DNA G-quadruplex structures mold the DNA methylome

Shi-Qing Mao1, Avazeh T. Ghanbarian1,4, Jochen Spiegel1, Sergio Martínez Cuesta1,2, Dario Beraldí1,5, Marco Di Antonio2, Giovanni Marsico1, Robert Hänsel-Hertsch1, David Tannahill1 and Shankar Balasubramanian1,2,3*

Control of DNA methylation level is critical for gene regulation, and the factors that govern hypomethylation at CpG islands (CGIs) are still being uncovered. Here, we provide evidence that G-quadruplex (G4) DNA secondary structures are genomic features that influence methylation at CGIs. We show that the presence of G4 structure is tightly associated with CGI hypomethylation in the human genome. Surprisingly, we find that these G4 sites are enriched for DNA methyltransferase 1 (DNMT1) occupancy, which is consistent with our biophysical observations that DNMT1 exhibits higher binding affinity for G4s as compared to duplex, hemi-methylated, or single-stranded DNA. The biochemical assays also show that the G4 structure itself, rather than sequence, inhibits DNMT1 enzymatic activity. Based on these data, we propose that G4 formation sequesters DNMT1 thereby protecting certain CGIs from methylation and inhibiting local methylation.

Methylation of cytosine at C-5 is a key DNA modification in development and disease1,2. In mammals, cytosine methylation occurs predominantly at CpG dinucleotides and is installed and maintained by three DNA methyltransferase enzymes (DNMT1, DNMT3A and DNMT3B) that are essential for development3–5. CpGs occur less frequently than expected in the mammalian genome and show a bimodal distribution with respect to methylation6,7. Sparsely distributed CpGs (~90%), found in genic and intergenic regions, tend to be highly methylated, while CpGs found in dense GC-rich (guanine-cytosine rich) regions, so-called CpG islands (CGIs), are largely depleted of methylation and are prevalent at the promoters of house-keeping and developmental genes8,9. Outside of embryonic development, gross methylation patterns are generally stable across different tissues10,11. Nonetheless, during key cellular events, methylation can be dynamic at specific loci to modulate gene expression, such as de novo methylation of some promoter CGI hypomethylation. In mouse embryonic stem cells, the CXXC finger protein 1, Cfp1, is believed to promote CGI hypomethylation through binding unmethylated CpG and recruitment of H3K4 methyltransferases to promote H3K4me312,13. However, Cfp1 binding and/or H3K4me3 are not required for the “protection” of CGI from DNA methylation since Cfp1 knockout results in a dramatic loss of H3K4me3 at CGIs without increasing DNA methylation14. This suggests that Cfp1 binding and/or H3K4me3 are not required to prevent CGIs from DNA methylation and thus there may be other factors that are fundamental to impart the hypomethylated state.

Alternative DNA secondary structures, known as G-quadruplexes (G4s), are found within certain G-rich sequences and arise through the self-association of guanine bases to form stacked tetraads (Fig. 1a)15,16. G4s are increasingly being recognized as important features in the genome and over 700,000 G4s have been biophysically mapped in purified human genomic DNA by high-throughput sequencing17. G4 structures have been observed in human using immunofluorescence with a G4-specific antibody (BG4)18 and linked with transcriptional regulation, and they are enriched in gene promoters including many oncogenes19,20. Recently, G4-chromatin immunoprecipitation sequencing (G4-ChIP-seq) has been developed to map G4 structures in human chromatin21,22. Corroborating a link with transcription, the majority of G4-ChIP-seq sites were found predominantly in regulatory, nucleosome-depleted chromatin, particularly in gene promoters23,24. As both G4s and hypomethylated CGIs are associated with actively transcribed genes25,26, this raises the question of whether there is an interplay between G4 formation and DNA methylation.

Herein we present evidence that most G4 structures, as detected by G4-ChIP-seq, are formed in regions comprising unmethylated CGIs in the human genome. We also uncover a striking co-localization of G4 structures and DNMT1 docked at CGIs, and we demonstrate that DNMT1 methylation activity is inhibited by DNA G4 structures. Our data suggest a mechanism for the “protection” of CGIs from methylation by G-quadruplex structures that locally sequester and inhibit local methylation.

Results

G4 structures in active chromatin are found within hypomethylated CGIs. To explore any potential relationship between G4

1Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, UK. 2Department of Chemistry, University of Cambridge, Cambridge, UK. 3School of Clinical Medicine, University of Cambridge, Cambridge, UK. 4Present address: European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, UK. 5Present address: Institute of Cancer Sciences, University of Glasgow, Glasgow, UK. *e-mail: sb10031@cam.ac.uk
Fig. 1 | G4 formation is associated with hypomethylation at CGIs. a, A G-tetrad stabilized by Hoogsteen hydrogen bonding and a central monovalent cation (left). Schematic representations of a three-tetrad G4 structure (right). b, Venn diagram illustrating the overlap of G4 structure formation (BG4 peaks) and CGIs. c, Violin plot showing size distribution of BG4 peaks and CGIs. d, Count of BG4 peaks overlapping a CGI. e, Box-and-whisker plot showing the average methylation for BG4 peaks (n = 8,210), DHSs (n = 142,115), and CGIs (n = 27,073). The center line represents the median value separating the upper and lower quartiles in the box, whiskers indicate 1.5× interquartile range (IQR), and points are actual values of outliers. Note that the methylation level at CpG sites with less than 5× coverage is considered unreliable and discarded. f, Histogram showing the distribution of BG4 peaks and CGIs relative to the percentage of GC. g, Histogram showing the distribution of BG4 peaks and CGIs relative to the percentage of CpGs per 100 bp. h, Box-and-whisker plot showing the methylation levels for BG4 peaks and CGIs at different CpG densities. Note that by definition there are no CGIs at a CpG density <5 CpGs per 100 bp and that at >20 CpGs per 100 bp there are few CGIs (1) and BG4 (36) peaks to consider. The number of CGI regions and BG4 peaks in each category are presented on the top of the plot. i, An IGV (Integrative Genomics Viewer) screen shot illustrating the coincidence of BG4 peaks (blue) with hypomethylated promoter CGIs (green) and DHSs (orange) for a representative genome region from Chr 7. Shown are the normalized signals. Whole genome bisulfite sequence tracks are in black (top). RefGene tracks are in grey (bottom). Source data for graphs shown in panels d, e, and h are available online.
structures in chromatin and methylation levels, we employed human K562 chronic myelogenous leukemia cells in which methylation has been comprehensively characterized at single base resolution using whole genome bisulfite sequencing (WGBS) by the ENCODE project\(^9\). We generated a genome-wide dataset for G4 structures by G4 ChIP-seq\(^11\) using the G4-specific antibody BG4\(^{48}\) and compared the BG4 peak overlap with CGIs\(^8\). Strikingly, we found that the majority of BG4 peaks (79\%, 7,111/8,952) overlapped with a CGI (covering 23\% of all CGIs) (Fig. 1b). The majority of CpG island regions span 200 to 1,000 base pair (bp) (median/mean, 569/775 bp), while BG4 peak regions span 100 to 400 bp (median/mean, 205/226 bp) (Fig. 1c).

Then 83\% (5,935/7,111) of these CGIs overlap with one BG4 peak (Fig. 1d). Furthermore, when the level of methylation at BG4 peaks was considered, we noted that there was a dramatic absence of methylation at BG4 peaks (mean 1\%, median 0.5\%), compared with average genome methylation (28.4\%) (Fig. 1e). To rule out any effect of the cytosine methylation state on the ability of the BG4 antibody to recognize a G4 structure, an enzyme-linked immunosorbent assay (ELISA) was used to show that BG4 can bind to G4-structured DNA with equal affinity irrespective of the presence of cytosine methylation (Supplementary Fig. 1). DNase I hypersensitive sites, which mark open chromatin, are also mainly hypomethylated (mean 11\%, median 2.5\%, Fig. 1e). Confirming our previous observations\(^27\), the majority of BG4 peaks are found in open chromatin (DHSs, 97\%, 8,655/8,952), and it is notable that these sites have the lowest methylation levels (Fig. 1e). Overall CGI methylation (mean 27\%, median 8\%, Fig. 1e) shows a broader distribution than BG4 regions, since some CGIs are associated with active hypomethylated promoters while others are associated with inactive genes or gene bodies, and thus are more heavily methylated\(^27\). This prominent association between BG4 peaks, hypomethylation, and particular CGIs is suggestive of a functional link between G4 secondary structures and the establishment and/or maintenance of low methylation status at these CGIs in active chromatin.

It has recently been concluded from work in mouse embryonic stem cells that both high CpG density and high GC content are required to establish the hypomethylated state at CGIs\(^15\). It is therefore notable that BG4 peaks have a similar level of GC content to CGIs (Fig. 1f) with most (79\%) being located in regions of CpG density comparable to that seen in CGIs (Fig. 1g). It has been suggested that CpG density alone is only a minor determinant of the unmethylated state, as dense CpG sequences embedded within an AT-rich (adenine-thymine rich) context are invariably highly methylated when inserted into the mouse genome\(^15\). Indeed, when we compare the average methylation of CGIs (Fig. 1b) to that of BG4 peaks at different CpG dinucleotide densities, it is noteworthy that across a wide range of CpG densities the BG4 peak regions are always largely devoid of methylation (Fig. 1h). We also confirmed these observations using an alternative CGI definition set generated by CpGCluster algorithm\(^36\) (Supplementary Fig. 2a,b). Furthermore, when methylation at CGIs is considered with respect to the presence or absence of a BG4 peak, it is notable that there is almost a total lack of methylation at CGIs with BG4 compared with those without (Supplementary Fig. 2c). This strongly suggests that CGIs associated with the physical presence of a G4 structure generally have particularly low methylation.

To explore low methylation in different CGI contexts, we calculated methylation levels relative to BG4 presence in CGIs containing (i) no promoter or DHS site, (ii) a promoter alone, (iii) a DHS site alone, and (iv) both a promoter and DHS site. It is apparent that CGIs containing a BG4 peak always have lower methylation in open (DHS+) or closed (DHS−) chromatin, or in the presence or absence of a promoter (Supplementary Fig. 2d). CGIs with a DHS site and promoter but without a BG4 peak (4,500 sites) have higher methylation (mean 2\%, median 2\%) than those CGIs (5,567 sites) with a BG4 peak plus promoter and DHSs (mean 1\%, median 0.5\%) (Supplementary Fig. 2d, right two panels). The lowest observed methylation states are found therefore at sites carrying a G4 structure, suggesting that the physical presence of a G4 structure within CGI is an important feature with respect to the hypomethylation state. This is illustrated in Fig. 1i which shows the coincidence of BG4 peaks with hypomethylated promoter CGIs for a representative genome region.

In earlier work, we found that treatment of human epidermal keratinocytes (HaCaT) cells with the HDAC (histone deacetylase) inhibitor entinostat led to increased BG4 binding signal primarily located in open chromatin promoter regions\(^3\). We therefore generated WGBS datasets to examine DNA methylation changes with respect to BG4 signal. Consistent with our observation in K562 cells, BG4 peaks in HaCaT cells have lower methylation compared with open chromatin and CGI regions (Supplementary Fig. 2e).

In open–chromatin promoter CGI regions, 307 had a significant effect in BG4 signal (BG4 increase, >1.5-fold change in signal and FDR (false discovery rate) <0.05, see Methods). No change in BG4 signal was seen in 3,261 CGI promoter regions before and after treatment (BG4 constant), or for 1,504 G-rich CGI promoter regions that do not have a BG4 peak (BG4 negative) but have the potential to form a G4 in vitro\(^3\). Despite open–chromatin promoter CGI regions already being predisposed to low methylation, we see a statistically significant additional drop in methylation levels at CGIs where BG4 peak size increases after HDAC inhibition (Supplementary Fig. 2f). Overall, these data support the idea that formation of G-quadruplex structures in CGIs is linked to lower methylation.

DNMT1 is sequestered at G4 structures associated with low methylation. Given that regions where G4 structures marked by BG4 peaks are generally observed to be hypomethylated, we considered that the DNA methyltransferases might have some form of physical interaction with G4 structures in the chromatin context. We focused on DNMT1 since DNMT1 knockout is lethal, causing global DNA methylation loss in all dividing somatic cells and human embryonic stem cells (ESCs)\(^3,5,38,39\), whereas DNMT3A/B knockouts mainly affect non-CpG methylation in human ESCs\(^3\). When we considered the distribution of DNMT1 binding sites in K562 cells, downloaded from ENCODE\(^33\) (516,483 peaks in total across both biological replicates), we found that 52\% (4,611/8,952, Monte Carlo simulation’s P value 1.25×10\(^{-10}\) of the G4 structures mapped by G4-ChIP (BG4 peaks) overlapped with at least one DNMT1 binding site. Of the remaining 4,341 BG4 peaks, 4,003 were within 1 kilobase (kb) of a DNMT1 binding site. The proximity of BG4 peaks to DNMT1 recruitment sites is illustrated graphically in Fig. 2a for a representative genome region. Intriguingly, when the distribution of DNMT1 binding is plotted relative to high, intermediate, or low methylated CGIs, we observe a prominent enrichment of DNMT1 binding at lowly methylated CGIs, which overlaps with those regions with the highest BG4 peak density as well as DHS sites (Fig. 2b).

A similar profile is also seen using an alternative CGI definition set generated by CpGCluster\(^36\) (Supplementary Fig. 3). The observation that DNMT1 enrichment at G4 regions that lack methylation is, at first glance, somewhat unexpected and counter-intuitive, given that DNMT1 installs methylation. We therefore considered the possibility of a mechanism whereby DNMT1 protein is sequestered at these sites in active chromatin but prevented from methylating CpGs in that locality.

DNMT1 selectively binds to and is inhibited by G4 structures. To address whether DNMT1 binds G4 structures directly, we carried out biophysical measurements using an ELISA to measure the binding of recombinant human FLAG-tagged full-length DNMT1 protein to immobilized target DNA structures (see Methods). Biotinylated single-stranded oligonucleotides of a sequence based
we found that DNMT1 binds to all three G4 structures with low thermal melting spectroscopic analysis (Supplementary Fig. 4a–f). The relative high binding affinity and selectivity of DNMT1 for G4 structures than known DNMT1 substrates such as a hemi-methylated duplex DNA (BCL2, 107 nM, Fig. 3d) or poly(dI-dC) (K_d = 2 nM, Fig. 3f). We indeed found that each of three G4 structures resulted in significant inhibition of DNMT1 methyltransferase activity whereas the mutated control oligonucleotides did not (Fig. 3h–i).

Gratifyingly, the potency of inhibition by each G4 was related to the binding affinity for DNMT1, as determined by ELISA with G4-structured oligonucleotides or mutated non-G4 controls, where the presence or absence of G4 structure had been confirmed by circular dichroism spectroscopy (Supplementary Fig. 4g–i). We found that DNMT1 binds to all three G4 structures with low thermal melting spectroscopic analysis (Supplementary Fig. 4a–f). We indeed found that each of three G4 structures resulted in significant inhibition of DNMT1 methyltransferase activity whereas the mutated control oligonucleotides did not (Fig. 3h–i).

DNMT1 is recruited to BG4 peaks associated with low methylation. a An IGV screen shot showing the coincidence (blue-masked) of BG4 peaks (blue) with DNMT1 ChIP-seq peaks (red) and CGIs (green) at the hypomethylated region from Chr 6. Orange-masked regions are hypermethylated and enriched with DNMT1 presence, but not BG4 signal. Whole genome bisulfite sequence tracks are in black (top). b Binding profile of DNMT1 shown across CGIs with low (less than 20%, n = 16,523), intermediate (between 20% and 80%, n = 6,042), and high (more than 80%, n = 4,266) methylation. The y-axis shows the number of reads in the ChIP normalized to 1 of the sequencing depth (also known as the RPGC, further details can be found in Methods). Replicates 1 and 2 are indicated in red and blue, respectively. Above each plot is a heat map showing the enrichment of BG4 peaks and DHSs across the respective regions. The heat maps show the RPGC of active chromatin marks (DHSs) and BG4 peaks on these three classes of CGIs.
concentration. C-rich oligonucleotides complementary to the G4 sequences (BCL2-CCC, KIT2-CCC, and MYC-CCC) or corresponding duplex DNA also had no effect on DNMT1 activity (Supplementary Fig. 4j–l). We also tested G4 oligonucleotides that were able to fold into a G4 structure but carried a reduced number of CpGs (BCL2, KIT) or had a number of artificially introduced CpGs (MYC). In all cases, changes in the number of CpG sites only had minor effects on DNMT1 inhibition (Supplementary Fig. 4j–l). Taken together, these results indicate a novel and unexpected feature of G4 structures as potential genomic regions that promote the unmethylated state through recruitment and inhibition of DNMT1 activity.

Recruitment of DNMT to G4 structures shapes the methylome. The above data suggest that there is a striking lack of methylation (Fig. 1e,h and Supplementary Fig. 2a–d) in chromatin regions where G4 structure formation is observed. To rigorously question whether this observation was related to the detectable formation of a G4 structure in chromatin (that is, a BG4 peak), or merely the G-rich sequences per se with potential to form a G4 structure, the methylation profile for BG4 peak regions was compared with those of G-rich sequences that can physically form a G4 structure in an in vitro sequencing assay (here called sequences with the potential to form G4s, Fig. 4a). As the majority of BG4 peaks (8,210) are

Fig. 3 | DNMT1 selectively binds and is inhibited by G4 structures. a–f, ELISA assays testing the binding of recombinant DNMT1 to G4 structure and control oligonucleotides. Binding curves for: a, BCL2 G4 and non-G4-forming control (BCL2-mut); b, KIT2 G4 and non-G4-forming control (KIT2-mut); c, MYC G4 and non-G4-forming control (MYC-mut); d, BCL2 duplex DNA; e, BCL2 hemi-methylated duplex DNA; f, (dI-dC)50. Absorbance was measured at 450 nm, a.u., arbitrary unit. Sequences of oligonucleotides are given below the graphs. g, Binding curve of BCL2 G4 in the presence of different concentrations of BCL2 duplex or poly(dI-dC). h–j, Relative methylation activity of recombinant DNMT1 in the presence of G4 structure and control oligonucleotides: h, BCL2 G4 and BCL2-mut; i, KIT2 G4 and KIT2-mut; j, MYC G4 and MYC-mut. Shown are mean ± s.d., n = 3 independent experiments in all plots but g (n = 2). Source data for the graphs are available online.
found in open chromatin, only sequences with potential to form G4s located in open chromatin (36,015) were considered. The mean and median lengths are 226/205 bp for BG4 peaks, and 383/285 bp for the latter. G-rich sequences with the potential to form a G4 are largely hypomethylated (12%), with methylation levels rising in the flanking regions (45%), whereas BG4 peaks have substantially lower methylation (down to 1%) with the flanking regions being largely hypomethylated (12%), with methylation levels rising in the flanking regions (45%), whereas BG4 peaks have substantially lower methylation (down to 1%) with the flanking regions being marked by a further dramatic loss of methylation due to the physical presence of a G4 structure, and also tin are depleted in methylation but carry a G4 structure, and also the surrounding flanking regions display higher than average methylation. This suggests that G4s may impart a previously unknown and important function in the establishment of epigenome.

We propose a model (Fig. 4b) in which G4 formation, together with transcription factor binding\(^\text{23,24}\), contributes to loss of methylation at key genomic loci by sequestering DNMT1, via G4 recognition, and locally inhibiting DNMT1 function at CpG islands. It is noteworthy that this mechanism resembles a recently proposed model for recruitment and inhibition of PRC2 complex by a RNA G-quadruplex present in the HOTAIR IncRNA\(^\text{48,49}\). This suggests that there may be other mechanisms for epigenetic regulation that operate by the sequestration and inhibition of epigenetic modifiers mediated by high-affinity interactions with nucleic acid secondary structures.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at [https://doi.org/10.1038/s41594-018-0131-8](https://doi.org/10.1038/s41594-018-0131-8).

Received: 19 December 2017; Accepted: 8 August 2018; Published online: 1 October 2018

**References**

1. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
2. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat. Rev. Cancer* **4**, 143–153 (2004).
3. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
4. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
5. Liao, J. et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).
6. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
7. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
8. Illingworth, R. S. & Bird, A. P. CpG islands – A rough guide. *FEBS Lett.* **583**, 1713–1720 (2009).
9. Deaton, A. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
10. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).
11. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662–673 (2002).
12. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
13. Long, H. K., King, H. W., Patient, R. K., Odom, D. T. & Klose, R. J. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res.* **44**, 6693–6706 (2016).
14. Lienert, F. et al. Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* **43**, 1091–1097 (2011).
15. Wachter, E. et al. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife* **3**, 1–16 (2014).
16. Krebs, A. R., Dessus-Babas, S., Burger, L. & Schübeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* **3**, e04094 (2014).
17. Takahashi, Y. et al. Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells. *Science* **356**, 503–508 (2017).
18. Quante, T. & Bird, A. Do short, frequent DNA sequence motifs mould the epigenome? *Nat. Rev. Mol. Cell Biol.* **17**, 257–362 (2016).
19. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).
20. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
21. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
22. Ooi, S. K. T. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
23. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. 16, 519–532 (2015).

24. Clouaire, T. et al. Cpf1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. 26, 1714–1728 (2012).

25. Thomson, J. P. et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082–1086 (2010).

26. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. Nat. Rev. Mol. Cell Biol. 18, 279–284 (2017).

27. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. 33, 1–7 (2015).

28. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182–186 (2013).

29. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 43, 8627–8637 (2015).

30. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. 13, 551–564 (2018).

31. Hänsel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. Nat. Genet. 48, 1267–1272 (2016).

32. De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. Nat. Struct. Mol. Biol. 18, 950–955 (2011).

33. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

34. Gardiner-Garden, M. & Frommer, M. CpG Islands in vertebrate genomes. J. Mol. Biol. 196, 261–282 (1987).

35. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492 (2012).

36. Hackenberg, M. et al. CpGcluster: A distance-based algorithm for CpG-island detection. BMC Bioinformatics 7, 1–13 (2006).

37. Chen, L. et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol. Cell 68, 745–757.e5 (2017).

38. Fan, G. et al. DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J. Neurosci. 21, 788–797 (2001).

39. Jackson-Grusby, L. et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. 27, 31–39 (2001).

40. Dai, J. et al. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. J. Am. Chem. Soc. 128, 1096–1098 (2006).

41. Kuryavyi, V., Phan, A. T. & Patel, D. J. Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. Nucleic Acids Res. 38, 6757–6773 (2010).

42. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. Biochemistry 44, 2048–2058 (2005).

43. Biffi, G., Tannahill, D. & Balasubramanian, S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. J. Am. Chem. Soc. 134, 11974–11976 (2012).

44. Giraldo, R. & Rhodes, D. The yeast telomere-binding protein RAPI binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J. 13, 2411–2420 (1994).

45. Bacolla, A., Pradhan, S., Roberts, R. J. & Wells, R. D. Recombinant human DNA (cytosine-5) methyltransferase. J. Biol. Chem. 274, 33006–33010 (1999).

46. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell 63, 167–178 (2016).

47. Ginno, P. A., Lott, P. L., Christensen, H. C., Korfi, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell 45, 814–825 (2012).

48. Wang, X. et al. Targeting of Polycomb Repressive Complex 2 to RNA by short repeats of consecutive guanines. Mol. Cell 65, 1056–1067.e5 (2017).

49. Wang, X. et al. Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat. Struct. Mol. Biol. (2017). https://doi.org/10.1038/natmb.3487

Acknowledgements
This work is supported by a core CRUK award (C14303/A17197). S.B. is a Senior Investigator of the Wellcome Trust (grant no. 099232/z/12/z). J.S. is a Marie Curie Fellow of the European Union (747297-QAPs-H2020-MSCA-IF-2016).

Author contributions
The project was conceived by S.M. and S.B. S.M. designed and carried out all the experiments with discussions with D.B., J.S., R.H.H., and S.B. S.M. carried out the analysis studies with input from A.G., D.T., and S.B. J.S. performed the G4-ChIP-seq experiments. A.G. and S.M.C. carried out all the computational analysis with discussions with S.M., D.T., R.H.H., and G.M. M.D. carried out the circular dichroism spectroscopy and ultraviolet melting experiments. All authors interpreted the results.

Competing interests
S.B. is an advisor and shareholder of Cambridge Epigenetix Limited.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41594-018-0131-8.

Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Oligonucleotide annealing. All oligonucleotides were PAGE purification quality (Sigma). For G4 formation, 10 μM DNA oligonucleotide was annealed in 10 mM Tris HCl, pH 7.4, 100 mM KCl by heating at 95 °C for 5 min followed by gradually cooling to 21 °C. For double-stranded DNA, 10 μM forward and reverse strand oligonucleotides were mixed and annealed in 10 mM Tris HCl, pH 7.4, 100 mM NaCl in the same manner. 20 μM (di-dC)20 was annealed as for double-stranded DNA.

Enzyme-linked immunosorbent assay. ELISAs for binding affinity and specificity were performed as described previously with minor modifications. Briefly, biotinylated oligonucleotides were bound to Pierce Streptavidin Coated High Capacity Plates (ThermoFisher) followed by blocking with 3% BSA and incubation with recombinant full-length human FLAG-tagged DNMT1 protein (Active Motif, Cat. No: 31404) in ELISA buffer (100 mM KCl, 50 mM KH2PO4, pH 7.4). After three washes with ELISA buffer, detection was achieved with an anti-FLAG horseradish peroxidase (HRP)-conjugated antibody (ab12388, Abcam) and 3,3′,5,5′-tetramethylbenzidine (TMB) ELISA substrate (Fast Kinetic Rate, ab171524, Abcam). Signal intensity was measured at 450 nm on a PHERAstar microplate reader (BMG Labtech). Dissociation constants (Kd) were calculated from saturation binding curves assuming one-site binding using Prism (GraphPad Software Inc.). Standard error of mean (s.e.m.) values were calculated from three replicates. Poly(dI-dC) was purchased from Sigma (catalog number P4929).

Differential methylation and binding analysis of entinostat treated HaCaT cells. HaCaT cells were treated with 10 μM entinostat for 48 h as we previously described.

Genomic DNA from untreated and treated cells were extracted with phenol/ chloroform. 50 ng of DNA were used to generate whole genome bisulfite sequencing libraries using Pico Methyl-Seq Library Prep Kit from Zymo research. Libraries were sequenced using the pair-end 150-bp high-output kit on the Illumina Next-seq platform. Data from four runs were pooled together. After quality assessment using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), reads were processed to remove adaptors and low-quality bases using cutadapt. Quality options -u 6 -u -1 -U 6 -U -1 were used to trim the initial six and 1 reads. Then 200 μl of oligonucleotides were annealed prior to measurement by warming up to 90 °C and slowly cooling down at room temperature.

Bioinformatic analysis and scripts. Bioinformatic analyses and processing were performed using Perl, Bash, Python, and R programming languages. The following tools were used: cutadapt (1.15), BWA (0.7.15), Picard (2.8.3), http://broadinstitute.github.io/picard), MACS (2.1.1), bedtools (2.26.0), http://bedtools.readthedocs.io/en/latest/content/overview.html, deepTools (2.5.1), and Bismark (v0.19.0). All scripts and software developed have been released on the Github page (https://github.com/slab-bioinformatics/dna-g4-methylation-dnmt1).

Differential methylation and binding analysis of entinostat treated HaCaT cells. HaCaT cells were treated with 10 μM entinostat for 48 h as we previously described.

Genomic DNA from untreated and treated cells were extracted with phenol/ chloroform. 50 ng of DNA were used to generate whole genome bisulfite sequencing libraries using Pico Methyl-Seq Library Prep Kit from Zymo research. Libraries were sequenced using the pair-end 150-bp high-output kit on the Illumina Next-seq platform. Data from four runs were pooled together. After quality assessment using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), reads were processed to remove adaptors and low-quality bases using cutadapt. Quality options -u 6 -u -1 -U 6 -U -1 were used to trim the initial six and 1 reads. Then 200 μl of oligonucleotides were annealed prior to measurement by warming up to 90 °C and slowly cooling down at room temperature.

Bioinformatic analysis and scripts. Bioinformatic analyses and processing were performed using Perl, Bash, Python, and R programming languages. The following tools were used: cutadapt (1.15), BWA (0.7.15), Picard (2.8.3), http://broadinstitute.github.io/picard), MACS (2.1.1), bedtools (2.26.0), http://bedtools.readthedocs.io/en/latest/content/overview.html, deepTools (2.5.1), and Bismark (v0.19.0). All scripts and software developed have been released on the Github page (https://github.com/slab-bioinformatics/dna-g4-methylation-dnmt1).

G4-ChIP-seq analysis. Raw fastq reads from G4-Chip-seq in S562 cells were trimmed with cutadapt to remove adapter sequences and low-quality reads (mapping quality <10). Reads were aligned to the human genome (version hg19)

with BWA and duplicates were removed using Picard. Peaks were called by MACS2 (P < 10−4) following our previous work: https://github.com/slab-bioinformatics/dna-secondary-struct-chrom-labs/blob/master/methods.md.

Peaks were merged from different replicates with bedtools multiregint. Only peaks overlapping in three out five replicates were considered high confidence. K562 datasets for DHSs (ENCSR000EPC), DNMT1 ChIP-seq (ENCSR987PBI), and whole genome bisulphite sequencing (ENCSR765PJC) were downloaded from ENCODE. Promoters were defined as 1 kb (−/+5) from the transcription start site of 31,239 hg19 transcripts. Methylation levels at CpG sites with less than 5X coverage were discarded. If not otherwise specified, CGIs were downloaded using the University of California Santa Cruz (UCSC) table browser and then ported to human genome release hg38 using the batch coordinate conversion (liftover) tool of the UCSC. The alternative CGI sets were generated using CpGCluster.

Enrichment analysis. ENCODE DHS and ChIP-seq datasets were normalized to a sequencing depth of 1 (that is, Reads Per Genomic Content, RPGC). Sequencing depth is defined as: (total number of mapped reads × fragment length)/effective genome size. The effective genome size was set to be 3,209,286,105 and enrichment values for DHSs and G4 peaks over CGIs and their flanking sequences were visualized in R using the ggplot2 library. Enrichment values for DNMT1 over CGIs and their flanks were visualized with deepTools.

Monte Carlo simulation. Monte Carlo simulation was used to calculate the significance of overlap between G4 peaks and high-confidence DNMT1 peaks, determined from single (di) double stranded DNA (SSD) G4-ChIP-seq. We first counted how many G4 peaks overlapped with sequences with the potential to form G4s in open chromatin. Such sequences were defined previously as those causing polymerase stalling in potassium and/or pyridostatin conditions (749,339 sequences), and overlap with at least one DHS region (43,506 sequences). We then randomly selected the same number of sequences with potential to form G4s from all in open chromatin and counted how many overlapped with at least one high-confidence DNMT1 peak. The Monte Carlo P value was then calculated as (N + 1)/(M + 1), where M is the number of iterations and N is the number of times the same or more overlaps were observed between randomized sequences with potential to form G4s and high-confidence DNMT1 peaks (compared to the number of overlaps observed between G4 peaks and high-confidence DNMT1 peaks). Randomization was repeated 8,000 times and on average the number of overlaps between the shuffled sequences with potential to form G4s and DNMT1 peaks were 2-fold less than those observed between the G4 and DNMT1 peaks.

Differential methylation and binding analysis of entinostat treated HaCaT cells. HaCaT cells were treated with 10 μM entinostat for 48 h as we previously described.

Genomic DNA from untreated and treated cells were extracted with phenol/ chloroform. 50 ng of DNA were used to generate whole genome bisulfite sequencing libraries using Pico Methyl-Seq Library Prep Kit from Zymo research. Libraries were sequenced using the pair-end 150-bp high-output kit on the Illumina Next-seq platform. Data from four runs were pooled together. After quality assessment using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), reads were processed to remove adaptors and low-quality bases using cutadapt. Quality options -u 6 -u -1 -U 6 -U -1 were used to trim the initial six and 1 reads. Then 200 μl of oligonucleotides were annealed prior to measurement by warming up to 90 °C and slowly cooling down at room temperature.

Bioinformatic analysis and scripts. Bioinformatic analyses and processing were performed using Perl, Bash, Python, and R programming languages. The following tools were used: cutadapt (1.15), BWA (0.7.15), Picard (2.8.3), http://broadinstitute.github.io/picard), MACS (2.1.1), bedtools (2.26.0), http://bedtools.readthedocs.io/en/latest/content/overview.html, deepTools (2.5.1), and Bismark (v0.19.0). All scripts and software developed have been released on the Github page (https://github.com/slab-bioinformatics/dna-g4-methylation-dnmt1).

G4-ChIP-seq analysis. Raw fastq reads from G4-Chip-seq in S562 cells were trimmed with cutadapt to remove adapter sequences and low-quality reads (mapping quality <10). Reads were aligned to the human genome (version hg19)
ENCODE. G4-ChIP-seq datasets for K562 and WGBS datasets for entinostat-treated and untreated HaCaT cells are available at the GEO repository under accession GSE107690. G4-ChIP-seq data in entinostat-treated and untreated HaCaT cells were taken from GSE76688. Source data for Figs. 1d,e,h and 3 are available online.

References
50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWAMEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
52. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
53. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
54. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
55. Peat, J. R. et al. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep.* **9**, 1990–2000 (2014).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| Item                                                                 | Confirmed |
|----------------------------------------------------------------------|-----------|
| n/a                                                                 | Confirmed |
| □ ✔ The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement |           |
| □ ✔ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |           |
| □ ✔ The statistical test(s) used AND whether they are one- or two-sided |           |
| □ ✔ Only common tests should be described solely by name; describe more complex techniques in the Methods section. |           |
| □ ✔ A description of all covariates tested |           |
| □ ✔ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |           |
| □ ✔ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |           |
| □ ✔ For null hypothesis testing, the test statistic (e.g. \( F, t, r \)) with confidence intervals, effect sizes, degrees of freedom and \( P \) value noted |           |
| Give \( P \) values as exact values whenever suitable. |           |
| □ ✔ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |           |
| □ ✔ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |           |
| □ ✔ Estimates of effect sizes (e.g. Cohen's \( d \), Pearson's \( r \)), indicating how they were calculated |           |
| □ ✔ Clearly defined error bars |           |
| State explicitly what error bars represent (e.g. SD, SE, CI) |           |

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

No software was used.

Data analysis

ELISA data were processed with Prism 7 (GraphPad Software Inc.). Bioinformatic data analyses and processing were performed using in-house Perl, Bash and R scripts and the following tools: cutadapt (1.15), BWA (0.7.15), Picard (2.8.3), (http://broadinstitute.github.io/picard), MACS (2.1.1), Bedtools (2.26.0), (http://bedtools.readthedocs.io/en/latest/content/overview.html), Deeptools (2.5.1) and Bismark (v0.19.0).

The pipeline used to analyze BG4 ChIP data is available at the following web page https://github.com/sblab-bioinformatics/dna-secondary-struct-chrom-lands/blob/master/Methods.md

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

K562 datasets for DHS (ENCSR000EPC), DNMT1 ChiP-seq (ENCSR987PBI) and whole genome bisulfate sequencing (ENCSR765JPC) were downloaded from ENCODE. G4-ChIP-seq data sets for K562 and WGBS datasets for entinostat-treated and untreated HaCaT cells are available at the NCBI GEO repository under accession number GSE107690. G4-ChIP-seq data in entinostat-treated and untreated HaCaT cells were taken from GSE76688.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size: ChIP-seq guidelines of the ENCODE consortia was followed. BG4 ChIP was performed in two biological replicates. In vitro experiments were performed in triplicates.

Data exclusions: No data was excluded from the analysis.

Replication: BG4 ChIP were performed in two biological replicate. The ELISA and in vitro DNA methylation assays were successfully replicated multiple times with recombinant full-length DNMT1 proteins from different providers (Abcam, Active Motif or NEB).

Randomization: No randomization was applied.

Blinding: No blinding techniques were used.

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
- n/a Unique biological materials
- n/a Antibodies
- n/a Eukaryotic cell lines
- n/a Palaeontology
- n/a Animals and other organisms
- n/a Human research participants

Methods

- n/a Involved in the study
- n/a ChIP-seq
- n/a Flow cytometry
- n/a MRI-based neuroimaging

Antibodies

Antibodies used: BG4 (Nat. Chem. 5, 182–6 (2013)), anti-FLAG horseradish peroxidase (HRP)-conjugated antibody (ab1238, Abcam)

Validation: BG4 antibody was established and validated in the following papers:
Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182–186 (2013).
Hänsel-hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. Nat. Genet. 48, (2016).
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) ATCC

Authentication Cell line was obtained directly from ATCC.

Mycoplasma contamination All cell lines are tested negative for mycoplasma.

Commonly misidentified lines (See ICLAC register) No misidentified lines are used.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107690

Files in database submission

B_REP1.SLX-12318_9.K562_asynch_a_701_504.tdf
B_REP1.SLX-12318_9.K562_asynch_c_703_504.tdf
B_REP1.SLX-12319.K562_asynch_input_704_504.tdf
B_REP2.SLX-12320.K562_P9_Async_a_701_517.tdf
B_REP2.SLX-12320.K562_P9_Async_b_701_502.tdf
B_REP2.SLX-12320.K562_P9_Async_c_701_503.tdf
B_REP2.SLX-12320.K562_P9_Async_input_702_504.tdf
B_REP1.SLX-12318_9.K562_asynch_a_701_504_peaks.narrowPeak
B_REP1.SLX-12318_9.K562_asynch_c_703_504_peaks.narrowPeak
B_REP2.SLX-12320.K562_P9_Async_a_701_517.peaks.narrowPeak
B_REP2.SLX-12320.K562_P9_Async_b_701_502.peaks.narrowPeak
B_REP2.SLX-12320.K562_P9_Async_c_701_503.peaks.narrowPeak
B_REP2.SLX-12320.K562_P9_Async_input_702_504.peaks.narrowPeak
K562_High_confidence_peaks.bed
B_REP1.SLX-12318_9.K562_asynch_a_701_504.bam
B_REP1.SLX-12318_9.K562_asynch_c_703_504.bam
B_REP1.SLX-12318_9.K562_asynch_IP_704_504.bam
B_REP2.SLX-12320.K562_P9_Async_a_701_517.rmdup.clean.bam
B_REP2.SLX-12320.K562_P9_Async_b_701_502.rmdup.clean.bam
B_REP2.SLX-12320.K562_P9_Async_c_701_503.rmdup.clean.bam
B_REP2.SLX-12320.K562_P9_Async_IP_702_504.rmdup.clean.bam
B_REP2.SLX-12320.K562_P9_Async_a_701_504.bw
B_REP2.SLX-12320.K562_P9_Async_c_703_504.bw
B_REP2.SLX-12320.K562_P9_Async_IP_704_504.bw
B_REP2.SLX-12320.K562_P9_Async_a_701_517.rmdup.clean.bw
B_REP2.SLX-12320.K562_P9_Async_b_701_502.rmdup.clean.bw
B_REP2.SLX-12320.K562_P9_Async_c_701_503.rmdup.clean.bw
B_REP2.SLX-12320.K562_P9_Async_IP_702_504.rmdup.clean.bw
B_REP1.SLX-12318.N701_N504.H7V58GXY.s_1.r_1.fq.gz
B_REP1.SLX-12318.N703_N504.H7V58GXY.s_1.r_1.fq.gz
B_REP1.SLX-12319.N701_N504.H7V28GXY.s_1.r_1.fq.gz
B_REP1.SLX-12319.N703_N504.H7V28GXY.s_1.r_1.fq.gz
B_REP1.SLX-12319.N704_N504.H7V28GXY.s_1.r_1.fq.gz
B_REP1.SLX-12319.N705_N504.H7V28GXY.s_1.r_1.fq.gz
B_REP2.SLX-12320.N701_N502.Async_b_HTYNCBGX.s_1.r_1.fq.gz
B_REP2.SLX-12320.N701_N503.Async_c_HTYNCBGX.s_1.r_1.fq.gz
B_REP2.SLX-12320.N701_N517.Async_a_HTYNCBGX.s_1.r_1.fq.gz
B_REP2.SLX-12320.N702_N504.Async_ip_HTYNCBGX.s_1.r_1.fq.gz

Genome browser session (e.g. UCSC)

Representative screen shots of all analyzed datasets are in Figure 1 & 2.

DNMT1 dataset from ENCODE could be visualized at:

https://www.encodeproject.org/experiments/ENCSR987PBI/

Methodology

Replicates

Two biological replicates comprising two & three technical replicates; peak overlap between these five replicates is 68 - 83%.
**Sequencing depth**

| Library          | Total reads | Uniquely mapped | Read length | Mode       |
|------------------|-------------|-----------------|-------------|------------|
| K562_BG4-ChIP-rep_1a | 24453052    | 18077431        | 75          | single-end |
| K562_BG4-ChIP-rep_1c | 25003857    | 17030377        | 75          | single-end |
| K562_BG4-ChIP-rep_1_input | 52086216  | 43107213        | 75          | single-end |
| K562_BG4-ChIP-rep_2a | 102690859   | 59904802        | 75          | single-end |
| K562_BG4-ChIP-rep_2b | 38524858    | 29961410        | 75          | single-end |
| K562_BG4-ChIP-rep_2c | 25242910    | 19693045        | 75          | single-end |
| K562_BG4-ChIP-rep_2_input | 48077620  | 38383341        | 75          | single-end |

**Antibodies**

BG4 antibody is reported in previous publications:

Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182–186 (2013).

Hänsel-hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. Nat. Genet. 48, (2016).

BG4 was prepared in BL21(DE3) E. coli using the expression vector pSANG10-3F-BG4 (Addgene, plasmid #55756). Prior to G4 ChIP, the binding affinity and specificity of BG4 for G4 structures were confirmed by standard enzyme-linked immunosorbent assay (ELISA).

**Peak calling parameters**

Reads were aligned to the human genome (version hg19) with BWA and duplicates were removed by Picard. Peaks were called by MACS2 (P < 1e-05) with the appropriate input control for each pull-down library following the pipeline described at the following web page https://github.com/sblab-bioinformatics/dna-secondary-struct-chrom-lands/blob/master/Methods.md, with the only difference being the merging of peaks from different replicates. For that, the Bedtools were used (command multiIntersect), and only peak regions overlapping in 3 out 5 replicates were considered as high-confidence peaks.

**Data quality**

| Library   | # peaks called | # peaks with fdr <= 0.05 and fold >= 5 |
|-----------|----------------|--------------------------------------|
| REP1_a    | 12,495         | 10,637                               |
| REP1_c    | 10,079         | 8,493                                |
| REP2_a    | 11,111         | 8,319                                |
| REP2_b    | 9,026          | 7,171                                |
| REP2_c    | 8,059          | 6,992                                |

Only peak regions overlapping in 3 out 5 replicates were considered as high-confidence peaks.

**Software**

Bioinformatic data analyses and processing were performed using in-house Perl, Bash and R scripts and the following tools: cutadapt (1.15), BWA (0.7.15), Picard (2.8.3), (http://broadinstitute.github.io/picard), MACS (2.1.1), Bedtools (2.26.0), (http://bedtools.readthedocs.io/en/latest/content/overview.html) and Deeptools (2.5.1).

The pipeline used to analyze BG4 ChIP data is available at the following web page https://github.com/slab-bioinformatics/dna-secondary-struct-chrom-lands/blob/master/Methods.md