Ectopic methylation of a single persistently-unmethylated CpG in the promoter of the vitellogenin gene abolishes its inducibility by estrogen through attenuation of USF binding.

Lia Kallenberger[1,2,3], Rachel Erb[1,2,3,5], Lucie Kralickova[3], Andrea Patrignani[4], Esther Stöckli[2], Josef Jiricny[1,2,3 *]

[1]*Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

[2]*Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

[3]*Institute of Biochemistry, Swiss Federal Institute of Technology (ETH), Otto-Stern-Weg 3, CH-8093 Zurich, Switzerland*

[4]*Functional Genomics Center of the University of Zurich and the ETH Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

[5]*Present address: Rachel Erb, Sandoz Pharmaceuticals AG, Suurstoffi 14, CH-6343 Rotkreuz, Switzerland*

Running title: E-box methylation abolishes VTG estrogen inducibility

Funding: This work was funded by the Swiss National Science Foundation (grants no. 31003A-149989 and 31003B-170267 to J.J.).

***** To whom correspondence should be addressed. Tel: +41 44 633 6260; Email: jjiricny@ethz.ch

ABSTRACT

The enhancer/promoter of the vitellogenin II (*VTG*) gene has been extensively studied as a model system of vertebrate transcriptional control. While deletion mutagenesis and *in vivo* footprinting identified the transcription factor (TF) binding sites governing its tissue specificity, DNase hypersensitivity- and DNA methylation studies revealed the epigenetic changes accompanying its hormone-dependent activation. Moreover, upon induction with estrogen (E_2), the region flanking the estrogen-responsive element (ERE) was reported to undergo active DNA demethylation. We now show that although the *VTG* ERE is methylated in embryonic chicken liver and in LMH/2A hepatocytes, its induction by E_2 was not accompanied by extensive demethylation. In contrast, E_2 failed to activate a *VTG* enhancer/promoter-controlled luciferase reporter gene methylated by *Sss*I. Surprisingly, this inducibility difference could be traced not to the ERE, but rather to a single CpG in an E-box (CACGTG) sequence upstream of the *VTG* TATA box, which is unmethylated *in vivo*, but methylated by *Sss*I. We demonstrate that this E-box binds the upstream stimulating factor USF1/2. Selective methylation of the CpG within this binding site with an E-box-specific DNA methyltranferase *Eco*72IM was sufficient to attenuate USF1/2 binding *in vitro* and abolish the hormone-induced transcription of the *VTG* gene in the reporter system.

INTRODUCTION

In vertebrates, the DNA methyltransferases DNMT1, DNMT3a and DNMT3b convert around 90% of cytosines in the CpG sequence context to 5-methylcytosines (Eckhardt et al., 2006; Gruenbaum et al., 1981; Rakyan et al., 2004). DNMT3a/b are believed to be the major *de novo* methyltransferases that modify unmethylated DNA, while DNMT1, often termed “maintenance methylase”, is believed to be the enzyme that copies the methylation pattern of the template strand onto the newly-synthesised strand following DNA replication or repair (Bacolla et al., 1999; Bestor,

1992; Hsieh, 1999; Okano et al., 1998; Pradhan et al., 1999). All three enzymes are essential for survival, as demonstrated by the fact that DNMT knock-out mice show early lethality (Li et al., 1992; Okano et al., 1999).

DNA methylation is largely erased during fertilisation, but the DNMTs lay down a new methylation pattern during early embryogenesis that will control the subsequent stages of development and differentiation. In general, gene bodies become densely-methylated, while gene regulatory sequences are methylated sparsely and in a highly-divergent manner. For example, many housekeeping genes are flanked by the so-called CpG islands. Although these regions are CpG-rich, they are generally unmethylated and the genes they control are constitutively-active (Bird et al., 1985; Cooper et al., 1983; Gardiner-Garden and Frommer, 1987). In contrast, CpG islands associated with imprinted genes or retroviral sequences are methylated, as are genes on the inactive X chromosome (Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997) and some become methylated during development (De Smet et al., 1996; De Smet et al., 1999), which leads to transcriptional silencing. Once established, DNA methylation patterns remain largely stable and unprogrammed changes such as the aberrant methylation of CpG islands are often linked to aging or tumorigenesis (Baylin et al., 1986; Gama-Sosa et al., 1983; Goelz et al., 1985; Noreen et al., 2014; Toyota et al., 1999). While the latter phenomena have been extensively studied, less attention has been paid to the dynamic changes of DNA methylation taking place outside of CpG islands (Eden and Cedar, 1994). These changes are often triggered by exogenous stimuli in a highly tissue-specific manner and are directly involved in the regulation of gene expression (Amenya et al., 2016; Kangaspeska et al., 2008; Metivier et al., 2008; Thomassin et al., 2001; Toker et al., 2013) by altering the binding affinity of TFs such as c-Myc/Myn (Prendergast and Ziff, 1991), E2F (Campanero et al., 2000), AP2 (Comb and Goodman, 1990), NF-κB (Kirillov et al., 1996) or USF1/2 (Fujii et al., 2006) for their cognate sequences.

One well-studied example of an inducible tissue-specific gene that is also regulated by DNA methylation is vitellogenin II (*VTG*). The gene encodes a precursor of egg yolk protein and is

present in all oviparous species. It is expressed exclusively in the female liver, but can be induced in males by estrogen (Saluz et al., 1986). This property brought it into recent limelight, because its expression in males can be used as a measure of estrogenic endocrine disruptive chemicals (EDCs) in the environment (Diamanti-Kandarakis et al., 2009). As in other species, the chicken *VTG* gene is expressed in the liver of mature hens, but not roosters. This difference was explained by the silencing of the *VTG* gene by sex-specific DNA methylation, because its transcriptional activation in rooster liver by a single β-estradiol (E2) injection was accompanied by demethylation of a *Hpa*II site within the estrogen response element (ERE) (Wilks et al., 1984; Wilks et al., 1982) and the appearance of DNaseI hypersensitive sites in the enhancer and promoter (Burch and Weintraub, 1983). Subsequent Church & Gilbert sequencing of the genomic DNA showed that the transcription was activated already after 6 hours and that this event coincided with the demethylation of four CpGs (**a-d**) in the non-transcribed strand flanking the ERE (Fig. 1A). Because loss of methylation through replication (the so-called "passive demethylation") could be excluded, this phenomenon was hailed as the first example of active demethylation (Saluz et al., 1986).

We set out to study the above phenomenon in greater detail, because we wanted to learn whether the demethylation was an obligate step in the activation of *VTG* expression and, if so, whether it involved the recently-discovered machinery of active DNA demethylation that makes use of the ten-eleven-translocation methylcytosine dioxygenase (TET) enzymes (Hassan et al., 2017; He et al., 2011; Tahiliani et al., 2009) and TDG (Neddermann and Jiricny, 1993; Wiebauer and Jiricny, 1989). We made use of chicken embryos (Burch and Weintraub, 1983), or of a chicken hepatoma LMH/2A cell line stably-expressing the estrogen receptor (ERα), (Binder et al., 1990; Philipsen et al., 1988; Seal et al., 1991; Sensel et al., 1994), both of which had been used to study *VTG* expression in the past. We also made use of a reporter plasmid that was devoid of CpGs and in which the expression of the luciferase gene was under the sole control of the *VTG* enhancer/promoter.

We now show that estrogen-dependent *VTG* activation in these experimental systems was not accompanied by significant demethylation. More importantly, we show that the ability of the gene to reactivate from methylation-dependent silencing is controlled by an unmethylated E-box element distal to the ERE sequences. Methylation of this E-box abolished the ability of the silenced gene to be reactivated by estrogen. We further show that the E-box binds the upstream-stimulating factor USF1/2.

RESULTS

The *VTG* enhancer/promoter is methylated and the gene is silenced, but exposure to β-D-estradiol induces transcription independently of DNA replication.

We first wanted to reproduce the phenomenon described by Saluz *et al.* (Saluz et al., 1986). However, due to restrictions on animal experimentation, we had to search for alternative systems. We were also interested in identifying an experimental set-up that would be amenable to manipulation. To this end, we decided to test whether the *VTG* gene is silenced and inducible in the liver of 10 day-old chicken embryos. We made small windows in freshly-fertilised eggs, sealed them with a microscope coverslip and hot wax (Baeriswyl and Stoeckli, 2006) and incubated them at 39˚C for 9 days. Eggs containing live embryos were then treated with an ethanol solution of E_2 or with ethanol alone. 24 hours later, the embryos were sacrificed, the livers were excised, immersed in RNA*later* and nucleic acids and proteins were immediately isolated. In parallel, we treated chicken LMH/2A cells, a Leghorn rooster hepatocellular carcinoma cell line stably expressing estrogen receptor alpha (ERα) (Sensel et al., 1994) in a similar manner, but nucleic acids and proteins were isolated after 6 and 24 hours.

Bisulphite sequencing of DNA isolated from mock-treated liver and LMH/2A cells revealed that CpGs a/2, c and d in the *VTG* enhancer (Fig. 1A) were ~90% methylated in both strands, whereas CpG b/1 appeared to be undermethylated in both LMH/2A cells (Fig. 1B) and embryonic liver (Fig. S1A). This is in line with the results of Saluz *et al.*, where CpGs a/2, c and d in rooster

liver DNA were reported to be fully-methylated, whereas CpG b/1 was hemimethylated (Saluz et al., 1986).

Using RT-qPCR, we could show that the *VTG* gene was transcribed neither in LMH/2A cells nor in embryonic liver prior to exposure to estrogen, but that it was induced upon a 24 hours E_2 treatment of chicken embryos (Fig. S1B) and even more efficiently in LMH/2A cells, where substantial transcription was detected already 6 hours after the addition of 100 nM E_2 (Fig. 1C). Its expression increased further after 24 hours and the increase continued up to 72 hours (Fig. S1C), possibly due to a positive feedback leading to increased transcription of ERα (Fig. S1D). This activation occurred in a replication-independent manner, as inhibition of DNA replication by the addition of the B-family polymerase inhibitor aphidicolin did not impede transcription, but rather enhanced it (Fig. 1D). A possible explanation for this observation is that interference of replication with the transcription process is inhibited by the addition of aphidicolin. That DNA synthesis was indeed inhibited was confirmed by a lack of 5-ethynyl-2'-deoxyuridine (EdU) incorporation into nuclear DNA during the course of the experiment (Fig. S1E).

Estrogen-dependent induction of *VTG* transcription in our system was not accompanied by demethylation of the enhancer/promoter region.

We next set out to investigate whether the estrogen-dependent transcriptional activation of the *VTG* gene was accompanied by demethylation as reported (Saluz et al., 1986). We therefore performed PCR on the bisulphite-converted DNA to distinguish cytosines (C) from 5-methylcytosines (mC). Because active demethylation is believed to involve the dioxygenases TET1-3, we also wanted to detect the intermediate of the oxidation, 5-hydroxymethylcytosine (hmC). To distinguish between mC and hmC, the bisulphite conversion was preceded by an oxidation step using $KRuO_4$, which selectively oxidizes hmC to fC, resulting in its conversion to uracil following bisulphite treatment (Booth et al., 2012). The PCR products were then sequenced using PacBio, which provides information on single molecules. HmC levels were assessed as the difference in the amount of

converted cytosines between bisulphite- and oxidized-bisulphite converted DNA. Using this approach, we could show that the *VTG* promoter/enhancer was not demethylated during the course of the induction, but that a small percentage of mCs were oxidized to hmCs, indicating a possible involvement of the TET enzyme(s) (Fig. S1F).

ERα binds to the *VTG* estrogen response element upon treatment with β-estradiol *in vivo*.

As mentioned in the Introduction, DNA methylation can attenuate transcription by interfering with the binding of TFs to their respective recognition sequences. Because we failed to detect demethylation in our system, ERα, the key activator of the *VTG* gene, would have had to bind to the methylated estrogen response element (ERE) in the enhancer upon E_2 treatment. In order to test this hypothesis, we performed a chromatin immunoprecipitation experiment using an ERα antibody or, as a control, a flag antibody. We could retrieve ERα in the chromatin fraction only with the ERα antibody and upon E_2 treatment (Fig. 1E, upper panel). Moreover, we could confirm its binding to the *VTG* ERE using RT-qPCR on the recovered DNA (Fig. 1E, lower panel). Only very little signal was apparent without E_2 treatment, confirming the widely-accepted model that estrogen receptors bind DNA *in vivo* only in response to hormone treatment (Klinge, 2001). Moreover, this showed that the receptor is able to bind to its recognition sequence also in methylated chromatin.

Binding of ERα to the ERE is insensitive to different cytosine modifications and hormone treatment.

To confirm that the binding of ERα to its cognate recognition sequence was indeed unaffected by methylation, we carried out a series of electrophoretic mobility shift assays (EMSAs), using synthetic oligonucleotides containing the *VTG* ERE (WT) or a variant with a single base pair deletion in the three-nucleotide spacer between the palindromic repeats (ΔG, Fig. 2A) that has been reported to abolish ERα binding *in vitro* (Klinge, 2001). Recombinant ERα was able to bind the ERE sequence with similar affinity in the presence or absence of E_2, the only notable difference

being the slightly-increased mobility of the shifted band in the presence of the hormone (Fig. 2A, lanes 2,3 and Fig S2A,B). The specificity of the protein/DNA complex was confirmed by a competition assay; addition of a 100-fold excess of the unlabelled ERE oligo duplex significantly diminished the intensity of the shifted band (lane 4, spec), while a scrambled sequence failed to do so (lane 5, unspec). In addition, addition of an ERα antibody to the reaction substantially retarded (supershifted) the mobility of the specific band (lane 7) as compared to the addition of the same amount of BSA (lane 6). In the EMSA assays, the affinity of ERα for a substrate symmetrically-methylated at the two CpGs within the ERE (CpGs c and d, mC/mC) was similar to that seen with the unmethylated substrate C/C (Fig. 2B) and the same was true for hemi- and fully-hydroxymethylated substrates, as well as substrates containing formyl- or carboxycytosine (Fig S2C,D).

The ERE is bound by a factor other than ERα in nuclear extracts of LMH/2A cells.

We were interested to learn whether the oligonucleotide substrates were bound with similar selectivity and affinity also by ERα present in nuclear extracts of LMH/2A cells. In order to limit non-specific binding, we used shorter substrates than in the previous experiments. As seen in Fig. 2C, the ERE substrates were efficiently bound by recombinant ERα irrespective of methylation (lanes 1, 3), whereas the ΔG oligonucleotide duplex failed to bind the receptor (lane 5). Unexpectedly, the mobility of the shifted band generated by the unmethylated oligo (C/C) upon incubation with LMH/2A nuclear extracts (NE) was faster than that seen with ERα (lane 2) and the factor that bound the unmethylated oligo bound only weakly to the methylated (mC/mC) one (lane 4). In contrast, the ΔG substrate was bound with very high affinity (lane 6). In order to ensure that the factor binding the oligo substrates in the extracts was distinct from ERα, we titrated increasing amounts of the recombinant protein into the NE. As shown in Fig. 2D (whole gel is shown in Fig S2E), ERα outcompeted the nuclear factor (referred to as Factor X) in reactions containing the C/C

substrate and even more efficiently in reactions containing the mC/mC oligo, but not the ΔG substrate.

In order to gain information regarding the substrate preference of Factor X, we carried out an EMSA assay using eight oligonucleotide substrates carrying sequences flanking the ten CpGs in the *VTG* enhancer/promoter (Fig. 1A; the two CpGs in the ERE were in a single duplex, as were CpGs 5 and 6, because of their proximity). Because the sequence was reported by Saluz *et al.* (Saluz et al., 1986) to be demethylated in a strand-specific manner, we also included hemimethylated substrates. Surprisingly, in addition to oligos ERE and ΔG, Factor X bound also to oligos containing CpG b/1, CpG a/2 and CpG7, but not CpG3, CpG4, CpG5-6 and CpG8 (Fig. 2E and Fig. S2F-H). Moreover, its affinity for hemimethylated substrates varied, which suggested that its sequence- and methylation specificity was unusually relaxed.

To characterize Factor X further, we carried out the EMSA assays with oligos containing one BrdU residue on each strand (Fig. 3A, Table S2). Upon incubation with the NE, half of the mixture was used for the EMSA experiment, while the other half was UV-crosslinked and the proteins were then resolved by SDS-PAGE. During the cross-linking, the radiolabelled oligo becomes covalently attached to the protein that binds it and the protein size can thus be estimated from the position of the radioactive band on the SDS-PAGE, once the molecular weight of the oligo is subtracted. As shown in Fig. 3B, the recombinant ERα-oligo complex migrated at the expected size of ~80 kDa (lane 5) and a band of similar size was seen in LMH/2A NE cross-linked to the C/C, mC/C or mC/mC substrates (lanes 1-4). In addition, a second, prominent band migrated at around 50 kDa, irrespective of substrate. After subtracting the molecular weight of the single-stranded oligo (~15 kDa), the size of Factor X was predicted to be around 35-40 kDa. When the recombinant receptor was titrated into the reaction with NE, we saw a weak, but reproducible, competition with the smaller protein (Fig. 3C). The ~80 kDa band could be outcompeted with a specific (lane 8, spec), but not with an unspecific (lane 7, unspec) competitor, further confirming that it was a complex of the oligo with ERα. The addition of E_2 did not alter the binding affinities or

ratios of the different proteins to the substrates (Fig. S3A). The migration of the ERα band was also unchanged in the presence of the hormone, confirming that the mobility shift seen in the EMSA assay represented a conformational change of the receptor, which is eliminated upon denaturation.

Factor X is the E-box-binding heterodimer of upstream stimulating factors USF1/2.

We wanted to learn whether Factor X was a chicken-specific protein, or whether it was present also in man. We therefore studied the binding properties of the eight oligonucleotide duplexes containing the ten CpGs in the *VTG* enhancer/promoter (see above) in EMSA experiments using NEs of the ERα-positive breast cancer cell line MCF7 and the ERα-negative cervical carcinoma cell line HeLa. As shown in Fig. S3B,C, the electrophoretic mobilities of the protein/DNA complexes formed in HeLa (ERE, ΔG, CpG4 and 7) and MCF7 (CpG4, 5-6, 7 and 8) extracts with the indicated oligos were similar to those seen in LMH/2A extracts. We therefore assigned them to Factor X. A distinct mobility shift was seen with oligo CpG3 (Fig. S2G), but the CpG4, 5/6 and 8 substrates failed to form protein/DNA complexes (Fig. 2E, Fig. S2H, Fig S3B,C). Interestingly, Factor X binding to the different oligos displayed distinct methylation sensitivities; thus, while its binding to CpG2 was unaffected by methylation, the binding to oligos ERE, CpG1 and in particular CpG7 was substantially attenuated by methylation (Fig. 2 E). The similarity of the proteins binding the oligo substrates ERE and CpG7 was further confirmed in UV-crosslinking experiments (Fig. S3D), thus implying that Factor X is a generic DNA binding protein with an ambiguous sequence- and methylation specificity.

In an attempt to identify Factor X, we subjected nuclear extracts of HeLa cells to affinity chromatography. We first generated the DNA substrates by ligating the ERE (C/C or mC/mC), CpG7 (C/C or mC/mC) or ΔG oligonucleotide duplexes end-to-end so as to obtain molecules of 100-300 base pairs in length. The sticky ends were then filled-in with dATP and Bio-dUTP and the tagged molecules were allowed to attach to streptavidin Dynabeads. We then preincubated HeLa NEs with non-specific competitor poly (dI-dC) and allowed them to incubate with the beads as

described in Material and Methods. Following extensive washing, the bound proteins were eluted with high salt and the tryptic digests were analysed by mass spectrometry (Fig. 4 A). We searched for peptide sequences present preferentially in the CpG7 and ΔG but not in the methylated CpG7 elutions and originating from proteins with molecular size between 25 and 50 kDa. Beads-only and CpG8 were used as negative controls. The peptides that fulfilled these exclusion criteria to the greatest extent were the TFs DEC1, MAX, MLX, Myc, USF1 and USF2, firstly because they were identified with elevated frequencies (Table 1) and, second, because their preferred recognition sequences overlapped to a large extent with those of our affinity probes.

We immunodepleted HeLa nuclear extracts with antibodies against these possible Factor X candidates and performed EMSA assays. The USF1- or USF2-immunodepleted extracts (Fig. S4A) largely lost the ability to bind the CpG7 and ΔG duplexes and the fraction of the shifted oligos in the USF2-depleted extracts was clearly reduced (Fig 4B). The same could be shown with the unmethylated- and methylated ERE substrates (Fig. S4B). For a direct comparison, Fig. S4C shows an EMSA experiment in which the four oligo substrates were incubated with the USF1-depleted and IgG-depleted (negative control) extracts. To confirm the depletion specificity, we performed EMSAs with HeLa NE that were pre-incubated with antibodies raised against the above proteins. As shown in Fig. 4C, the antibody against USF1 caused a supershift of the complex, while the antibody against USF2 attenuated the binding of Factor X to the two duplexes. To confirm the specificity of the depletion further, we carried out a UV-crosslinking experiment with the USF1- or USF2-depleted extracts, as well as with proteins eluted from the beads with high salt. The effect of the depletion was not very pronounced (Fig. 4D, lanes 1-4), probably because EMSA assays detect complexes with relatively long life times, whereas UV-cross-links even transient complexes. However, a band corresponding in size to Factor X was present in the eluates from the USF1 and USF2 beads, but not from the control, IgG beads (Fig. 4D). This evidence strongly suggests that Factor X is a heterodimer of USF1 and USF2, which is known to be methylation-sensitive (Chen et al., 2012; d'Adda di Fagagna et al., 1995; Fujii et al., 2006). However, our data provide strong

evidence that this heterodimer is rather promiscuous in its sequence specificity, contrary to reports that describe its preference for the perfect E-box consensus CACGTG (d'Adda di Fagagna et al., 1995; Giacca et al., 1989) present e.g. in the CpG7 duplex.

Transcriptional activity of the *VTG* enhancer/promoter is controlled by E-box methylation and USF binding.

In the Saluz *et al*. study (Saluz et al., 1986), the silenced, methylated *VTG* gene was activated by estrogen and this event was accompanied by demethylation of CpGs **a-d** in the enhancer. These findings led to the assumption that the E_2-activated estrogen receptor triggered a series of events that brought about a demethylation of the ERE and the downstream sequences, and that this demethylation was a prerequisite for transcriptional activation. However, in our *in vivo* system, induction of *VTG* transcription with E_2 did not require demethylation (Fig. 1C and Fig S1F). We therefore had to consider the possibility that the gene was silent in rooster only because of the lack of the hormone and that its methylation was simply a mark of inactive chromatin. In order to elucidate this phenomenon, we generated a reporter vector based on the CpG-free pCpGL-basic plasmid, in which the luciferase gene is driven by the *VTG* enhancer/promoter. We first introduced into the *VTG* sequence unique *Kpn*I and *Hin*dIII sites on either side of the ERE and then ligated this enhancer/promoter into pCpGL-basic to generate VTG-CpGL (Fig S5A). Upon transfection of this reporter into LMH/2A cells, luciferase expression could be efficiently induced with E_2 (Fig 5A, ERE). Convertion of the wild type ERE sequence to ΔG by site-directed mutagenesis resulted in similar levels of basal transcription, but substantially lower inducibility (Fig. 5A), caused by the significantly lower affinity of ERα for the ΔG ERE lacking the spacer deoxyguanosine (Fig. 2B,C). *In vitro* methylation of these two reporters with *Sss*I largely abolished transcriptional inducibility (Fig. 5A). When the *Kpn*I/*Hin*dIII fragment was replaced with the wild type sequence, unmethylated, methylated or hydroxymethylated at the two CpGs, estrogen inducibility was unchanged (Fig. 5B). This result extends the *in vitro* findings (Fig. 2B, Fig. S2C) showing that the

311 binding of ERα is unaffected by methylation and shows that its transcriptional activity is also
312 methylation-independent. (Hydroxymethylation was included, because we detected low levels of
313 this modification in the *in vivo* activation experiments shown in Fig. S1F.)

314 Having shown that methylation of the ten CpGs in the *VTG* enhancer/promoter by *Sss*I
315 substantially attenuated E_2 inducibility of the reporter (Fig. 5A), but that methylation of the ERE
316 was without effect (Fig. 5B), we wanted to learn which CpGs were responsible for the
317 transcriptional silencing. We therefore converted CpGs 1-8 (Fig. 1A) to TpGs by site-directed
318 mutagensis. We could show that the C to T transition mutations at CpGs 1-6 and 8 failed to affect
319 the basal transcriptional activity of the reporter and reduce its inducibility by E_2. Indeed, the
320 inducibility was increased in some cases (Fig. S5B). In contrast, C to T mutations in CpGs **c** and **d**
321 in the ERE (mut ERE) and CpG7 (C594T) attenuated the inducibility upon E_2 treatment (Fig.
322 S5B,C). To learn whether the mutation in CpG7 affected USF1/2 binding, we tested the mutant
323 oligo in an EMSA assay and found that the affinity of the protein for the mutated sequence was
324 intermediate between that of the wt unmethylated CpG7 duplex and the methylated one (Fig. S5D).
325 This translated directly to luciferase expression, where the mutant CpG7 (C594T) showed
326 intermediate expression between the wt mock-methylated (wt C) and *Sss*I-methylated (wt mC)
327 reporters (Fig. S5B). The expression of all the mutant reporters was inhibited upon methylation
328 with *Sss*I, showing that the individual sites had no major influence on transcription of the reporter,
329 irrespective of whether they were methylated or unmethylated (Fig. S5E).

330 In order to exclude the role of CpG1-6 and 8 in the transcriptional regulation of the VTG
331 enhancer/promoter, we mutated them sequentially, such that the final mutant contained only CpG **c**
332 and **d** in the ERE and CpG7. We tested the single and multiple mutants to exclude the possibility
333 that interactions between different sites had additional influence on transcription, but this was not
334 the case. All combinations tested were active when unmethylated and inhibited when methylated
335 with *Sss*I (data not shown). Analysis of the final plasmid (ERE/7 wt) unmethylated and methylated
336 confirmed that the silencing of the *VTG* enhancer/promoter was mediated by methylation of CpG7

in combination with methylation of the ERE (Fig. 5C), even though methylation of ERE alone had no effect on reporter activity (Fig. 5B).

In order to demonstrate the key importance of CpG7 in the control of *VTG* transcription, we methylated the wt reporter with *Eco72I* methylase (Rimseliene et al., 1995), which modifies solely the CACGTG E-box sequence. As shown in Fig. 5D, methylation of this site substantially attenuated basal luciferase expression, as well as its inducibility by E_2, if not to the same extent as modification of all ten CpG sites by *Sss*I methylase. In contrast, methylation of the reporter with *Hpa*II methylase, which modifies CpG **d** in the ERE, had no appreciable effect on luciferase expression.

In a converse experiment, we wanted to reproduce the methylation pattern seen in the chicken embryo and in LMH/2A cells. To this end, we cleaved the reporter vector with *Pml*I, which cuts the E-box sequence and, as control, with *Hpa*II that cuts the ERE or *Xmn*I that cuts the vector backbone. We then methylated the linear DNA with *Sss*I and recircularised it with T4 DNA ligase. In the *Pml*I-cut vector, all CpGs with the exception of CpG7 were methylated, whereas only the *Hpa*II site in the ERE remained unmethylated in the *Hpa*II-cut vector and all CpGs were methylated in the *Xmn*I-cut vector. Transfection into LMH/2A cells showed that expression of the reporter luciferase gene could be efficiently induced by E_2 solely when CpG7 was unmethylated (Fig. 5E, *Pml*I). In contrast, the reporter gene in the *Hpa*II- or *Xmn*I-cleaved vectors, as well as a plasmid that was treated with *Sss*I without prior cleavage (undigested) were refractory to induction with E_2.

Finally, we wanted to confirm that the transcriptional control of *VTG* expression rests not only with CpG7, but also with the protein that binds the E-box: USF1/2. We therefore knocked down USF1 or luciferase (control) with siRNA (Fig. S5F-H) and transfected the cells with the reporter plasmids 48 hours later. Simultaneously, we assessed the induction of endogenous *VTG* in the siRNA-treated cells. Knock-down of USF1 reduced luciferase inducibility and the induction of endogenous *VTG* by more than 50% (Fig. 5F, G) or by a third in the case of luciferase inducibility with the second siRNA (Fig. S5H, left panel), which shows that USF1/2 plays a key role in the

control of *VTG* transcription. (The incompleteness of the silencing could be explained by residual levels of the factor in the siRNA-treated cells.) This prompted us to conclude that USF1/2 binding to CpG7 is key to the regulation of expression of *VTG* and that its likely role in the prevention of methylation of its binding site is essential to ensure that the gene remains poised for ready hormone activation.

DISCUSSION

We could confirm the findings of earlier studies and show that the chicken *VTG* gene was heavily-methylated already in early embryos, as well as in LMH/2A cells, and that it could be induced by a single dose of β-estradiol. The CpGs in the *VTG* enhancer/promoter seen to be demethylated in hormone-treated rooster liver (Saluz et al., 1986; Wilks et al., 1984; Wilks et al., 1982) remained resistant to bisulphite conversion in our experimental systems, however, we were able to detect low levels of the TET oxidation product hmC already 6 hours after E_2 treatment that increased further with longer exposure to E_2. The TET enzymes and TDG are present in both chicken embryonic and adult liver, as well as in the LMH2A cell line (data not shown), and it is therefore possible that the hormone-activated *VTG* gene is indeed targeted for demethylation, but that, unlike in the rooster, the demethylation machinery in our experimental set-up is not fully-functional. The finding that hmC proportion increased at the four CpGs upon hormone treatment at longer time points and that *VTG* induction was higher upon TET2 overexpression (data not shown) implied that the demethylation machinery can act on the *VTG* locus and that it has a positive effect on transcription. This could be further confirmed in cells treated with 5-azadeoxycytidine prior to induction with E_2. In these cells, *VTG* expression was several-fold higher than even in the TET2-overexpressing cells (data not shown).

Puzzlingly, we failed to find evidence of hmC oxidation to fC or caC, steps necessary for the TDG-initiated base excision repair process that completes the multistep demethylation process. It is conceivable that the TET-mediated cascade terminated in our system after a single oxidation

step, but this clearly wasn't the case in the rooster experiments, where the genomic DNA became susceptible to cleavage with *Hpa*II that does not cleave hydroxymethylated DNA (Wilks et al., 1984; Wilks et al., 1982). Another possibility is that fC and caC excision and the subsequent BER steps were extremely rapid and were followed by immediate remethylation. Although such cyclic DNA demethylation/methylation events have been described in the pS2 promoter upon hormone treatment, it is unlikely that this was so in our system, because the cycling process supresses transcription (Kangaspeska et al., 2008; Metivier et al., 2008).

Although the above phenomenon requires further study, our data demonstrate that while global DNA demethylation facilitates transcription of the endogenous *VTG* gene upon hormone activation, site-specific demethylation of the ERE region is clearly not necessary. This agrees with our *in vitro* data (Fig. 2B) showing that the ERα binding to the ERE was unaffected by methylation. Indeed, the hormone-activated receptor needs to be able to bind to its cognate sequence irrespective of its methylation status, in order to recruit the chromatin-remodelling- and possibly also the demethylation machinery that are needed to facilitate transcription.

Given the requirement for ERα binding in E_2-dependent *VTG* activation, the weak inducibility of the ΔG ERE reporter was puzzling, because the ΔG oligonucleotide did not detectably bind the receptor *in vitro* (Fig. 2B, C). However, it is possible that the observed induction was mediated by a second, imperfect ERE sequence, which lies around position -350 from the transcription start site (Fig. 1A).

As mentioned in the Introduction, the *VTG* enhancer/promoter has been studied extensively in the past and it could be shown that it contains several TF binding sites that control its basal- and liver-specific transcription, as well as its hormone-inducibility. Linker-scanning mutagenesis (Seal et al., 1991) revealed that deletion of the sequence between -113 and -335 eliminated the hepatocyte specificity, permitting the reporter contruct to be expressed also in fibroblasts. Importantly, insertion of a linker around position -50 abolished estrogen inducibility of the reporter. This mutation disrupted the E-box sequence, which contains CpG7 identified in this study as being

415 critical for hormone induction of the silenced gene. Clearly, the interaction of the E-box binding
416 factor (in our case the USF1/2 heterodimer) with the estrogen receptor is essential for hormone
417 activation, given that mutational inactivation (Seal et al., 1991) or methylation (this study) of the E-
418 box attenuates USF1/2 binding *in vitro* and the inducibility of the gene *in vivo*, as does mutation
419 (but not methylation) of the ERE (Fig. S5B, C).

420 The interaction of different TF combinations bound at promoters and enhancers potentially
421 provides cells with an enormous flexibility of gene expression, and DNA methylation extends this
422 range still further by altering the affinity of some, but not all, TFs for their respective recognition
423 sites. Moreover, the E-box hexamer consensus sequence CANNTG has been shown to bind a large
424 family of basic helix-loop-helix leucine zipper (b-HLH-LZ) proteins that can bind as homo- or
425 heterodimers, and often in methylation-sensitive manner (Fujii et al., 2006; Hou et al., 2012; Perini
426 et al., 2005; Prendergast et al., 1991; Prendergast and Ziff, 1991). Yet, in spite of the multitude of
427 proteins that could potentially bind to the *VTG* E-box (see Table 1) and the plethora of possible
428 protein/protein interactions (both activating and repressing) that this could generate, *VTG*
429 expression is strictly-controlled and appears to require exclusively USF1/2. This was unexpected.
430 USF1/2 has been reported to interact with ERα and to bind preferentially to the perfect E-box
431 sequence CACGTG and less so to CATGTG (Corre and Galibert, 2005; deGraffenried et al., 2004).
432 This could explain why the C to T mutation in CpG7 had such a deleterious effect on estrogen
433 inducibility (Fig. S5B, C). However, it does not explain why the factor bound so efficiently to the
434 different sequences in our EMSA assays (Fig. 2D,E; Fig. S2F,G,H). Similarly, methylation of the
435 E-box sequence would also lower the E_2 inducibility of the VTG enhancer/promoter, because the
436 affinity of USF1/2 for the methylated sequence is low (Fig. 2D, E). As noted above, however, there
437 are a number of TFs that could potentially bind to the E-box in the absence of USF1/2 and activate
438 VTG transcription, particularly as some – notably c-Myc – have been reported to interact with ERα
439 (Cheng et al., 2006). These questions require clarification in the future, as does the reason why the
440 *VTG* E-box sequence is the only CpG site out of the ten that remains unmethylated *in vivo*. It is

possible that its methylation is actively prevented, possibly due to the fact that it is rapidly bound by USF1/2 or other E-box proteins and thus that the access of methyl transferases to this site is hindered. This has been proposed as a possible mechanism that protects CpG islands from *de novo* methylation (Brandeis et al., 1994; Macleod et al., 1994), but whether inducible enhancers/promoters where only a single site is affected are subject to similar protection from methylation remains to be demonstrated.

DNA methylation changes, particularly hypermethylation, have been associated with a long list of pathologies and disorders, ranging from cancer, immune system dysegulation (Crohn's disease, Grave's disease, rheumatoid arthritis, type 1 diabetes), neurological disorders (Alzheimer's disease, Fragile X syndrome, schizophrenia, epilepsy, depression) and atherosclerosis to osteoporosis (Ehrlich, 2019). In the vast majority of the studies concerned with the human methylome, focus was placed on CpG islands and other CpG-rich regions, the hypermethylation (or indeed hypomethylation) could be seen to be associated with the given pathology. Our study highlights an important caveat of these studies, namely, that the alteration of the methylation status of a single CpG can have equally dramatic consequences as the hypermethylation of a CpG island. The outcome of both events can be the silencing – or at least attenuation - of transcription. Interestingly, the two main protagonists of this study, ERα and USF1/2, may be implicated (separately and together) in a subset of the above syndromes. USF1/2 plays a major role in cell proliferation control, as well as in the control of expression of the *TP53* (Hale and Braithwaite, 1995; Reisman and Rotter, 1993) and *BRCA2* genes but it has been reported to be transcriptionally inactive in three of six transformed breast cancer cell lines (Ismail et al., 1999), the widely-used MCF7 line among them, which has a hypermethylator phenotype (Xing and Archer, 1998). The two TFs interact in the transcriptional activation of the Cathepsin D gene, the dysregulation of which is linked to a number of the above-listed disorders (Dubey and Luqman, 2017). It will be interesting to learn whether some of these dysregulation events might be linked to the aberrant methylation of USF1/2/ E-boxes within the transcriptional control regions of genes involved in these pathologies.

MATERIALS AND METHODS

Cell culture

LMH/2A cells (ATCC CRL-2118) were grown in Williams' E medium (GIBCO) without phenol red, supplemented with 10% charcoal-stripped FCS, streptomycin/penicillin (100 U/ml) and L-glutamine (2.4 mM). MCF-7 cells were grown in DMEM (GIBCO) with phenol red, supplemented with 10% FCS, streptomycin/penicillin (100 U/ml) and insulin (Sigma-Aldrich, 10 µg/ml). Cell lines were grown at 37°C in a 6% CO_2 humidified atmosphere. Where indicated, cells were treated with 100 nM β-estradiol (E_2, ethanol solution, Sigma-Aldrich) or an equivalent volume of ethanol, or with 8 µM aphidicolin (aph, DMSO solution, Sigma-Aldrich) or an equivalent volume of DMSO. E_2 and/or aphidicolin were added fresh every 12 h, because of instability under cell culture conditions. RNA interference was carried out using Lipofectamine RNAiMAX (ThermoFisher) according to the manufacturer's instructions. SiRNA against chicken 1-USF1: $^{5'}$CCCAAUAUCAAAUAUGUCUUC$^{3'}$; 2-USF1: $^{5'}$UAUGUCUUCCGCACAGAGAUU$^{3'}$; against luciferase (control): $^{5'}$CGUACGCGGAAUACUUCGAdTdT$^{3'}$.

Handling and treatment of fertilized eggs

Eggs were prepared as described (Baeriswyl and Stoeckli, 2006) and kept at 39°C. At day 9, the eggs were treated with ethanol or 100 nM E_2 by pipetting ~55 µl (depending on weight of egg) of a 100 µM solution (10% EtOH) into the egg, through the window prepared at day 3 of development. 24 hours later, the embryos were sacrificed, the livers were excised and immersed in RNA*later* (Sigma). Livers were homogenized using a Tissue Ruptor (Qiagen) in RIPA buffer for proteins and in RLT buffer for RNA. Following homogenization, protein samples were processed further according to the Western Blot protocol and RNA was extracted using an RNAeasy Kit (Qiagen).

Western Blotting (WB)

Cells were collected using trypsin, washed in phosphate-buffered saline (PBS) and lysed in RIPA buffer (50 mM Tris pH 8, 1 mM EDTA, 1% NP-40, 0.5% deoxycholate (DOC), 0.1% SDS, 150 mM NaCl). Lysates were sonicated (18 s, 50 cycles, 70% amplitude; in a Bandelin Sonoplus GM70) and protein concentrations were measured using the Bradford assay. 30-80 µg of protein were boiled in 1x SDS loading buffer (50 mM Tris, pH 6.8, 10% glycerol, 1.6% SDS, 0.1 M DTT, 0.01% bromophenol blue) and separated on polyacrylamide gels (6-10%) in 10% SDS-running buffer at 130 V. Proteins were transferred onto activated PVDF membranes (Amersham Pharmacia Biotech) which were incubated with primary antibody overnight at 4°C in 5% milk, washed 3x with TBS-T and incubated for 1 hours at RT with the secondary antibody (horseradish peroxidase-conjugated sheep anti-mouse or donkey anti-rabbit IgG, GE Healthcare). Membranes were analyzed with a Fusion Solo (Vilber Lourmat). Antibodies used were mouse Anti-VTG (Abcam, ab36794), rabbit Anti-TFIIH p89 (Santa Cruz, sc-293), mouse Anti-ERα (ThermoFisher, MA5-13065), rabbit Anti-USF1 (GeneTex GTX16396), mouse Anti-USF2 (Santa Cruz, sc-293443).

Real-time quantitative PCR (RT-qPCR)

Cells were harvested using trypsin and RNA was extracted according to manufacturer's instructions (RNeasy, Qiagen). 2 µg of the extracted RNA were reverse-transcribed using the high capacity cDNA reverse transcription kit (Applied Biosystems) according to manufacturer's instructions. 125 ng of the cDNA were used for the PCR using the platinum SYBR green qPCR superMix-UDG kit (Invitrogen) according to manufacturer's instructions, except that the reaction was scaled down from 50 µl to 20 µl. The standard cycling program for ABI instruments was used (50°C for 2 min, 95°C for 2 min, 40x (95°C for 15 s, 60°C for 30 s) and melting curve analysis at 65°C for 15s, heating to 97°C with continuous acquisition per 5°C, 40°C for 30 s). The RT-qPCR was run on a LightCycler 480 (Roche). GAPDH primers were used as an internal control. Technical triplicates were made for every sample and primer. The primers used are listed in Table S1 (SI).

519

520

521 *Chromatin immunoprecipitation*

522 75 Mio LMH/2A cells were harvested after 24 hours treatment with 100 nM E_2 or EtOH. Cells
523 were washed twice with PBS and fixed for 10 min in 1% formaldehyde at RT. Fixation was
524 quenched with 125 mM glycine at RT for 10 min. Cells were collected by centrifugation at 300 x g
525 for 5 min at 4°C, washed twice with PBS and the pellet was resuspended in cell lysis buffer (10 mM
526 Tris, pH 8, 1 mM EDTA, 0.5% IGEPAL, protease inhibitors) and incubated on ice for 10 min. The
527 lysate was centrifuged at 2'000 x g at 4°C for 5 min. The pellet was resuspended in nuclear lysis
528 buffer (10 mM Tris, pH 8, 1 mM EDTA, 0.5 M NaCl, 1% Triton X-100, 0.5% sodium
529 deoxycholate, 0.5% lauroylsarcosine, protease inhibitors) and incubated on ice for 10 min with
530 repeated vortexing. Lysate was centrifuged at 3'000 x g for 5 min at 4°C. The pellet was
531 resuspended in PBS, split for different immunoprecipitations and centrifuged at 3'000 x g for 10
532 min at 4°C. Pellets were resuspended in 300 µl lysis buffer (10 mM Tris, pH 8, 1 mM EDTA, 150
533 mM NaCl, 0.1% sodium deoxycholate, 0.1% SDS, protease inhibitors) and sonicated at maximum
534 power for 10 min with 30 s intervals in a Bioruptor (Diagnode). Samples were centrifuged at 16'000
535 x g for 10 min at 4°C and the supernatant was collected. Input samples (15%) were frozen and the
536 rest was made up to 1 ml with IP buffer (16.7 mM Tris, pH 8, 1.2 mM EDTA, 300 mM NaCl, 1.1%
537 Triton X-100, protease inhibitors). 5 µg of the respective antibodies (Anti-ERα; ThermoFisher
538 MA5-13065 or Anti-Flag; Sigma F3165) were added to the lysate and incubated on a rotating wheel
539 at 4°C overnight.
540 The beads (G protein Sepharose, GE Healthcare) were washed twice in IP buffer and centrifuged
541 for 2 min at 2'700 x g before blocking with 1 mg/ml BSA in IP buffer for 1 h. After one additional
542 wash, 50 µl of beads were added to 1 ml lysate with the antibody. The mixture was incubated for 3
543 hours on a rotating wheel at 4°C. Mixtures were centrifuged for 2 min at 2'000 x g and the
544 supernatant was removed. The beads were washed successively in 1 ml of four different wash

buffers for 5 min on the rotating wheel (wash buffer 1: 20 mM Tris, pH 8, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 300 mM NaCl; wash buffer 2: like wash buffer 1, but with 500 mM NaCl; wash buffer 3: 10 mM Tris, pH 8, 1 mM EDTA, 1% sodium deoxycholate, 500 mM LiCl, 1% NP-40; wash buffer 4: 10 mM Tris, pH 8, 1 mM EDTA). One additional wash with wash buffer 4 was performed, during which the beads were divided for DNA and proteins. For proteins: the beads were taken up in 2x SDS loading dye (100 mM Tris, pH 6.8, 20% glycerol, 3.2% SDS, 0.2 mM DTT, 0.04% bromophenol blue), the mixture was vortexed vigorously and incubated for 30 min at 100°C (denaturing and de-cross-linking). The beads were spun down and the supernatant was loaded on an SDS-PAGE together with inputs. Gels were processed further according to general WB protocol. For DNA: beads and inputs were taken up in 100 µl crosslink reversal buffer (20 mM Tris, pH 8, 0.5 mM EDTA, 0.1 M $NaHCO_3$, 1% SDS, RNase at 10 µg/ml (Roche)) and incubated at 65°C overnight. The next day, the DNA was purified using the PCR purification kit from Qiagen and eluted in 10 µl of H_2O. RT-PCR was performed on 3.5 µl of DNA using the platinum SYBR green qPCR superMix-UDG kit (Invitrogen) according to manufacturer's instructions, except that the reaction was scaled down to 20 µl and 1 mM of $MgCl_2$ were added. The following protocol was used for amplifications: 50°C for 2 min, 95°C for 2 min, 50x (95°C for 15 s, 57°C for 30 s) and melting curve analysis at 65°C for 15 s, heating to 97°C with continuous acquisition per 5°C, 40°C for 30 s). The qPCR was run on a LightCycler 480 (Roche) with the ChIP primers listed in table S1.

Labelling of oligos

Fill-in reaction (3'-labelling): 2 pmoles of oligo duplex with a 5'-overhang were incubated for 15 min at RT with 0.2 µl of the respective [α-^{32}P]-dNTP (2 µCi, Hartman Analytic), 1 mM of each of the other three dNTPs and 5 U Klenow fragment (3'—>5' exo-, NEB) in the corresponding buffer. The enzyme was heat-inactivated for 20 min at 75°C and the reaction was allowed to slowly cool down to re-anneal the strands. The free dNTPs were removed on a Sephadex G-25 column (GE Healthcare).

Kinase reaction (5'-labelling): 2 pmoles of ssDNA were phosphorylated with 0.7 µl of [γ-^{32}P]-ATP (7 µCi, Hartman Analytic) using T4 polynucleotide kinase (10 U, NEB). The reaction was supplemented with 10 mM DTT and incubated for 30 min at 37°C. The reaction was heat-inactivated for 5 min at 95°C and the free nucleotides were removed on a Sephadex G-25 column (GE Healthcare). The oligo was annealed to a 1.5x excess of the unlabelled complementary strand by heating for 5 min to 80°C and slowly cooling to RT in annealing buffer (50 mM Hepes, pH 7.5, 100 mM NaCl).

Nuclear extracts

All nuclear extracts were prepared according to Dignam et al. (Dignam et al., 1983), with or without pre-treatment of the cells with E_2.

Electrophoretic Mobility Shift Assay (EMSA)

All EMSA reactions were carried out in binding buffer (10 mM Tris-HCl, pH 7.6, 50 mM NaCl, 1 mM DTT, 1 mg/ml BSA, 5% glycerol (Hyder et al., 1999)) in a volume of 5 µl using the unspecific competitor poly(dI-dC) (Sigma-Aldrich). 20 ng of poly(dI-dC) were used for 100 ng of recombinant ERα (ThermoFisher, RP-310) and 1 µg for 10 µg of nuclear extracts. For all EMSAs, 10 fmol of [α-^{32}P]-dNTP labelled, or [γ-^{32}P]-ATP phosphorylated oligos were used. Where indicated, an excess of unlabelled competitor DNA or 100 nM of E_2 were added. For the supershift assays, 200 ng of antibody (Anti-ERα; ThermoFisher MA5-13065) were added together with the proteins. Proteins, and competitor or antibodies where indicated, were incubated for 20 min on ice in binding buffer. Labelled DNA was added and the mixture was left at RT for another 20 min before loading on a 5% polyacrylamide gel (Acryl/ Bis 29:1, Amresco) eluted with 1x TAE (40 mM Tris, 20 mM acetate, 1 mM EDTA). The gels were run at 200 V for 1 hours in 1x TAE and dried in a gel dryer for 1 h. The gels were exposed to phosphor screens and the autoradiographs were

596 developed in a Typhoon FLA 9500 (GE Healthcare Life Sciences). The oligonucleotides used are
597 listed in Table S2 (SI).
598
599 *SDS-PAGE of cross-linked binding reactions*
600 EMSAs were performed as described and cross-linked for 5 min (~720 mJ) in a UV Stratalinker
601 1800 (Stratagene). Samples were taken up in 2x SDS loading buffer and boiled for 5 min before
602 being loaded on a 7.5 % denaturing polyacrylamide gel and run for 1-2 hours in 10 % SDS-running
603 buffer at 130 V. Gels were dried for 1 h, exposed to phosphor screens and developed in a Typhoon
604 FLA 9500 (GE Healthcare).
605
606 *CpGL cloning and Luciferase assay*
607 The enhancer/promoter region of the chicken vitellogenin II gene spanning nucleotides -638 to -1
608 was cloned into a CpG-free firefly luciferase plasmid (pCpGL-basic, InvivoGen) using *Hin*dIII and
609 *Nco*I. The enhancer/promoter region was amplified from genomic DNA of LMH/2A cells using the
610 primers 'enhancer fwd and rev' (Table S1, SI). The plasmid was further cut with *Hin*dIII and *Hpa*II
611 to excise the intervening sequence and the annealed upper
612 $^{5'}$AGCTTAAAAATATTCCTGGTCAGCGTGAC$^{3'}$ and lower
613 $^{5'}$CGGTCACGCTGACCAGGAATATTTTTA$^{3'}$ ds-oligo was ligated into the plasmid to translocate
614 the *Hin*dIII site closer to the ERE for further ligations of the ERE. On the other side of the ERE a
615 *Kpn*I site was created by mutagenesis and another *Kpn*I site in the plasmid had to be destroyed by
616 mutagenesis (primers in Table 1, SI).
617 The resulting VTG-CpGL plasmid was co-transfected with a Renilla luciferase-expressing plasmid
618 (pRL-SV40, Promega) into LMH/2A cells using Lipofectamine 2000 (Invitrogen) according to
619 manufacturer's instructions. Cells were re-seeded the next day in technical triplicates both for E_2
620 and ethanol treatment and treated for 24 h. Cells were subsequently lysed directly in firefly
621 luciferase substrate from Promega (Dual-Glo Luciferase kit) for 30 min, with vigorous shaking.

Cell lysates were processed further according to manufacturer's instructions and luminescence was measured on the SpectraMax i3 (Molecular Devices) plate reader. Relative luciferase units (RLU) were calculated as the ratio between the firefly and *Renilla* signals.

Mutagenesis of VTG-CpGL

200 ng of vector DNA were mixed with 50 pmoles forward and reverse primers, 250 µM dNTPs, 2% DMSO and 1 U Phusion high fidelty polymerase (NEB) in 1x buffer provided by the manufacturer. Extension was performed in 50 µl according to the following protocol: 95 °C, 2 min; 30x (95 °C, 1 min, 55 °C, 1 min, 63 °C, 30 min); 68 °C, 20 min; 15 °C hold. Extension reaction was digested twice with 20 U *Dpn*I for 1 hours at 37°C followed by heat inactivation of the enzyme for 20 min at 80°C. The reaction mix was then transformed into electro-competent Pir1 bacteria (R6K gamma ORI, Invitrogen) after being desalted by incubating for 15 min on a 0.025 µm Millipore nitrocellulose filter on H_2O. Bacteria were shaken at 37°C for 40 min and plated onto Zeocin (25 µg/ml, InvivoGen) agar plates. Clones were picked the next day, grown in liquid culture overnight and plasmids were extracted using the NucleoSpin Plasmid Kit (Macherey-Nagel) according to manufacturer's instructions. Plasmids were sequenced at Microsynth to check for successful mutagenesis. The primers for mutagenesis are listed in Table S1 (SI).

Ligation of modified ERE into VTG-CpGL

50 µg of the plasmid DNA were cut with *Hin*dIII-HF (NEB) in Cut Smart buffer with 100 U of enzyme for 1.5 hours at 37°C in 50 µl. Linearization was verified on a 1% agarose gel and enzyme was heat-inactivated for 20 min at 80°C. For the first ligation step, 10 µg of linearized DNA and 20x excess of annealed oligo with different modifications at the CpGs (ERE insert, Table S1, SI) were incubated for 2 hours at RT in T4 ligase buffer with 400 U of T4 ligase (NEB) in a total volume of 100 µl. Ligase was inactivated by heating to 65°C for 10 min and efficient ligation was verified on an agarose gel. Reaction volume was increased to 200 µl with 1x Cut Smart buffer, 40 U

of *Kpn*I-HF (NEB) were added and incubation was continued for additional 1.5 hours at 37°C. 20 U of *Hin*dIII-HF were then added, the mixture was incubated for an additional hour and subsequently cleaned-up on MinElute columns (5µg per column, Qiagen) and eluted in 20 µl H_2O. 20 reactions of the first ligation were pooled for the second ligation. 12.5 µg of DNA from the first ligation were incubated in 10 ml of 1x T4 ligase buffer with 4000 U of ligase for 1 hours at RT. Recircularization was verified on an agarose gel. The resulting mix was concentrated by EtOH precipitation and the supercoiled form of the plasmid was extracted on a CsCl gradient and purified further as described (Baerenfaller et al., 2006).

Expression and purification of Eco72IM

The *Eco*72IM sequence was PCR-amplified from the *Eco*72IRM plasmid (Thermo Fisher) and recombined with the pDONR 221 (Gateway, Thermo Fisher) to form an entry vector. The resulting entry vector was recombined with pDest15 (Gateway) adding an N-term GST tag to the methyltransferase. BL21 pLysS *E. coli* (Promega) were transformed with the expression vector. A starter culture (5 ml) was grown from a single colony and subsequently used to inoculate a 1 l culture that was grown at 37°C to an OD of 0.65. Culture was cooled down to 22°C, a sample of uninduced culture was taken for SDS gel and protein expression was induced with the addition of 250 µM IPTG with shaking overnight. The next day, bacteria were washed with cold PBS by centrifugation at 4'000 x g for 10 min at 4 °C. 14 ml lysis buffer (100 µg/ml Lysozyme, 1 % Triton, 1 mM PMSF, 1 x protease inhibitor (cOmplete, Roche), 10 mM DT, 1 x PBS) were added to pellet and lysate was incubated for 30 min in a beaker stirring at 4 °C. Lysate was sonicated twice for 1 min on ice (ampl 70%, 50% cycle), together with uninduced sample for SDS gel and centrifuged at 18'000 x g for 30 min at 4 °C. GSH beads (GE Healthcare, 600 µl) washed in lysis buffer were added to the supernatant and incubated on a rotating wheel for 2 hours at 4 °C. The beads were then washed three times in washing buffer (10% glycerol, 10 mM DTT, 1 mM PMSF, 1 x PBS) by inverting the tube several times followed by centrifugation at 500 x g for 5 min. Flow-through and

674 first wash were retained for the SDS gel. The beads were then incubated with 500 µl elution buffer
675 (washing buffer with freshly-added 20 mM glutathione, pH adjusted to 8 with NaOH) on rotating
676 wheel at 4 °C. The first elution was aliquoted and snap frozen.
677
678 *In vitro methylation*
679 1 µg of plasmid DNA was incubated in 1x NEB buffer 2 (*SssI*) or Eco72IM buffer (10 mM Tris
680 HCl, 50 mM NaCl, 1 mM DTT, 10 mM EDTA), 160 µM freshly-diluted SAM (NEB) and 4 U of
681 *Sss*I (NEB) or 4 µl purified *Eco*72IM (diluted 1:100) for 1 hours at 37°C. The reaction was stopped
682 by heating to 65°C for 20 min.
683
684 *EdU incorporation and Click-iT Reaction*
685 The cells were grown on coverslips and treated with 8 µM aphidicolin or DMSO, 100 nM E_2 or
686 ethanol and 10 µM of 5-ethynyl-2'-deoxyuridine (EdU, Invitrogen) for 24h. Medium was removed,
687 the cells were washed once with PBS and 1 ml of 3.7% formaldehyde in PBS was added. The cells
688 were fixed for 15 min at RT, then washed twice with 1 ml of 3% BSA in PBS. 1 ml of 0.5% Triton
689 X-100 in PBS was added and the cells were incubated for 20 min at RT. Click-iT reaction master
690 mix was prepared. [For 5 slides: 129 µl 1x Click-iT reaction buffer (freshly diluted 1:10 in H_2O), 6
691 µl $CuSO_4$, 0.36 µl AlexaFluor azide, 15 µl reaction buffer additive (freshly diluted 1:10 in H_2O)].
692 Permeabilization buffer was removed and the cells were washed twice with 1 ml 3% BSA in PBS.
693 30 µl of Click-iT reaction mix were pipetted onto Parafilm and the coverslips were put cells-down
694 onto the mix and incubated for 30 min at RT, protected from light. The coverslips were washed
695 once with 1 ml of 3% BSA in PBS and then washed well with 1x PBS to remove BSA. After one
696 final wash with H_2O, the cells were fixed with mounting media containing 4',6-diamidino-2-
697 phenylindole (DAPI, VectaShield). The slides were analyzed on an Olympus IX81 fluorescence
698 microscope.
699

DNA extraction, oxidation of hmC and bisulphite conversion

Genomic DNA was extracted from cells using the Wizard Genomic DNA Purification Kit (Promega). DNA was eluted with water and digested with *Eco*RV overnight. EtOH precipitation was performed and DNA was additionally cleaned up on Micro Bio-Spin 6 chromatography columns in SSC (BioRad) as purity was essential for oxidation of hmC. Samples were split and one half was subjected to oxidation. Selective oxidation of hmC to fC was achieved using potassium perruthenate ($KRuO_4$, Sigma-Aldrich) as described (Booth et al., 2012; Booth et al., 2013). In short, 0.5-2 µg of DNA were incubated in 50 mM NaOH in 24 µl for 30 min at 37°C after vigorous vortexing to denature DNA. 1 µl of 15 mM $KRuO_4$ solution in 50 mM NaOH was added and oxidation was incubated on ice for 1 hours with vortexing every 5 min. Reaction was cleaned up on polyacrylamide columns (89849, ThermoScientific) and processed further with the rest of the sample for bisulphite conversion. Bisulphite conversion was achieved using the EZ DNA Methylation-Gold Kit from Zymo Research according to manufacturer's instructions. The cycling protocol was adapted because of the slightly less efficient conversion of fC as compared to unmodified cytosine (95°C for 5 min; 2x (60°C for 25 min, 95°C for 5 min, 60°C for 85 min, 95°C for 5 min, 60°C for 175 min, 95°C for 5 min); 20°C hold.

Polymerase chain reaction on bisulphite-converted DNA

100 ng of bisulphite-converted DNA was used as template for the PCR using the ZymoTaq DNA polymerase (Zymo Research). The reaction was carried out in a total volume of 40 µl in the reaction buffer provided by the manufacturer supplemented with of 1 mM dNTPs, 1.5 mM $MgCl_2$, 500 mM each primer and 2 U of polymerase. The cycling protocol was as follows: 10 min 95°C; 5x (30 s 94°C; 30 s 52°C; 90 s 72°C); 5x (30 s 94°C; 30 s 52°C; 90 s 72°C); 35x (30 s 94°C; 30 s 55°C; 90 s 72°C); 7 min 72°C. The primers used for the amplification of bisulphite converted DNA are listed in Table S1 (SI). The forward primers were only added after the first 5 cycles to reduce the formation of primer dimers. The PCR fragments were purified using a BluePippin (Sage Science)

on a 2% agarose gel according to manufacturer's instructions, before being bar-coded for PacBio sequencing.

PacBio sequencing and analysis

The PacBio single-molecule real-time (SMRT) sequencing technology works by a strand-displacement mechanism on circularised single molecules. The highly-processive polymerase copies the circle multiple times to generate a long read containing many repeats of the same sequence. The deconvolution to single reads generates a consensus sequence that corrects the high error-rate of the polymerase, resulting thus in very accurate sequencing results. The sequencing and data evaluation were carried out in collaboration with the Functional Genomic Center Zurich (www.fgcz.ch).

Affinity purification and mass spectrometry

Different CpG-containing oligonucleotides with 5'-TTAA overhangs were annealed and end-to-end ligated overnight at 16°C with T4 ligase (NEB). Oligos were ethanol-precipitated and filled-in with 1 mM biotinylated dUTP (Thermo Fisher) using Klenow fragment (NEB). Reactions were passed twice through a Sephadex G-25 desalting column (GE Healthcare) to get rid to free dUTP and subsequently bound to Dynabeads M-280 Streptavidin (Invitrogen) by rotating 20 min at RT. 1.5 mg of nuclear extracts were pre-incubated with 30 µg of poly(dI-dC), before an equal volume was added to the beads. Binding was allowed to take place in 1x EMSA buffer without BSA for 30 min on a rotating wheel at 4 °C. Beads were washed 2x with 60 µl 1x EMSA buffer without BSA. Bound proteins were eluted with 60 µl of elution buffer (10 mM Tris-HCl pH 7.6, 5 % glycerol, 1M NaCl**)** 30 min rotating at 4°C. Elution was desalted on 0.025 µm VSWP membranes (Merck) against water for EMSAs.

Shotgun LC-MS/MS

The protein mixture was digested with trypsin. After cleaning up, the peptide mixture was applied to a reversed phase high-performance liquid chromatography (HPLC) column and separated prior to ionization and analysis by the mass spectrometer. The peptide ions selected by an instrument algorithm for fragmentation were recorded as a peptide signature, which was analyzed by a sequence algorithm (MASCOT). The peptides were scored according to the average probability using total spectra counts. The complete list of the MS results can be found online in Supplementary Information.

Immunodepletion of extracts

50 µg Dynabeads Protein G (Invitrogen), washed in PBS-T (0.1% Tween 20), were incubated with 0.4 µg of the respective antibody for 20 min at RT on rotation wheel. Beads were washed twice in PBS-T by inverting the tube several times and putting it on the magnet for 2 min. 100 µg of nuclear extracts were added and binding was allowed for 2 hours rotating at 4 °C. Supernatant was aliquoted and snap frozen as immunodepleted extract. Samples were taken for control Western Blots and EMSAs.

Statistical analysis

All experiments were performed at least three times. Results are shown as means +/- SD. Statistical significance was determined byStudent's *t*-tests. $P \leq 0.05$ was considered statistically significant.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Hanspeter Saluz for providing hen and rooster DNA samples, to Peter Hunziker for the proteomic analysis and Giancarlo Russo for assistance with

the bioinformatics, to Thermo Fischer for the generous gift of the Eco72I DNA, to Beat Kunz for assistance with the egg manipulations and to Maite Olivera Harris for constructive discussions.

REFERENCES

Amenya, H.Z., Tohyama, C., and Ohsako, S. (2016). Dioxin induces Ahr-dependent robust DNA demethylation of the Cyp1a1 promoter via Tdg in the mouse liver. Scientific reports *6*, 34989.

Bacolla, A., Pradhan, S., Roberts, R.J., and Wells, R.D. (1999). Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated DNA. The Journal of biological chemistry *274*, 33011-33019.

Baerenfaller, K., Fischer, F., and Jiricny, J. (2006). Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro. Methods in enzymology *408*, 285-303.

Baeriswyl, T., and Stoeckli, E.T. (2006). In ovo RNAi opens new possibilities for temporal and spatial control of gene silencing during development of the vertebrate nervous system. J RNAi Gene Silencing *2*, 126-135.

Baylin, S.B., Hoppener, J.W., de Bustros, A., Steenbergh, P.H., Lips, C.J., and Nelkin, B.D. (1986). DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer research *46*, 2917-2922.

Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J *11*, 2611-2617.

Binder, R., MacDonald, C.C., Burch, J.B., Lazier, C.B., and Williams, D.L. (1990). Expression of endogenous and transfected apolipoprotein II and vitellogenin II genes in an estrogen responsive chicken liver cell line. Molecular endocrinology *4*, 201-208.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

802 Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S.
803 (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-
804 base resolution. Science *336*, 934-937.
805 Booth, M.J., Ost, T.W., Beraldi, D., Bell, N.M., Branco, M.R., Reik, W., and Balasubramanian, S.
806 (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine.
807 Nature protocols *8*, 1841-1851.
808 Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin,
809 A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature
810 *371*, 435-438.
811 Burch, J.B., and Weintraub, H. (1983). Temporal order of chromatin structural changes associated
812 with activation of the major chicken vitellogenin gene. Cell *33*, 65-76.
813 Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a
814 mechanism for the regulation of E2F activity. Proc Natl Acad Sci U S A *97*, 6481-6486.
815 Chen, B., Hsu, R., Li, Z., Kogut, P.C., Du, Q., Rouser, K., Camoretti-Mercado, B., and Solway, J.
816 (2012). Upstream stimulatory factor 1 activates GATA5 expression through an E-box motif.
817 Biochem J *446*, 89-98.
818 Cheng, A.S., Jin, V.X., Fan, M., Smith, L.T., Liyanarachchi, S., Yan, P.S., Leu, Y.W., Chan, M.W.,
819 Plass, C., Nephew, K.P., *et al.* (2006). Combinatorial analysis of transcription factor partners
820 reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. Mol Cell *21*,
821 393-404.
822 Comb, M., and Goodman, H.M. (1990). CpG methylation inhibits proenkephalin gene expression
823 and binding of the transcription factor AP-2. Nucleic Acids Res *18*, 3975-3982.
824 Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA.
825 Nucleic Acids Res *11*, 647-658.
826 Corre, S., and Galibert, M.D. (2005). Upstream stimulating factors: highly versatile stress-
827 responsive transcription factors. Pigment Cell Res *18*, 337-348.

828 d'Adda di Fagagna, F., Marzio, G., Gutierrez, M.I., Kang, L.Y., Falaschi, A., and Giacca, M.
829 (1995). Molecular and functional interactions of transcription factor USF with the long terminal
830 repeat of human immunodeficiency virus type 1. J Virol *69*, 2765-2775.
831 De Smet, C., De Backer, O., Faraoni, I., Lurquin, C., Brasseur, F., and Boon, T. (1996). The
832 activation of human gene MAGE-1 in tumor cells is correlated with genome-wide
833 demethylation. Proc Natl Acad Sci U S A *93*, 7149-7153.
834 De Smet, C., Lurquin, C., Lethe, B., Martelange, V., and Boon, T. (1999). DNA methylation is the
835 primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich
836 promoter. Molecular and cellular biology *19*, 7327-7335.
837 deGraffenried, L.A., Hopp, T.A., Valente, A.J., Clark, R.A., and Fuqua, S.A. (2004). Regulation of
838 the estrogen receptor alpha minimal promoter by Sp1, USF-1 and ERalpha. Breast Cancer Res
839 Treat *85*, 111-120.
840 Diamanti-Kandarakis, E., Bourguignon, J.P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M.,
841 Zoeller, R.T., and Gore, A.C. (2009). Endocrine-disrupting chemicals: an Endocrine Society
842 scientific statement. Endocr Rev *30*, 293-342.
843 Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA
844 polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-
845 1489.
846 Dubey, V., and Luqman, S. (2017). Cathepsin D as a Promising Target for the Discovery of Novel
847 Anticancer Agents. Curr Cancer Drug Targets *17*, 404-422.
848 Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V.,
849 Davies, R., Down, T.A., *et al.* (2006). DNA methylation profiling of human chromosomes 6, 20
850 and 22. Nature genetics *38*, 1378-1385.
851 Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. Current
852 opinion in genetics & development *4*, 255-259.

853 Ehrlich, M. (2019). DNA hypermethylation in disease: mechanisms and clinical relevance.
854 Epigenetics, 1-23.
855 Fujii, G., Nakamura, Y., Tsukamoto, D., Ito, M., Shiba, T., and Takamatsu, N. (2006). CpG
856 methylation at the USF-binding site is important for the liver-specific transcription of the
857 chipmunk HP-27 gene. Biochem J *395*, 203-209.
858 Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and
859 Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids
860 Res *11*, 6883-6894.
861 Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. Journal of
862 molecular biology *196*, 261-282.
863 Giacca, M., Gutierrez, M.I., Demarchi, F., Diviacco, S., Biamonti, G., Riva, S., and Falaschi, A.
864 (1989). A protein target site in an early replicated human DNA sequence: a highly conserved
865 binding motif. Biochemical and biophysical research communications *165*, 956-965.
866 Goelz, S.E., Vogelstein, B., Hamilton, S.R., and Feinberg, A.P. (1985). Hypomethylation of DNA
867 from benign and malignant human colon neoplasms. Science *228*, 187-190.
868 Gruenbaum, Y., Naveh-Many, T., Cedar, H., and Razin, A. (1981). Sequence specificity of
869 methylation in higher plant DNA. Nature *292*, 860-862.
870 Hale, T.K., and Braithwaite, A.W. (1995). Identification of an upstream region of the mouse p53
871 promoter critical for transcriptional expression. Nucleic Acids Res *23*, 663-669.
872 Hassan, H.M., Kolendowski, B., Isovic, M., Bose, K., Dranse, H.J., Sampaio, A.V., Underhill,
873 T.M., and Torchia, J. (2017). Regulation of Active DNA Demethylation through RAR-Mediated
874 Recruitment of a TET/TDG Complex. Cell reports *19*, 1685-1697.
875 He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.*
876 (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian
877 DNA. Science *333*, 1303-1307.

878 Hou, Y., Yuan, J., Zhou, X., Fu, X., Cheng, H., and Zhou, R. (2012). DNA demethylation and USF
879 regulate the meiosis-specific expression of the mouse Miwi. PLoS genetics *8*, e1002716.
880 Hsieh, C.L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b.
881 Molecular and cellular biology *19*, 8211-8218.
882 Hyder, S.M., Chiappetta, C., and Stancel, G.M. (1999). Interaction of human estrogen receptors
883 alpha and beta with the same naturally occurring estrogen response elements. Biochem
884 Pharmacol *57*, 597-601.
885 Ismail, P.M., Lu, T., and Sawadogo, M. (1999). Loss of USF transcriptional activity in breast
886 cancer cell lines. Oncogene *18*, 5582-5591.
887 Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P.,
888 Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA.
889 Nature *452*, 112-115.
890 Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T., and Bergman, Y. (1996). A role for
891 nuclear NF-kappaB in B-cell-specific demethylation of the Igkappa locus. Nature genetics *13*,
892 435-441.
893 Klinge, C.M. (2001). Estrogen receptor interaction with estrogen response elements. Nucleic Acids
894 Res *29*, 2905-2919.
895 Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene
896 results in embryonic lethality. Cell *69*, 915-926.
897 Liu, W.M., Maraia, R.J., Rubin, C.M., and Schmid, C.W. (1994). Alu transcripts: cytoplasmic
898 localisation and regulation by DNA methylation. Nucleic Acids Res *22*, 1087-1095.
899 Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene
900 promoter are required to prevent methylation of the CpG island. Genes & development *8*, 2282-
901 2292.

902 Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson,
903 D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a
904 transcriptionally active promoter. Nature *452*, 45-50.
905 Neddermann, P., and Jiricny, J. (1993). The purification of a mismatch-specific thymine-DNA
906 glycosylase from HeLa cells. The Journal of biological chemistry *268*, 21218-21224.
907 Noreen, F., Roosli, M., Gaj, P., Pietrzak, J., Weis, S., Urfer, P., Regula, J., Schar, P., and Truninger,
908 K. (2014). Modulation of age- and cancer-associated DNA methylation change in the healthy
909 colon by aspirin and lifestyle. Journal of the National Cancer Institute *106*.
910 Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and
911 Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.
912 Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian
913 DNA (cytosine-5) methyltransferases. Nature genetics *19*, 219-220.
914 Perini, G., Diolaiti, D., Porro, A., and Della Valle, G. (2005). In vivo transcriptional regulation of
915 N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci U S A *102*, 12117-
916 12122.
917 Philipsen, J.N., Hennis, B.C., and Ab, G. (1988). In vivo footprinting of the estrogen-inducible
918 vitellogenin II gene from chicken. Nucleic Acids Res *16*, 9663-9676.
919 Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA
920 (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and
921 maintenance methylation. The Journal of biological chemistry *274*, 33002-33010.
922 Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of
923 max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell
924 *65*, 395-407.
925 Prendergast, G.C., and Ziff, E.B. (1991). Methylation-sensitive sequence-specific DNA binding by
926 the c-Myc basic region. Science *251*, 186-189.

927 Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe,
928 K.L., Otto, T., Olek, A., *et al.* (2004). DNA methylation profiling of the human major
929 histocompatibility complex: a pilot study for the human epigenome project. PLoS biology *2*,
930 e405.
931 Reisman, D., and Rotter, V. (1993). The helix-loop-helix containing transcription factor USF binds
932 to and transactivates the promoter of the p53 tumor suppressor gene. Nucleic Acids Res *21*, 345-
933 350.
934 Rimseliene, R., Vaisvila, R., and Janulaitis, A. (1995). The eco72IC gene specifies a trans-acting
935 factor which influences expression of both DNA methyltransferase and endonuclease from the
936 Eco72I restriction-modification system. Gene *157*, 217-219.
937 Saluz, H.P., Jiricny, J., and Jost, J.P. (1986). Genomic sequencing reveals a positive correlation
938 between the kinetics of strand-specific DNA demethylation of the overlapping
939 estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA
940 synthesis. Proc Natl Acad Sci U S A *83*, 7167-7171.
941 Seal, S.N., Davis, D.L., and Burch, J.B. (1991). Mutational studies reveal a complex set of positive
942 and negative control elements within the chicken vitellogenin II promoter. Molecular and
943 cellular biology *11*, 2704-2717.
944 Sensel, M.G., Binder, R., Lazier, C.B., and Williams, D.L. (1994). Reactivation of apolipoprotein II
945 gene transcription by cycloheximide reveals two steps in the deactivation of estrogen receptor-
946 mediated transcription. Molecular and cellular biology *14*, 1733-1742.
947 Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer,
948 L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-
949 hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.
950 Thomassin, H., Flavin, M., Espinas, M.L., and Grange, T. (2001). Glucocorticoid-induced DNA
951 demethylation and gene memory during development. EMBO J *20*, 1974-1983.

952 Toker, A., Engelbert, D., Garg, G., Polansky, J.K., Floess, S., Miyao, T., Baron, U., Duber, S.,
953 Geffers, R., Giehr, P., *et al.* (2013). Active demethylation of the Foxp3 locus leads to the
954 generation of stable regulatory T cells within the thymus. Journal of immunology *190*, 3180-
955 3188.
956 Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG
957 island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A *96*, 8681-8686.
958 Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses
959 is constrained by cytosine methylation. Nature genetics *20*, 116-117.
960 Wiebauer, K., and Jiricny, J. (1989). In vitro correction of G.T mispairs to G.C pairs in nuclear
961 extracts from human cells. Nature *339*, 234-236.
962 Wilks, A., Seldran, M., and Jost, J.P. (1984). An estrogen-dependent demethylation at the 5' end of
963 the chicken vitellogenin gene is independent of DNA synthesis. Nucleic Acids Res *12*, 1163-
964 1177.
965 Wilks, A.F., Cozens, P.J., Mattaj, I.W., and Jost, J.P. (1982). Estrogen induces a demethylation at
966 the 5' end region of the chicken vitellogenin gene. Proc Natl Acad Sci U S A *79*, 4252-4255.
967 Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1997).
968 Asymmetric methylation in the hypermethylated CpG promoter region of the human L1
969 retrotransposon. The Journal of biological chemistry *272*, 7810-7816.
970 Xing, W., and Archer, T.K. (1998). Upstream stimulatory factors mediate estrogen receptor
971 activation of the cathepsin D promoter. Molecular endocrinology *12*, 1310-1321.
972

973 **FIGURE LEGENDS**

974 **Fig 1. Sequence of the *VTG* enhancer/promoter, its methylation and inducibility *in vivo.* A**,
975 Sequence of the enhancer/promoter region of the chicken *VTG* gene. The ERE binding site (violet),
976 the CpGs (green) and the translation start site (yellow) are highlighted. The four CpGs (**a**-**d**)
977 analyzed in Saluz *et al.* (Saluz et al., 1986) are indicated, as well as the additional six CpGs (3-8) in

the enhancer/promoter region. **B**, Bisulphite sequencing of CpGs **a-d** in LMH/2A cells. **C**,**D** *VTG* mRNA levels measured by RT-qPCR after 6 and 24 hours of 100 nM E_2 treatment (**C**) and upon additional treatment with 8 µM aphidicolin (aph) or DMSO for 24h (**D**). Data are represented as mean ± SD. Significance was assessed using Sidak's multiple comparisons test. *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001. **E**, Upper panel: Western Blot of chromatin extracts immunoprecipitated with Flag– or ERα antibodies (- and + β-estradiol treatment). Lower panel: Ratio of RT-qPCR signal of immunoprecipitated versus input DNA.

Fig 2. Electrophoretic mobility shift analysis of ERE-binding proteins. A, EMSA with 60 ng recombinant ERα and unmethylated ERE wt oligo (C/C), with (+, lane 3) or without (-, lanes 1, 2, 4-7) E_2 or competitor. The specific competitor (spec, lane 4) was the unlabelled ERE wt duplex C/C, the unspecific competitor (unspec, lane 5) was an unrelated duplex of similar length. An antibody against ERα (MA5-13065, lane 7) or BSA (lane 6) were used in the supershift experiment with the unmethylated oligo C/C. The oligonucleotides used for the EMSAs are depicted below the autoradiograph. Purple, consesus ERE; green, CpGs. The ΔG duplex contains a single base pair deletion (white) in the spacer region between the two dyad-symmetry elements of the ERE. **B**, Upper panel: representative EMSAs with increasing concentrations of recombinant ERα (0, 5, 10, 20, 30, 40, 80 and 200 ng) and unmethylated (C/C, left panel) or methylated (mC/mC, right panel) ERE wt or ΔG oligo with 40 ng of ERα. (In the mC/mC duplex both CpGs were symmetrically-methylated.) Lower panel: Quantification of percentage bound oligo in three independent experiments. Data are represented as mean ± SD. **C**, EMSA comparing the binding of recombinant ERα protein or LMH/2A NEs to shorter duplexes (ERE short shown below the autoradiograph) unmethylated (C/C) and methylated (mC/mC) or a shorter ΔG oligo (ΔG short). **D**, Left panel: EMSA with the indicated duplexes and LMH/2A NE supplemented with the indicated amounts of recombinant ERα. Right panel: Quantification of the ratio of oligo bound by ERα or factor X in three independent experiments. Data are represented as mean ± SD. Significance was assessed

using the Holm-Sidak test. **E**, Top panel: EMSA with LMH/2A nuclear extracts and five duplexes containing the indicated CpGs from the *VTG* enhancer/promoter (Fig. 1A), bearing different combinations of cytosines and methylcytosines. Bottom panel: Quantification of percentage bound oligo in three independent experiments. Data are represented as mean ± SD. Significance was assessed using the Tukey test for multiple comparisons. The panels show autoradiographs of non-denaturing 6% polyacrylamide gels eluted with TAE buffer. *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001.

Fig 3. Identification of Factor X by UV cross-linking. A, Experimental set-up showing the position of the BrdU residues in the indicated duplex substrates. The binding reactions contained either 60 µg of LMH/2A NE or recombinant ERα. **B,** Proteins cross-linked to the oligo substrates shown in A (unmethylated or methylated as indicated). **C**, As in B, but the cross-linking reactions were supplemented with the indicated amounts of recombinant ERα or competitor oligo. The cross-linked complexes were resolved by 10% SDS-PAGE and visualised by autoradiography.

Figure 4. Identification of Factor X by affinity chromatography/MS. A, Experimental set-up of affinity pull-down and mass spectrometric analysis. The oligonucleotide duplexes were ligated end-to-end, tailed with bio-dUMP and bound to streptavidin Dynabeads. The beads were incubated with the extracts and the eluted proteins were analysed by MS as described in Materials and Methods. **B**, EMSA with oligo duplex CpG7 (left panel) or ΔG (right panel) and HeLa NE preincubated with the indicated antibodies. **C**, Supershift using oligo duplex CpG7 and antibodies specific for USF1 or USF2. **D,** UV cross-linking reactions. Depleted extracts (lanes 1-4) and proteins eluted from the Dynabeads (lanes 5-8) were bound to the unmethylated ERE oligo, UV-cross-linked, separated on SDS-PAGE and visualized by autoradiography.

Figure 5. Estrogen inducibility of VTG-CpGL luciferase reporter vector transfected into LMH/2A cells. A, Luciferase expression before and after *in vitro* methylation of the ERE C/C (wt) or ΔG vectors with *Sss*I. **B**, Same as A, but the ERE sequence was replaced with oligonucleotides carrying the indicated cytosine modifications only in CpGs **d** and **c**. **C**, Same as A, but either with wt VTG-CpGL or the mutated reporter in which all CpGs except for the ERE and CpG7 were substituted for TpGs, +/-*Sss*I. **D**, Luciferase expression from the reporter before and after methylation with *Eco*72IM, *Hpa*II.M or *Sss*I. *Eco*72IM without S-adenosylmethionine (-SAM) was used as control, the other methylation reactions were performed in the presence of SAM (+SAM). Significance was assessed using the Tukey test for multiple comparisons. *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001. **E**, Luciferase assay using VTG-CpGL linearised with the indicated enzymes, methylated with *Sss*I and religated. The purified circular DNA was then transfected into LMH/2A cells. In these substrates, the cleaved restriction site remained unmethylated after the circularisation. The ratio between methylated and unmethylated is shown. **F**, Luciferase assay using unmethylated VTG-CpGL in LMH/2A cells in which USF1 was depleted with siRNA; siLuc was used as control [NB: this siRNA does not recognise either of the luciferases expressed from our vectors]. Relative luciferase units (RLU) are defined as ratio between firefly and *Renilla* signal. The graphs show the mean ± SD of three independent experiments. **G**, RT-qPCR of *VTG* mRNA isolated from cells treated with siLuc or siUSF1 for 96 hours and EtOH or E_2 for the last 24 h. The graph shows the mean ± SD of three independent experiments. Significance was assessed using Sidak's multiple comparisons test. *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001.

Table 1. Peptides identified by LC-MS/MS in fractions eluted from affinity chromatography on the indicated oligonucleotide matrices. To subtract proteins that bound unspecifically to the matrix , beads only were used. As a second filter, we used an oligonucleotide containing the sequence around CpG8 that showed no binding of factor X in the EMSA experiments. Significance was

1054 assessed with an ANOVA multiple comparison t-test. Only hits identified with >95% confidence

1055 are shown. The false discovery rate (FDR) threshold was set to 1%.

1056

Figure 1

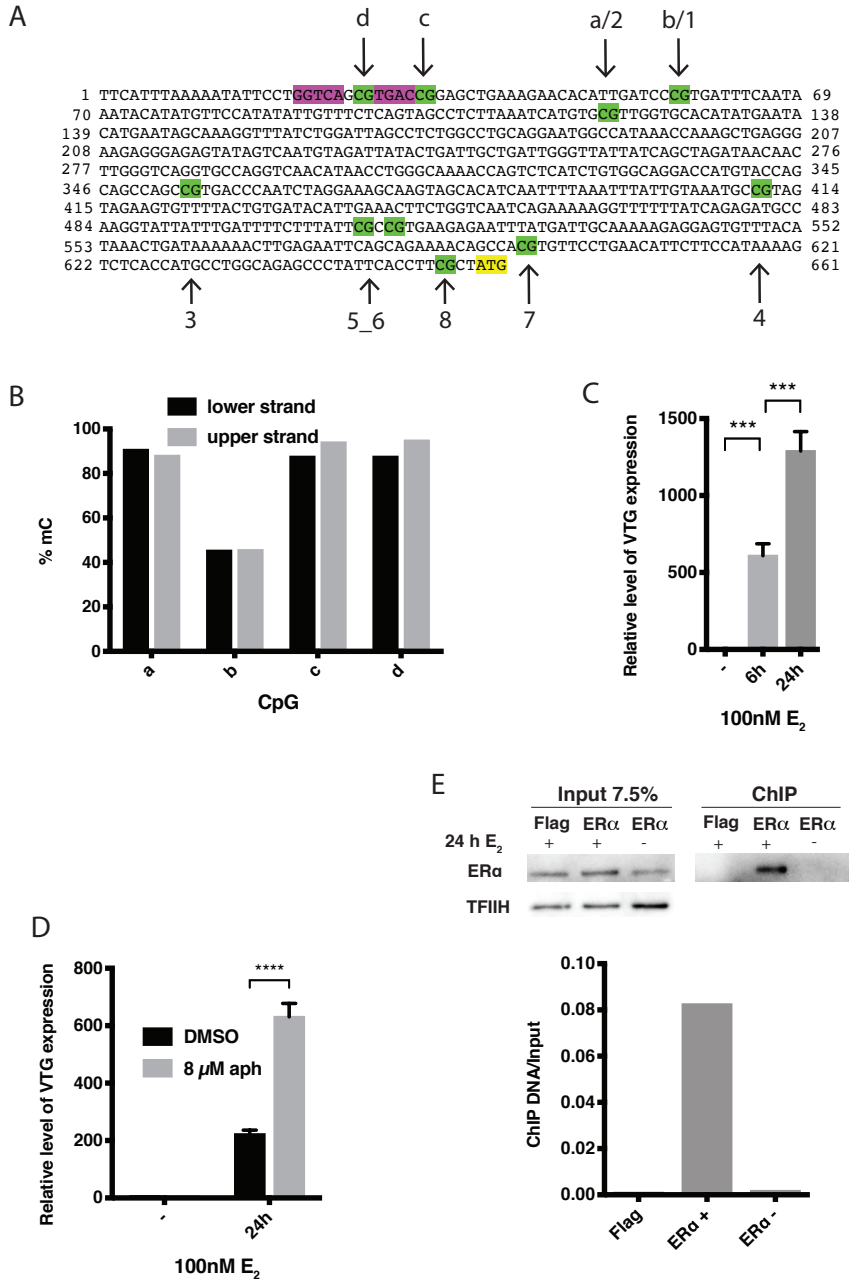

A
d
c
a/2
b/1
1 TTCATTTAAAAATATTCCTGGTCAGCGTGACCGGAGCTGAAAGAACACATTGATCCCGTGATTTCAATA 69
70 AATACATATGTTCCATATATTGTTTCTCAGTAGCCTCTTAAATCATGTGCGTTGGTGCACATATGAATA 138
139 CATGAATAGCAAAGGTTTATCTGGATTAGCCTCTGGCCTGCAGGAATGGCCATAAACCAAAGCTGAGGG 207
208 AAGAGGGAGAGTATAGTCAATGTAGATTATACTGATTGCTGATTGGGTTATTATCAGCTAGATAACAAC 276
277 TTGGGTCAGGTGCCAGGTCAACATAACCTGGGCAAAACCAGTCTCATCTGTGGCAGGACCATGTACCAG 345
346 CAGCCAGCCGTGACCCAATCTAGGAAAGCAAGTAGCACATCAATTTTAAATTTATTGTAAATGCCGTAG 414
415 TAGAAGTGTTTTACTGTGATACATTGAAACTTCTGGTCAATCAGAAAAAGGTTTTTTATCAGAGATGCC 483
484 AAGGTATTATTTGATTTTCTTTATTCGCCGTGAAGAGAATTTATGATTGCAAAAAGAGGAGTGTTTACA 552
553 TAAACTGATAAAAAACTTGAGAATTCAGCAGAAAACAGCCACGTGTTCCTGAACATTCTTCCATAAAAG 621
622 TCTCACCATGCCTGGCAGAGCCCTATTCACCTTCGCTATG 661
3
5_6
8
7
4
B
lower strand
upper strand
% mC
CpG
C
Relative level of VTG expression
100nM E2
6h
24h
E
Input 7.5%
ChIP
Flag
ERα
24 h E2
ERα
TFIIH
D
DMSO
8 μM aph
Relative level of VTG expression
100nM E2
ChIP DNA/Input
Flag
ERα +
ERα -

Figure 2

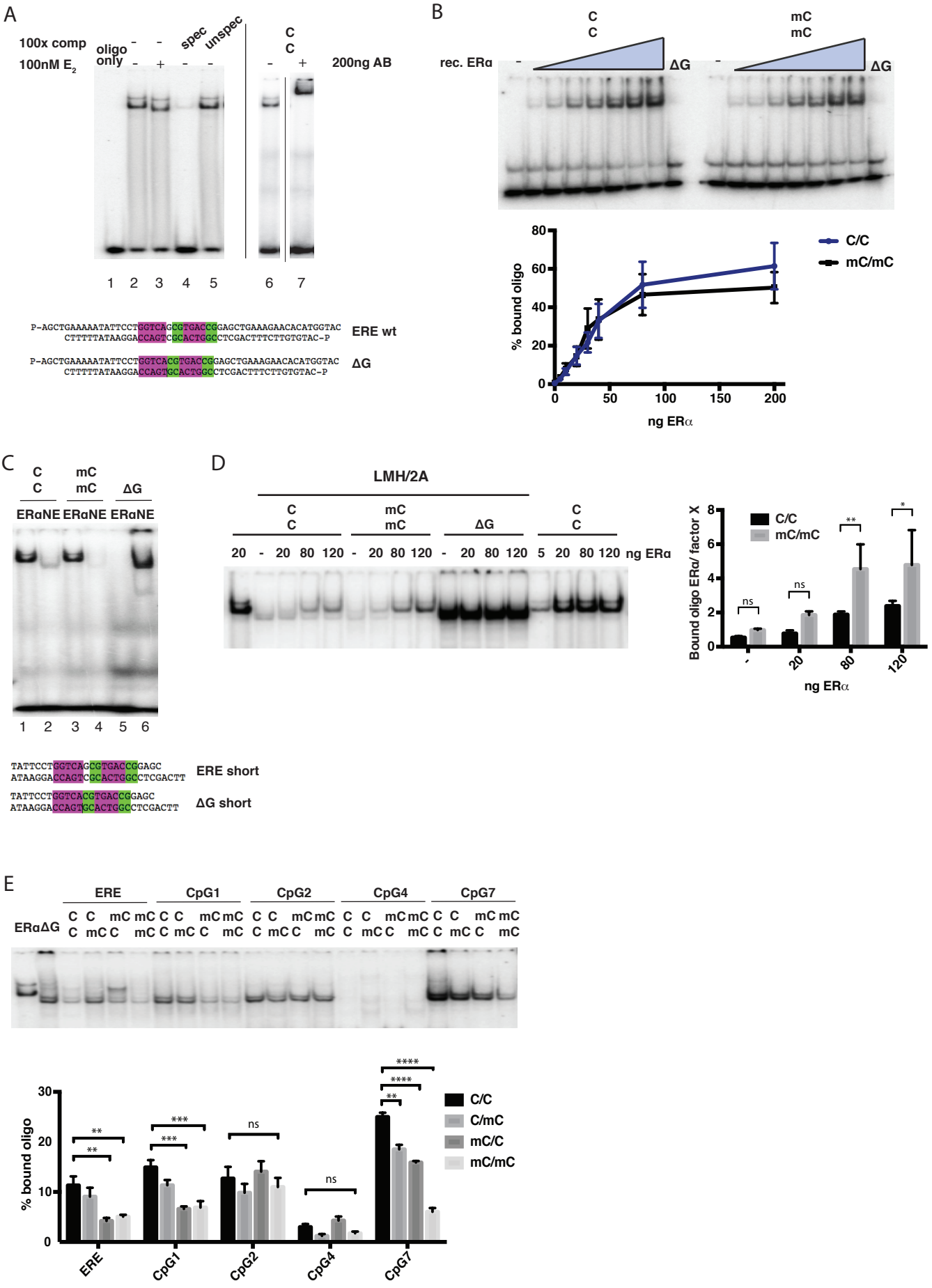
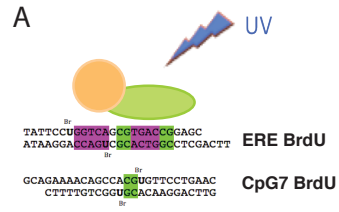
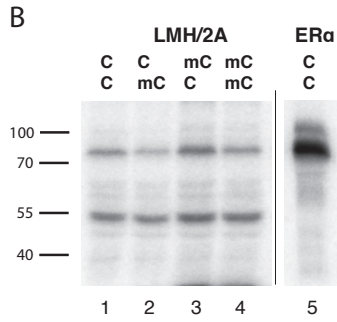


Figure 3


A
UV
TATTCCUGGTCAGCGTGACCGGAGC
ATAAGGACCAGUCGCACTGGCCTCGACTT
ERE BrdU
GCAGAAAACAGCCACGUGTTCCTGAAC
CTTTTGTCGGUGCACAAGGACTTG
CpG7 BrdU


B
LMH/2A
ERa
C C mC C mC C mC mC
C C
100
70
55
40
1 2 3 4 5

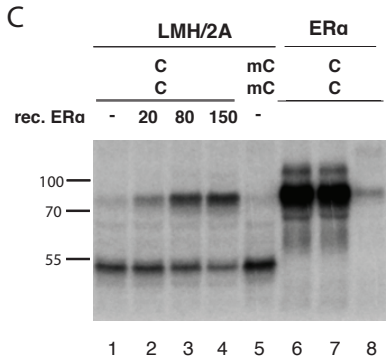
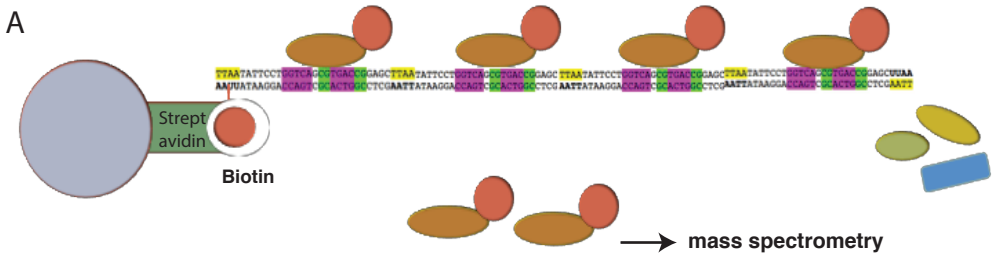
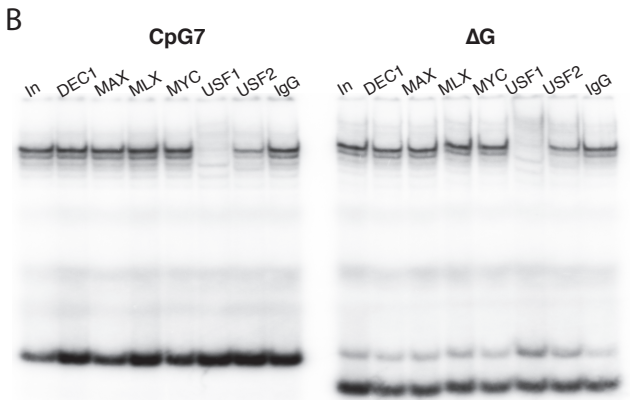
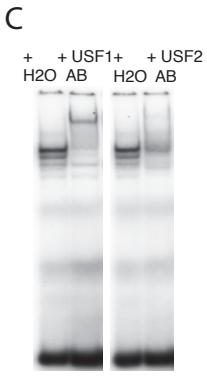

C
LMH/2A
ERa
C C mC mC
C C
rec. ERa - 20 80 150 -
100
70
55
1 2 3 4 5 6 7 8

Figure 4







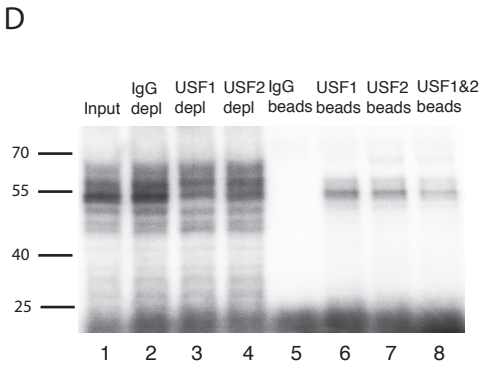
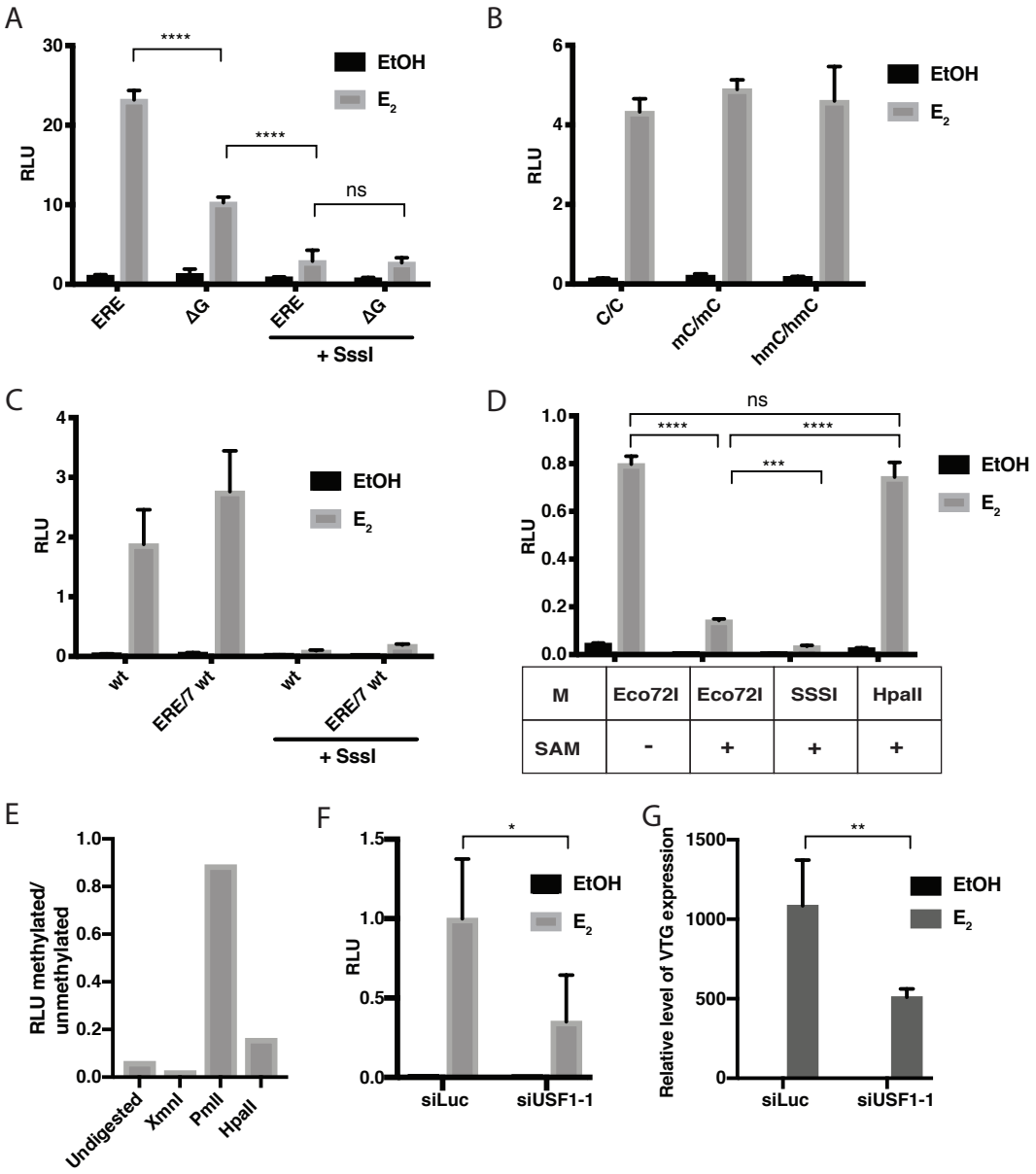


Figure 5



Gene	MW (kDa)	CpG 7	CpG 8	ERE	beads	ΔG	mCpG 7	mERE	Quantitative profile ANOVA
MAX	18	6 7 6		5 5 2		5 5 5	5 4 5	2 2 2	
MAFF	18	2		2 2				3 3 3	
MAFK	18	5 7 3		8 6 2			4 2	9 9 6	
SAP30	23	3 6 2				2 2	2		
BRMS1	28	7 7 4		3 3		3 2	2		
MLX	33	10 9 9		8 6		12 13 12		4	
USF1	34	6 5 6		7 4		7 5 6	3 3 3	4 5 5	
FOSL2	35	8 8 6	7 6 5	7 6 2	2	5 4 5	2 2	9 5 7	
USF2	37	9 9 10		8 10 2		8 10 10	6 3 2	6 4 3	
BRMS1L	38	5 6 3				2	2		
SDS3	38	6 7 6		3 2		3 3	4		
TFAP4	39	21 23 17	5 5 5	11 8	3 3	13 11 6	19 8 11	11 7 5	
BHLHE40	46	12 17 8		9 7 2		14 18 15		2 2	
MYC	49	3 3 3							
ZBTB26	50	2 2 2					2 3		
ZBTB9	51	8 10 8				8 8 5	11 4 7	2	
ATF7	52	7 4 4	7 6 4	9 5		9 7 6		7 5 3	
MeCP2	52			8 3		3 2	13 6 8	18 16 13	
FOXC1	57	20 16 8					20 8 12		
UBP1	60	3 3 3		10 6		3 4	7 3 4	11 6 3	
MNT	62	17 14 9		6 2		10 5 3	2		
MBD4	66						7 5 6	11 7 8	
FOXK2	69	17 19 14				3 2	10 2 4