Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation

Ethan Ford[1], Matthew R. Grimmer[2,3,4], Sabine Stolzenburg[5], Ozren Bogdanovic[1,6,7], Alex de Mendoza[1], Peggy J. Farnham[2], Pilar Blancafort[5,8], Ryan Lister[1,8]

Affiliations

[1]Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.
[2]Department of Biochemistry and Molecular Medicine, University of Southern California, 1450 Biggy St, Los Angeles, CA 90089, USA
[3]Integrated Genetics and Genomics, University of California, Davis, 451 Health Sciences Dr, Davis, CA 95616, USA
[4]Department of Neurological Surgery, University of California, San Francisco, 1450 3rd St, San Francisco, CA 94158, USA.
[5]School of Anatomy, Physiology and Human Biology, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.
[6]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, New South Wales, Australia.
[7]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia.
[8]Harry Perkins Institute of Medical Research, 6 Verdun St, Nedlands, WA 6009, Australia.

Abstract

It is widely assumed that the addition of DNA methylation at CpG rich gene promoters silences gene transcription. However, this conclusion is largely drawn from the observation that promoter DNA methylation inversely correlates with gene expression. The effect of forced DNA methylation on endogenous promoters has yet to be comprehensively assessed. Here, we conducted artificial methylation of thousands of promoters in human cells using an artificial zinc finger-DNMT3A fusion protein, enabling assessment of the effect of forced DNA methylation upon transcription and histone modifications, and the durability of DNA methylation after the removal of the fusion protein. We find that DNA methylation is frequently insufficient to transcriptionally repress promoters. Furthermore, DNA methylation deposited at promoter regions associated with H3K4me3 is rapidly erased after removal of the zinc finger-DNMT3A fusion protein. Finally, we demonstrate that induced DNA methylation can exist simultaneously on promoter nucleosomes that possess the active histone modification H3K4me3. These findings suggest that promoter DNA methylation is not generally sufficient for transcriptional inactivation, with implications for the emerging field of epigenome engineering.

Introduction

The mammalian genome is almost entirely methylated at the 5th position of cytosine in the CpG dinucleotide context (mCG)[1]. Small regions of the genome remain in a low- or un-methylated state[2,3], mostly associated with transcriptional enhancers, transcriptional start sites, and the binding sites of sequence-specific DNA binding proteins. Most commonly, DNA methylation of the promoter regions of genes inversely correlates with transcriptional activity[2,4,5,6]. For example, treatment of cells with demethylating drugs, such as 5-azacytidine, an inhibitor of DNA methyltransferases, causes global demethylation and activation of many silenced genes[7]. Furthermore, loss of DNA methyltransferase 1 (DNMT1) in mouse fibroblasts results in global losses in DNA methylation and global increases in gene expression[8]. These relationships have led to the conclusion that DNA methylation of CpG island promoters or enhancer elements results in gene silencing. However, these observations are correlative and challenging to interpret because genome-wide demethylation could result in complex changes to chromatin structure and gene expression throughout the genome. Furthermore, in early development, transcription can occur from genes with methylated promoters[9–11].

To more precisely address the relationship between DNA methylation and promoter activity, targeted methods to methylate or demethylate specific loci in the genome have been devised. The first attempts to induce targeted DNA methylation at endogenous loci used arrays of six zinc finger (ZF) domains, engineered to recognize 18- base pair specific genomic sequences, fused to the catalytic domain of the human DNA methyltransferase DNMT3A (ZF-DNMT3A)[12–16] or the bacterial methyltransferase M.SssI[17]. Because of the difficulty in synthesizing customized ZF proteins and questions regarding their *in vivo* specificity[18], subsequent studies have utilized transcription activator-like effectors (TALE) proteins or a catalytically inactive version of the Cas9 protein (dCas9) fused to DNMT3A[19–24]. Collectively, these customized DNA binding proteins fused to artificial epigenetic modifiers, which are referred to as artificial epigenetic modifiers, now allow the investigation into whether DNA methylation merely correlates with transcriptional repression or if it is a causative epigenetic mark. Indeed, the results of such experiments have demonstrated that DNA methylation deposited on promoters by engineered DNA binding proteins can sometimes be sufficient to silence or reduce transcription. However, a broader examination of the susceptibility of gene promoters to DNA methylation and the extent to which DNA methylation can induce silencing of transcription has yet to be reported.

Unlike the maintenance of histone modifications, the heritability of mCG during DNA replication has a reasonably well established mechanism. After DNA replication, a hemi-methylated CpG dinucleotide is formed, with an unmethylated cytosine on the newly synthesized strand and a methylated cytosine on the template strand. Hemi-methylated CpG dinucleotides can be rapidly methylated, primarily by the maintenance DNA methyltransferase DNMT1[25]. For this reason, DNA methylation has been implicated in the long-term silencing of genes[26]. For example, we have shown that targeting a ZF-DNMT3A fusion protein to the promoter of the *SOX2* gene resulted in the silencing of *SOX2* for at least 8 days after the withdrawal of the ZF-DNMT3A fusion protein[16]. In contrast, the same ZF fused to the Krupple-Associated Box (KRAB) domain, which recruits repressive histone modifiers, induces transcriptional repression that is transient and dependent on the continued presence of the ZF-KRAB fusion protein[27]. However, not all

promoters respond to enforced DNA methylation in the same way. It has been reported that a ZF-DNMT3A fusion protein targeted to the *VEGF-A* promoter or a TALE-DNMT3A targeted to the *B2M* promoter results in only transient silencing of the respective gene[13,22]. Thus, the context in which DNA methylation is sufficient to establish transcriptional repression and the stability of induced DNA methylation has not been characterized throughout the genome[28].

We designed a set of experiments to gain a broad understanding of the effects on gene expression and chromatin structure upon artificial epigenetic modifier-mediated DNA methylation at regulatory regions throughout the genome. To accomplish this, we utilized a human cancer cell line in which the programmable overexpression of an artificial ZF-DNMT3A fusion protein results in widespread addition of DNA methylation at unmethylated regions throughout the genome. Using a suite of genome-wide analyses to characterize these cells, including whole genome bisulfite sequencing, ChIP-seq, TAB-seq and RNA-seq, we have undertaken a broad assessment of the consequences of forced methylation at promoter regions upon gene expression and chromatin state. We observe only a modest negative correlation between the induction of DNA methylation and transcriptional repression and demonstrate that forced DNA methylation is not sufficient to evict active chromatin marks from promoter regions. In addition, induced DNA methylation is lost in the absence of the ZF-DNMT3A fusion protein, indicating the activity of an efficient proofreading system to protect active promoters from DNA methylation.

Results

ZF-DNMT3A methylates the *SOX2* promoter and represses *SOX2* transcription

To assess the effects of forced DNA methylation, we took advantage of a previously described MCF-7 derived cell line that overexpress an artificial epigenetic modifier composed of 6 zinc finger domains linked to the catalytic domain of human DNMT3A upon treatment with doxycycline[16], herein referred to as ZF-DNMT3A (Fig. 1A). The ZF arrays were designed to bind to a GC-rich 18 bp sequence in the *SOX2* promoter, but was found to have widespread off-target binding throughout the genome[18]. We performed ChIP-seq of the HA-epitope-tagged ZF-DNMT3A protein to determine its binding sites throughout the genome. As expected, we observed a strong ChIP peak centered on the intended binding site in the SOX2 promoter (Fig. 1B).

To determine the changes in DNA methylation induced by the ZF-DNMT3A construct, we next performed whole genome bisulfite sequencing (WGBS, by MethylC-seq) of MCF-7 cells using cells harboring an integrated empty vector construct and grown in the presence of doxycycline for 3 days (empty vector dox-induced) to establish baseline methylation levels, ZF-DNMT3A cells induced with doxycycline for 3 days (ZF-DNMT3A dox-induced), and ZF-DNMT3A cells induced with doxycycline for 3 days and subsequently grown in doxycycline-free media for 9 days (ZF-DNMT3A dox-withdrawn). We observed an increase in the level of mCG (difference

between samples in the number of C base calls divided by the number of total base calls for all cytosine positions in the CpG context in the defined region; ΔmCG) of 0.34 at the *SOX2* promoter region (defined as the region of unmethylated DNA in empty upstream and downstream of the transcriptional start site of the *SOX2* gene vector dox-induced cells; chr3:181429226-181431783) in the ZF-DNMT3A dox-induced cells when compared to the empty vector dox-induced cells. However, 53% of the induced methylation was lost in the ZF-DNMT3A dox-withdrawn cells (Fig. 1B).

Finally, we measured steady state RNA levels by RNA-seq in empty vector dox-induced cells, ZF-DNMT3A cells without doxycycline induction (ZF-DNMT3A no-dox), ZF-DNMT3A dox-induced cells, and ZF-DNMT3A dox-withdrawn cells. The raw read data from the empty vector dox-induced cells and ZF-DNMT3A no-dox cells were combined to create a data set herein referred to as “no-ZF”. From this data it was observed that expression of ZF-DNMT3A resulted in a 2.5 fold decrease in SOX2 mRNA levels in comparison to the no-ZF cells, and the decrease in SOX2 mRNA levels was maintained after withdrawal of doxycycline for 9 days (Fig. 1B). Thus, ZF-DNMT3A methylates its target region and results in long-term repression of the *SOX2* gene. Interestingly, transcriptional repression is maintained in dox-withdrawn cells despite the loss of over half of the induced methylation.

ZF-DNMT3A displays widespread off-target binding and methylation

Given the G/C-rich characteristics of the intended *SOX2* promoter target binding site, the propensity of ZF domains to bind sites with similar sequences to their intended target[29] as well as G/C rich sequences in general[30], and the prevalence of CpG-rich regulatory regions (e.g. gene promoters and enhancers), the ZF-DNMT3A protein could potentially bind to and deposit DNA methylation at many regions throughout the genome. Indeed, further analysis of the ZF-DNMT3A ChIP-seq data identified 25,142 off-target binding sites throughout the genome (FDR < 0.05). These off-target binding sites were found in gene promoters, active enhancers, gene bodies, and intergenic regions (Fig. 1C, Supplemental Table S1).

The measured global methylation levels of the empty vector dox-induced and the ZF-DNMT3A dox-induced cells in the CpH (H = A,T or C) sequence context was equal to the bisulfite non-conversion rate of 0.5%, indicating that the fusion construct was not broadly altering CpH methylation in the genome. In contrast, global methylation levels of the empty vector dox-induced and the ZF-DNMT3A dox-induced cells in the CpG sequence context were 0.63 and 0.682 respectively, constituting a 0.052 increase in CpG methylation in the ZF-DNMT3A dox-induced cells. Consequently, we focused further analyses on DNA methylation in the CpG context.

Comparison of the empty vector dox-induced and ZF-DNMT3A dox-induced WGBS data revealed 10,356 differentially methylated regions (DMRs) in the CpG dinucleotide context. 38% of the identified DMRs overlapped with gene promoters (+/- 2 kb of a transcriptional start site), and 27% were located in gene bodies, a small fraction (6%) of the DMRs were found in

transcriptional enhancers (H3K4me1/H3K27ac modified chromatin), and 30% were in intergenic regions not corresponding to enhancers (Fig. 1D, Supplementary Table S2). Notably, only 35% of the DMRs intersected with the ZF-DNMT3A ChIP-seq peaks (Fig. 1F). While there was an enrichment of ZF-DNMT3A ChIP-seq signal in the in a majority of DMRs, many DMRs showed no detectable binding of ZF-DNMT3A (Fig. 1G), suggesting that ZF-DNMT3A may induce off-target methylation in the absence of apparent strong and persistent binding by the protein or that some changes in methylation occurred as a response of phenotypic changes in the cells.

While the boundaries of a DMR are normally defined by a genomic interval of induced methylation, an alternative approach is to define the boundaries of a region of interest by an unmethylated genomic interval flanked by methylated DNA in empty vector dox-induced cells, such as the promoter region of a gene. This allows for the measurement of the induction DNA methylation in ZF-DNMT3A dox-induced cells across an entire gene promoter as opposed to just a portion of it as in the case of DMRs. To perform this analysis, we first identified 15,911 unmethylated regions (UMRs) in the empty vector dox-induced cells. UMRs are defined as unmethylated regions that contain a high CpG dinucleotide density and at least 30 CpG dinucleotides[31]. As expected, most (77%) of the UMRs are found within 2 kb of an annotated transcriptional start site and 82% overlapped with CpG islands (CGIs, Fig. 1E, Supplementary Table S3). These identified UMRs, along with the identified DMRs, were used to analyze the effects of the widespread binding and induction of DNA methylation by ZF-DNMT3A.

ZF-DNMT3A broadly induces methylation at DMRs and UMRs

To gain more detailed insights into the nature and extent of DNA methylation induced by ZF-DNMT3A, we compared the fraction of methylated CpG dinucleotides in each DMR in empty vector dox-induced, ZF-DNMT3A dox-induced, and ZF-DNMT3A dox-withdrawn cells (Fig. 2A and 2C). An increase in DNA methylation is observed in 99.9% of DMRs identified in the ZF-DNMT3A dox-induced cells, with only 12 DMRs losing DNA methylation. To assess whether the retention of DNA methylation correlated with the baseline level of methylation in the absence of ZF-DNMT3A, we focused upon DMRs with low (< 0.1) and high (> 0.3) initial levels of methylation in empty vector dox-induced cells. Upon doxycycline induction and subsequent withdrawal in ZF-DNMT3A cells, 66% of DMRs that had low initial methylation level returned to their baseline levels after doxycycline withdrawal (ΔmCG <0.1; Fig 2B and 2C). In contrast, only 8.2% of DMRs that had initial DNA methylation levels of greater than 0.3 in the empty vector dox-induced sample were returned to their initial levels (ΔmCG <0.1) after doxycycline withdrawal. This illustrates that DMRs with a higher initial DNA methylation level in the empty vector dox-induced cells more frequently retained DNA methylation after the withdrawal of ZF-DNMT3A.

In a similar fashion, we also assessed the ZF-DNMT3A induced DNA methylation in the UMRs identified in empty vector dox-induced cells (Fig. 2D-F). As expected, nearly all UMRs have very

low levels of DNA methylation in the empty vector dox-induced cells. Upon expression of ZF-DNMT3A, 67.6% of UMRs exhibited a ΔmCG > 0.1. However, nearly all of these UMRs lost the induced DNA methylation in the ZF-DNMT3A dox-withdrawn cells, suggesting that an inherent property of UMRs leads to the loss of DNA methylation in the absence of ZF-DNMT3A.

Given the observation that UMRs lose DNA methylation in the absence of ZF-DNMT3A, and that most UMRs are located in the promoters of genes, we wanted to assess the association of DMRs with different active chromatin states, and in particular, to determine whether the presence of H3K4me3 was associated with DMRs that were more prone to the loss of methylation in the ZF-DNMT3A dox-withdrawn cells. We therefore performed ChIP-seq of H3K4me3 in ZF-DNMT3A no-dox cells and ZF-DNMT3A dox-induced cells. As observed with forced DNA methylation in UMRs, DMRs at sites enriched for H3K4me3 in ZF-DNMT3A dox-induced cells are largely lost upon doxycycline withdrawal (Fig. 2G, top DMRs), whereas DMRs not enriched for H3K4me3 (Fig. 2G, bottom DMRs) exhibit higher retention of DNA methylation. Specifically, 78% of the top quartile of DMRs in dox-withdrawn cells sorted by H3K4me3 signal returned to similar methylation levels (ΔmCG < 0.1 compared to the empty vector dox-induced) in dox-withdrawn cells, whereas in the bottom quartile of H3K4me3 marked DMRs,only 14% returned to ΔmCG < 0.1 in dox-withdrawn cells.

To determine whether putative enhancers were also associated with the loss of DNA methylation, we ranked the DMRs not associated with H3K4me3 by the level of H3K4me1 that they exhibited (Fig. 2H). The correlation between H3K4me1 and the loss of induced DNA methylation is less pronounced than for H3K4me3. Specifically, 9.6% of DMRs in the highest quartile of H3K4me1 binding returned to near baseline levels of DNA methylation (ΔmCG < 0.1 compared to empty vector dox-induced) in dox-withdrawn cells versus 3.2% in the bottom quartile. Thus, we do not observe any evidence that the presence of H3K4me1 correlates with loss of DNA methylation after removal of ZF-DNMT3A from the cells.

ZF-DNMT3A methylates DNA associated with active chromatin marks and transcription

Previous studies have reported that mC and H3K4me3 are, for the most part, mutually exclusive epigenetic marks in CpG-rich regions[32], and that H3K4me3 may protect against cytosine methylation[33]. However, it is not known whether forced mC deposition is sufficient to induce the loss of H3K4me3. Analysis of H3K4me3 ChIP-seq read density within ±1 kb of all DMRs with a ΔmCG > 0.2 between ZF-DNMT3A no-dox and dox-induced cells revealed only a very slight decrease in H3K4me3 signal after induction of ZF-DNMT3A (Fig. 3A-B, Supplemental Fig. S1). The two data sets were normalized by peak height in H3K4me3 peaks that had a ΔmCG < 0.2 between the no-dox and dox-induced cells (Fig. 3A, bottom). While DMRs only contain regions that have at least a ΔmCG of >0.2 upon doxycycline induction, UMRs display a broad range of induced methylation with many that are not methylated by ZF-DNMT3A. For this reason, we looked at the correlation between H3K4me3 binding and DNA methylation in UMRs. However, the decrease in H3K4me3 signal did not correlate with the level of induced methylation (Fig.

3C), suggesting that DNA methylation has minimal, if any, effect on H3K4me3 modification in the UMRs.

Because of the limited sensitivity of ChIP-seq and the possibility that there could be a heterogeneous population of loci with methylated and unmethylated DNA molecules, we sought to directly measure the level of DNA methylation in H3K4me3-modified chromatin using ChIP-bisulfite-sequencing[34–36]. For this analysis, UMRs were favored over DMRs in order to incorporate DMR-adjacent unmethylated sequences which, in theory, could bind factors ejected from DMRs. We identified 6,392 H3K4me3 ChIP-seq peaks from both ZF-DNMT3A dox-induced cells and ZF-DNMT3A no-dox cells that overlapped with UMRs that had a ΔmCG > 0.2 across the entire UMR region upon induction of ZF-DNMT3A. In DNA purified from K4me3 ChIP upon induction of ZF-DNMT3A, 21% of cytosines within these 6,392 peaks were methylated, demonstrating that these bivalent modifications can exist simultaneously at the same site upon forced methylation (Fig. 3D).

We also sought to determine if the initiated form of RNA polymerase II (phospho-ser5) was able to bind DNA methylated by ZF-DNMT3A. ChIP-bisulfite-sequencing[11] was performed with an anti-phospho-ser5 RNA polymerase II antibody. Analysis of UMRs that exhibited a ΔmCG > 0.2 between the ZF-DNMT3A no-dox and ZF-DNMT3A dox-induced cells and intersected with a phospho-ser5 RNA polymerase II ChIP-seq peak revealed that, similar to H3K4me3, there was a small decrease in the median DNA methylation level from 0.29 in genomic DNA to 0.23 in the phospho-ser5 ChIP-purified DNA. Importantly there was a clear direct association of initiated RNA polymerase II with ZF-DNMT3A methylated DNA (Fig. 3E). These data show that forced DNA methylation is not sufficient to prevent the deposition of H3K4me3 or binding of phospho-ser5 RNA polymerase II to the methylated DNA.

ZF-DNMT3A induced methylation does not require DNA replication for its removal

A time course experiment was performed to gain insights into the durability of forced DNA methylation after removal of ZF-DNMT3A. Cells were harvested every 24 hours during the induction period and subsequent withdrawal of doxycycline, and HA-tagged ZF-DNMT3A protein levels were quantitated over time by western blotting. Prior to doxycycline induction, a band is visible at the molecular weight of ZF-DNMT3A, which could either be from leaky expression of the ZF-DNMT3A construct or a non-specific protein (Fig. 4A). The level of any potential leaky expression has minimal effect on methylation levels in DMRs, as observed by comparing the methylation levels in DMRs between ZF-DNMT3A cells treated with and without doxycycline (Supplemental Fig. S2A). In addition, PCA analysis of RNA-seq data from empty-vector dox-induced, ZF-DNMT3A no-dox, and ZF-DNMT3A dox-induced cells also shows that empty vector dox-induced and ZF-DNMT3A no-dox cells cluster more closely to each other than to ZF-DNMT3A dox-induced cells (Supplemental Fig. S2B). ZF-DNMT3A protein is strongly expressed 1 day after the addition of doxycycline and returns to undetectable/baseline levels 1 day after the doxycycline withdrawal (Fig. 4A).

In order to measure the methylation dynamics of 7 representative gene promoters which, upon dox induction, showed strong local increases in methylation, we performed targeted bisulfite sequencing across daily timepoints through dox induction and withdrawal (Fig 4B, Supplemental Fig. S3). All 7 promoters showed a robust increase in methylation under ZF-DNMT3a induction and a return to near baseline methylation levels 5 days after the removal of doxycycline. Thus, both WGBS (Fig. 2A,B) and targeted bisulfite sequencing analyses demonstrate that forced DNA methylation by ZF-DNMT3A is at every site tested, not stably maintained in the absence of the ZF-DNMT3A protein.

The loss of the induced DNA methylation following removal of doxycycline and ZF-DNMT3A could be the result of passive loss of DNA methylation through the lack of its maintenance after DNA replication, or the result of active removal of DNA methylation by a mechanism such as TET dioxygenase-mediated demethylation[37]. To determine if loss of DNA methylation occurred through an active or passive process, we grew ZF-DNMT3A cells in the presence of doxycycline for 3 days, after which the media was changed to doxycycline-free media containing either Lovastatin, RO-336 or thymidine at concentrations previously shown to inhibit the cell cycle of MCF-7 cells (Lovastatin and thymidine) or HeLa cells (RO-336)[38–40]. To assess the effectiveness of Lovastatin, thymidine and RO-336 in the ZF-DNMT3A cell line, we treated ZF-DNMT3A cells with each drug for 48 hours and determined the percentage of cells in the G1, S and G2/M phases of the cell cycle by DNA staining with propidium iodide. Lovastatin treatment resulted in an increase in the percentage of cells in G1 from 44.4% to 71.5%, while RO-336 resulted in an increase from 24.2% to 57.1% of cells in G2/M. While thymidine caused a decrease in the percentage of cells in S-phase, a clear shift within S-phase to cells near the S-G2 transition as compared to the untreated cells was observed (Supplemental Fig. S4A,B). However, these compounds had very little effect on the loss of DNA methylation (Fig. 4C, Supplemental Fig S4C). For example, after 5 days of doxycycline withdrawal, 88-100% of the DNA methylation at the 7 loci analyzed was lost in the cells treated with the three cell cycle inhibitors, as compared to the untreated cells, indicating that the DNA methylation deposited by ZF-DNMT3A is actively removed.

The most thoroughly documented mechanism to actively remove DNA methylation operates via the TET family of dioxygenase proteins, which proceed through a 5-hydroxymethylcytosine (hmC) intermediate[41]. Therefore, we identified sites of hmC genome-wide by Tet-Assisted-Bisulfite-sequencing (TAB-seq)[42] in ZF-DNMT3A cells with and without doxycycline induction. We observed a significant increase (p-value $< 2.2 \times 10^{-16}$, Wilcoxon signed rank test) in hmC levels in the DMRs upon doxycycline induction of ZF-DNMT3A (Fig. 4D, Supplemental Fig. S1A). We also observed a significant enrichment (p-value $< 2.2 \times 10^{-16}$, Wilcoxon signed rank test) in hmC in a 150 bp window centered at the DMRs compared to a 150 bp window 2.5 kb upstream of the DMR center in ZF-DNMT3A dox-induced cells and no significant (p-value 0.29, Wilcoxon signed rank test) enrichment in the same regions in ZF-DNMT3A no-dox cells (Fig. 4E). Taken together with the observation that the loss of ZF-DNMT3A induced methylation does not require DNA replication, this strongly implicates the TET proteins in the active removal of ZF-DNMT3A induced methylation.

ZF-DNMT3A fails to repress the genes of many promoters that it methylates

Forced DNA methylation by artificial epigenetic modifiers fused to DNMT3A have previously been shown to downregulate transcription and have been proposed as systems to silence disease-associated genes[12,14,16]. However, the efficacy of such targeted DNA methylation-induced silencing and the fraction of genes at which forced DNA methylation is sufficient to induce silencing are unknown. We compared the increase in DNA methylation in UMRs found within 2 kb of a transcriptional start site with the steady state mRNA levels in cells expressing or not expressing ZF-DNMT3A. Of the 700 genes with associated UMRs that had a ΔmCG > 0.4 between the empty vector dox-induced and ZF-DNMT3A dox-induced cells, 80% exhibited a decrease in mRNA levels. Conversely, of the 1,758 genes with associated UMRs that had a ΔmCG < 0.1, only 37% had a decrease in mRNA levels, showing a clear relationship between DNA methylation and repression of the associated gene (Fig. 5A, Supplementary Table S4). However, of the repressed genes, the decreases in transcript abundance were mostly small (<2-fold), and furthermore, of the 1,985 genes with an associated UMR with a ΔmCG > 0.3, 420 showed either no decrease or a gain in mRNA abundance. (Fig. 5B-C, Supplemental Fig. S5). We also did not observe any correlation between the ability of ZF-DNMT3A to repress transcription with the expression level of the gene (Supplemental Fig. S6A-B, upper panels), the number of CpG dinucleotides in the UMR (Supplemental Fig. S6A-B, middle panels), or the CpG dinucleotide density in the UMR (Supplemental Fig. S6A-B, lower panels). Overall, we observe a modest inverse correlation between the induction of DNA methylation and RNA transcription, suggesting that DNA methylation on its own is frequently insufficient to silence gene transcription.

Because ZF-DNMT3A does not induce 100% methylation throughout the UMRs, there exists the possibility that there are two distinct populations of cells: one in which a promoter remains unmethylated and the transcription of the associated gene is not repressed and another population where the promoter of the gene is methylated and the gene is silenced. To address this issue, we performed RNA-fluorescence in-situ hybridization (RNA-FISH) with probes to the *GAPDH* and *PSME1* mRNAs. These two genes were chosen for closer scrutiny because they had UMRs that were strongly methylated by ZF-DNMT3A and showed no reduction in mRNA levels upon ZF-DNMT3A expression (Fig. 5C). The cells were then sorted by FACS into high and low subpopulations according to the expression of *GAPDH* or *PSME1*, as measured by the RNA-FISH signal. Notably, in both cases the unsorted cells showed a normal distribution, as opposed to two distinct subpopulations (Fig. 6A and 6D). FACS reanalysis of the sorted cells (Fig. 6A and 6D), in addition to RT-qPCR (*GAPDH*) or RNA-seq (*PSME1*), confirmed that RNA-FISH was effective for isolating subpopulations of cells exhibiting high and low levels of the targeted mRNA (Fig. 6B and 6E). Analysis of DNA methylation levels in the promoter regions by targeted bisulfite sequencing of the different sorted populations revealed that, for both promoters, DNA methylation levels were very similar in low expressing and high expressing cells (Fig 6C,F,G). The only exception to this was two adjacent cytosines in the PSME1 promoter region showing higher (ΔmCG > 0.2) methylation in the PSME1 low expressing cells

(indicated with red arrowheads). Overall this data further supports the model whereby DNA methylation in gene promoters often does not result in transcriptional repression.

To explore the possibility that there is a subpopulation of DNA molecules that are resistant to methylation by ZF-DNMT3A we analyzed the distribution of methylation in single DNA molecules. To this end, we analyzed the distribution of DNA methylation in WGBS reads from ZF-DNMT3A dox-induced cells that contained >10 cytosines in the CpG context and aligned to DMRs. We further sub-selected for reads aligned to regions that had aggregate methylation levels of 0.4-0.6 (mCG/CG) in ZF-DNMT3A dox-induced cells and <0.1 (mCG/CG) in empty-vector dox-induced cells. Of 16,159 reads analyzed from the ZF-DNMT3A dox-induced cells, only 20% (3,224 reads) were unmethylated or had a low fraction of methylated cytosines (mCG/CG < 0.15). In contrast, 80% of the reads (12,935 reads) from ZF-DNMT3A dox-induced cells were modestly to completely methylated (mCG/CG 0.15-1), with most exhibiting a robust gain of methylation (mCG/CG > 0.4 and < 0.8, Fig. 6H, red line). The genomic coordinates of each read were used to calculate the aggregate methylation in ZF-DNMT3A dox-induced cells of each read interval (Fig. 6H, blue line). While there is a broader distribution of methylation on single DNA molecules than the average, this observation does not support the hypothesis that there is a large population of DNA molecules that were not methylated by ZF-DNMT3A. Thus, the inability of ZF-DNMT3A to induce transcriptional silencing at a large number of promoters is due to the inability of DNA methylation on its own to induce transcriptional silencing.

Discussion

In recent years there have been substantial efforts to alter the epigenomic state of promoters in order to stably silence their associated genes. Targeted approaches to alter DNA methylation, which is commonly considered a stable epigenetic mark, are being explored for both research and potential clinical applications[16]. However, very little is known about the stability of a change in DNA methylation state when it is artificially induced at a locus in the genome. In fact, reports that look at the durability of artificially induced DNA methylation show different outcomes[13,16,22]. Several studies using ZF, TALE, or dCas9-based targeting of DNA methylation have reported moderate transcriptional repression by artificially induced DNA methylation[12,14–16,19,21–23,43], but it remains far from clear whether DNA methylation alone can generally induce chromatin compaction and transcriptional silencing of nearby genes. Thus, investigations are required to broadly assess the stability of induced DNA methylation and its effects upon chromatin and transcription.

In this study we performed a series of high-throughput genome-wide experiments to profile the effects of forced DNA methylation on thousands of regions in the human genome. We observed that after methylating the DNA at gene promoters, there is very little, if any, change in the association of H3K4me3 with methylated DNA, showing that DNA methylation on its own is not sufficient to reconfigure DNA into a stable heterochromatinized state. Furthermore, we observed that DNA methylation is quickly lost at nearly all promoters in the absence of the artificial methyltransferase.

The classical model of how the DNA methylation landscape is broadly established and maintained is that the genome is methylated in early development by the *de novo* DNA methyltransferases (DNMT3A/B), and these patterns are maintained by the maintenance methyltransferase DNMT1, which methylates hemi-methylated CpG sites after DNA replication. The discovery of the TET proteins, which actively demethylate 5-methylcytosine through a 5-hydroxymethylcytosine intermediate, prompted the refinement of the classical model. Nevertheless, it is not entirely clear how the CpG-rich regions associated with transcriptional start sites establish and retain their unmethylated state in a genome that is almost entirely methylated in the CpG context[5,44]. It has been proposed that the *de novo* DNA methyltransferases are excluded from these regions[33,45]. However, this cannot be the only process, because aberrant methylation deposited in CpG islands would require a proofreading mechanism to prevent the entire genome from drifting towards a fully methylated state.

Our observations that DNA methylation is actively removed from the promoters of genes and the presence of increased levels of 5-hydroxymethylcytosine in areas forcibly methylated by ZF-DNMT3A support a model whereby CpG islands are also protected from DNA methylation by the surveillance of the TET proteins[46,47]. This is in agreement with the observation that the CxxC domains of the TET1 and TET3 proteins specifically bind to CpG dinucleotides with a slight preference for the unmethylated state[48–50]. Another important observation is that while induction of the ZF-DNMT3A protein resulted in large increases in DNA methylation, it never resulted in full methylation of an entire promoter region. The broad distribution of methylation states that we observed in individual reads suggests that while the ZF-DNMT3A protein is methylating a locus, TET proteins could simultaneously be catalyzing its demethylation.

Another important aspect of this work is the lack of effective gene silencing by ZF-DNMT3A induced DNA methylation, even at genes that accumulate high levels of DNA methylation in their promoters. While DNA methylation is clearly an important epigenetic mark, there is very little direct evidence that on its own it is sufficient to block transcription or force DNA into a compacted, heterochromatinized state. It has been observed that DNA methylation follows transcriptional repression during development, suggesting that it is a reinforcing mark[51]. Notably, Amabile *et al.* recently observed that a TALE-DNMT3A fusion protein targeting the *B2M* gene had minimal effects upon transcription in transfected human cells[22]. However, when DNMT3A, DNMT3L and KRAB domain proteins fused to TALE DNA binding domains were targeted simultaneously to the *B2M* gene, a durable and long lasting silencing of the gene was achieved. The KRAB domain recruits several epigenetic repressors including histone modifying proteins and the nucleosome remodelling complex NuRD[52]. The requirement of the combination of the KRAB domain and DNA methylation to stably repress transcription, taken together with our observation that DNA methylation is actively removed when deposited on promoters and the recognition that DNA methylation often follows gene silencing in development, suggests that active chromatin is protected against DNA methylation in part by the TET proteins, and that more extensive changes to multiple DNA and histone modifications are required simultaneously to affect changes in chromatin structure and gene expression.

Here we show that the methylation of thousands of promoters in the human genome had, at the most, moderate effects on gene transcription and frequently resulted in no detectable repression of gene expression. Furthermore, DNA methylation did not promote the loss of the active histone mark H3K4me3 and methylation deposited by an artificial epigenetic modifier was actively removed, most likely by the TET proteins. Our results not only support a model by which the cell maintains the CpG-dense regions around transcription start sites in the unmethylated state, but also highlights potential technical barriers to the remodelling of the epigenome using artificial epigenetic modifiers.

Acknowledgements

This work was funded by the Australian National Health and Medical Research Council (GNT1069830), the Australian Research Council (ARC) Centre of Excellence program in Plant Energy Biology (CE140100008), the National Institutes of Health (R01CA170370 and 1R01DA036906), the Raine Medical Research Foundation, the National Cancer Institute (P30CA014089), and the National Human Genome Research Institute (R21HG006761). RL was supported by an ARC Future Fellowship (FT120100862) and Sylvia and Charles Viertel Senior Medical Research Fellowship. O.B. was supported by an Australian Research Council Discovery Early Career Researcher Award (DE140101962). PB was supported by an ARC Future Fellowship (FT130101767) and by a Cancer Council Western Australia Research fellowship. We would like to thank Andrea Holme at the University of Western Australia Centre for Microscopy, Characterisation and Analysis and Ji Kevin Li at the Harry Perkins Institute of Medical Research Flow Cytometry Facility for performing the FACS and flow cytometry analysis.

Materials and Methods

Cell culture

The MCF-7 ZF-DNMT3A and MCF-7 empty vector cell lines are previously described[16]. Cells were grown in MEM α nucleosides (Life Technologies) supplemented with 0.075% sodium bicarbonate and 10% tetracycline-free fetal bovine serum (Clonetech). Doxycycline inductions were performed by adding doxycycline to a final concentration of 100 ng/ml.

Propidium iodide staining

Ten centimeter cell culture dishes with cells at 30% confluency were grown for 48 hours with or without cell cycle inhibitors and harvested by incubation in PBS with 5 mM EDTA and spinning at 900 x g in a swinging bucket rotor. The cell pellet was loosened by vortexing and ice cold 70% ethanol was added dropwise while vortexing and then incubated on ice for 1 hours. The cells were spun at 900 x g for 5 minutes in a swinging bucket rotor and washed twice by resuspending in PBS and spinning at 900 x g for 5 minutes. Finally, the cells were resuspended in 100 ul PBS and 5 ul 100 mg/ml RNase A and incubated for 5 minutes at room temperature. 100 ul of 100 ug/ml propidium iodide was added and the cells were analyzed on a FACScanto.

Whole Genome Bisulfite Sequencing by MethylC-seq

Genomic DNA was extracted from cells using the Qiagen Blood and Tissue Kit according to the manufacturer's instructions. 1 ug of genomic DNA was spiked with 0.5% (w/w) of unmethylated lambda phage DNA for the calculations of the non-conversion rate and sheared with a Covaris S2 sonicator to an average length of 200 bp. The sheared DNA was end-repaired, A-tailed and ligated to methylated Illumina TruSeq adapters and subjected to 4 cycles of PCR amplification using KAPA HiFi Uracil+ DNA polymerase (KAPA Biosystems). 100 bp single-end sequencing was performed on an Illumina HiSeq 1500. Reads were mapped to the human genome (hg19) with the Bowtie alignment algorithm[53] with the following parameters: -e 1120 –O 20 –n 1 as previously reported[54].

TAB-seq

Genomic DNA was isolated as described for MethylC-seq. TAB-seq libraries were generated using the 5hmC TAB-seq kit (WiseGene) according to the manufacturer's instructions. 5-hydroxymethylated pUC19 DNA (WiseGene) was used to estimate the protection of 5hmC by β-glucosyltransferase and unmethylated lambda phage DNA was used to estimate the bisulfite conversion rate. Single-end 100bp sequencing was performed on a HiSeq1500. Reads were mapped as described for MethylC-seq.

RNA-seq

RNA was extracted from cells using the Qiagen RNeasy kit. 330 ng of total RNA was used to generate libraries using the Illumina TruSeq Stranded mRNA kit according to the manufacturer's instruction, except one-third of all reaction volumes were used and the final amplification used 10 cycles of PCR. Reads were mapped to the human genome assembly hg19 and transcriptome using TopHat2[55]. Uniquely mapped reads were assigned to genes with HTseq-count and differentially expressed genes and fold-change of expression was called with DESeq2 [56]. FPKM values were calculated with CuffDiff2[57]. For all analyses only genes with greater than 20 uniquely mapped reads were used.

ChIP-seq

Chromatin immunoprecipitation followed by next-generation DNA sequencing (ChIP-seq) for the HA-tag (for HA-tagged ZF-DNMT3A localization), as well as for the H3K4me3 histone modification was performed as described in[29].

ChIP-bisulfite-sequencing

Two 15 cm plates of cells were grown and doxycycline induced for three days. Cells were washed two times with 10 ml of PBS and crosslinked for 5 minutes in 50 mM HEPES-KOH, pH 7.5, 100 mM NaCl, 1mM EDTA, 1% formaldehyde. The crosslinking reaction was quenched by the addition of glycine to a final concentration of 125 mM and washed two times with phosphate buffered saline (PBS). All subsequent solutions were supplemented with a protease inhibitor cocktail from Sigma (Cat. # P8340). Cells were scraped off the plates with a rubber policeman in 10 ml of PBS and spun at 3,000 rpm for 5 minutes in a swinging bucket rotor. The cell pellets

were resuspended in 10 ml of 50 mM HEPES-KOH, pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, incubated on ice for 10 minutes and centrifuged at 3,000 rpm for 10 minutes in a swinging bucket rotor. Cell pellets were washed two times by adding gently adding 10 mM Tris-HCl, pH 8.1, 200 mM NaCl, 1 mM EDTA to the cell pellets trying not to disturb the pellets and centrifuged at 3,000 rpm for 5 min. Finally the cell pellets were resuspended in 0.1% SDS, 1 mM EDTA and transferred to a Covaris TC12x12 tube. The chromatin was sheared with a Covaris S2 sonicator with the following settings: Time 12 min, duty cycle 5%, intensity 4, cycles per burst 200, temperature 4˚C, power mode frequency sweeping. Triton X-100 and NaCl were added to a final concentration of 1% and 150 mM respectively. The sheared chromatin was centrifuged at maximum speed in a microfuge for 15 minutes at 4˚C and the supernatant was transferred to a new tube. 2 ul of anti-H3K4me3 (Diagenode, Cat. # C15410003) or 4 ul of anti-phospho-Ser5 RNA polymerase antibody (Active Motif, Cat. # 39233) was added and incubated overnight at 4˚C. 30 ul of Protein G Dynabeads (Life Technologies) was added and incubated on a tube rotator for 90 minutes at 4˚C. The beads were washed two times with 20 mM HEPES-KOH, pH 7.9, 0.1% SDS, 150 mM NaCl, 1% Triton X-100 2 mM EDTA, two times with 20 mM HEPES-KOH, pH 7.9, 0.1% SDS, 500 mM NaCl, 1% Triton X-100 2 mM EDTA, one time with 100 mM Tris-HCl pH 7.5, 0.5 M LiCl, 1% NP-40, 1% Sodium Deoxycholate and one time with 10 mM Tris-HCl, pH 8.0, 1 mM EDTA. The DNA was eluted twice by incubating for 30 minutes in 25 ul of 20 mM HEPES-KOH, pH 7.9, 1 mM EDTA, 0.5% SDS, 0.5 mg/ml Proteinase K. To the 50 ul of eluted DNA, 3 ul of 3M Sodium Acetate, pH 5.3 and 0.5 ul 30 mg/ml RNase A was added and incubated overnight at 65˚C in a hybridization oven. 1.5 ul of 20 mg/ml proteinase K was added and incubated for 1 hour at 50˚C and the DNA was purified with a 2X volume of a homemade version of AMPure XP beads and eluted in 20 ul Tris-HCl, pH 8.0, 0.1 mM EDTA. Libraries were made with the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the manufacturer's instructions. Reads were aligned and methylation was called as it was described for MethylC-seq above except after adapter trimming 10 nt were hard cropped off the 3' end of the read.

RNA fluorescence-in-situ-hybridization and FACS sorting

Cells were labeled with fluorescent probes to the *GAPDH* gene or the *PSME1* gene using the PrimeFlow RNA kit (eBiosciences) according to the manufacturer's instructions. Cells were sorted into high and low expression populations on a FACS Aria2 by the Harry Perkins Cell Sorting Facility. The sorted cell populations were aliquoted into two tubes for RNA and DNA extraction.

RNA isolation and quantitation from FACS-sorted cells

RNA was extracted using the Qiagen RNeasy FFPE kit according the manufacturer's instructions. Exon spanning primers were designed to the *GAPDH* gene and *RPL13A* gene as an internal control. RNA was reverse transcribed with gene specific primers using SuperScriptII (Life Technologies) and qPCR was performed on an LC480 thermocycler using KAPA SYBR FAST DNA qPCR Master Mix (Kapa Biosystems). For the PSME1 high and low expressing population, the RNA was quantitated by RNA-seq. Ribosomal RNA was depleted from the sample using the RiboCop rRNA Depletion Kit (Lexogen) and libraries were generated with the

TruSeq Stranded mRNA Kit (Illumina). The libraries were sequenced on an Illumina HiSeq 1500 with single-end 100 bp sequencing. The reads were adapter and quality trimmed with cutadapt [58] and transcripts per million were calculated using Kallisto[59].

Bisulfite-PCR amplicon sequencing

DNA was extracted from FACS sorted cells and bisulfite converted with the EZ DNA Methylation-Direct Kit (Zymo Research). PCR amplicons to the *GAPDH* and *PSME1* promoters were designed with methprimer and genomic DNA was amplified with 40 cycles of PCR using EpiMark Hot Start Taq DNA Polymerase (New England Biolabs). PCR reactions were pooled and purified with a homemade version of AMPureXP beads. 1 ug of pooled PCR products in 14.5 ul were phosphorylated by adding 15 ul 2X Quick Ligase Buffer (New England Biolabs) and 0.5 ul T4 polynucleotide kinase (New England Biolabs) and incubating at 37˚C for 30 minutes. Illumina TruSeq adapters synthesized by IDT and annealed by heating to 99˚C and slow cooling to 20˚C were ligated to the phosphorylated PCR products by adding 3.75 ul 10 uM annealed TruSeq adapters, 10 ul 2X Quick Ligase Buffer (New England Biolabs) and 6 ul water. The ligation reactions were incubated at 25˚C for 20 minutes and stopped by adding 2 ul 0.5 M EDTA. The DNA was purified by adding 20.8 ul (0.4 volumes) of homemade AMPure XP beads. The libraries were subjected to single-end 300 bp sequencing on the Illumina MiSeq. The reads were adapter trimmed with cutadapt and aligned to the human genome (hg19) with BS-Seeker2 [60] and bowtie2 [61]. Methylation was called using bs-seeker2.

DNA methylation data analysis

DMRs between MCF-7 empty vector doxycycline induced cells and MCF-7 ZF-DNMT3A doxycycline induced cells were called using the software package DSS[62]. DSS was run on the data with a low p-value cutoff to obtain the DMR boundaries with the following parameters: smoothing=TRUE, smoothing.span=500, delta=0.1, p.threshold=0.05, minCG=4, dis.merge=1000, minlen=100, pct.sig=0. To obtain a more stringent set of DMRs, DSS was run a second time with the same parameters, except the p-value threshold was decreased to 0.005. DMRs from the less stringent set that were also found in the more stringent set were used as the final set of DMRs for subsequent analyses.

UMRs in MCF-7 empty vector doxycycline induced cells were called using the software package methylSeekR[31] with the FDR cutoff set to 5. To filter UMRs that were called in large hypomethylated regions, UMRs that were within 40 kb were merged into single genomic intervals. The resulting set of genomic intervals were filtered for regions that were larger than 20kb and had an average methylation of less than 30%. UMRs form the original set that did not overlap with the UMRs from large unmethylated regions were used for further analyses.

Methylation levels in DMRs and UMRs were calculated by dividing the sum of C base calls in the genomic interval by the sum of C + T base calls in the genomic interval, for all C bases in the reference genome in the interval. Methylation heatmaps were made by dividing DMRs into 100 bp bins and calculating sum of C base calls in the genomic interval by the sum of C + T base calls in each bin. The resulting matrix was plotted with MeV[63]. Smooth scatter plots were

made with the smooth.scatter R function. All other plots were made with ggplot2 in R. Replicate methylomes were compared by selecting DMRs that had at least 20 base calls in all samples and plotting the resulting heatmap with the R function p.heatmap using the default parameters.

UMRs were assigned to genes by calculating the distance of the UMRs to the nearest transcriptional start site annotated by UCSC[64] with the closestBed function of the BEDTools software package [65]. UMRs that were fell within 2 kb of a transcriptional start site were assigned to the respective gene.

ChIP-seq and ChIP-bisulfite-sequencing data analysis

For ChIP-seq reads were mapped with the bowtie software package[53] and non-uniquely mapped reads were discarded. For ChIP-bisulfite-seq reads were mapped and methylation was called as described for MethylC-seq. Peaks were called using MACS1.4[66]. H3K4me3 peaks from ZF-DNMT3A no-dox and ZF-DNMT3A dox-induced samples were combined and overlapping peaks were merged to the outer boundaries of each peak. Heatmaps were generated by dividing DMRs into bins of 50 bp and counting the reads in each bin. The resulting matrices were plotted with MeV [63]. ChIP-seq reads for H3K4me1 and H3K27ac in MCF-7 cells were obtained from the GEO database (GSE38447)[67].

Data access

DNA methylation, TAB-seq, ChIP-seq, and RNA-seq data for this publication can be accessed at the Gene Expression Omnibus (GEO) under the accession number (GSE102395) and viewed in the following UCSC genome browser session (https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=ethanford&hgS_otherUserSessionName=FordEtAlBioRxiv2017).

Figure Legends

Figure 1. ZF-DNMT3A induces widespread DNA methylation. (A) Schematic representation of the ZF-DNMT3A doxycycline inducible system. **(B)** Genome browser screenshot of the *SOX2* locus. **(C)** Genomic distribution of identified ZF-DNMT3A binding sites. **(D)** Genomic distribution of differentially methylated regions (DMRs) between empty vector dox-induced and ZF-DNMT3A dox-induced cells. **(E)** Genomic distribution of unmethylated regions (UMRs) in empty vector dox-induced cells. **(F)** Venn diagram showing the intersection of ZF-DNMT3A binding sites, DMRs and UMRs. **(G)** Heatmap showing ΔmCG (left) and ZF-DNMT3A ChIP signal in the 10,356 DMRs (right).

Figure 2. ZF589-DNMT3A induced DNA methylation is lost in the absence of the artificial epigenetic modifier. Comparison of methylation levels in DMRs between **(A)** empty vector dox-induced cells and ZF-DNMT3A dox-induced cells, and **(B)** between empty vector dox-induced cells and ZF-DNMT3A dox-withdrawn cells. **(C)** Methylation levels of DMRs in empty vector dox-induced, ZF-DNMT3A dox-induced and ZF-DNMT3A dox-withdrawn cells. Comparison of methylation levels in UMRs between **(D)** empty vector dox-induced cells and

ZF-DNMT3A dox-induced cells, and **(E)** between empty vector dox-induced cells and ZF-DNMT3A dox-withdrawn cells. **(F)** Methylation levels of UMRs in empty vector dox-induced, ZF-DNMT3A dox-induced and ZF-DNMT3A dox-withdrawn cells. **(G)** Heatmap of DMRs showing DNA methylation, H3K4me3 read density, H3K4me1 read density and H3K27ac read density ordered by H3K4me3 read density. **(H)** Heatmap of DMRs showing DNA methylation, H3K4me3 read density, H3K4me1 read density and H3K27ac read density ordered by H3K4me1 read density.

Figure 3. DNA methylated by ZF-DNMT3A is bound by H3K4me3 and phospho-ser5-RNA polymerase II. (A) Heat map showing H3K4me3 read density in DMRs with ΔmCG > 0.2 (top panel) and ΔmCG < 0.2 (bottom panel). **(B)** Scatter plot with trend line of normalized read counts in DMRs with a ΔmCG > 0.2 from H3K4me3 ZF-DNMT3A no-dox ChIP-seq and ZF-DNMT3A dox-induced ChIP-seq experiments. **(C)** Scatter plot comparing ΔmCG with the change in normalized read counts from H3K4me3 ChIP-seq in ZF-DNMT3A no-dox and ZF-DNMT3A dox-induced cells. **(D)** ChIP-bisulfite-sequencing in ZF-DNMT3A no-dox cells and ZF-DNMT3A dox-induced cells with an anti-H3K4me3 antibody. **(E)** ChIP-bisulfite-sequencing in ZF-DNMT3A no-dox cells and ZF-DNMT3A dox-induced cells with an anti-phospho-ser5 RNA polymerase II antibody (right panel).

Figure 4. DNA methylation is lost in the absence of ZF-DNMT3A by an active process. (A) Western blot showing the induction and loss of the ZF-DNMT3A protein upon doxycycline induction and withdrawal. **(B)** DNA methylation levels in the *DACH1* promoter region showing the induction and loss of DNA methylation upon doxycycline induction and withdrawal. **(C)** The effect of the cell cycle inhibitors Lovastatin, R0886 and thymidine on DNA methylation levels in the *DACH1* promoter region during doxycycline induction and withdrawal. **(D)** mC (left) and hmC (right) levels in DMRs ordered by the change in fraction of hydroxymethylated CpG sites between ZF-DNMT3A no-dox and ZF-DNMT3A dox-induced cells. **(E)** Heat map of hmC levels +/- 2 kb of DMRs ordered by hmC density in ZF-DNMT3A dox-induced cells (left panel) or ZF-DNMT3A no-dox cells (right panel).

Figure 5. Forced DNA methylation of gene promoters correlates with a minor reduction in mRNA abundance. (A) Correlation between the change in DNA methylation levels in UMRs and the change in mRNA abundance of the associated gene upon induction of ZF-DNMT3A. **(B)** Distribution of fold-change in mRNA levels in genes with an associated UMR with a ΔmCG > 0.3. **(C)** Genome browser screenshots of three genes with heavily methylated promoter regions that have no reduction in mRNA abundance of the respective genes.

Figure 6. Cells sorted into high and low expressing subpopulations show similar levels of DNA methylation. (A) Fluorescence distribution of cells stained with FISH probes targeting the GAPDH mRNA before and after sorting by FACS. **(B)** RT-qPCR quantitation of GAPDH mRNA after FACS sorting. **(C)** DNA methylation levels of single cytosines at the *GAPDH* promoter region in the low (horizontal axis) and high (vertical axis) expressing GAPDH sorted cell populations. **(D)** Fluorescence distribution of cells stained with FISH probes targeting the

PSME1 mRNA before and after sorting by FACS. **(E)** RNA-seq quantitation of PSME1 mRNA after FACS sorting. **(F)** Methylation levels of single cytosines at the *PSME1* promoter region in the low (horizontal axis) and high (vertical axis) expressing PSME1 sorted cell populations. **(G)** Genome browser screenshots of the *GAPDH* (top) and *PSME1* (bottom) promoter regions showing methylation levels in the low and high expressing GAPDH sorted cell populations. **(H)** Per-read methylation levels of reads with ten or more cytosines in the CpG context that occur in DMRs with an average fraction of methylated cytosines in the CpG context between 0.4 and 0.6 (blue line). Aggregate methylation levels from WGBS data of ZF-DNMT3A dox-induced cells for read intervals (red line).

Supplementary Figure Legends

Supplemental Figure S1. (A) Genome browser screenshots of a representative locus showing off-target DNA methylation by ZF-DNMT3A in the TMC6/TMC7 promoter regions.

Supplemental Figure S2. (A) Methylation levels in DMRs of biological replicates. **(B)** PCA plot of biological replicate RNA-seq datasets.

Supplemental Figure S3. Kinetics of DNA methylation by ZF-DNMT3A expression and withdrawal in the promoter region of six selected genes.

Supplemental Figure S4. **Cell cycle inhibitors do not affect the kinetics of DNA methylation by ZF-DNMT3A expression and withdrawal. (A)** FACS analysis of cells stained with propidium iodide with and without treatment with cell cycle inhibitors. **(B)** Cell cycle distribution before and after drug treatment. **(C)** Shown are methylation levels during the time course of doxycycline induction and withdrawal in the presence of the cell cycle inhibitors Lovastatin (G1 block), thymidine (S-phase block), RO-336 (G2/M block) and untreated control.

Supplemental Figure S5. Genome browser screenshots showing DNA methylation levels and RNA expression levels for genes with promoters that are methylated by ZF-DNMT3A but display no decrease in steady state mRNA levels upon expression of ZF-DNMT3A.

Supplemental Figure S6. (A) Repression of transcription by ZF-DNMT3A does not correlate with expression level (top), number of CpG dinucleotides in UMR (middle) or CpG density in UMR (bottom). **(B)** UMRs with a ΔmCG > 0.25 were divided into repressed and non-repressed genes. Genes in the repressed and non-repressed categories were required to have a log2 fold change of less than -1 or greater than -0.1 mRNA abundance between the no ZF and ZF-DNMT3A dox-induced data sets respectively.

1. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).

2. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
3. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
4. Boyes, J. & Bird, A. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* **11,** 327–333 (1992).
5. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39,** 457–466 (2007).
6. Rhee, I. *et al.* DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416,** 552–556 (2002).
7. Yang, X. *et al.* Gene Reactivation by 5-Aza-2′-Deoxycytidine–Induced Demethylation Requires SRCAP–Mediated H2A.Z Insertion to Establish Nucleosome Depleted Regions. *PLoS Genet.* **8,** e1002604 (2012).
8. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* **27,** 31–39 (2001).
9. Bogdanovic, O. *et al.* Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Res.* **21,** 1313–1327 (2011).
10. Hontelez, S. *et al.* Embryonic transcription is controlled by maternally defined chromatin state. *Nat. Commun.* **6,** 10148 (2015).
11. Hammoud, S. S. *et al.* Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* **15,** 239–253 (2014).
12. Nunna, S., Reinhardt, R., Ragozin, S. & Jeltsch, A. Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells.

PLoS One **9,** e87703 (2014).

13. Kungulovski, G. *et al.* Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenetics Chromatin* **8,** 12 (2015).
14. Siddique, A. N. *et al.* Targeted Methylation and Gene Silencing of VEGF-A in Human Cells by Using a Designed Dnmt3a–Dnmt3L Single-Chain Fusion Protein with Increased DNA Methylation Activity. *J. Mol. Biol.* **425,** 479–491 (2013).
15. Rivenbark, A. G. *et al.* Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7,** 350–360 (2012).
16. Stolzenburg, S. *et al.* Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. *Oncogene* **34,** 5427–5435 (2015).
17. Xu, G. L. & Bestor, T. H. Cytosine methylation targetted to pre-determined sequences. *Nat. Genet.* **17,** 376–378 (1997).
18. Grimmer, M. R. *et al.* Analysis of an artificial zinc finger epigenetic modulator: widespread binding but limited regulation. *Nucleic Acids Res.* **42,** 10856–10868 (2014).
19. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44,** 5615–5628 (2016).
20. Bernstein, D. L., Le Lay, J. E., Ruano, E. G. & Kaestner, K. H. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J. Clin. Invest.* **125,** 1998–2006 (2015).
21. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233–247.e17 (2016).
22. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167,** 219–232.e14 (2016).
23. McDonald, J. I. *et al.* Reprogrammable CRISPR/Cas9-based system for inducing

site-specific DNA methylation. *Biol. Open* **5,** 866–874 (2016).

24. Lo, C.-L., Choudhury, S. R., Irudayaraj, J. & Zhou, F. C. Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. *Sci. Rep.* **7,** 42047 (2017).
25. Bashtrykov, P. *et al.* Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. *Chem. Biol.* **19,** 572–578 (2012).
26. Bestor, T. H., Edwards, J. R. & Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 6796–6799 (2015).
27. Stolzenburg, S. *et al.* Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. *Nucleic Acids Res.* **40,** 6725–6740 (2012).
28. Bestor, T. H., Edwards, J. R. & Boulard, M. Reply to Wilkinson: Minor role of programmed methylation and demethylation in mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E2117 (2015).
29. Grimmer, M. R. *et al.* Analysis of an artificial zinc finger epigenetic modulator: widespread binding but limited regulation. *Nucleic Acids Res.* **42,** 10856–10868 (2014).
30. Maeder, M. L. *et al.* Rapid 'Open-Source' Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Mol. Cell* **31,** 294–301 (2008).
31. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41,** e155 (2013).
32. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).
33. 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).
34. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–258 (2012).

35. Statham, A. L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* **22,** 1120–1127 (2012).

36. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22,** 1128–1138 (2012).

37. Piccolo, F. M. & Fisher, A. G. Getting rid of DNA methylation. *Trends Cell Biol.* **24,** 136–143 (2014).

38. Javanmoghadam-Kamrani, S. & Keyomarsi, K. Synchronization of the cell cycle using lovastatin. *Cell Cycle* **7,** 2434–2440 (2008).

39. McFarlane, C. *et al.* The deubiquitinating enzyme USP17 is highly expressed in tumor biopsies, is cell cycle regulated, and is required for G1-S progression. *Cancer Res.* **70,** 3329–3339 (2010).

40. Ma, H. T. & Poon, R. Y. C. Synchronization of HeLa Cells. *Methods Mol. Biol.* **1524,** 189–201 (2017).

41. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

42. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **7,** 2159–2170 (2012).

43. Cui, C. *et al.* P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. *Genome Biol.* **16,** 252 (2015).

44. 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).

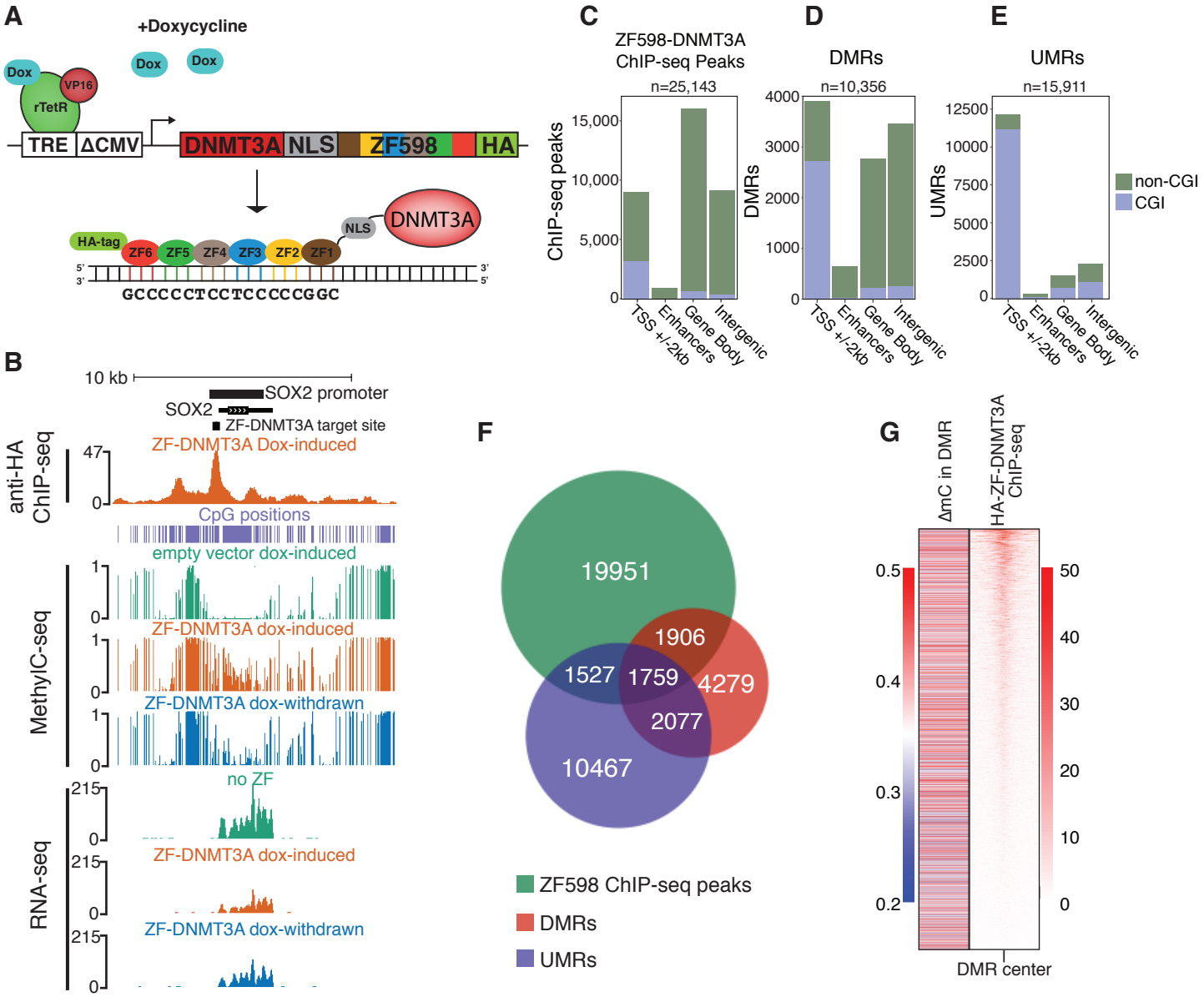
45. Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517,** 640–644 (2015).
46. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins—guardians of CpG islands? *EMBO Rep.* **13,** 28–35 (2012).
47. Kong, L. *et al.* A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. *Nucleic Acids Res.* **44,** 8682–8692 (2016).
48. Zhang, H. *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20,** 1390–1393 (2010).
49. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42,** 451–464 (2011).
50. Xu, Y. *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151,** 1200–1213 (2012).
51. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).
52. Lupo, A. *et al.* KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Curr. Genomics* **14,** 268–278 (2013).
53. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
54. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
55. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for

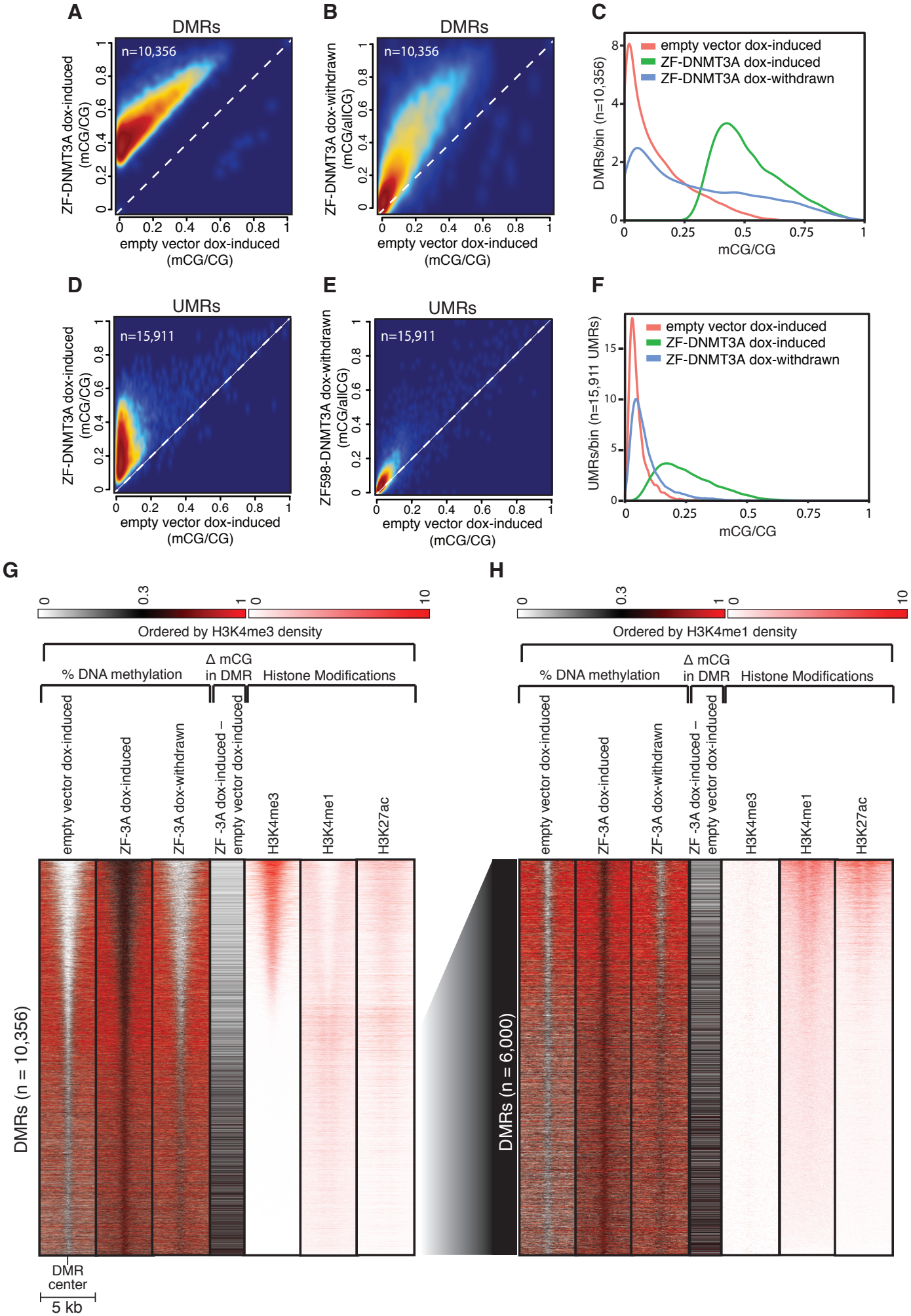
RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

57. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31,** 46–53 (2013).
58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
59. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).
60. Guo, W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14,** 774 (2013).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
62. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42,** e69 (2014).
63. Howe, E. A., Sinha, R., Schlauch, D. & Quackenbush, J. RNA-Seq analysis in MeV. *Bioinformatics* **27,** 3209–3210 (2011).
64. Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43,** D670–81 (2015).
65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
66. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Curr. Protoc. Bioinformatics* **Chapter 2,** Unit 2.14 (2011).
67. Frietze, S. *et al.* Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* **13,** R52 (2012).

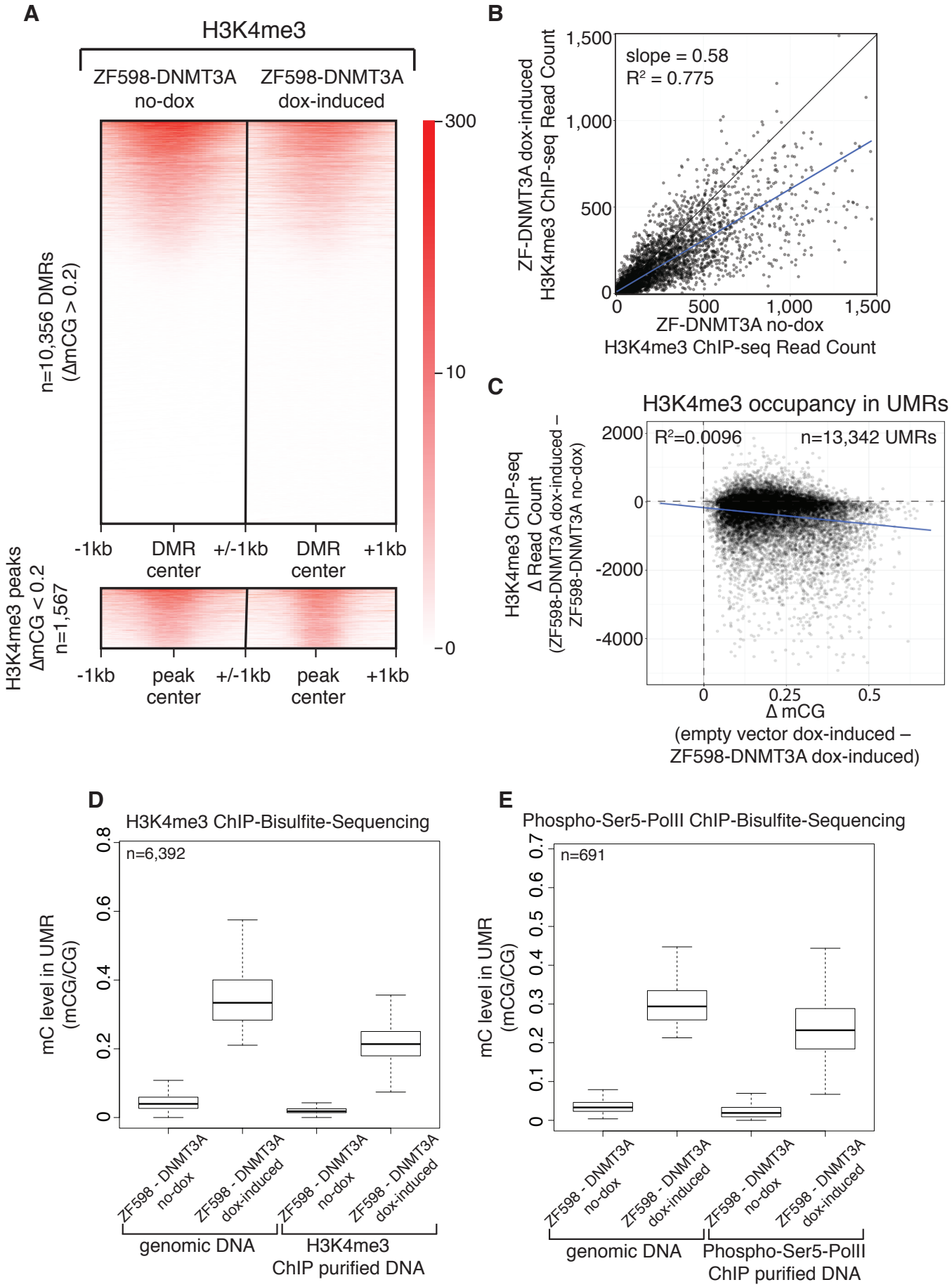
67. Frietze, S. *et al.* Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* **13,** R52 (2012).

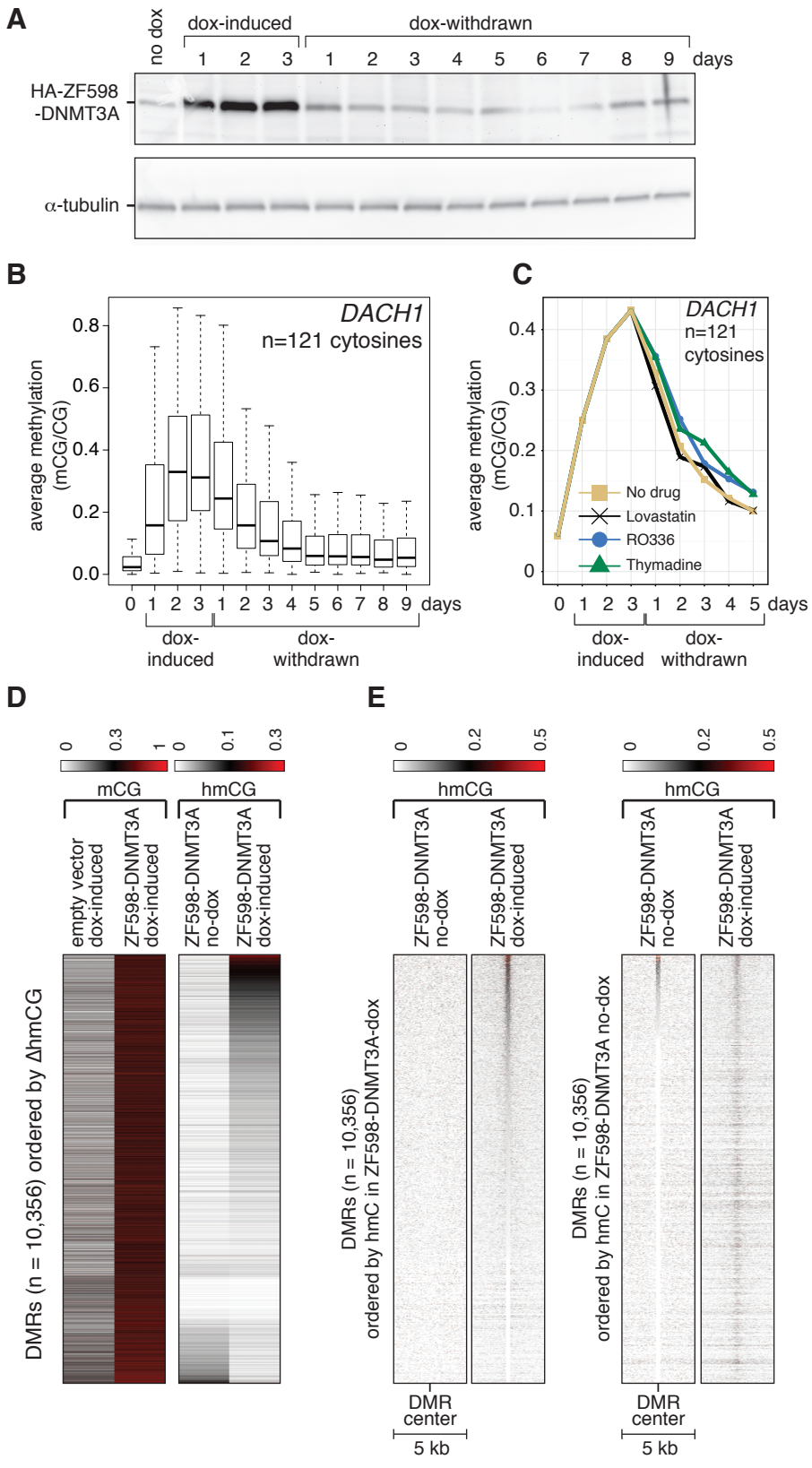
Ford et al. Figure 1

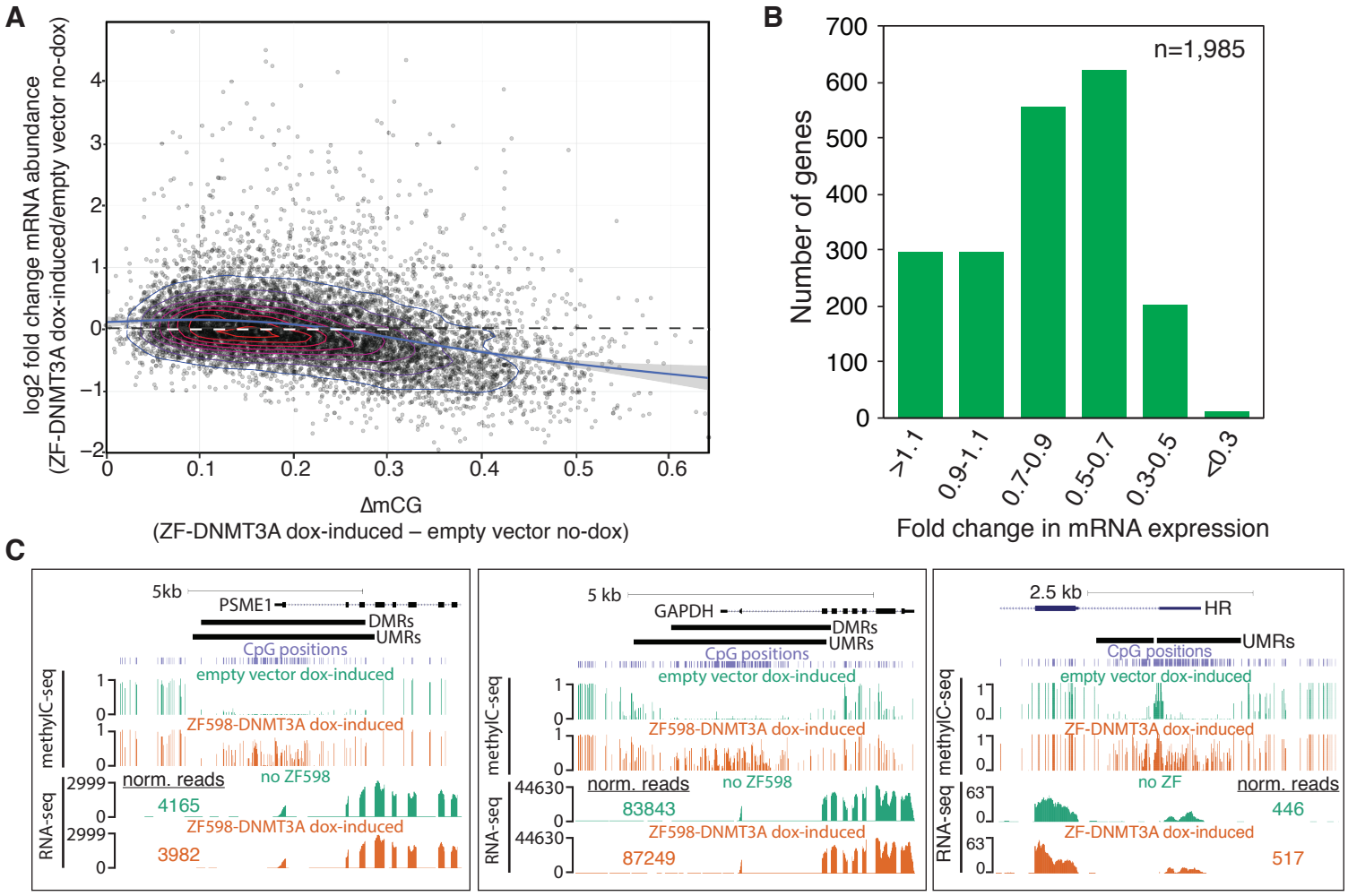




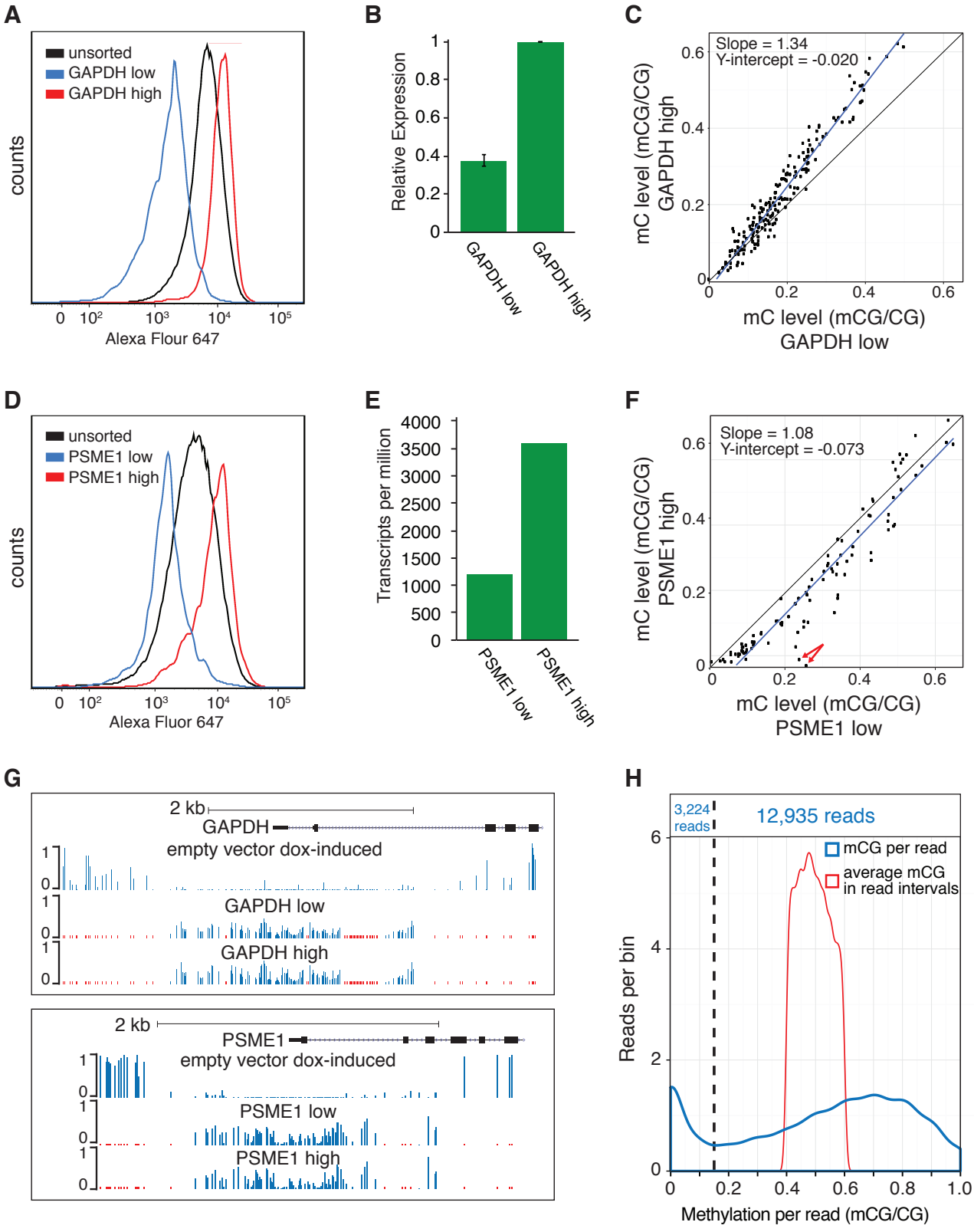
Ford et al. Figure 3



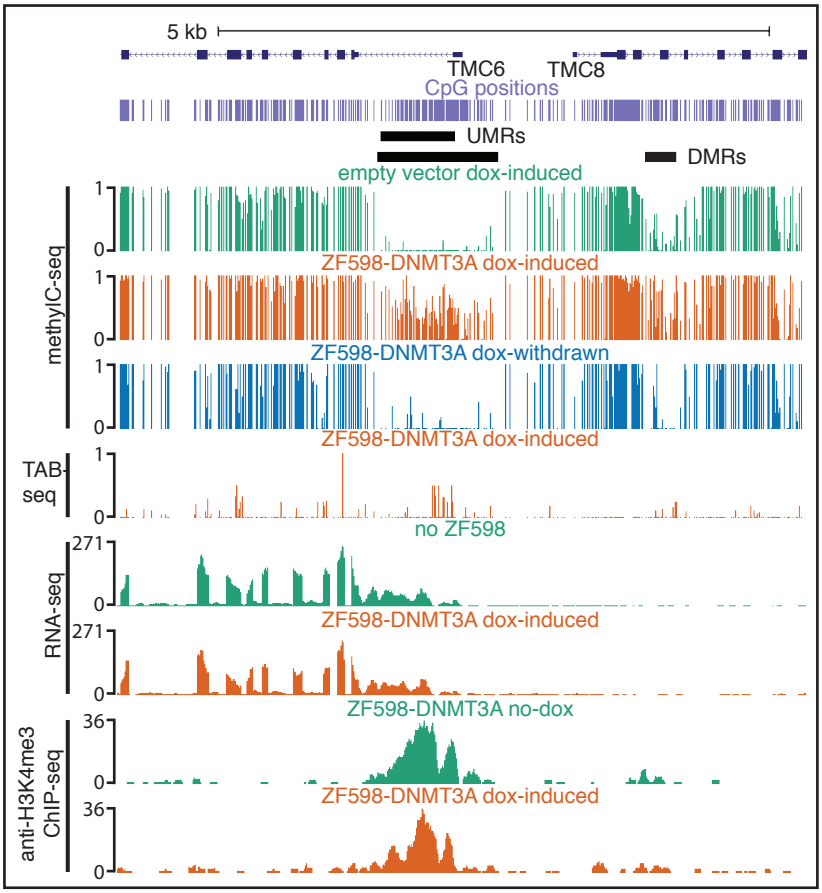




Ford et al. Figure 6

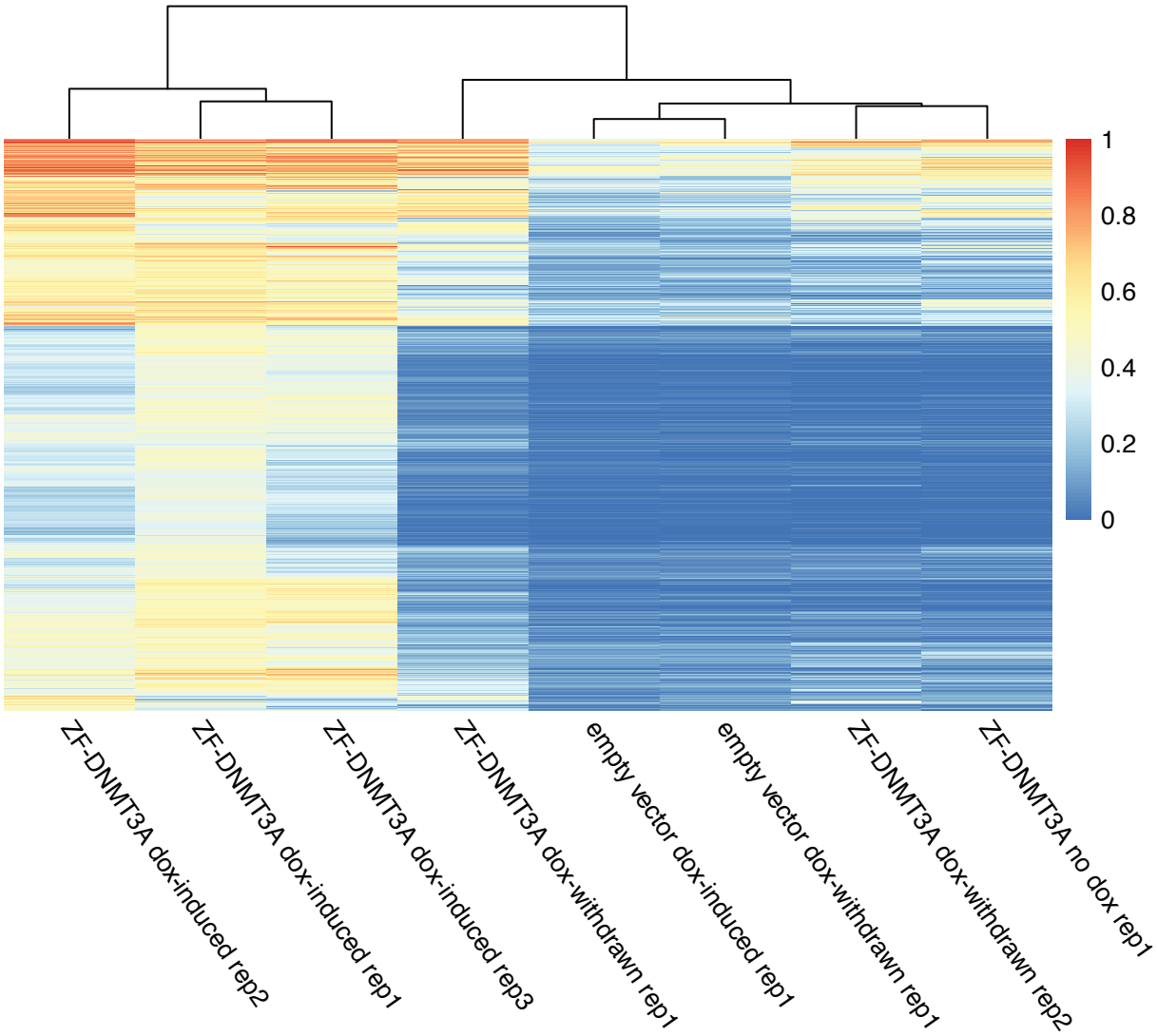


Ford et al.
Supplemental Figure S1

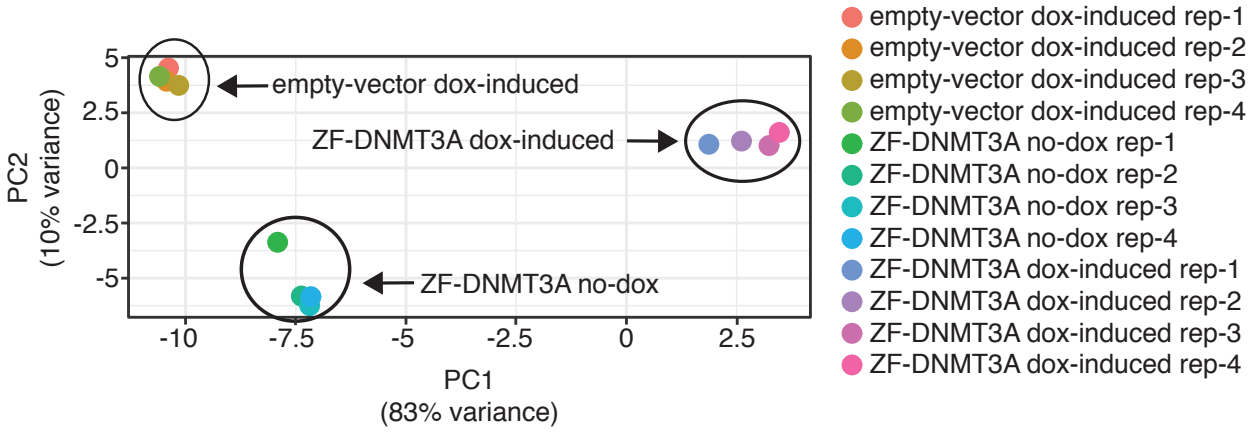


Ford et al.
Supplemental Figure S2

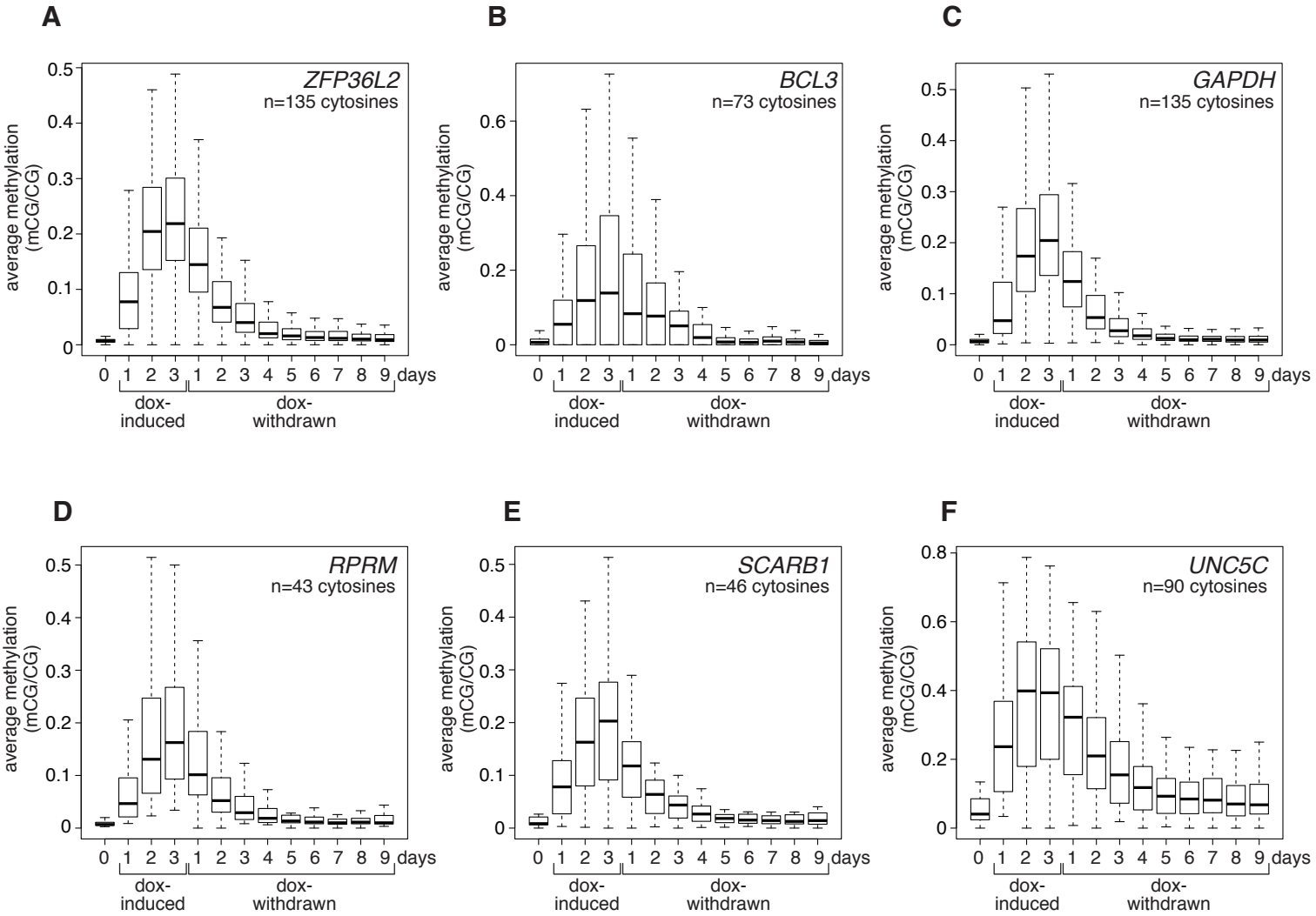
A



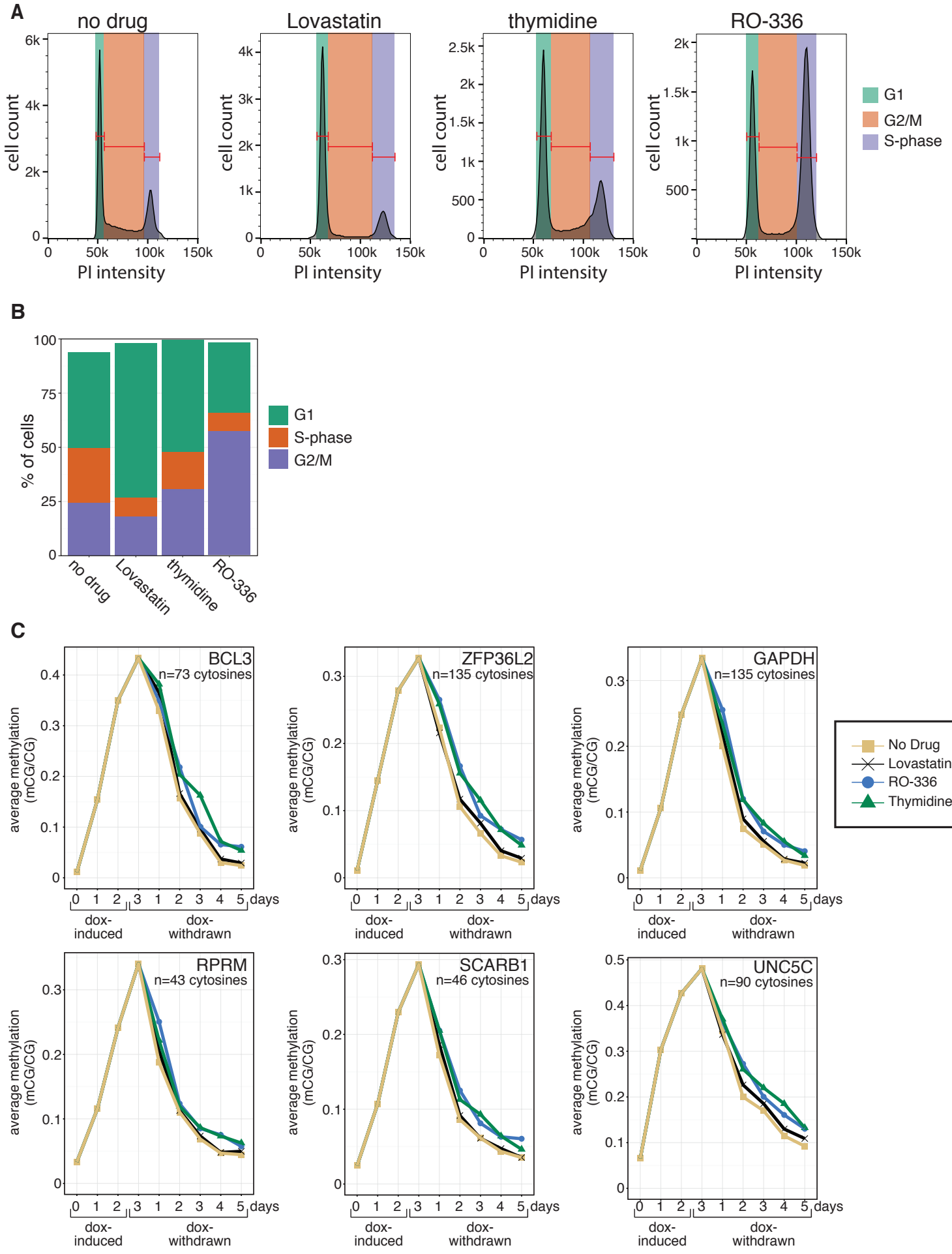
B



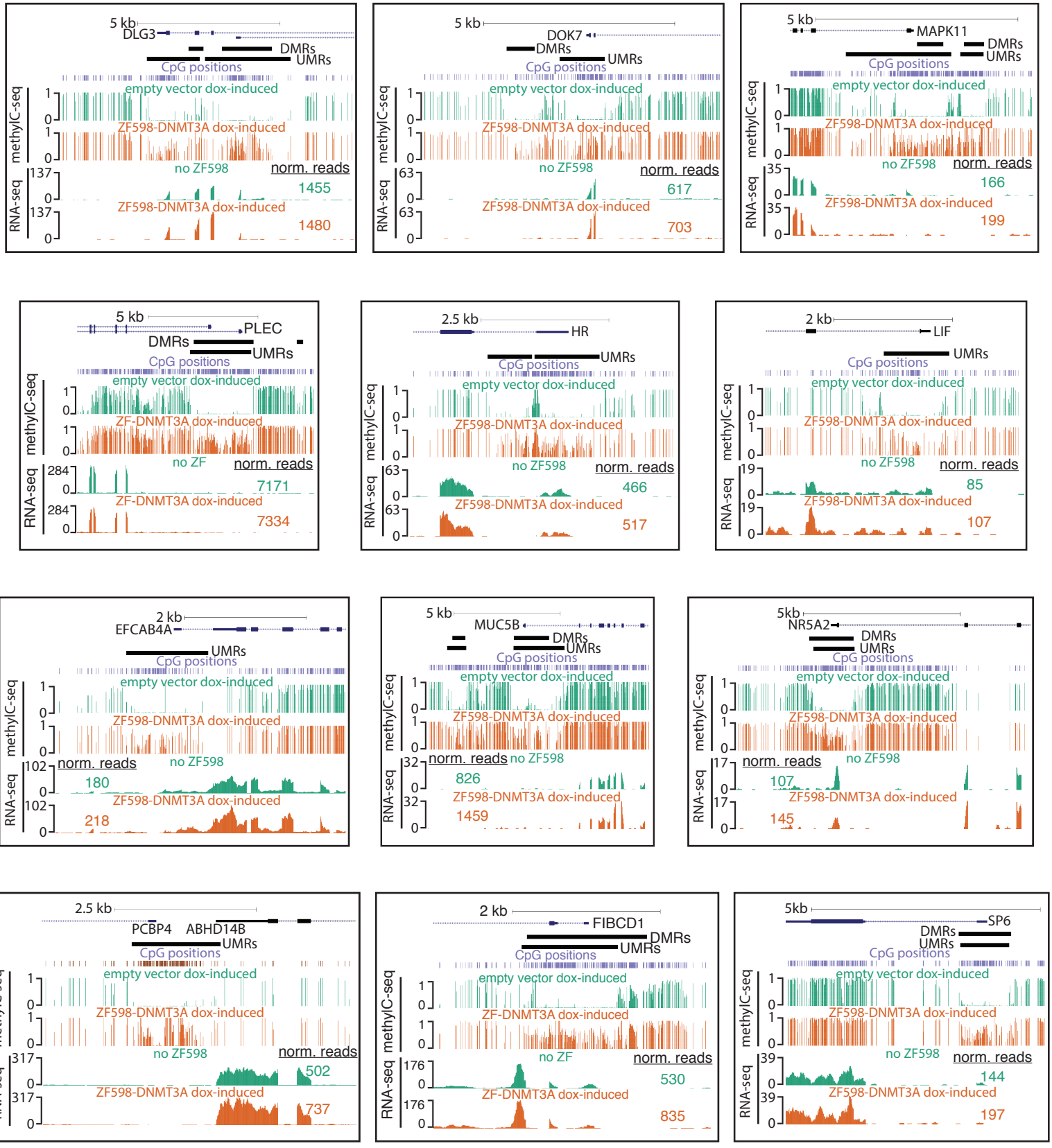
Ford et al.
Supplemental Figure S3



Ford et al.
Supplemental Figure S4

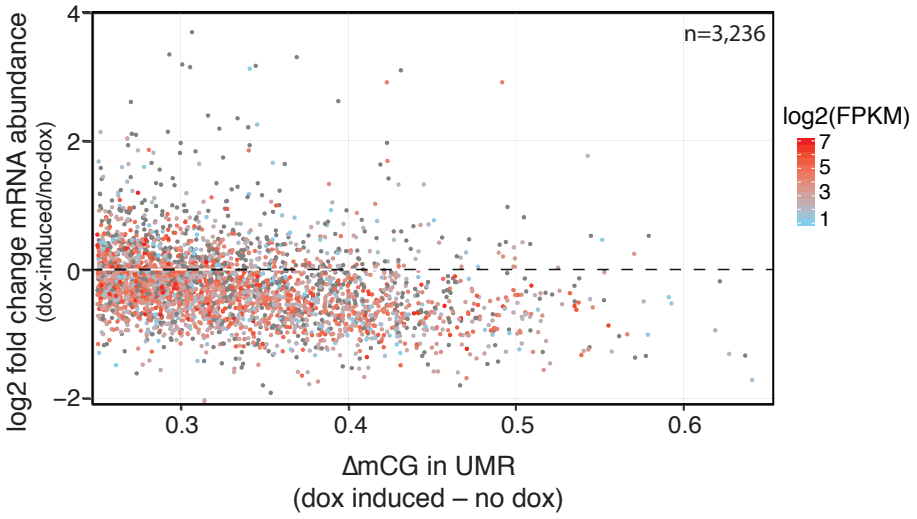


Ford et al.
Supplemental Figure S5



Ford et al.
Supplemental Figure S6

A



B