



Article

# Direct and Base Excision Repair-Mediated Regulation of a GC-Rich *cis*-Element in Response to 5-Formylcytosine and 5-Carboxycytosine

Nadine Müller <sup>1</sup>, Eveliina Ponkkinen <sup>2</sup>, Thomas Carell <sup>2</sup> and Andriy Khobta <sup>1,3,\*</sup>

<sup>1</sup> Institute of Toxicology, University Medical Center Mainz, 55131 Mainz, Germany; nadgeier@uni-mainz.de

<sup>2</sup> Department of Chemistry and Biochemistry, Ludwig-Maximilians University Munich, 81377 Munich, Germany; eveliina.ponkkinen@cup.uni-muenchen.de (E.P.); thomas.carell@cup.uni-muenchen.de (T.C.)

<sup>3</sup> Institute of Nutritional Sciences, Friedrich Schiller University Jena, 07743 Jena, Germany

\* Correspondence: a.khobta@uni-jena.de

**Abstract:** Stepwise oxidation of the epigenetic mark 5-methylcytosine and base excision repair (BER) of the resulting 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) may provide a mechanism for reactivation of epigenetically silenced genes; however, the functions of 5-fC and 5-caC at defined gene elements are scarcely explored. We analyzed the expression of reporter constructs containing either 2'-deoxy-(5-fC/5-caC) or their BER-resistant 2'-fluorinated analogs, asymmetrically incorporated into CG-dinucleotide of the GC box *cis*-element (5'-TGGGCGGAGC) upstream from the RNA polymerase II core promoter. In the absence of BER, 5-caC caused a strong inhibition of the promoter activity, whereas 5-fC had almost no effect, similar to 5-methylcytosine or 5-hydroxymethylcytosine. BER of 5-caC caused a transient but significant promoter reactivation, succeeded by silencing during the following hours. Both responses strictly required thymine DNA glycosylase (TDG); however, the silencing phase additionally demanded a 5'-endonuclease (likely APE1) activity and was also induced by 5-fC or an apurinic/aprimidinic site. We propose that 5-caC may act as a repressory mark to prevent premature activation of promoters undergoing the final stages of DNA demethylation, when the symmetric CpG methylation has already been lost. Remarkably, the downstream promoter activation or repression responses are regulated by two separate BER steps, where TDG and APE1 act as potential switches.

**Keywords:** DNA demethylation; 5-formylcytosine; 5-carboxycytosine; thymine DNA glycosylase (TDG); base excision repair (BER); gene regulation; epigenetic marks



**Citation:** Müller, N.; Ponkkinen, E.; Carell, T.; Khobta, A. Direct and Base Excision Repair-Mediated Regulation of a GC-Rich *cis*-Element in Response to 5-Formylcytosine and 5-Carboxycytosine. *Int. J. Mol. Sci.* **2021**, *22*, 11025. <https://doi.org/10.3390/ijms222011025>

Academic Editor: Ennio Prospero

Received: 7 September 2021

Accepted: 7 October 2021

Published: 13 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Epigenetic regulation of the genome function is crucial for concerted realization of gene expression programs during development and for maintenance of the lineage-specific gene expression patterns in adulthood. During these processes, functions of the whole genome and of the individual genes are dynamically regulated by deposition and removal or maintenance of epigenetic marks, which include specific chromatin components as well as chemically modified DNA bases [1,2]. 5-methylcytosine (5-mC) at CpG dinucleotides is a major DNA modification in vertebrates, and the only inheritable one. It regulates numerous cellular processes, including tissue-specific gene expression, genomic imprinting and X-chromosome inactivation [3]. Depending on the nature of a specific gene regulatory element (and perhaps cell lineage), a methylated cytosine base can function to promote or preclude recruitment of regulatory proteins by two different mechanisms [4]. Symmetric methylation of CpG sites can be read by the methyl-CpG binding protein (MBP) family members [5,6], which mediate transcriptional repression by recruitment of chromatin-modifying enzymes [7–9]. In addition, the methyl mark sometimes acts by directly preventing transcription factor binding to their target sites [10–12].

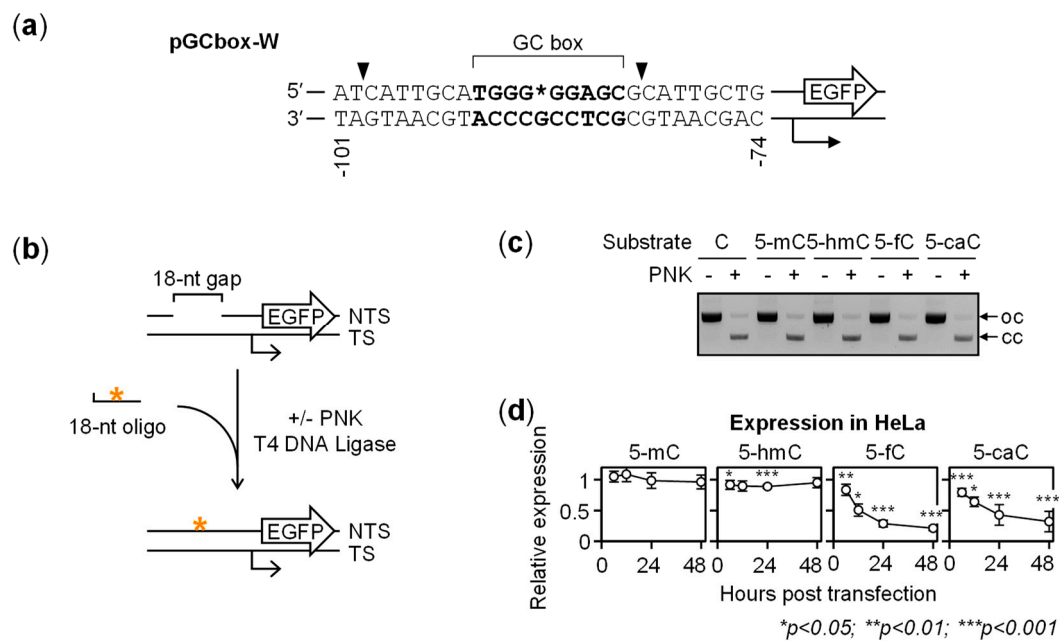
Recent progress in the field of epigenetic reprogramming has revealed that 5-mC can undergo enzymatic oxidation by ten-eleven translocation (TET) family dioxygenases to 5-hydroxymethylcytosine (5-hmC) and further to 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) [13–15]. The latter two modifications are recognized by the TDG DNA N-glycosylase and cleared from DNA by the base excision repair (BER) pathway [16–18]. Newly developed whole-genome sequencing techniques revealed enrichment of 5-fC and 5-caC in promoters of transcribed genes and active enhancers, indirectly suggesting their roles in transcriptional activation [19–21]. Still, investigation of the dynamics of the oxidation products of 5-mC at specific genomic loci and its significance for the regulation of transcription is hampered by low abundances of these modifications in the genome and a troublesome discrimination between different cytosine modifications by sequencing techniques [22,23]. It remains a matter of controversy whether 5-hmC should be regarded as a functionally autonomous epigenetic mark, similar to 5-mC, or barely a demethylation intermediate which does not have a dedicated function [22–24]. Potential biological functions of the downstream products of the TET-mediated oxidation pathway, 5-fC and 5-caC, are even harder to characterize, as these modifications are much scarcer in genomic DNA [23].

To directly address functional impacts of 5-hmC and its derivatives on the regulation of gene expression, these modifications can be site-specifically incorporated into a suitable reporter vector. We previously used such an approach to investigate the effects of the 5-mC oxidation products in a cyclic adenosine monophosphate response element (CRE). There, 5-mC, 5-hmC, 5-fC and 5-caC all negatively affected the promoter activity by inhibition of the cognate transcription factor CREB binding [25], thus corroborating the mechanism earlier described for 5-mC in the central CpG dinucleotide of the CRE sequence [10]. In addition, 5-fC and 5-caC initiated an indirect silencing mechanism attributed to the base excision and DNA strand cleavage [25]. Thereby, the outcomes of 5-fC and 5-caC in CRE largely recapitulated responses to common types of DNA damage processed by the BER [26–28]. It should be kept in mind, however, that binding of regulatory proteins to their cognate target elements in DNA, and consequently the outcomes of these interactions, can be highly context-specific [11,12,29]. Here, we systematically investigated the effects of synthetic 5-mC, 5-hmC, 5-fC and 5-caC incorporated into a GC-rich regulatory element (GC box), on the promoter activity. To model potential DNA demethylation stages, we placed defined cytosines modifications asymmetrically in either strand of the GC box, opposite to the non-methylated CpG dinucleotide. This led to identification of distinct functional states of the promoter conferred by defined steps of the DNA demethylation pathway.

## 2. Results

### 2.1. Gene Repression in HeLa Cells Induced by 5-fC and 5-caC in the GC Box CpG Dinucleotide

To investigate the impact of 5-mC oxidation products in a GC-rich promoter on the gene expression, we generated enhanced green fluorescence protein (EGFP) reporter constructs containing single 5-mC, 5-hmC, 5-fC or 5-caC residues at the CpG site of an artificial promoter containing the 5'-TGGGCGGAGC-3' GC box sequence as the only *cis*-regulatory element (Figure 1a). In the pGCbox-W and pGCbox-C plasmid vectors, this GC box enhances the basal gene expression of the downstream *EGFP* gene by the factor of  $\geq 2$ , compared to a scrambled DNA sequence [30]. Synthetic oligonucleotides containing all types of cytosine modifications were incorporated into the pGCbox-W vector, with efficiencies closely approaching 100% (Figure 1b,c). Moreover, the fractions of correctly ligated vector DNA were >92% for all modifications, which warranted that DNA topology or misalignment of the inserted synthetic oligonucleotides would not affect subsequent measurement of the promoter activity in cells.

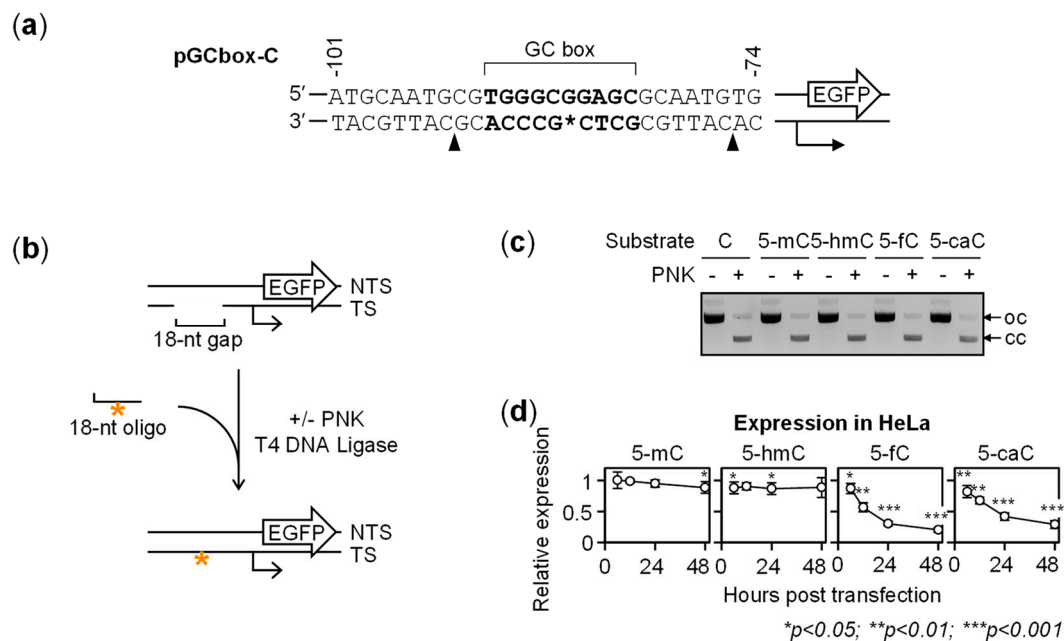


**Figure 1.** Effects of 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) in the purine-rich strand on the GC box activity. (a) Scheme of the reporter enhanced green fluorescence protein (*EGFP*) gene under the control of a GC box as the only upstream activating element. Synthetic oligonucleotide containing cytosine modifications at the unique CpG site (position indicated with an asterisk) were incorporated into the gap generated with the *Nb.BsrDI* nicking endonuclease (cleavage sites indicated with arrowheads). (b) Procedure for the incorporation of synthetic oligonucleotides containing C/5-mC/5-hmC/5-fC/5-caC (\*) into the gap generated by depletion of the purine-rich strand of the targeted GC box. Aliquots of the same annealing reactions were incubated with or without T4 polynucleotide kinase (PNK) to validate full replacement of the native DNA strand. DNA strand labeling denotes transcribed strand (TS) of the *EGFP* gene and the non-transcribed strand (NTS); broken arrow indicates the transcription start site and direction. (c) Agarose gel electrophoresis of the reporter constructs generated by targeted incorporation of C/5-mC/5-hmC/5-fC/5-caC into plasmid DNA. Arrows indicate the open circular (oc) and covalently closed (cc) forms. (d) Expression time course of constructs containing 5-mC/5-hmC/5-fC/5-caC in transfected HeLa cells. All values (mean  $\pm$  SD) are calculated relative to the expression of the control construct harboring synthetic oligonucleotide containing cytosine for  $n = 6$  independent experiments (the 12-h point was skipped in three of the experiments). Representative flow cytometry data is shown in Supplementary Materials Figure S1.

Time course analyses of pGCbox-W constructs containing different cytosine modifications at the CpG site in the purine-rich strand showed that 5-mC and 5-hmC did not appreciably influence the *EGFP* gene expression levels in transfected HeLa cells (Figure 1d). In contrast, constructs containing 5-fC and 5-caC showed considerably decreased *EGFP* expression levels. These effects were present already 6 h after transfection (the earliest time point when the expression could be reproducibly measured) and grew stronger in the course of time. At the initial time point, the decrease of the gene expression was more significant for 5-caC than for 5-fC, whereas subsequent dynamics was somewhat faster for 5-fC, leading to a stronger impairment of the gene expression at the later time points. Thus, relative to the reference construct containing unmodified cytosine, the expression levels decreased from  $83.4 \pm 6.5\%$  to  $21.0 \pm 3.7\%$  for 5-fC between 6 and 48 h post-transfection. For 5-caC, the promoter activity decreased from  $79.4 \pm 6.5\%$  to  $32.0 \pm 16.7\%$  in the same experiment series.

Also in the pyrimidine-rich DNA strand, the incorporation rates of all cytosine modifications were very close to 100%, with >88% synthetic strands fully and correctly ligated (Figure 2). As in the opposite strand, neither 5-mC nor 5-hmC caused considerable changes of the gene expression levels over the entire time interval, whereas 5-fC and 5-caC caused the gene expression to decline strongly and in a time-dependent manner. Once again, the inhibition of gene expression was initially stronger in case of 5-caC, whereas the 5-fC

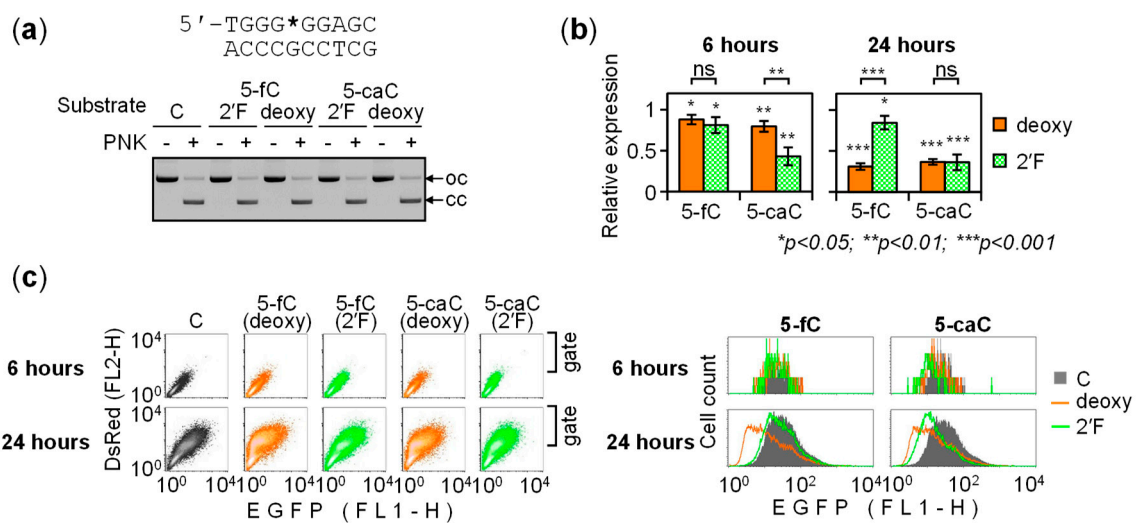
construct demonstrated a faster decline of the gene expression in the course of time. The 6- and 48-h expression values decreased from  $87.2 \pm 8.1\%$  to  $20.3 \pm 3.5\%$  for 5-fC and from  $82.0 \pm 9.8\%$  to  $29.3 \pm 6.3\%$  for 5-caC. Combined, the results show that 5-hmC and 5-mC in either strand of the GC box did not tangibly affect the promoter activity over a period of at least 48 h. In contrast, 5-fC and 5-caC clearly perturbed the gene expression in a time-dependent fashion.



**Figure 2.** Effects of 5-mC, 5-hmC, 5-fC and 5-caC in the pyrimidine-rich strand of the GC box. (a) Promoter sequence with the modified cytosine position (asterisk) and Nb.BsrDI nicking sites (arrowheads). (b) Scheme of the incorporation procedure of C/5-mC/5-hmC/5-fC/5-caC (\*) into the pyrimidine-rich strand of the GC box. (c) Agarose gel electrophoresis of the reporter constructs containing C/5-mC/5-hmC/5-fC/5-caC. The open circular (oc) and covalently closed (cc) forms are indicated by arrows. (d) The expression time course in transfected HeLa cells (mean  $\pm$  SD) for  $n = 6$  independent experiments (the 12-h point was skipped in three of them). Representative flow cytometry data are shown in Supplementary Materials Figure S1.

## 2.2. BER-Resistant 5-caC and, to Some Extent, 5-fC Directly Diminish the GC Box Activity

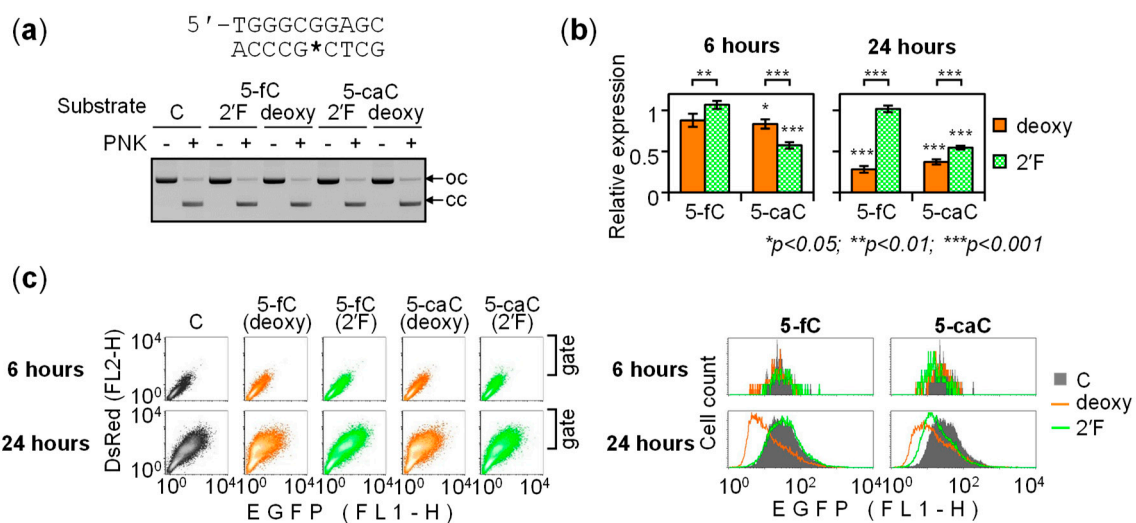
The observed dynamic changes of the expression of constructs carrying 5-fC and 5-caC, in contrast to 5-mC or 5-hmC, could likely be attributed to removal of these modifications from DNA. Therefore, to measure direct impacts of 5-fC and 5-caC on the GC box activity, it was necessary to eliminate their repair. Activities of DNA N-glycosylases towards their substrates, including 5-fC and 5-caC, can be efficiently inhibited by deoxyribose fluorination at the C2' position [25,31]. This motivated us to generate GC box constructs carrying the 2'-(R)-fluorinated derivatives of 5-fC and 5-caC. Both modifications were incorporated into the purine-rich DNA strand as efficiently as the respective 2'-deoxy compounds (Figure 3a). Quantification of the EGFP expression levels in transfected HeLa cells showed that 2'-(R)-fluorination entirely reversed the repressory effect of 5-fC, as the difference between the 6- and 24-h points disappeared. The expression levels were  $81.2 \pm 9.8\%$  (at 6 h) and  $84.6 \pm 8.4\%$  (at 24 h post-transfection), relative to the construct containing cytosine. The respective values for 2'-deoxy 5-fC were  $88.0 \pm 5.7\%$  and  $31.0 \pm 4.1\%$  (Figure 3b). The repression of the promoter activity caused by 5-fC between 6 and 24 h post-transfection was statistically highly significant ( $p = 8.6 \times 10^{-6}$ , Student's two tailed heteroscedastic t-test), whereas the effect of 2'-(R)-fluorinated 5-fC was not ( $p = 0.62$ ). In the case of 5-caC, the differences between the 6- and 24-h time points were, again, highly significant for deoxynucleotide ( $p = 1.6 \times 10^{-4}$ ) but not for the 2'-fluorinated analog ( $p = 0.38$ ). We thereby conclude that dynamic changes of the promoter activity induced by 5-fC and 5-caC are the consequence of excision of the modified bases by a DNA N-glycosylase mechanism.



**Figure 3.** Effects of BER-resistant analogs of 5-fC and 5-caC in the purine-rich strand on the GC box activity. **(a)** Efficient incorporation of 5-fC and 5-caC deoxyribonucleotides (deoxy) and their 2'-(R)-fluorinated analogs (2'F) into the purine-rich strand of the GC box. Agarose gel electrophoresis of the fully ligated constructs and the respective “no PNK” controls. The open circular (oc) and covalently closed (cc) forms are indicated by arrows. **(b)** Quantification of the EGFP expression driven by GC box containing 5-fC and 5-caC with or without 2'-(R)-fluorination, relative to C, 6 and 24 h post-transfection (mean  $\pm$  SD,  $n = 4$  independent experiments). **(c)** A representative flow cytometry experiment. Cells were gated based on the expression of the transfection marker DsRed, as shown on two-dimensional fluorescence scatter-plots (left panels), and the resulting EGFP signal distributions (right panels, samples overlaid) analyzed to determine the median EGFP fluorescence.

As 2'-(R)-fluorinated 5-fC and 5-caC in the GC box constructs showed stable levels of the reporter gene expression, it was now possible to quantify their direct effects on the promoter activity, based on the EGFP expression levels in transfected HeLa cells. We used 24-h values to estimate these effects, since fluorescent cell counts as well as average fluorescence intensity per cell were higher at this time point (Figure 3c). Reduction of the gene expression by 2'-(R)-fluorinated 5-fC was minimal yet statistically significant ( $p = 0.035$ ). The effect of 2'-(R)-fluorinated 5-caC was much stronger ( $36.2 \pm 9.6\%$  residual gene expression relative to the cytosine control) and statistically highly significant ( $p = 9.3 \times 10^{-4}$ ). Considering that GC box enhances the gene expression by a factor of 2 to 2.5 over the basal expression level [30], the results indicate that the presence of a single 5-caC in the purine-rich strand abolished the activation attributable to GC box completely or almost completely.

We also analyzed the effects of the 2'-(R)-fluorinated 5-fC and 5-caC analogs in the pyrimidine-rich GC box strand (Figure 4). As in the purine-rich strand, 2'-(R)-fluorination abolished time-dependent repression by both cytosine modifications, as judged by disappeared differences between the 6- and 24-h expression values. The respective  $p$ -values were  $8.1 \times 10^{-5}$  (5-fC), 0.19 (2'-fluoro 5-fC),  $9.1 \times 10^{-4}$  (5-caC) and 0.30 (2'-fluoro 5-caC). The effects of BER-resistant modifications differed slightly between the strands. Thus, 2'-(R)-fluorinated 5-fC in the pyrimidine-rich strand did not at all inhibit the GC box activity, showing the 24-h relative expression level of  $102.1 \pm 4.1\%$  ( $p = 0.38$ ). The inhibitory effect of 2'-(R)-fluorinated 5-caC remained highly significant in the pyrimidine-rich strand, with relative expression of  $54.7 \pm 2.3\%$  ( $p = 3.5 \times 10^{-5}$ ).



**Figure 4.** Effects of BER-resistant analogs of 5-fC and 5-caC in the pyrimidine-rich strand on the GC box activity. (a) Agarose gel electrophoresis of the fully ligated deoxy and 2'F constructs and the respective “no PNK” controls. The open circular (oc) and covalently closed (cc) forms of the vector are indicated by arrows. (b) Relative expression of the 5-fC and 5-caC (deoxy versus 2'F) constructs 6 and 24 h post-transfection (mean  $\pm$  SD,  $n = 4$  independent experiments). (c) A representative flow cytometry experiment: two-dimensional fluorescence scatter-plots (left) and the derived EGFP fluorescent distribution plots for the indicated transfection condition (right panels, samples overlaid).

In summary, apart from subtle quantitative differences between the purine- and pyrimidine-rich strands of the GC box, the results indicate that processing of 5-fC and 5-caC by BER plays a key role in the regulation of the gene expression. Over an extended time interval (24 h), in HeLa cells, the role of BER is manifested by dynamic repression of the promoter activity, in contrast to steady expression levels observed in the presence of the repair-resistant analogs. The potent negative effect of 2'-(R)-fluorinated 5-caC on the promoter activity strongly suggests that 5-caC functions as a negative regulatory mark in the GC box. The function of 5-fC seems to be different, as this modification has little or no effect on the GC box activity, depending on the DNA strand.

### 2.3. BER of 5-caC in the GC Box Induces a Transient Promoter Activation

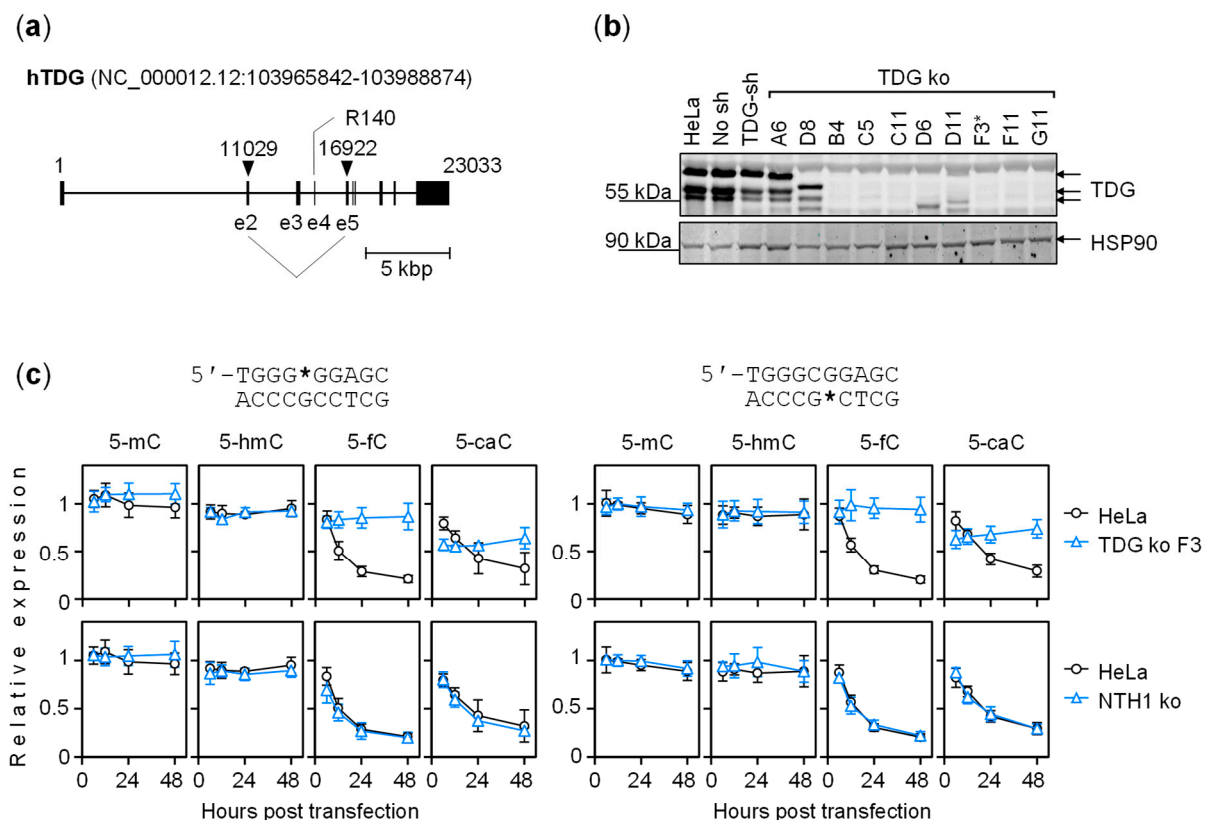
The evidence implicating the base excision in the transcriptional repression overall corroborated the idea that the time-dependent decline of the gene expression observed earlier (Figures 1 and 2) was caused by gradual BER-mediated clearance of 5-fC and 5-caC from DNA. However, even though the cumulative outcomes of the excision of 5-fC and 5-caC were manifested after 24 h as repression of the reporter gene, we noticed that the initial effects of 5-caC were the opposite (Figures 3b and 4b). In particular, the 6-h EGFP expression values showed an inverse relationship between the activities of the GC box constructs containing BER-resistant versus BER-sensitive modifications, wherein the expression was significantly enhanced by 2'-deoxy 5-caC compared to its 2'-fluorinated counterpart in both the purine-rich ( $p = 1.2 \times 10^{-3}$ ) and the pyrimidine-rich DNA strands ( $p = 7.2 \times 10^{-4}$ ). These results indicate that removal of 5-caC reactivated the GC box promoter at least transiently, whereas subsequent gradual decrease of the EGFP expression (reported in Figures 1 and 2) suggests that activation was followed by transcriptional silencing within the next few hours. Thereby, the response to 5-caC in GC box is different from reported previously in CRE, where TDG-mediated silencing occurred without a preceding activation phase [25].

If the pulse of promoter activation during the first hours is specific to 5-caC, it would provide an explanation to the ostensibly slower silencing kinetics in comparison to the 5-fC constructs (Figures 1 and 2). Indeed, the 5-fC constructs did not display a similar early activation by the BER mechanism. Rather on the contrary, BER-sensitive 2'-deoxy 5-fC in

the pyrimidine-rich strand slightly but significantly decreased the expression with respect to its BER-resistant counterpart already after 6 h ( $p = 6.4 \times 10^{-3}$ ), while the respective values for the purine rich strand did not differ significantly ( $p = 0.28$ ).

#### 2.4. The Dynamics of Transcriptional Regulation by 5-fC and 5-caC Is Entirely TDG-Dependent

We next aimed at determining whether the effects described for the 2'-(R)-fluorinated analogs of 5-fC and 5-caC hold true also for the modifications in the context of the 2'-deoxyribose backbone. To eliminate BER of 5-fC and 5-caC in HeLa cells, we targeted the TDG locus by CRISPR-Cas9 (Figure 5a). This resulted in deletions encompassing the exons 2 to 5 (Supplementary Materials Figure S2) and conferred complete elimination of the TDG protein expression in eight of the selected clones (Figure 5b).



**Figure 5.** Effect of TDG knockout on the activity of GC box containing 5-mC/5-hmC/5-fC/5-caC. **(a)** Strategy for TDG knockout in HeLa cells. The TDG gene was targeted simultaneously at two different sites by a pair of single guide RNAs enclosing the codon for the catalytic R140 residue. The Cas9 cut sites are indicated by arrowheads. **(b)** Validation of TDG negative clones by Western blotting. Clone F3 further used as a transfection host for the gene expression analyses is marked (\*). **(c)** Expression time course of constructs containing 5-mC/5-hmC/5-fC/5-caC in the purine-rich (left group of plots) or the pyrimidine-rich GC box strand (right). Isogenic clonal TDG knockout (upper row) and NTH1 knockout (lower row) cell lines were compared with the parental HeLa cell line (overlaid in plots). All cell lines were transfected in parallel with the same sets of reporter constructs. Results of  $n = 3$  independent experiments (mean  $\pm$  SD).

We next performed time-course expression analyses of GC box constructs containing the whole spectrum of cytosine modifications in a TDG knockout cell line derived from clone F3. The results showed that dynamic EGFP expression changes characteristic for the parental HeLa cells were completely abrogated by TDG knockout (Figure 5c). In the absence of TDG, 5-caC steadily inhibited promoter activity, in the same manner as its 2'-(R)-fluorinated analog (Figure 5c and Supplementary Materials Figure S3). As this result also closely recapitulated the effect of the 2'-(R)-fluorinated analog in the parental HeLa cell line (Figures 3 and 4), we conclude that 2'-fluorination does not influence biological

properties of 5-fC and 5-caC beyond stabilization of the N-glycosidic bond. In the case of 5-fC, a rather mild inhibition of transcriptional activation was observable only when the modification was present in the purine-rich strand. This outcome was, again, consistent with results obtained for the 2'-(R)-fluorinated analog (Figures 3 and 4; Supplementary Materials Figure S3).

In an independent TDG knockout clone C11 (not shown), responses to 5-fC and 5-caC were the same as in F3. In contrast, responses in the isogenic NTH1 knockout cell line displayed the same pattern as in the parental HeLa cell line (Figure 5c). We thereby conclude that the phenotype displayed by the TDG knockout clones was specific to the targeted locus. Finally, the expression of constructs carrying 5-mC or 5-hmC essentially did not differ between the cell lines, as expected based on the universally stable character of these modifications (Figure 5c). In summary, the results in TDG knockout cells corroborate the conclusion that unrepaired 5-caC prevents or counteracts the GC box activation in the absence of repair. Besides, the steady character of the effect of 5-caC on the gene expression in the absence of TDG strongly suggests that TDG-dependent BER is by far the most efficient pathway for 5-caC removal in the chosen cell model.

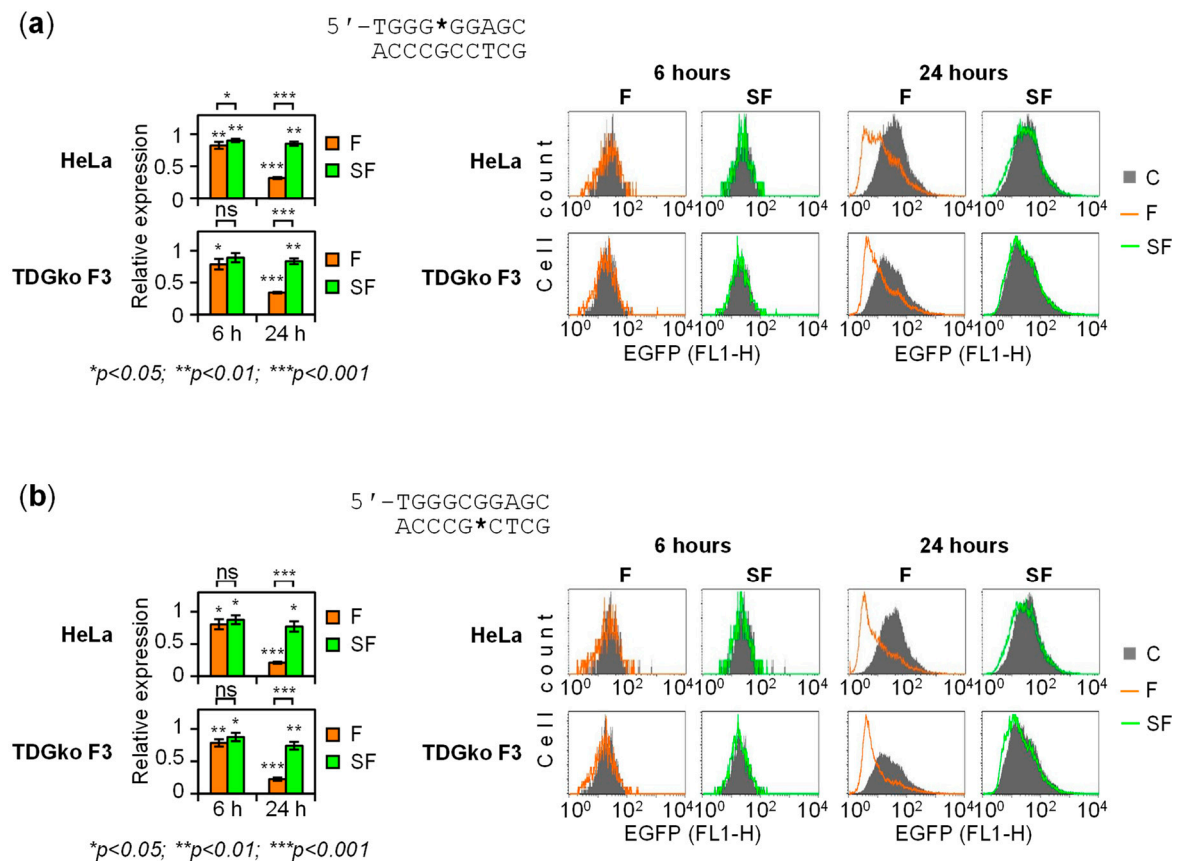
Of note, the results confirmed our previous conclusion that BER is accountable not only for long-term gene silencing induced by 5-fC/5-caC but also for transient activation by 5-caC at the beginning of the time course. Moreover, by comparison of expression levels of the same constructs between cell lines with different TDG states, we now could attribute the activation specifically to TDG, as the 6-h expression values were significantly higher in the parental HeLa cell line than in the derived TDG knockouts (Figure 5c and data not shown). For 5-caC in the purine-rich GC box strand, relative expression level was  $57.0 \pm 5.8\%$  in the TDG knockout clone F3 versus  $84.3 \pm 5.6\%$  in HeLa ( $p = 4.2 \times 10^{-3}$ ). In the pyrimidine-rich strand, it was  $62.6 \pm 9.4\%$  versus  $84.7 \pm 7.2\%$  in HeLa ( $p = 3.2 \times 10^{-2}$ ). The respective expression levels in clone C11 also lied significantly below the HeLa values:  $66.9 \pm 5.3\%$  ( $p = 1.7 \times 10^{-2}$ ) for 5-caC in the purine-rich strand and  $65.2 \pm 5.9\%$  ( $p = 2.2 \times 10^{-2}$ ) in the pyrimidine-rich strand.

### 2.5. Regulation of the GC Box Promoter by Acytosinic Sites and the Effect of Strand Cleavage

TDG is a monofunctional DNA-N glycosylase that leaves deoxyribose as a reaction product in the DNA strand [17]. Intriguingly, analogous apurinic/apyrimidinic (AP) lesions were implicated in the regulation of GC-rich DNA *cis*-elements previously [30,32]. Therefore, we constructed reporters carrying acytosinic lesions in either of the GC box strands. This was achieved by using oligonucleotides, where 2'-deoxy cytosines at the CpG sites were replaced by two types of tetrahydrofuran AP lesions (Supplementary Materials Figure S4). Being a close structural analog of the natural AP lesion, tetrahydrofuran (F) is efficiently excised by APE1; however, a combination of F with a phosphorothioate linkage on the 5' side (SF) yields an APE1-resistant lesion [26,33].

The expression levels of GC box constructs containing APE1-resistant acytosinic sites (SF) were slightly decreased with respect to the counterparts containing unmodified C (Figure 6). In both HeLa and TDG knockout cells, they were in the range of 87–90% at the 6-h time point, regardless of the strand, indicating that AP lesion has a mild inhibitory effect on the promoter activity. The respective values for F lesions were somewhat lower (78–83%), which is probably attributable to their incision by APE1. The difference between the F and SF expression levels grew highly significant by the 24-h points: whilst expression of F constructs intensely declined, the expression levels of both SF constructs decreased only slightly. Consequently, the potent gene silencing response to F is attributable to the phosphodiester bond cleavage (Figure 6).





**Figure 6.** Effect of acytosinic sites on the GC box activity. **(a)** Expression analyses of constructs containing a synthetic abasic site in the purine-rich strand of the GC box. Relative EGFP expression (mean  $\pm$  SD) for  $n = 4$  independent experiments (left) and the representative fluorescent distribution plots (right). Acytosinic modifications used were tetrahydrofuran (F) or a tetrahydrofuran with nuclease resistant phosphorothioate 5'-linkage (SF). **(b)** Analogous analyses of constructs containing F or SF acytosinic sites in the pyrimidine-rich strand.

Endonucleolytic strand cleavage seems to be a rather common inducer of transcriptional repression, as previously reported for AP lesions in other contexts [28,30,32]. In contrast, endonuclease-resistant aguaninic lesions can occasionally cause promoter activation, presumably via a non-processive APE1 binding [30,32]. Based on present results, we deduce that acytosinic lesions at the GC box CpG dinucleotide do not cause a full promoter activation. Nonetheless, they enable substantially higher expression levels than observed in the presence of the unreparable 5-caC analog (Figures 3 and 4).

### 3. Discussion

Both active and passive DNA demethylation mechanisms must necessarily involve a step when only one cytosine in a given double-stranded CpG dinucleotide is modified (methylated, hydroxymethylated, formylated or carboxylated) at the C5 position. Recognition of CpG sites by methyl-CpG-binding domain (MBD) proteins, including DNMT1, is strongly inhibited in the presence of 5-hmC [34–36]. This causes impaired methylation maintenance and leads to hemi-hydroxymethylated sites upon replication [37]. An even more pronounced DNMT1 inhibition was reported by 5-fC and 5-caC, implying that replication of DNA containing these modifications would generate hemi-formylated and hemi-carboxylated CpG sites [38]. Hence, we propose that the impacts of defined cytosine modifications on the GC box activity that were reported here can be extrapolated to model functional outcomes of a range of potential DNA demethylation intermediates. Our results suggested that GC box containing a hemi-methylated CpG dinucleotide (with 5-methyl group present in either DNA strand) is as active as in the absence of any modification

(Figures 1, 2 and 5). This is in agreement with biochemical evidence that methylation does not inhibit transcription factor binding to various GC box consensus motifs [39,40]. We further found that GC box remained fully active in the presence of single 5-hmC or 5-fC in the pyrimidine-rich strand, whereas the modifications in the pyrimidine-rich strand seemed to cause only a very minor decrease of the activity (Figures 1–5). In contrast, 5-caC appears to be the only cytosine modification within the TET pathway, which causes a strong direct impairment of the GC box activity (Figures 3–5). Our results thus suggest that the nature of the modification present at asymmetrically modified CpG dinucleotides could be of a critical functional importance.

The inhibitory effect of 5-caC on the GC box can be relieved by a TDG-dependent mechanism. Thus, our results show that expression of constructs carrying 5-caC is reactivated early upon their delivery to cells, but only if TDG is available (Figure 5) and base excision is unhindered (Figures 3 and 4). In HeLa cells, a short period of promoter activation was followed by the onset of a repressed state, which was TDG-dependent as well. However, this response should be regarded as separate from previous activation, since it required, besides base excision, the presence of a labile phosphodiester linkage 5' to the target nucleotide (Figure 6). A strand cleavage reaction at the critical position is most likely catalyzed by APE1, which is by far the most important AP endonuclease in human cells, whereas the downstream gene silencing mechanism remains to be elucidated. Similar silencing responses were seen with various BER substrates previously and were attributed to adoption of a repressive chromatin structure after the completion of BER [28,41]. It is intriguing to speculate that TDG and APE1 may be subjected to regulation in particular cell lineages and genomic contexts or, perhaps, by endogenous or exogenous signals. This would allow diversification of transcriptional responses and enable plasticity of the epigenetic states in cells undergoing TET- and TDG-dependent DNA demethylation.

To understand the effects of single intermediates arising during the stepwise DNA demethylation pathway on the promoter activity, the expression levels should be related to the amount of a given cytosine modification present. For DNA modifications undergoing repair in cells, quantification of residual modifications specifically in transcription-competent DNA poses a serious technical challenge, because significant fractions of vector DNA distribute to non-nuclear compartments or undergo dynamic changes of expression due to chromatinization [41]. Nevertheless, under the assumption that the effects of BER-resistant 2'-fluorinated synthetic analogs of 5-fC and 5-caC remain steady, it was possible to derive conclusions about direct effects of these modifications on the GC box activity. Indeed, the expression levels in the presence of these modifications, relative to cytosine controls, were found constant over time (Figures 3 and 4). The same holds for natural (2'-deoxy) modifications, when delivered to repair-deficient host cells (Figure 5). Similarly, the effect of AP intermediate could be inferred based on the results obtained with APE1-resistant SF lesion, which also displayed steady expression levels (Figure 6).

#### 4. Conclusions

In summary, our results revealed several levels of GC box regulation by cytosine modifications generated within the active DNA demethylation pathway. GC box upstream from RNA polymerase II core promoter retains its full activity when the CpG dinucleotide is hemi-methylated. Oxidation of 5-mC to 5-hmC and 5-fC is well tolerated; however, the ultimate oxidation product 5-caC behaves as a stable repressory mark in the absence of TDG or if protected from the N-glycosylase activity. TDG restores the GC box activity almost completely. It can be assumed that promoter reactivation occurs as soon as an acytosinic lesion is generated. In support of this notion, modeling of the post-excision step, with the help of AP site analogs, led to very similar effects on the promoter activity (Figure 6). Reactivation by the excision of 5-caC lasted only for a few hours in HeLa cells before getting overturned by a concurrent silencing response, which was elicited by strand cleavage at the AP lesion. Although our data provide no evidence of a more permanent promoter activation by the TDG pathway, such a scenario may take place in another cell type or

promoter context. For instance, AP lesion can be protected from endonucleolytic processing by a non-canonical DNA structure [42–45] or shielded by a specific binding protein [46,47]. Alternatively, signaling downstream from the single strand break generation could be modulated towards a different functional state of the promoter.

## 5. Materials and Methods

### 5.1. Synthetic Oligonucleotides Carrying Cytosine Modifications

Deoxyribo-oligonucleotides containing the specified cytosine modifications were 5'-CATTGCATGGG[C\*]GGAGCG and 5'-CATTGCGCTC[C\*]GCCACG (where C\* is C, 5-mC, 5-hmC, 5-fC or 5-caC). DNA CE-phosphoramidites Bz-dA, Bz-dC, iBu-dG, dT and Bz-mdC were obtained from Glen Research (Sterling, VA, USA) or Link Technologies (Bellshill, Scotland, UK). Syntheses of the 5-hmC, 5-fC and 5-caC phosphoramidites [48], along with the 2'-(R)-fluorinated derivatives of 5-fC and 5-caC [31], were performed as described previously. The solid-phase synthesis, HPLC-purification and MALDI/MS quality-control procedures of the 18-mer deoxyribo-oligonucleotides were performed by using the standard protocols described previously [31]. Synthetic apurinic/aprimidinic (AP) lesions were tetrahydrofuran with either the phosphodiester (F) or the APE1-resistant phosphorothioate 5'-linkage (SF). Oligonucleotides 5'-CATTGCATGGG[AP]GGAGCG and 5'-CATTGCGCTC[AP]GCCACG (where AP is F or SF) were purchased from BioSpring GmbH (Frankfurt am Main, Germany).

### 5.2. Generation of Reporter Constructs Containing Cytosine Modifications in the GC Box CpG Dinucleotide

Vectors pGCbox-W and pGCbox-C, allowing substitution of the selected GC box strand with synthetic oligonucleotides, were described previously [30]. Both vectors contain a common GC box motif 5'-TGGGCGGAGC as the only *cis*-regulatory element upstream from the RNA polymerase II transcription initiation site and sustain equivalent levels of the reporter *EGFP* gene expression. Defined modifications targeting cytosines of the CpG dinucleotide were introduced into the purine-rich GC box strand of pGCbox-W or into the pyrimidine-rich strand of pGCbox-C, using the available sites by the Nb.BsrDI nicking endonuclease. The procedure was described in detail previously for introducing 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxoG) into the same vectors [30]. Plasmid DNA was cut at two tandem sites by the Nb.BsrDI nicking endonuclease, and the excised native DNA strand fragments were substituted for synthetic oligonucleotides containing the modifications of choice, as described previously for a different promoter context [25]. The efficient incorporation of synthetic DNA strands was verified by inhibition of ligation in the absence of T4 polynucleotide kinase [49]. Percentages of covalently closed DNA in the vector preparations were determined by agarose gel electrophoresis in the presence of 0.5 mg/L ethidium bromide, followed by band quantification, using a GelDoc™ EZ imager and the ImageLab™ software (Bio-Rad Laboratories, GmbH, Munich, Germany), as described previously [28]. The presence of AP lesions was verified by excision analysis, using endonuclease IV (NEB GmbH, Frankfurt am Main, Germany). Constructs were incubated with endonuclease IV (8 U/200 ng plasmid DNA) 1 h at 37 °C in 15 µL buffer composed of 10 mM HEPES (pH 7.5), 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid and 0.1 mg/mL nuclease free bovine serum albumin (NEB). The enzyme was heat-inactivated for 20 min, at 85 °C.

### 5.3. Quantitative Analyses of EGFP Expression in Transfected Cells

HeLa cells used in experiments were clones descending from HELA cervical carcinoma cell line (German Collection of Microorganisms and Cell Cultures No. ACC 57). The derived TDG and NTH1 knockout cell lines were generated in our lab. Cells exponentially growing in 6-well plates were transfected with mixtures containing 400 ng GC box reporter vector (pGCbox-W or pGCbox-C) and 400 ng tracer pDsRed-Monomer-N1 vector (Clontech, Saint-Germain-en-Laye, France), using the Effectene reagent (QIAGEN, Hilden,

Germany), as described previously [25,30]. GC box constructs containing C or the specified modifications at the respective site were transfected in parallel. Cells were split 6 h after transfection and either fixed immediately with 1% of formaldehyde or seeded into separate wells to be fixed after the indicated time intervals, as described previously. Harvested formaldehyde-fixed cells were analyzed by using a FACSCalibur™ flow cytometer and the CellQuest™ Pro software (Beckton Dickinson GmbH, Heidelberg, Germany). The EGFP expression was quantified as median FL1-H fluorescence over the population of transfected cells, defined by the DsRed expression, as described previously [41]. Relative expression levels were calculated for each modification type in the individual experiments based on the expression of the control construct harboring cytosine.

#### 5.4. TDG Gene Knockout in HeLa Cells

HeLa-derived clonal cell lines with deletion of a critical portion of the *TDG* gene were generated by a CRISPR-Cas9-mediated gene-editing procedure, using the pX330-sgCas9-HF1 vector (Addgene, Watertown, MA, USA) according to the supplier's instructions [50], with minor adjustments. Single guide RNA (sgRNA) sequences targeting human *TDG* locus were designed by using CHOPCHOP online tool and subcloned into the BbsI sites [51]. A pair of sgRNAs targeting the exons 2 (sg11029) and 5 (sg16922) was identified as the most efficient, based on screening of four different sgRNAs. The inserts used for cloning of these sgRNAs were obtained by pairwise annealing of synthetic oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany): 5'-CACCGACGAAATATGGACGTTCAAG (sg11029, forward) with 5'-AAACCTTGAACGTCCATATTTTCGTC (sg11029, reverse) and 5'-CACCGCTACCAGGGAAGTATGGTAT (sg16922, forward) with 5'-CACCGCTACCAGGGAAGTATGGTAT (sg16922, reverse). Exponentially growing HeLa cells were co-transfected in 6-well plates with the combination of both sgRNA/Cas9 expression vectors (300 ng each) and 50 ng pZAJ vector [52] as a transfection marker. Transfected cells were sorted after 40 h into two 96-well plates, based on the top 5 percentile of the EGFP expression. Single cell sorting was performed in the Flow Cytometry Core Facility of the Institute for Molecular Biology gGmbH (IMB Mainz) under supervision of scientific staff. After two weeks, growing clones were transferred to 25 cm<sup>2</sup> flasks and screened by PCR for the presence of a deletion spanning the sequence between the sgRNA-targeted sites in the exons 2 and 5 (including the catalytic R140 codon). The primers were 5'-TCCTCTGTAATCCACTCTAA (forward) and 5'-AGCTCAGCTTGAAGTAGATA (reverse). Preselected clones were next screened for the presence of non-rearranged *TDG* alleles to eliminate the positives. The primer pair detecting the non-rearranged exon 2 was 5'-TCCTCTGTAATCCACTCTAA (forward) and 5'-ATGTCCCTACTCTGATCTTT (reverse). The remaining clones were expanded and the *TDG* knockout was validated by Western blotting of protein extracts with a 1:5000 dilution of the *TDG* rabbit polyclonal antibody #PA5-29140 (Thermo Fisher Scientific Inc.) and a 1:10,000 dilution of the IRDye® 800CW donkey anti-rabbit IgG #926-32213 (LI-COR Biosciences GmbH, Bad Homburg, Germany). Blots were analyzed by using the Odyssey 9120 infrared imaging system (LI-COR). Stripped membranes were subsequently probed with a 1:10,000 dilution of the mouse monoclonal antibody AC88 to HSP90 # ADI-SPA-830 (Enzo Life Sciences GmbH, Lörrach, Germany) and a 1:10,000 dilution of the IRDye® 800CW donkey anti-mouse secondary antibody # 926-68072 (LI-COR).

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms222011025/s1>.

**Author Contributions:** Conceptualization, A.K. and T.C.; methodology, A.K. and T.C.; validation, N.M., E.P. and A.K.; formal analysis, N.M. and A.K.; investigation, N.M.; resources, T.C. and E.P.; data curation, N.M.; writing—original draft preparation, A.K. and N.M.; writing—review and editing, A.K., N.M., T.C. and E.P.; visualization, N.M. and A.K.; supervision, A.K.; project administration, A.K.; funding acquisition, A.K. and T.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): project-ID 393547839—SFB 1361, project-ID 321812289—GRK 2338, grant numbers KH263/2-2 and KH263/5-1. This project received funding from the European Union’s Horizon 2020 Research and Innovation Programme, under the Marie Skłodowska-Curie grant agreement No. 765266.

**Acknowledgments:** We thank the following persons and organizations for help with experiments and reagents: IMB Flow Cytometry Core Facility for help with cell sorting; Oliver H. Krämer and Mandy Beyer at the Institute of Toxicology (UMC Mainz) for advice and shared Western blotting reagents; Martin Rossa for preparing the first batch of synthetic DNA strands; Marta Rodriguez-Alvarez for help with handling the TDG knockout clones and Nataliya Kitsera for constructive comments on the manuscript draft.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Bernstein, B.E.; Meissner, A.; Lander, E.S. The mammalian epigenome. *Cell* **2007**, *128*, 669–681. [[CrossRef](#)] [[PubMed](#)]
- Berger, S.L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23*, 781–783. [[CrossRef](#)] [[PubMed](#)]
- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21. [[CrossRef](#)] [[PubMed](#)]
- Tate, P.H.; Bird, A.P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **1993**, *3*, 226–231. [[CrossRef](#)]
- Boyes, J.; Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **1991**, *64*, 1123–1134. [[CrossRef](#)]
- Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bird, A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **1992**, *69*, 905–914. [[CrossRef](#)]
- Jones, P.L.; Veenstra, G.J.; Wade, P.A.; Vermaak, D.; Kass, S.U.; Landsberger, N.; Strouboulis, J.; Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **1998**, *19*, 187–191. [[CrossRef](#)]
- Nan, X.; Ng, H.H.; Johnson, C.A.; Laherty, C.D.; Turner, B.M.; Eisenman, R.N.; Bird, A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **1998**, *393*, 386–389. [[CrossRef](#)]
- Fuks, F.; Burgers, W.A.; Brehm, A.; Hughes-Davies, L.; Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* **2000**, *24*, 88–91. [[CrossRef](#)]
- Iguchi-Arigo, S.M.; Schaffner, W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **1989**, *3*, 612–619. [[CrossRef](#)]
- Spruijt, C.G.; Gnerlich, F.; Smits, A.H.; Pfaffeneder, T.; Jansen, P.W.; Bauer, C.; Munzel, M.; Wagner, M.; Muller, M.; Khan, F.; et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **2013**, *152*, 1146–1159. [[CrossRef](#)]
- Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P.K.; Kivioja, T.; Dave, K.; Zhong, F.; et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*. [[CrossRef](#)] [[PubMed](#)]
- Tahiliani, M.; Koh, K.P.; Shen, Y.; Pastor, W.A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L.M.; Liu, D.R.; Aravind, L.; et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935. [[CrossRef](#)]
- Ito, S.; Shen, L.; Dai, Q.; Wu, S.C.; Collins, L.B.; Swenberg, J.A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300–1303. [[CrossRef](#)]
- Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C.A.; Hagemeyer, C.; Carell, T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7008–7012. [[CrossRef](#)] [[PubMed](#)]
- He, Y.F.; Li, B.Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011**, *333*, 1303–1307. [[CrossRef](#)] [[PubMed](#)]
- Maiti, A.; Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **2011**, *286*, 35334–35338. [[CrossRef](#)] [[PubMed](#)]
- Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.L.; Luo, C.; Jiang, H.; He, C. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012**, *8*, 328–330. [[CrossRef](#)] [[PubMed](#)]
- Neri, F.; Incarnato, D.; Krepelova, A.; Rapelli, S.; Anselmi, F.; Parlato, C.; Medana, C.; Dal Bello, F.; Oliviero, S. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* **2015**, *10*, 674–683. [[CrossRef](#)] [[PubMed](#)]
- Iurlaro, M.; McInroy, G.R.; Burgess, H.E.; Dean, W.; Raiber, E.A.; Bachman, M.; Beraldi, D.; Balasubramanian, S.; Reik, W. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biol.* **2016**, *17*, 141. [[CrossRef](#)] [[PubMed](#)]
- Licyte, J.; Gibas, P.; Skardziute, K.; Stankevicius, V.; Ruksenaite, A.; Kriukiene, E. A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell Rep.* **2020**, *32*, 108155. [[CrossRef](#)]
- Cadet, J.; Wagner, J.R. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutat. Res. Genet. Toxicol. Environ. Mutagen* **2014**, *764–765*, 18–35. [[CrossRef](#)] [[PubMed](#)]

23. Carell, T.; Kurz, M.Q.; Muller, M.; Rossa, M.; Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angew. Chem.-Int. Ed.* **2018**, *57*, 4296–4312. [[CrossRef](#)] [[PubMed](#)]
24. Branco, M.R.; Ficz, G.; Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **2012**, *13*, 7–13. [[CrossRef](#)] [[PubMed](#)]
25. Kitsera, N.; Allgayer, J.; Parsa, E.; Geier, N.; Rossa, M.; Carell, T.; Khobta, A. Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. *Nucleic Acids Res.* **2017**, *45*, 11033–11042. [[CrossRef](#)]
26. Kitsera, N.; Rodriguez-Alvarez, M.; Emmert, S.; Carell, T.; Khobta, A. Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res.* **2019**, *47*, 8537–8547. [[CrossRef](#)]
27. Luhnsdorf, B.; Epe, B.; Khobta, A. Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.* **2014**, *289*, 22008–22018. [[CrossRef](#)]
28. Allgayer, J.; Kitsera, N.; Bartelt, S.; Epe, B.; Khobta, A. Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Res.* **2016**, *44*, 7267–7280. [[CrossRef](#)]
29. Kribelbauer, J.F.; Lu, X.J.; Rohs, R.; Mann, R.S.; Bussemaker, H.J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *J. Mol. Biol.* **2020**, *432*, 1801–1815. [[CrossRef](#)]
30. Muller, N.; Khobta, A. Regulation of GC box activity by 8-oxoguanine. *Redox Biol.* **2021**, *43*, 101997. [[CrossRef](#)]
31. Schroder, A.S.; Kotljarova, O.; Parsa, E.; Iwan, K.; Raddaoui, N.; Carell, T. Synthesis of (R)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides. *Org. Lett.* **2016**, *18*, 4368–4371. [[CrossRef](#)] [[PubMed](#)]
32. Fleming, A.M.; Ding, Y.; Burrows, C.J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2604–2609. [[CrossRef](#)] [[PubMed](#)]
33. Wilson, D.M., 3rd; Takeshita, M.; Grollman, A.P.; Demple, B. Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *J. Biol. Chem.* **1995**, *270*, 16002–16007. [[CrossRef](#)] [[PubMed](#)]
34. Valinluck, V.; Tsai, H.H.; Rogstad, D.K.; Burdzy, A.; Bird, A.; Sowers, L.C. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **2004**, *32*, 4100–4108. [[CrossRef](#)] [[PubMed](#)]
35. Valinluck, V.; Sowers, L.C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **2007**, *67*, 946–950. [[CrossRef](#)]
36. Hashimoto, H.; Liu, Y.; Upadhyay, A.K.; Chang, Y.; Howerton, S.B.; Vertino, P.M.; Zhang, X.; Cheng, X. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **2012**, *40*, 4841–4849. [[CrossRef](#)]
37. Giehr, P.; Kyriakopoulos, C.; Ficz, G.; Wolf, V.; Walter, J. The Influence of Hydroxylation on Maintaining CpG Methylation Patterns: A Hidden Markov Model Approach. *PLoS Comput. Biol.* **2016**, *12*, e1004905. [[CrossRef](#)]
38. Ji, D.; Lin, K.; Song, J.; Wang, Y. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* **2014**, *10*, 1749–1752. [[CrossRef](#)]
39. Harrington, M.A.; Jones, P.A.; Imagawa, M.; Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2066–2070. [[CrossRef](#)]
40. Holler, M.; Westin, G.; Jiricny, J.; Schaffner, W. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* **1988**, *2*, 1127–1135. [[CrossRef](#)]
41. Khobta, A.; Anderhub, S.; Kitsera, N.; Epe, B. Gene silencing induced by oxidative DNA base damage: Association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res.* **2010**, *38*, 4285–4295. [[CrossRef](#)]
42. Broxson, C.; Hayner, J.N.; Beckett, J.; Bloom, L.B.; Tornaletti, S. Human AP endonuclease inefficiently removes abasic sites within G4 structures compared to duplex DNA. *Nucleic Acids Res.* **2014**, *42*, 7708–7719. [[CrossRef](#)]
43. Burra, S.; Marasco, D.; Malfatti, M.C.; Antoniali, G.; Virgilio, A.; Esposito, V.; Demple, B.; Galeone, A.; Tell, G. Human AP-endonuclease (Ape1) activity on telomeric G4 structures is modulated by acetylable lysine residues in the N-terminal sequence. *DNA Repair* **2019**, *73*, 129–143. [[CrossRef](#)]
44. Fleming, A.M.; Zhu, J.; Howpay Manage, S.A.; Burrows, C.J. Human NEIL3 Gene Expression Regulated by Epigenetic-Like Oxidative DNA Modification. *J. Am. Chem. Soc.* **2019**, *141*, 11036–11049. [[CrossRef](#)]
45. Davletgildeeva, A.T.; Kuznetsova, A.A.; Fedorova, O.S.; Kuznetsov, N.A. Activity of Human Apurinic/Apyrimidinic Endonuclease APE1 toward Damaged DNA and Native RNA with Non-canonical Structures. *Front. Cell Dev. Biol.* **2020**, *8*, 590848. [[CrossRef](#)]
46. Mohni, K.N.; Wessel, S.R.; Zhao, R.; Wojciechowski, A.C.; Luzwick, J.W.; Layden, H.; Eichman, B.F.; Thompson, P.S.; Mehta, K.P.M.; Cortez, D. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **2019**, *176*, 144–153.e13. [[CrossRef](#)]
47. Thompson, P.S.; Amidon, K.M.; Mohni, K.N.; Cortez, D.; Eichman, B.F. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat. Struct. Mol. Biol.* **2019**, *26*, 613–618. [[CrossRef](#)]
48. Schroder, A.S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F.A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2'-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 315–318. [[CrossRef](#)] [[PubMed](#)]
49. Luhnsdorf, B.; Kitsera, N.; Warken, D.; Lingg, T.; Epe, B.; Khobta, A. Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal. Biochem.* **2012**, *425*, 47–53. [[CrossRef](#)] [[PubMed](#)]

- 
50. Bauer, D.E.; Canver, M.C.; Orkin, S.H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J. Vis. Exp.* **2015**, *95*, e52118. [[CrossRef](#)] [[PubMed](#)]
  51. Labun, K.; Montague, T.G.; Gagnon, J.A.; Thyme, S.B.; Valen, E. CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **2016**, *44*, W272–W276. [[CrossRef](#)] [[PubMed](#)]
  52. Allgayer, J.; Kitsera, N.; von der Lippen, C.; Epe, B.; Khobta, A. Modulation of base excision repair of 8-oxoguanine by the nucleotide sequence. *Nucleic Acids Res.* **2013**, *41*, 8559–8571. [[CrossRef](#)] [[PubMed](#)]