

Cite this: *Lab Chip*, 2012, **12**, 376

www.rsc.org/loc

## TECHNICAL NOTE

### Fabrication of duplex DNA microarrays incorporating methyl-5-cytosine

Christopher L. Warren,<sup>†,ab</sup> Jianfei Zhao,<sup>c</sup> Kimberly Glass,<sup>d</sup> Vikas Rishi,<sup>c</sup> Aseem Z. Ansari<sup>be</sup> and Charles Vinson<sup>\*c</sup>

Received 28th July 2011, Accepted 14th November 2011

DOI: 10.1039/c1lc20698b

We synthesized customized double-stranded DNA microarrays including methyl-5-cytosine at CpG dinucleotides and produced all 163,555 possible 8-mers (un-, hemi-, and di-methylated) to gain insight into how methylation affects transcription factor binding. An antibody to methyl-5-cytidine showed greater binding to the methylated DNA, demonstrating efficient incorporation of methyl-5-cytosine into the synthesized DNA. In contrast, binding of the transcription factor CREB was inhibited by CpG methylation. This platform represents a powerful new technology to evaluate the effect of DNA methylation on protein binding in any sequence context.

### Introduction

The binding of transcription factors (TFs) to specific DNA sequences is critical for the precise control of gene expression.<sup>1,2</sup> Determining the DNA binding specificity of TFs is an evolving process. Twenty years ago, PCR based selection procedures were developed that identified a small set of DNA sequences that were best bound by a TF.<sup>3</sup> More recently, investigators have used microarrays containing hundreds of thousands of double stranded DNA sequences to determine the binding preference of a TF to a multitude of DNA sequences, leading to the detection of many suboptimal DNA sequences that may be biologically important.<sup>4,5</sup>

In mammalian systems, the epigenetic differences between cell types and pathological states can be mediated by differences in methylation of the cytosine that occurs in CpG dinucleotides.<sup>6</sup> Methyl-5-cytosine, described by some as the fifth DNA base, is an epigenetic mark that regulates both gene activation and suppression.<sup>7</sup> However, the effect of CpG methylation on the binding affinity of TFs for all DNA sequences is unclear.

To determine how CpG methylation affects the DNA binding of TFs to multiple DNA sequences, we fabricated DNA microarrays containing methyl-5-cytosine only when it occurred in the CpG dinucleotide. These microarrays contain 163,555 double-stranded features which are all possible 8-mers including all

65,536 (4<sup>8</sup>) unmethylated 8-mers and 98,019 hemi-methylated and di-methylated versions of each 8-mer that contains one or more CpG dinucleotides.<sup>4,8</sup>

### Materials and methods

#### Microarray synthesis

SuperClean glass slides (Arrayit) were incubated in buffered silane (1.5% N-(3-triethoxysilylpropyl)-4-hydroxybutyramide (Gelest), 95% ethanol, 0.1% glacial acetic acid) with shaking for 4 h, according to current protocols.<sup>8</sup> After silane coating, slides were rinsed in wash solution (95% ethanol, 0.1% glacial acetic acid) with shaking for 20 min. Silanized slides were dried at 120 °C for 1 h and then baked in a vacuum oven at 120 °C for 12 h. Silanized slides were stored desiccated at room temperature until use for synthesis. DNA was synthesized on the silanized slides using MAS units connected to Expedite DNA synthesizers (Applied Biosystems). Two grams of photolabile NPPOC methyl-5-cytosine (Sigma-Aldrich) were used in conjunction with the other four photolabile phosphoramidites (NPPOC adenosine, NPPOC cytosine, NPPOC guanine, NPPOC thymine) (Nimblegen Systems). All phosphoramidites were diluted to 0.1M in acetonitrile and used with standard DNA synthesis-grade reagents (Sigma-Aldrich, Fisher Scientific, Nimblegen Systems) to synthesize the microarrays using standard protocols.<sup>9</sup> After synthesis, the base-protecting groups were removed by immersing arrays in a 1 : 1 v/v solution of ethylenediamine/ethanol (Sigma-Aldrich) for 2 h. The arrays were rinsed in water, dried, and stored desiccated at room temperature until use.

#### Protein purification

The CREB leucine zipper (B-ZIP) DNA binding domain was expressed in the *E. coli* BL21 (LysE) strain and purified as described previously.<sup>10</sup> The 9-amino acid HA epitope

<sup>a</sup>McArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health, Madison, WI, 53706, USA

<sup>b</sup>Department of Biochemistry, University of Wisconsin, Madison, WI, 53706, USA

<sup>c</sup>Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, USA. E-mail: vinsonc@mail.nih.gov; Fax: +301-494-8419; Tel: +301-496-8753

<sup>d</sup>Department of Physics, University of Maryland, College Park, MD, 20742, USA

<sup>e</sup>The Genome Center, University of Wisconsin, Madison, WI, 53706, USA

<sup>†</sup>C.L.W. has a financial interest in Illumavista Biosciences LLC.

(YPYDVPDYA) was added to the N-terminus of the B-ZIP domain for immuno-detection. HPLC using Vydac C18 reverse phase column was used for final protein purification, where a linear gradient from 0–100% acetonitrile containing 0.1% trifluoroacetic acid over 45 min with a flow rate of 1 ml min<sup>-1</sup> was used to elute the proteins.

### Electrophoretic mobility shift assay (EMSA)

The 28-mer oligonucleotides (Sigma-Aldrich) were PAGE purified. Top strand oligonucleotide was end-labeled with  $\gamma$ -<sup>32</sup>P ATP using T4 phage polynucleotide kinase. The labeled oligonucleotide was purified using a G-50 column (GE Healthcare) according to manufacturer instructions and annealed to the unlabeled bottom strand oligonucleotide. CREB was mixed with 7 pM <sup>32</sup>P-radiolabeled double-stranded oligonucleotides in the gel shift buffer (0.5 mg ml<sup>-1</sup> BSA, 10% glycerol, 2.5 mM DTT, 12.5 mM K<sub>2</sub>HPO<sub>4</sub>-KH<sub>2</sub>PO<sub>4</sub>, pH 7.4, 0.25 mM EDTA, 10 ng  $\mu$ l<sup>-1</sup> poly(dI-dC)). The final volume of the reaction was adjusted to 20  $\mu$ l, and incubated at 37 °C for 10 min, followed by cooling at room temperature for 5 min. 10  $\mu$ l samples were resolved on 7.5% PAGE at 150 V for 1.5 h in the 1x TBE buffer (25 mM Tris-boric acid, 0.5 mM EDTA). Sequences of oligonucleotides used for EMSA experiments were:

Top: 5'-GTCAGTCAGATGACGTCATATCGGTCAG-3'  
 Bottom: 5'-CTGACCGATATGACGTCATCTGACTGAC-3'  
 Underlined nucleotides are the consensus CREB binding site.

### Microarray experiments

**Methyl-5-cytidine antibody binding.** Arrays were blocked with 2.5% non-fat dried milk for 1.5 h prior to protein incubation. Methyl-5-cytidine antibody (Abcam ab10805) was diluted 1 : 1000 and mixed with a 1 : 2000 dilution of a fluorescently-labeled Cyanine 5 secondary antibody (Abcam) in mAb buffer (50 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM MgCl<sub>2</sub>, 0.5 mM EDTA). The antibody mixture was added to the hybridization chamber on the array and incubated for 1 h at room temperature with constant rotation. The arrays were washed with non-stringent wash buffer (6X SSPE pH 7.5, 0.01% Tween-20), dried, and visualized using an Axon 4000B 5  $\mu$ m scanner (Molecular Devices). Data was viewed using GenePix<sup>TM</sup> Pro 6.0 microarray analysis software (Molecular Devices).

**CREB binding.** Arrays were blocked with 2.5% non-fat dried milk for 1.5 h prior to protein incubation. CREB was heated at 65 °C for 15 min in CD buffer (12.5 mM phosphate buffer, pH 7.4, 0.25 mM EDTA, 1 mM DTT), and cooled at room temperature for 5 min. The protein was then diluted to a final concentration of 50 nM, and mixed with a directly-labeled fluorescent antibody to the HA tag of the protein, HA-TRITC (Sigma-Aldrich), in the binding buffer (10% glycerol, 12.5 mM phosphate buffer, pH 7.4, 0.25 mM EDTA, 2.5 mM DTT, 150 mM KCl, 0.5 mg ml<sup>-1</sup> BSA, 2 ng  $\mu$ l<sup>-1</sup> HA-TRITC, 0.25% milk, 0.005% Tween-20). The protein-antibody mixture was added to the hybridization chamber on the array and incubated for 1 h at room temperature with constant rotation. Subsequently, the arrays were washed first with the binding buffer, and then with the non-stringent wash buffer, dried, and visualized using an

Axon GenePix 4000B 5  $\mu$ m scanner. Data was viewed using the GenePix<sup>TM</sup> Pro 6.0 microarray analysis software.

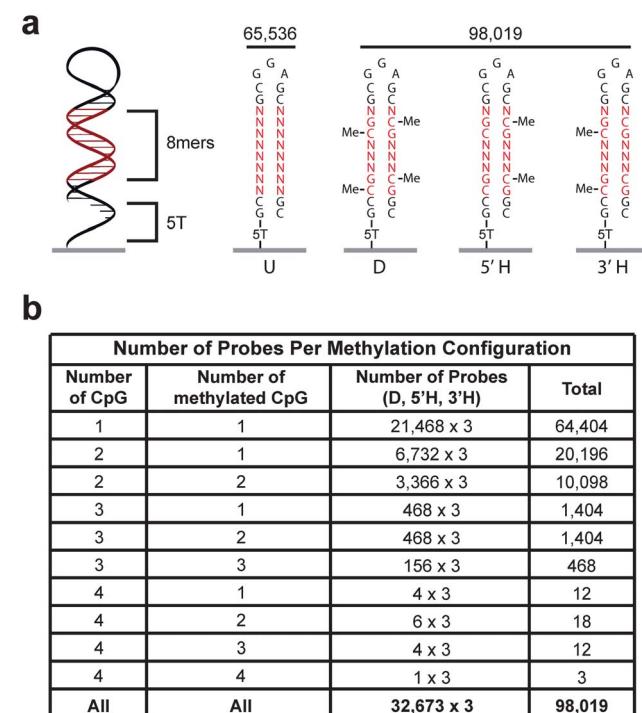
### Data normalization

For each array, local mean normalization<sup>11</sup> was used to ensure the intensity was evenly distributed throughout each sector of the microarray surface. Complementary unmethylated sequences were averaged together. Corresponding 3' (3'H) and 5' (5'H) hemi-methylated sequences were averaged together given that the methyl-5-cytidine antibody is unaffected by unmethylated sequence context (Fig. 2) and the CREB homodimer binds equivalently to 5'H or 3'H sequences (Fig. 3d). Non-specific binding for CREB was accounted for by subtracting aggregate single-stranded DNA intensities. Motif searching was conducted using MEME/MAST System Motif Discovery and Search (<http://meme.sdsc.edu/meme/intro.html>) on the 100 highest intensity/affinity sequences.<sup>12</sup>

## Results and discussion

### Fabrication of methyl-5-cytosine microarray

Microarrays were synthesized using a Maskless Array Synthesizer (NimbleGen Systems, Madison, WI).<sup>9</sup> 3'-nitrophenylpropyloxycarbonyl (NPPOC) methyl-5-cytosine was



**Fig. 1** Double-stranded methylated DNA microarray design. (a) Schematic of the 32 nucleotide double-stranded DNA probe. All 163,555 possible 8-mers including 98,019 CpG methylation permutations are on the array. For each 8-mer that contains at least one CpG dinucleotide, unmethylated (U), di-methylated (D, both 5' and 3' strands), and both hemi-methylated (3' H, or 5' H) configurations are synthesized on the array at unique features. (b) Table showing the number of probes on the array that contain each methylation configuration. CpG indicates the number of 5'-CpG-3' dinucleotides in the probe.

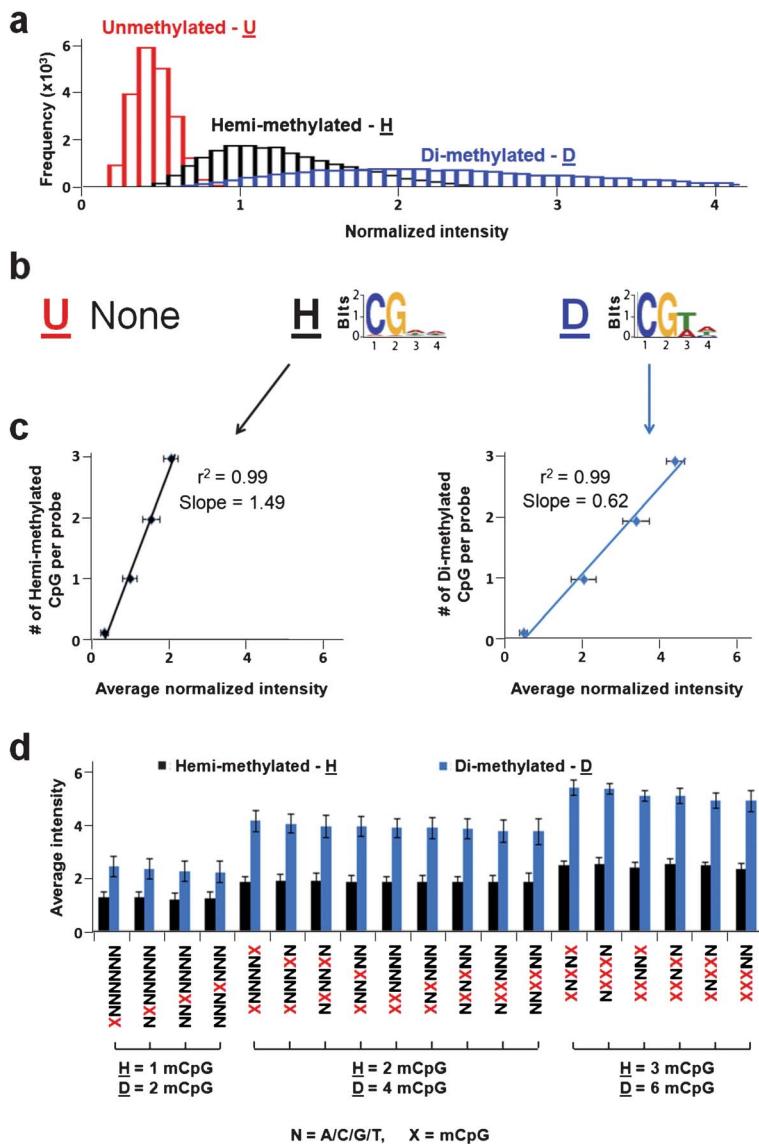
purchased from Sigma-Aldrich. Homopolymer ( $T_5$ ) linkers were covalently attached to monohydroxylsilane glass slides and oligonucleotides were synthesized on the homopolymers to create a high-density oligonucleotide microarray. Each feature on the microarray represented a different sequence and was composed of approximately  $10^6$  identical synthesized probes. The synthesized DNA was 32 nucleotides long and designed to become double stranded. Hairpins were induced by incubation

with 7M urea in phosphate buffered saline (PBS) for 30 min at 65 °C and then in PBS for 15 min at 65 °C<sup>4,8</sup> (Fig. 1).

### Experimental validation

To evaluate whether the methyl-5-cytosine was incorporated into DNA during the synthesis, we examined if an antibody to methyl-5-cytidine preferentially bound to the methylated DNA.

## Methyl Cytosine Antibody

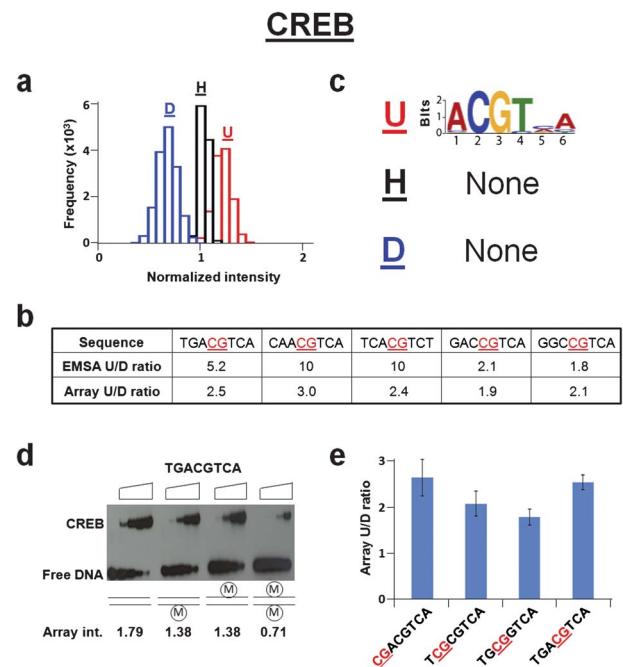


**Fig. 2** Effects of CpG methylation on the DNA binding of the methyl-5-cytidine antibody. (a) Histograms showing the binding intensity of the methyl-5-cytidine antibody to CpG containing DNA sequences that are unmethylated (red, U), hemi-methylated (black, H), and di-methylated (blue, D). Frequency is defined as the number of sequences included in the normalized intensity range spanned by each bar. Each bar spans 0.10 normalized intensity units. (b) Logos generated from the 100 probes representing the highest binding affinity of the methyl-5-cytidine antibody to unmethylated (left, U), hemi-methylated (middle, H), or di-methylated (right, D) DNA. Non-CpG containing sequences were included in all three motif analyses. (c) Binding intensity of the methyl-5-cytidine antibody to DNA containing 1, 2, or 3 hemi-methylated CpGs (left) or di-methylated CpGs (right). The intensities of all DNA probes with the specified number of methyl-5-cytosines were averaged and plotted. Error bars represent one standard deviation. Only one sequence contained four hemi- or di-methylated CpGs and was therefore excluded from the analysis. (d) Bar graph of the average intensities of all configurations of methylated CpG dinucleotides (mCpG) within an 8 bp double-stranded probe, where N = A/C/G/T and X = mCpG. Plotted intensities are an average of all probes matching the specified sequence. Error bars represent one standard deviation.

This antibody was raised against methyl-5-cytosine in an oligonucleotide context and discriminates against both unmodified cytidine and hydroxy-5-methylcytidine.<sup>13</sup> Binding of the antibody on the methyl-5-cytosine microarray was visualized using a fluorescently-labeled secondary antibody. Fig. 2a shows that the methyl-5-cytidine antibody preferentially bound to DNA containing di-methylated CpGs, followed by hemi-methylated CpGs with the least binding occurring to CpG containing DNA that is unmethylated. The antibody binding to the unmethylated DNA probes was low which we suggest represents non-specific binding. Position weight matrices (PWMs), shown as logos, were created from the best bound probes on the microarray using the MEME motif-finding algorithm. We produced a PWM for the antibody binding to three kinds of DNA sequences: unmethylated, hemi-methylated and di-methylated DNA 8-mers. No PWM was observed for the unmethylated probes. When the hemi-methylated or di-methylated probes were included with the unmethylated probes, the CpG dinucleotide was produced, thus confirming the specificity of the antibody (Fig. 2b). The binding intensity to methylated probes showed a range of intensities. We evaluated if the stronger binding represented binding to probes with a larger number of methyl CpGs. We observed a linear relationship between the binding intensities and the number of methylated cytosines per probe for both the hemi-methylated and di-methylated probes (Fig. 2c). Specifically, the antibody exhibited an increase in average normalized intensity of 0.67 per methylated cytosine for hemi-methylated DNA (inverse of the slope) and 0.81 for di-methylated DNA (inverse of the slope divided by two due to symmetrically methylated cytosines per di-methylated CpG). An analysis of the placement of the methyl CpG within the 8-mer showed no preference of the antibody to a particular CpG location (Fig. 2d). In addition, the equivalent binding of the antibody regardless of whether the methyl-5-cytosine was incorporated early or late in the synthesis (Fig. 2d) confirmed that methyl-5-cytosine incorporation does not compromise subsequent synthesis of the DNA probe. Finally, we observed that the methyl-5-cytidine antibody recognizes a single methyl-5-cytosine epitope<sup>14</sup> without steric occlusion between adjacent methylated CpG sites.

### Effect of cytosine methylation on CREB specificity

We next used these microarrays to examine the binding of the CREB DNA binding domain, a protein where CpG methylation is known to inhibit DNA binding.<sup>7,15</sup> By visualizing binding to the microarray using a fluorescently labeled antibody to an HA epitope tag on CREB, we observed that CREB preferentially binds to the unmethylated DNA compared to the di-methylated DNA with intermediate binding to the hemi-methylated DNA. This suggests that CpG methylation globally inhibits CREB binding (Fig. 3a). These data also confirmed a previous report examining specific DNA sequences<sup>7</sup> (Fig. 3b). The unmethylated consensus CREB binding site termed the CRE (TGACGTCA) was the best-bound sequence on the array. When the top 100 sequences were examined, a CRE-like logo was observed (Fig. 3c). However, when hemi- and di-methylated DNA were examined, no logo was observed suggesting that CpG methylation abolishes sequence specific DNA binding. To verify the microarray data, we performed electrophoretic mobility shift



**Fig. 3** Effects of CpG methylation on the binding of CREB DNA binding domain. (a) Histograms showing the binding intensity of CREB to CpG containing DNA sequences that are unmethylated (red, U), hemi-methylated (black, H), and di-methylated (blue, D). Frequency is defined as the number of sequences included in the normalized intensity range spanned by each bar. Each bar spans 0.10 normalized intensity units. (b) Table displaying CREB ratios of unmethylated (U) to di-methylated (D) binding affinities from EMSA and microarray intensities for five DNA sequences. The site of DNA methylation is underlined and in red. (c) Logos generated from the 100 probes representing the highest affinity binding to unmethylated (top, U), hemi-methylated (middle, H), or di-methylated (bottom, D) DNA. The unmethylated logos are generated from the full 8-mer (65,536) unmethylated probe set. (d) EMSA for CREB binding to a CREB response element (CRE, 5'-TGACGTCA-3') that is unmethylated, hemi-methylated or di-methylated. Triangles on top of the gel denote increasing CREB concentrations (0, 5, 15, 50, 150 nM dimer). Normalized microarray intensities corresponding to each methylation state of the CRE are shown below. (e) Bar graph displaying the effect of the position of the methylated CpG in the CREB consensus sequence on the ratio of unmethylated (U) to di-methylated (D) microarray intensities. The site of DNA methylation is underlined and in red.

assays (EMSA) using CREB. The results validated the microarray data that CREB preferentially binds to unmethylated DNA as compared to di-methylated DNA and binds hemi-methylated DNA with intermediate affinity (Fig. 3d). When we examined if the position of the methylated CpG in CREB binding sites affects CREB binding, we observed that methylation is universally detrimental to binding (Fig. 3e).

### Conclusions

This technology will allow for a global examination of the effect of CpG methylation on DNA binding of both naturally occurring DNA binding proteins and synthetic molecules with potentially important therapeutic properties. Though we have only examined the methylation of cytosine in the context of the CpG dinucleotide, this technology can also examine how

cytosine methylation in additional DNA sequence contexts, as occurs in embryonic stem cells and plants,<sup>16,17</sup> modulates protein binding.

## Acknowledgements

This research was supported by the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health. AZA was supported by grants from the National Institutes of Health (GM069420), March of Dimes, U.S. Department of Agriculture, Innovation and Economic Development Research Program, and Vilas Associate and Shaw scholar awards. CLW was supported by a National Institutes of Health/National Library of Medicine predoctoral fellowship (T15LM007359).

## Notes and references

- 1 K. E. Hauschild, C. D. Carlson, L. J. Donato, R. Moretti and A. Z. Ansari, *Wiley Encyclopedia of Chemical Biology*, ed. Begley T (John Wiley & Sons, Inc, New York), 2008, 4, 566–584.
- 2 J. E. Darnell Jr, *Nat. Rev. Cancer*, 2002, **2**, 740–749.
- 3 C. Tuerk and L. Gold, *Science*, 1990, **249**, 505–510.
- 4 C. L. Warren, N. C. Kratochvil, K. E. Hauschild, S. Foister, M. L. Brezinski, P. B. Dervan, G. N. Phillips Jr. and A. Z. Ansari, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 867–872.
- 5 C. D. Carlson, C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina and A. Z. Ansari, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4544–4549.
- 6 K. D. Robertson, *Nat. Rev. Genet.*, 2005, **6**, 597–610.
- 7 V. Rishi, T. Potter, J. Laudeman, R. Reinhart, T. Silvers, M. Selby, T. Stevenson, P. Krosky, A. G. Stephen, A. Acharya, J. Moll, W. J. Oh, D. Scudiero, R. H. Shoemaker and C. Vinson, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20311–20316.
- 8 M. S. Ozers, C. L. Warren and A. Z. Ansari, *Methods Mol. Biol.*, 2009, **544**, 637–653.
- 9 S. Singh-Gasson, R. D. Green, Y. Yue, C. Nelson, F. Blattner, M. R. Sussman and F. Cerrina, *Nat. Biotechnol.*, 1999, **17**, 974–978.
- 10 S. Ahn, M. Olive, S. Aggarwal, D. Krylov, D. D. Ginty and C. Vinson, *Mol. Cell. Biol.*, 1998, **18**, 967–977.
- 11 C. Colantuoni, G. Henry, S. Zeger and J. Pevsner, *Biotechniques*, 2002, **32**, 1316–1320.
- 12 T. L. Bailey and C. Elkan, *Proc. of the International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 1994, **2**, 28–36.
- 13 <http://www.abcam.com/5-Methyl-Cytidine-antibody-33D3-ChIP-Grade-ab10805.html>.
- 14 M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam and D. Schübler, *Nat. Genet.*, 2005, **37**, 853–862.
- 15 S. M. Iguchi-Ariga and W. Schaffner, *Genes Dev.*, 1989, **3**, 612–619.
- 16 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 and J. R. Ecker, *Nature*, 2009, **462**, 315–322.
- 17 R. Lister, R. C. O’Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar and J. R. Ecker, *Cell*, 2008, **133**, 523–536.