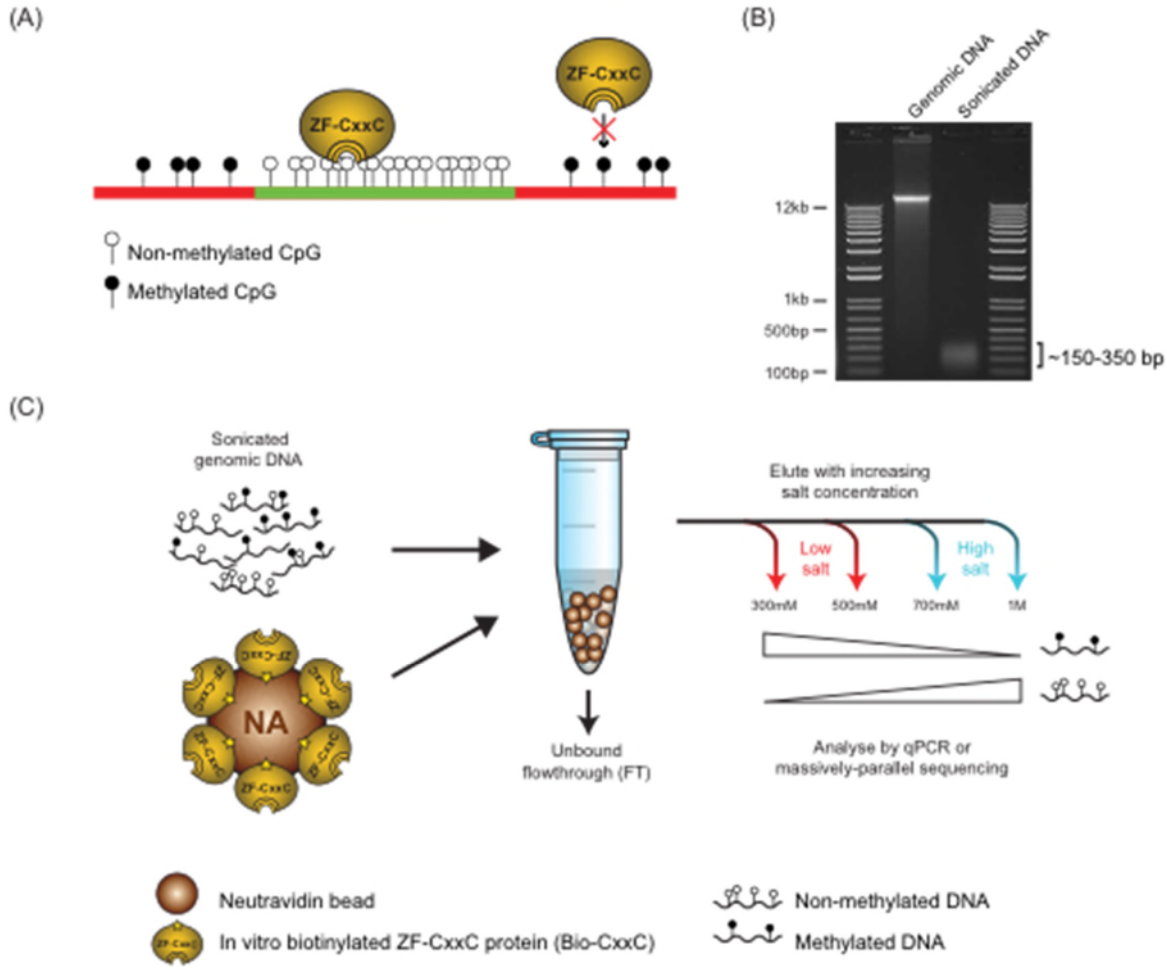
[41] H.K. Long, D. Sims, A. Heger, N.P. Blackledge, C. Kutter, M.L. Wright, F. Grützner, D.T. Odom, R. Patient, C.P. Ponting, R.J. Klose, Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates, Elife. 2013 (2013) 1–19. doi:10.7554/eLife.00348.

[42] M. Howarth, A.Y. Ting, Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin., Nat. Protoc. 3 (2008) 534–545. doi:10.1038/nprot.2008.20.

[43] A. Weyrich, Preparation of genomic DNA from mammalian sperm, Curr. Protoc. Mol. Biol. 1 (2012) 2–4. doi:10.1002/0471142727.mb0213s98.

Figure Legends

Figure 1. The BioCAP technique



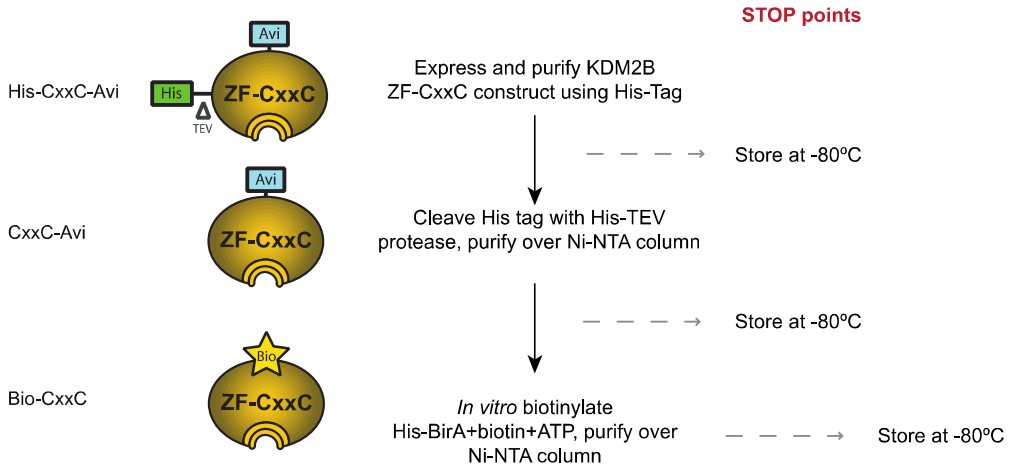
(A) The ZF-CxxC domain is able to bind to non-methylated CpG dinucleotides. ZF-CxxC DNA binding is abrogated by DNA methylation of cytosine at CpG dinucleotides. Non-methylated CpGs are depicted as open circles, and methylated CpGs as filled circles, a region corresponding to a CpG island is shown in green.

(B) For BioCAP experiments, genomic DNA was sonicated for two hours to yield DNA fragments of 150-350 bp typically. Genomic DNA shown was extracted from mouse testes.

(C) A schematic of the BioCAP procedure. *In vitro* biotinylated recombinant ZF-CxxC protein (Bio-CxxC) was immobilized onto neutravidin beads and incubated with sheared genomic DNA (B),

allowing the ZF-CxxC domain to bind to DNA. Unbound DNA was removed in the FT, a series of elutions then followed using increasing salt concentrations. Highly methylated DNA is removed in the early low salt elutions (300 mM and 500 mM) and non-methylated DNA was enriched in the late high salt elutions (700 mM and 1000 mM). DNA from each of these fractions was interrogated by qPCR or the 700 mM and 1000 mM elutions were combined for massively parallel sequencing library preparation.

Figure 2. Production of BioCAP protein



A schematic of the production of the affinity module used in BioCAP. Appropriate stopping points are indicated.