Regulation of the CRE and GC box activity by 5-formylcytosine, 5-carboxycytosine and 8-oxoguanine

Dissertation zur Erlangung des Grades

„Doktor der Naturwissenschaften“

im Promotionsfach Pharmazie

am Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften

der Johannes Gutenberg-Universität in Mainz

Nadine Müller

Geboren am 28.02.1992 in Baden-Baden

Mainz, 2021

1. Gutachter:

2. Gutachter:

Tag der mündlichen Prüfung: 30.09.2021

Table of contents

1 Abbreviation List

5-caC	5-carboxycytosine
5-caU	5-carboxyuracile
5-fC	5-formylcytosine
5-fU	5-formyluracile
5-hmC	5-hydroxymethylcytosine
5-hmU	5-hydroxymethyluracil
5-mC	5-methylcytosine
8-oxoG / 8oG	8-oxo-7,8-dihydroguanine
A	Adenine
A230/260/280	Absorption at 230 nm / 260 nm / 280 nm
AID	Activation-induced cytidine deaminase
AP site	Apurinic/apyrimidinic site
APE1	Apurinic/apyrimidinic endonuclease 1
APOBEC	Apolipoprotein B editing complex
ATM	Protein kinase ataxia-telangiectasia mutated
ATR	ATM- and Rad3-Related
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CAS9	Clustered regularly interspaced short palindromic repeats-associated protein 9
cc	Covalently closed
CCL20	C-C motif chemokine 20
CDK	Cyclin-dependent kinase
CGI	CpG Islands
CHD4	Chromodomain helicase DNA binding protein 4
CMV	Human cytomegalovirus
CoREST	Co-repressor for RE-1-silencing transcription factor
CpG	Deoxycytidine-phosphate-deoxyguanosine
CRE	Cyclic adenosine monophosphate-response element
CREB	Cyclic adenosine monophosphate-response element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
dAMP	Deoxyadenosine monophosphate
DDR	DNA Damage Response
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	Deoxyribonucleic acid methyltransferase
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-stranded deoxyribonucleic acid
dTMP	Deoxythymidine monophosphate
DTT	Dithiothreitol
dUMP	Deoxyuridine monophosphate
E.coli	*Escherichia Coli*

EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
Endo	Endonuclease
ERG1	Early growth response protein 1
EthBr	Ethidium bromide
F	Forward
F	Tetrahydrofuran
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
F-caC	2′-fluorinated 5-carboxycytosine
FEN1	Flap endonuclease 1
F-fC	2′-fluorinated 5-formylcytosine
Fpg	Formamidopyrimidine DNA glycosylase
G	Guanine
G4	G-quadruplex
GADD45	Growth arrest and DNA damage 45
HCR	Host cell reactivation
HDAC	Histone deacetylase
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethane sulfonic acid
HMT	Histone methyltransferase
HPLC	High performance liquid chromatography
HPSF	High purity salt free
HSP90	Heat shock protein 90
K	Lysine
kDa	Kilodalton
KRAS	Kirsten rat sarcoma 2 viral oncogene homolog
LB	Lysogeny broth
LIG	Ligase
LSD1	Lysine-specific demethylase 1
MAX	MYC Associated Factor X
MAZ	Myc-associated zinc finger protein
MBD4	Methyl-CpG-binding domain 4 protein
MBP	Methyl-CpG binding proteins
MeCP2	Methyl CpG binding protein 2
mESCs	Mouse embryonic stem cells
MMR	Miss-match repair
MUTYH	MutY homologue DNA glycosylase
NEIL	Nei-endonuclease VIII-like protein
NFkB	Nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
NP-40	Nonylphenol ethoxylate
nt	Nucleotide
NTHL1	Nth-endonuclease III-like protein 1
NTS	Non-transcribed DNA strand
oc	Open circular
OD	Optical density
OGG1	8-oxoguanine DNA glycosylase
PAM	Protospacer adjacent motif

PARP	Poly ADP-ribose polymerases
PAR	Poly ADP-ribose
PBS	Phosphate buffer saline
PCNA	Proliferating-cell-nuclear-antigen
PCR	Polymerase chain reaction
PIPES	1,4-Piperazinediethanesulfonic acid
PMSF	Phenylmethylsulphonyl fluoride
PNK	Polynucleotide kinase
POL	DNA polymerase
PQS	Potential quadruplex-forming sequence
PTH	Parathyroid hormone
R	Reverse
R	Purine-rich DNA strand
REST	RE-1-silencing transcription factor
RNA	Ribonucleic acid
RNAP2	RNA polymerase 2
ROS	Reactive oxygen species
rpm	Rounds per minute
SAH	S-adenosylhomocysteine
SAM	S-adenyl methionine
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SF	Phosphorothioate tetrahydrofuran
sgRNA	Single guide ribonucleic acid
shRNA	Short hairpin ribonucleic acid
SIRT1	Sirtuin-1
SMARCC2	SWI/SNF related, matrix associated, actin-dependent regulator of chromatin subfamily c member 2
SMUG1	Single-strand selective monofunctional uracil DNA glycosylase 1
SOB	Super optimal broth
SP	Specificity protein
SSB	Single-strand break
ssDNA	single-stranded DNA
SUMO	Small ubiquitine-like modifier
T	Thymine
TB	Tris-borate
TBS	Tris-buffer saline
TDG	Thymine DNA glycosylase
TE	Tris-EDTA
TET	Ten eleven translocation methylcytosine dioxygenase
TF	Transcription factor
Tg	Thymine glycol
TNF	Tumour necrosis factor
TRE	12-O-tetradecanoylphorbol 13-acetate-response element
Tris	Tris(hydroxymethyl)-aminomethane
TS	Transcribed DNA strand
TSS	Transcription start site
U	Uracil

U	Units
UHRF1	Ubiquitin like with PHD and ring finger domains 1
UNG1/2	Uracil DNA glycosylase 1/2
URE	Upstream regulatory element
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
WB	Western blot
WT1	Wilms tumour protein 1
XRCC	X-ray repair cross-complementing protein
Y	Pyrimidine-rich DNA strand
Δ / ko	Knockout

2 Summary

The oxidatively induced deoxyribonucleic acid (DNA) modifications 5-formylcytosine (5-fC), 5-carboxycytosine (5-caC) and 8-oxo-7,8-dihydroguanine (8-oxoG) are known to modulate the transcriptional activity when present in gene promoters. However, the knowledge about the functional consequences of the 5-methylcytosine (5-mC) oxidation products 5-fC and 5-caC remains limited, due to their low abundance in the human genome. Functional consequences of the major guanine oxidation product 8-oxoG have been investigated in more detail, but it remains unclear how 8-oxoG can induce both, activation and repression of gene promoters.

The focus of this investigation therefore was set on scrutinising the basic effects of 5-fC, 5-caC and 8-oxoG on the gene expression in human cells, using a plasmid-based reporter gene assay. The strand-exchange method was applied to generate enhanced green fluorescent protein (*EGFP)* reporters carrying a single 5-fC, 5-caC or 8-oxoG residue at the central CpG dinucleotide of the common upstream regulatory elements (UREs), cyclic adenosine monophosphate-response element (CRE) and GC box. Reporters were used for transfection of human cells and the transcriptional consequences of 5-fC, 5-caC and 8-oxoG were analysed by quantification of the cells' EGFP fluorescence. To distinguish functional consequences of the primary modifications from any repair-induced effects, base excision repair (BER) was inhibited using chemically stabilised 2′-fluorinated nucleotides or DNA-glycosylase deficient cells.

Quantitative expression analysis of minimal CRE reporters containing BER-resistant 5-fC or 5-caC, showed that the primary modifications negatively affect the activity of the CRE promoter to similar extends. In a minimal GC box promoter, 5-caC reduced transcription to a much stronger degree than 5-fC, verifying that the effects of 5-fC and 5-caC on promoter activity strongly depend on the affected UREs. Comparing the functional consequences of BER-resistant and BER-sensitive 5-fC and 5-caC on the CRE and GC box activity, it was shown that repair of both base modifications also reduced the gene expression by an independent mechanism. The transcriptional repression by BER of 5-fC and 5-caC intensified over time and was stronger in magnitude than the promoter inhibition by 5-fC and 5-caC per se. Knockout of the thymine DNA glycosylase (TDG) completely reverted transcriptional repression by

BER of 5-fC and 5-caC, demonstrating TDG's significance in the repression process. Intriguingly, BER-proficiency correlated with an initial expression increase of 5-caC containing GC box reporters, indicating that removal of the 5-caC base by TDG reactivated the promoter activity before inducing transcriptional repression. To investigate which repair intermediate eventually induced transcriptional repression, expression analysis was performed on reporters containing a single BER-resistant and BER-sensitive apurinic/apyrimidinic site (AP site) repair-intermediate. Transcriptional repression was only observed for BER-sensitive AP sites, verifying that AP endonuclease 1 (APE1)-mediated strand incision is essential to induce gene silencing. Expression analysis of cells transfected with 8-oxoG containing GC box reporters showed that the guanine oxidation negatively affected the promoter activity in a position dependent manner. Of the four investigated positions, 8-oxoG at the three positions in the purine-rich strand of the GC box consistently inhibited the promoter activity whilst 8-oxoG was transcriptionally neutral in the pyrimidine-rich DNA strand. Additionally, BER of 8-oxoG in the pyrimidine-rich strand and one position in the purine-rich strand of the GC box induced transcriptional repression. BER-dependent transcriptional repression was diminished by knockdown of the 8-oxoguanine DNA glycosylase (OGG1) and was only observed for positions with preferential in vitro cleavage by OGG1, indicating that OGG1 preferences drive the functional outcomes of 8-oxoG in cells. Investigations on BER-resistant and BER-sensitive apurinic sites revealed that APE1-mediated strand incision at the position of 8-oxoG is essential to induce gene silencing. Interestingly, BER of the unrelated base modifications thymine glycol (Tg) and 5-hydroxymthelyurail (5-hmU) also silenced CRE reporters, indicating that gene silencing is a common functional outcome of URE positioned BER substrates.

The presented project reveals a notable transcription regulation complexity by the oxidatively induced base modifications 5-fC, 5-caC and 8-oxoG even in the simplest CRE and GC box promoters. In summary it was shown, that 5-fC, 5-caC and 8-oxoG directly inhibit the CRE and GC box activity in a URE/position dependent manner. Furthermore, BER of 5-fC, 5-caC and 8-oxoG triggers an APE1-dependent gene silencing mechanism, during which 5-caC base removal from the GC box initially activates transcription.

3 Zusammenfassung

Die DNA-Modifikationen 8-Oxo-7,8-dihydroguanin (8-oxoG), 5-Formylcytosin (5-fC) und 5-Carboxycytosin (5-caC) entstehen durch die Oxidation von Guanin (8-oxoG) und 5-Methlycytosin (5-fC/5-caC) und können die Genexpression bereits ab einer Häufigkeit von einer Base pro Promotor stark beeinflussen. Über die generellen Auswirkungen von 5-fC und 5-caC ist allerdings nur sehr wenig bekannt, da die Cytosinmodifikationen im humanen Genom nur sehr selten auftreten. Die transkriptionellen Effekte von 8-oxoG als häufigstes Guanin-Oxidationsprodukt sind zwar umfangreich erforscht, jedoch ist nicht abschließend geklärt wie 8-oxoG die Promotoraktivität sowohl positiv als auch negativ beeinflussen kann.

Der Fokus dieses Projektes lag daher auf der Untersuchung der grundlegenden transkriptionellen Effekte von 5-fC, 5-caC und 8-oxoG in humanen Zellen. Dafür wurden zunächst *enhanced green fluorescent protein (EGFP)*-Reporter hergestellt, welche ein einzelnes 5-fC, 5-caC oder 8-oxoG innerhalb des zentralen CpG-Dinukleotides der häufig auftretenden *upstream regulatory elements (UREs)* cyclic adenosine monophosphate response element (CRE) und GC box enthalten. Humane Zellen wurden mit den Reportern transfiziert und die EGFP-Expression mittels durchflusszytometrischer Fluoreszenzanalyse quantifiziert. Um die primären Effekte der Basenmodifikationen von Reparatur-induzierten Effekten zu unterscheiden, wurden die Basenexzisionsreparatur (BER) inhibiert. Dies geschah einerseits durch die Verwendung chemisch stabilisierter, 2′-fluorinierter Nukleotide und andererseits durch die Verwendung DNA-Glycosylase-defizienter Zellen.

Die Expressionsanalyse minimaler CRE-Reporter ergab, dass BER-resistentes 5-fC und 5-caC die Aktivität des CRE-Promotors gleichermaßen negativ beeinflussen. In einem minimalen GC Box-Promotor reduzierte 5-caC die Transkription deutlich stärker als 5-fC, was zeigt, dass UREs die Auswirkungen von 5-fC und 5-caC stark beeinflussen. Unabhängig davon reduzierte auch die Reparatur von 5-fC und 5-caC die Expression der CRE- und GC Box-Reporter. Dies ließ sich daraus schließen, dass BER-sensitives 5-fC und 5-caC eine zeitabhängige Reduktion der Transkription verursachten, die bei BER-resistenten Basenmodifikationen nicht auftrat. Diese Transkriptionsreduktion war deutlich stärker als die Promotorhemmung durch die primären Basenmodifikationen und wurde durch den Knockout der Thymin-DNA-

Glykosylase (TDG) vollständig verhindert. Nicht nur die Basenexzision durch TDG, sondern auch der darauffolgende Einzelstrangbruch durch die Apurinische/Apyrimidinische Endonuklease 1 (APE1) ist essentiell für die Genstilllegung. Dies zeigt sich dadurch, dass nur BER-sensitive, nicht aber BER-resistente Apurinische/ Apyrimidinische (AP)-Stellen die Transkription reprimierten. Interessanterweise korrelierte der BER-Mangel mit einem anfänglichen Anstieg der Expression der 5-caC enthaltenden GC Box-Reporter, woraus sich schließen lässt, dass die Promotoraktivität durch die TDG-initiierte Basenexzision reaktiviert wurde bevor die Genstilllegung einsetzte.

Die Expressionsanalyse 8-oxoG-haltiger GC Box-Reporter zeigte, dass Guanin-Oxidation die Promotoraktivität in einer positionsabhängigen Weise negativ beeinflusst. Von den vier untersuchten Positionen, hemmte 8-oxoG an drei Positionen im purinreichen Strang der GC Box dauerhaft die Promotoraktivität, während es im pyrimidinreichen Strang transkriptionell neutral war. Die negativen Effekte von 8-oxoG wurden durch die BER-induzierte Genstilllegung deutlich übertroffen. Diese wurde nur von 8-oxoG im pyrimidinreichen Strang und einer Position im purinreichen Stranges der GC Box ausgelöst und wurde durch den Knockdown der 8-Oxoguanin-DNA-Glykosylase (OGG1) signifikant gemindert. Da eine transkriptionelle Repression durch 8-oxoG nur für Positionen mit präferierter in vitro Basenexzision durch OGG1 beobachtet wurde ist anzunehmen, dass OGG1 die Auswirkungen von 8-oxoGs auch in Zellen bestimmt. Die APE1-vermittelte Stranginzision an AP-Stellen war essentiell für die Induktion der Genstilllegung. Interessanterweise verursachte die BER der unabhängigen Basenmodifikationen Thyminglykol (Tg) und 5-Hydroxymethyluracil (5-hmU) ebenfalls eine Stilllegung des minimalen CRE-Promotors, was darauf hindeutet, dass dies ein genereller Effekt von URE positionierten BER-Substraten ist.

Die vorgelegte Arbeit zeigt eine bemerkenswerte Komplexität der Transkriptionsregulation durch die oxidativ induzierten Basenmodifikationen 5-fC, 5-caC und 8-oxoG selbst in einfachsten CRE- und GC-Box-Promotoren. Zusammenfassend konnte gezeigt werden, dass 5-fC, 5-caC und 8-oxoG die CRE und GC box Aktivität direkt in einer URE/positionsabhängigen Weise hemmen. Darüber hinaus löst die BER von 5-fC, 5-caC und 8-oxoG eine APE1-abhängigen Genstilllegung aus, während derer die Entfernung der 5-caC-Base aus der GC Box die Transkription zunächst aktiviert.

4 Introduction

4.1 Origin of modified nucleobases within the human genome

The human DNA is a remarkable molecule, which stores all essential information for life in a minimalistic code with only four variables; the canonical DNA bases adenine (A), cytosine (C), guanine (G) and thymine (T). Together they set up the genetic code. Interestingly, more than 33 modifications of these four bases have been discovered in genomic DNA[1,2], some generated by cellular enzymatic reactions at synthesised DNA, others were generated as DNA damage products. The first discovered DNA modification in the human genome was 5-methylcytosine[3,4], which due to its high genomic frequency and its essential role in transcriptional regulation, has been nicknamed the 5th base of the DNA.

Although abundant base modifications like 5-mC have been thoroughly investigated, the functional outcomes of many other modified nucleobases remain elusive. Main obstacles for a comprehensive investigation of these DNA modifications are their low abundance and dynamic character in genomic DNA, as well as the lack of direct functional readouts. Two of those unexplored DNA modifications are direct oxidation products of 5-mC: 5-formylcytosine and 5-carboxycytosine. Although generation and repair mechanisms of 5-fC and 5-caC have been stablished, their fundamental effects on gene expression remain mostly unknown. The same holds true for the widespread and long-known guanine oxidation product 8-oxoguanine. Its basic transcriptional outcomes remain under debate since both, activating and repressing consequences were described in literature. When it was demonstrated that already a single 5-fC, 5-caC or 8-oxoG residue at a critical gene position significantly affected the transcription of a gene[5–7], the need for understanding the base modifications' consequences for gene expression became apparent. This is even more pressing, since 8-oxoG and 5-mC as precursor of 5-fC and 5-caC are frequently found in cancer cells. The link to human health makes determining the functional consequences of 5-fC, 5-caC and 8-oxoG even more relevant, since it not only advances basic research but also opens up the possibility of base modification usage in disease diagnostic and treatment, for example as biomarkers or drug targets.

4.1.1 DNA damage as a major source of modified nucleobases

A significant proportion of modified nucleobases present in the human genome under physiological conditions is a product of DNA damage induced by exogenous and endogenous DNA damaging agents[1]. These agents change the chemical structure of DNA by randomly generating DNA lesions like DNA breaks, DNA adducts, inter- and intra-strand crosslinks and base modifications. DNA damage mainly occurs at the nucleobases due to their chemically active side chains, with an estimated daily rate of 120,000 incidents per 6.5 Gbp within human liver cells[1]. Nucleobase lesions pose a constant natural risk for genome stability, because they can immediately impair transcription and if left unrepaired can cause permanent genomic mutations[8,9]. Permanent mutations in the DNA pose an even higher risk for the affected organism since they are passed on to subsequent cell generations and are implicated in carcinogenesis, immunological defects and degenerative disorders[10–16]. One of the most frequent forms of base damage is nucleobase oxidation, occurring at all four canonical DNA bases. Pyrimidine (C and T) oxidation results in the generation of hydrate and glycol derivatives like 5-hydroxymethyluracil and thymine glycol[17,18], whilst purine oxidation generates 8-oxoguanine, ring-opened formamidopyrimidine derivatives and 2-hydroxyadenine[19,20].

In the DNA context, guanine has the lowest redox potential with respect to other bases and is therefore the most frequently oxidised base in human DNA[21]. Oxidation of guanine predominantly occurs at the C8 of the imidazole ring of G, resulting in the formation of 8-oxoG[22] as shown in the left panel of Figure 4-1. Main cause of guanine oxidation is G's reaction with reactive oxygen species (ROS)[23] such as the hydroxyl radical (•OH), the hydroperoxide radical (•OOH) and the rare singlet oxygen ($^1O2$)[21,24]. ROS are continuously generated upon exposure to exogenous DNA damaging agents such as industrial chemicals (e.g. asbestos)[25], xenobiotics (e.g. antibiotics)[26] and radiation (e.g. ultraviolet radiation)[27]. To a lesser degree, ROS are also generated by endogenous metabolism such as respiratory chain reactions and antioxidant reactions [26,28–30] and are critical components of cellular signalling pathways[31]. As a result, 8-oxoG lesions are present in the human genome under physiological conditions with a frequency of approximately $1x10^{-6}$ modifications per G residue[32], an amount which is still more elevated under conditions of oxidative stress[22,33].

Random base oxidation by DNA damaging agents also occurs at the abundant DNA modification 5-mC. Thus, exposure of 5-mC to ultraviolet light and ionising radiation

can result in the oxidation of the C5 position of 5-methylcytosine, forming 5-hydroxymethylcytosine (5-hmC) and the higher oxidation form 5-fC respectively[34–38], as shown in the right panel of Figure 4-1.

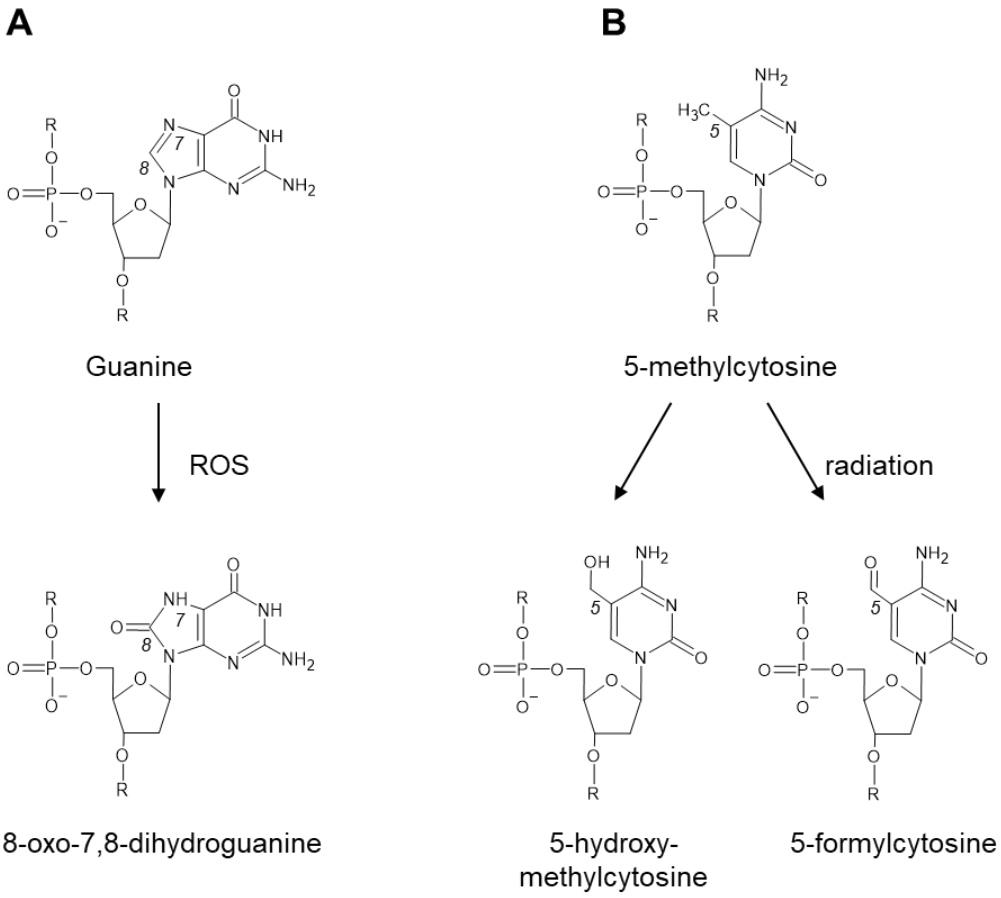


Figure 4-1:Generation of 8-oxoguanine, 5-hydroxymethylcytosine and 5-formylcytosine by DNA damaging agents

Schematic presentation of the generation of the oxidatively induced DNA modifications 8-oxoG, 5-hmC and 5-fC by DNA damaging agents in human cells, showing the nucleotide structures. 8-oxoG results from guanine oxidation at the C8 position by ROS arising from exogenous and endogenous DNA damaging agents (left panel). Exposure of 5-mC to DNA damaging agents such as ionising radiation results in base oxidation at position C5, generating the base modifications 5-hmC and 5-fC (right panel).

4.1.2 Enzymatic induction of base modifications in genomic DNA

In addition to randomly generated base lesions, genomic DNA also comprises altered nucleobases which are generated by specific DNA modifying enzymes in a location selectively manner. In mammals, 5-mC is the best studied enzymatically generated base modification, with an abundance between 1% and 5% of all cytosine residues in genomic DNA[39,40]. 5-mC regulates many cellular functions and is intergenerationally inheritable by daughter cells, classifying it not only as an enzymatically generated modification, but as regulatory, epigenetic mark[41]. Cytosine methylation in the human genome establishes a 5-mC code, which in concert with histone modification,

chromatin remodelling and the action of special non-coding ribonucleic acids (RNAs) defines the epigenetic information layer in human cells, which - in contrast to the genetic information layer - regulates the gene expression without changing the DNA sequence[42].

In human cells, 5-mC is introduced into the DNA by DNA methyltransferases (DNMTs), which transfer a methyl group from S-adenyl methionine (SAM or SAH-CH_3) to the C5 position of cytosine, generating 5-mC and S-adenosylhomocysteine (SAH)[43–47], as depicted in Figure 4-2. DNMT3a and DNMT3b establish new methylation patterns within unmodified genomic sites during *de novo* methylation, which mostly takes place throughout embryonic development[44,47,48]. The 5-mC pattern is maintained over cell divisions by maintenance methylation accomplished by the action of DNMT1[45,46,49]. This process is necessary, since DNA methylation is diluted during semiconservative replication, which uses canonical bases only. The hemi-methylated DNA product is recognised by the Ubiquitin like with PHD and ring finger domains 1 (UHRF1) protein, which recruits DNMT1[50–52]. DNMT1 copies the methylation pattern of the parental DNA strand to the unmethylated daughter strand, thereby maintaining the original 5-mC code[45,46,49]. This sophisticated mechanism of intergenerational epigenetic inheritance is unique for 5-mC as a DNA modification, which makes understanding its generation, function and removal even more important.

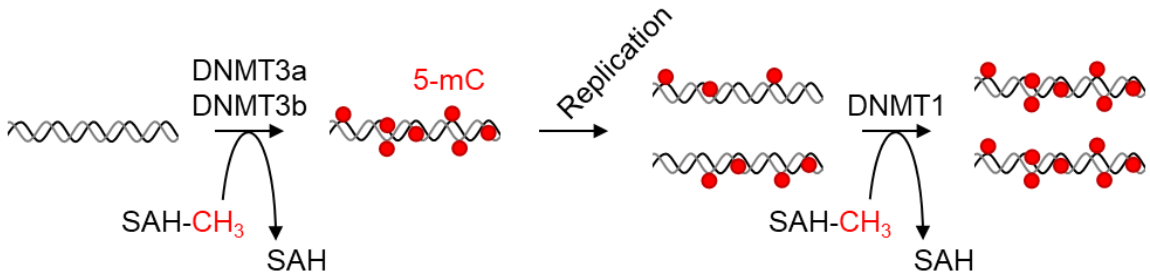


Figure 4-2: Enzymatic induction of 5-methylcytosine by DNA methyltransferases

Schematic presentation of the genomic generation of 5-mC by DNMT-mediated transfer of a methyl group from SAM to the C5 position of cytosine. DNMT3a and DNMT3b establish 5-mC in unmethylated DNA during *de novo* methylation (left), whilst DNMT1 maintains the established methylation pattern throughout replication (right). Semiconservative replication of the methylated DNA results in hemi-methylated DNA, which is targeted by DNMT1 to re-establish the 5-mC code in the unmethylated daughter strand.

DNMTs preferentially target cytosine residues within symmetric CpG dinucleotides, which consequently are methylated with a frequency of approximately 80% in mammalian genomes[53,54]. Exceptions are the rarely methylated CpG islands (CGIs), which are CpG-rich DNA stretches of roughly 1000 base pairs (bp) majorly found at mammalian promoters[55,56]. Whilst CGI methylation is mostly consistent across healthy tissues, the methylation pattern of 15%-21% of the CpGs found outside CGIs varies

between human tissues[57,58]. The dynamically methylated CpGs are commonly found in regions distal to transcription start sites (TSSs), which greatly overlap with enhancer sequences[58], indicating that dynamic promoter methylation at those sites might regulate transcription in a flexible manner.

Various other enzymatically generated DNA modifications originate from 5-mC, including its oxidation products 5-hmC, 5-fC and 5-caC, as shown in Figure 4-3. Ten-eleven-translocases (TETs) iteratively oxidise 5-mC a the C5 position to 5-hmC, 5-fC and 5-caC, while consuming oxygen and 2-oxoglutarate and releasing their reduced forms carbon dioxide and succinate[59–62]. Interestingly, 5-hmC, 5-fC and 5-caC are present at different levels in various tissues, with an uneven distribution pattern throughout the genome[39,59,63–65]. In mouse embryonic stem cells (mESCs), 5-hmC occurs with a frequency of $1.3x10^{-3}$ modifications per C residues[59], with specific enrichment in promoter and enhancer regions[66–69]. In contrast, 5-fC and 5-caC only occur with a frequency of $2x10^{-5}$ and $3x10^{-6}$ modifications per C residues in mESCs respectively[59]. Since TETs are essential for 5-hmC formation in mouse embryonic stem cells and 5-hmC, 5-fC and 5-caC are already present in undamaged genomic DNA[62,70], these data support the hypothesis that all three DNA modifications are rather enzymatically generated DNA modifications than DNA lesions.

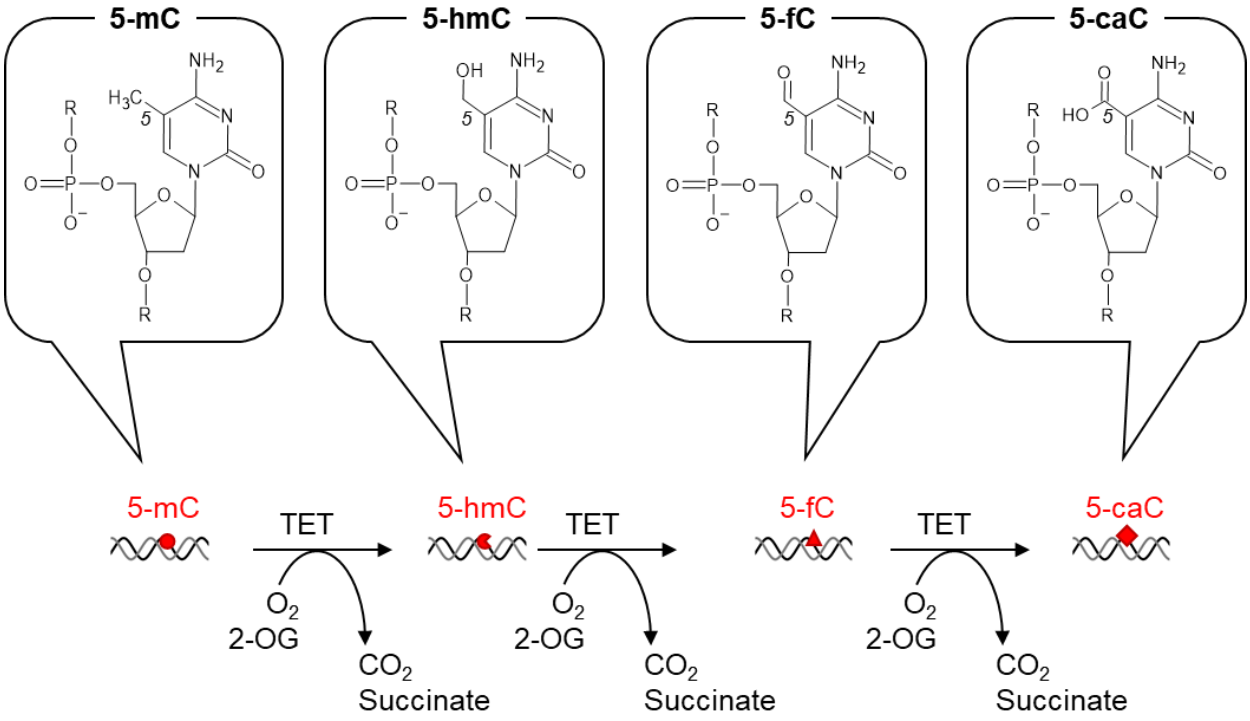


Figure 4-3: Enzymatic induction of 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine by Ten-eleven-translocases

Schematic presentation of the enzymatic generation of the DNA modifications 5-hmC, 5-fC and 5-caC by Ten-eleven-translocases. TETs iteratively oxidise 5-mC to 5-hmC, 5-fC and 5-caC. During this process, TETs consume oxygen and 2-oxoglutarate and produce carbon dioxide and succinate.

Interestingly, the DNA lesion 8-oxoG (discussed under chapter 4.1.1) can also be enzymatically generated in the human genome, as shown for the promoters of oestrogen-responsive genes (Figure 4-4). Oestrogen application activates the lysine-specific demethylase 1 (LSD1), which triggers lysine (K) demethylation at Histone 3-K9me2 at promoter and enhancer sites of oestrogen-responsive genes. ROS are generated as a co-product and locally induce 8-oxoG[71].

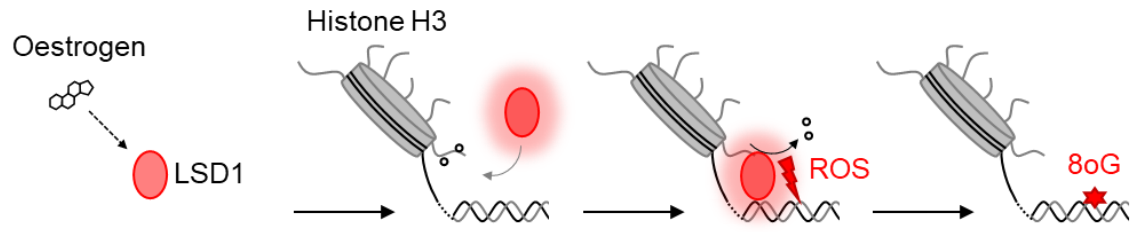


Figure 4-4: Enzymatic induction of 8-oxoguanine by lysine-specific demethylase 1

Schematic presentation of the genomic generation of 8-oxoG by lysine-specific demethylase 1. Oestrogen application activates LSD1, which triggers Histone 3 K9me2 demethylation at oestrogen-responsive gene promoters and enhancer sites (methyl group-white dots). This process generates ROS (red bolt) as a co-product, which locally induce 8-oxoG

4.2 Important functions and harmful consequences of modified nucleobases

4.2.1 Functional consequences of the epigenetic mark 5-methylcytosine and its oxidation products 5-hydroxymethlycytosine, 5-formylcytosine and 5-carboxycytosine

Alterations of genomic nucleobases can profoundly influence cellular functions and survival[1,72,73]. These drastic consequences can originate from the nucleobases' regulatory function in the genome or from its inherent destructive potential as a DNA lesion. One such influential DNA modification is the epigenetic mark 5-mC, whose presence and absence drastically influences human health. 5-mC is an essential transcriptional regulator in mammalian cells, which is implicated in genomic imprinting, X-chromosome-inactivation, transposon silencing and long-term gene silencing[41,56,74–76]. Due to the many important cellular functions of 5-mC, aberrant DNA methylation has severe outcomes for cell physiology and is linked to cancer[77], Alzheimer's disease[78,79] and Parkinson[79]. Although a causality between promoter methylation and gene silencing is broadly accepted[53,54,56,75,80], the mechanism of 5-mC mediated gene silencing is still not fully understood. Transmission of the 5-mC gene silencing signal is attributed to methyl-CpG binding proteins (MBPs), which together with DNMTs, recruit repressive histone modifying enzymes like histone deacetylases (HDACs) and histone methyltransferases (HMTs), as depicted in Figure 4-5. HDACs and HMTs modify histone tails to locally condense the chromatin and thereby abolish transcription[81,82]. For the large group of MBPs several different mechanisms of transcriptional repression have been reported[83]. The methyl-CpG binding protein 2 (MeCP2) for example, directly interacts with HDAC2 to remove activating histone acetylation marks and reduce transcription[84]. Furthermore, MeCP2 can induce promoter silencing by recruitment of the co-repressor for RE-1-silencing transcription factor (CoREST)[85], whilst it also directly remodels histone positioning to locally compact the chromatin[86].

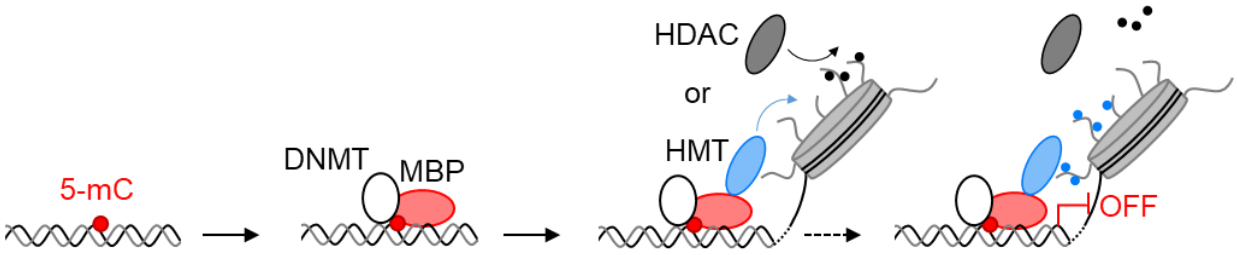


Figure 4-5: Potential mechanisms of epigenetic gene silencing by 5-methylcytosine in genomic DNA

Schematic presentation of the potential epigenetic gene silencing mechanisms by 5-mC. Methyl-CpG binding proteins and DNMTs bind to 5-mC in the DNA and recruit repressive histone modifying enzymes like HDACs and HMTs. HDACs and HMTs modify the histone tails by removing activating acetylation marks (black dots) and inducing repressive methylation marks (blue dots) respectively to locally condense the chromatin and silence the affected genes.

In addition to MBP-mediated chromatin condensation, DNA methylation modulates transcription by positively or negatively influencing the binding of methylation sensitive transcription factors (TFs)[87,88] including members of the essential homeobox protein family and common basic leucine zipper-containing proteins, such as specificity protein 1 (SP1), SP2 and the CRE-binding protein (CREB)[87,89]. TF binding inhibition can drastically modulate the gene expression, as demonstrated for the SP1 regulated expression of the forkhead box F2 transcription factor, the alpha adrenergic receptor, the ZNF132 suppressor and the human secretin receptor, which were drastically reduced upon 5-mC-mediated SP1 binding inhibition[90–93]. Interestingly, there is also published evidence that 5-mC enhances SP1 binding to its target site[87] or leave it unaffected[94].

Due the crucial role of 5-mC as a gene regulator, it is hypothesised that the oxidation and thus loss of 5-mC influences the activity of gene promoters in the human DNA. Indeed, 5-hmC, 5-fC and 5-caC at the position of 5-mC abolish the binding of methyl-CpG binding proteins such as MBP1, methyl-CpG-binding domain protein 4 (MBD4), MeCP2 and Krüppel-like factor 4[35,89]. Abolished MBP recruitment cuts the critical chain of 5-mC induced gene silencing, indicating that 5-mC oxidation alters epigenetic programming. 5-hmC, 5-fC and 5-caC probably abolish MBP binding by steric hindrance due to their bigger and electrostatically charged side groups. By analogy it is assumed, that the hydroxymethyl-, formyl and carboxy-moieties also prevent the binding of other proteins such as TFs. Little is known about the consequences of 5-hmC, 5-fC and 5-caC on TF binding, which has only been investigated for a hand full of proteins. It was shown, that all three 5-mC oxidation products influenced the binding of the two transcription factors early growth response protein 1 (ERG1) and Wilms

tumour protein 1 (WT1) to the DNA. The high affinity of ERG1 and WT1 for their methylated consensus sequences was highly reduced when either 5-hmC or 5-fC were present in the DNA[95,96]. Interestingly, 5-caC only reduced the binding of ERG1 but not of WT1[95,96], which was ascribed to different electrostatic interactions between 5-caC and charged protein groups. Electrostatic repulsion between ERG1's glutamate 354 and the negatively-charged carboxy group of 5-caC disfavours the DNA-protein binding, whereas WT1's corresponding glutamine interacts favourably with the carboxylate[95,96]. Analogous residues to ERG1's glutamate 354 are present in MBP proteins like MeCP2 and KAISO, as well as several zinc finger proteins including the Krüppel-like factor 4 transcription factor, indicating that genomic 5-caC may negatively affects the binding of these proteins under physiological conditions. Another protein whose DNA-binding ability is influenced by 5-hmC, 5-fC and 5-caC is the common TF CREB. As previously shown in our lab, single 5-mC, 5-hmC, 5-fC and 5-caC diminished CREB binding to its target CRE gene regulatory element[5,89]. Interestingly, the presence of any of the four DNA modifications also reduced the CRE promoter activity in a plasmid-based reporter gene assay[5]. The promoter inhibition correlated with the amount of 5-mC, 5-hmC, 5-fC and 5-caC residues within the CRE sequence, indicating that the cytosine modifications directly reduce transcription assumedly by inhibiting TF binding[97]. Furthermore, these results verify, that already a single 5-hmC, 5-fC and 5-caC residue in a gene promoter can influence transcription.

To be classified as true regulatory marks like their epigenetic precursor, 5-hmC, 5-fC and 5-caC not only need to be deliberately introduced into the DNA by specific writer proteins, but also need to possess modification selective reader proteins, which transduce their regulatory function within human cells. Although no proof for any epigenetic inheritance exists for any of the three cytosine modifications so far, lots of evidence for a distinct regulatory function of 5-hmC has been emerging recently. Thus, 5-hmC not only seems to be rather stable in the DNA[98,99], but also recruits specific reader proteins such as MeCP2 and UHRF2[89,100] and is linked to promoter activation[100]. There is less evidence for a regulatory function of the oxidatively induced nucleobase modifications 5-fC and 5-caC, although some modification selective reader proteins have been identified. 5-fC is at least partially stable in genomic DNA[64,101] and was shown to be specifically bound by DNMTs and tumour protein p53[89,102]. In mouse embryonic stem cells, 5-fC also directly interacts with histones by Schiff base interactions, thereby modulating the nucleosome organisation and ultimately

regulating gene expression[103]. Comparable to 5-fC, 5-caC also recruits specific proteins in a modification-dependent manner like the MYC Associated Factor X (MAX). MAX arginine 36 directly recognises 5-caC by electrostatic attraction between the protein's basic environment and the carboxy group of 5-caC. The DNA-protein binding is abolished by 5-mC, 5-hmC and 5-fC, indicating that MAX is a specific 5-caC reader[104]. Other proteins preferentially binding to 5-caC containing DNA are DNMT1 and SMARCC2 (Swi/Snf chromatin-remodelling complex related, matrix associated, actin-dependent regulator of chromatin subfamily c member 2)[89,104]. In general, the readers of the biochemically different base modifications 5-hmC, 5-fC and 5-caC only show limited overlap[89], furthering the notion that all three DNA modifications may be autonomous regulatory marks in the human genome.

In addition to direct effects of 5-fC and 5-caC on DNA-protein binding, both base modifications modify the gene expression by a potentially repair-dependent mechanism as demonstrated by Julia Allgayer in our lab. Thus, 5-fC and 5-caC in the CRE-gene regulatory element reduced promoter activity to a much stronger degree than 5-mC and 5-hmC in HeLa cells[97]. Repression of the gene expression by 5-fC and 5-caC intensified with time, indicating that it was not caused by TF binding inhibition but by a transcription regulation mechanism which only was established over the course of time. Interestingly, reduction of the gene expression by 5-fC and 5-caC was significantly less pronounced in cells with stable knockdown of the thymine DNA glycosylase[97]. Since TDG initiates base excision repair of 5-fC and 5-caC in human cells, it is tempting to suggest a causal connection between the repair of 5-fC and 5-caC and their negative effect on transcription of the affected gene. Although transcriptional repression by 5-fC and 5-caC has only been described in one gene regulatory element so far, the consequences are of special interest, because CREB binding contributes to the activation of 25% of all eukaryotic promoters[105,106]. Since a significant portion of these promoters drives the expression of essential pro-neuron survival proteins[107,108], it is not surprising that CREB dysregulation is associated with central nervous system disorders such as Chorea Huntington[108,109].

The presence of specific readers for 5-fC and 5-caC in combination with their targeted enzymatic generation and drastic effects on CRE driven gene expression support the hypothesis that both base modifications possess a regulatory function in human cells, a hypothesis that has already found great support for 5-hmC. However, since extensive

in vivo investigations on the biological consequences of 5-fC and 5-caC are lacking, it remains elusive whether the progressive transcriptional repression of CRE represents a rare URE-specific phenomenon or a general consequence of 5-fC and 5-caC in DNA.

4.2.2 Functional consequences of the major guanine oxidation product 8-oxoguanine

The major guanine oxidation product 8-oxoG is historically regarded as a DNA lesion, generated by endogenous and exogenous DNA damaging agents as described in chapter 4.1.1. Genomic 8-oxoG poses a constant natural risk for genome stability, since it can compromise the integrity of regulatory elements and impede error-free DNA and RNA synthesis. If 8-oxoG is present during transcription, RNA polymerase 2 (RNAP2) can incorporate an incorrect A opposite to 8-oxoG, thereby generating mutant transcripts. Cause of 8-oxoG's mutagenic potential is its ability to form non-canonical (Hoogsteen-) base pairs with A instead of C (Watson-Crick pairing), which not only occurs during transcription but also during replication[110]. The presence of 8-oxoG during replication poses a great risk for the cell, as DNA-polymerases can introduce A into the daughter strand in a first round of replication. In a second round of replication this mutation is fixed in the genome inducing permanent G:C → T:A transversion mutations[14,110–115], which are suspected to promote cancer development[115–117]. Cells counteract the deleterious effects of persisting 8-oxoG by efficient base excision repair (described in chapter 4.3.1)[118,119], which rapidly removes the modified nucleobase from the genome and restores the canonical DNA sequence[119–121].

Despite highly efficient BER, significant amounts of 8-oxoG are constantly present in the human genome[32], with an unexpected enrichment in specific regions[122] including gene promoters[123,124]. Interestingly, 8-oxoG contents in gene promoters of oxidatively stressed mice correlate with the transcriptional activity and number of active genes[125]. The positive correlation between 8-oxoG occurrence and transcription activity was also observed in rats, where the levels of hypoxia-induced 8-oxoG in promoters of hypoxia-inducible genes correlated with an increase in gene expression[126], indicating that guanine oxidation has the potential to regulate transcription.

Regulation of the gene expression by 8-oxoG can for example be accomplished by modulating TF-binding [127–129], as previously demonstrated for CREB and SP1. It was

demonstrated that a single 8-oxoG residue sufficed to decrease the binding of CREB and SP1 to their target gene regulatory elements, the CRE and GC box respectively[127,128]. It should be noted, that both TFs modulate the transcription of essential human genes. Whilst 25% of all eukaryotic promoters contain important activating CRE elements[105,106], the GC box gene regulatory element has been shown to control the housekeeping gene dihydrofolate reductase[130,131] and several tumour suppressors like the cyclin-dependent kinase 5 (cdk5) activator p35[132], the transcription factor forkhead box F2[90] and the protooncogene c-Ha-ras1[133,134]. Unsurprisingly, disruption of GC box driven promoter regulation is associated with cancer progression[92], highlighting the importance of the GC box integrity for general promoter regulation. Not only the mere presence but also the position of 8-oxoG within a TF binding site influences protein binding. 8-oxoG at the central positions of the GC box sequence completely inhibits SP1 binding, whereas 8-oxoG at the outer positions of the regulatory element only partially reduces SP1 binding[127]. 8-oxoG within the target sequence of nuclear factor 'kappa-light-chain-enhancer' of activated B-cells (NFkB) can even repress and activate the binding of NFkB, depending on the position of guanine oxidation within the gene regulatory element[129].

A more drastic transcriptional effect was reported for the repair of 8-oxoG by OGG1-initiated base excision repair, which can immensely enhance and completely abolish the gene expression (summarised in Figure 4-6). It was for example shown that, BER of 8-oxoG in a promoter positioned CREB site, as well as 5'-untranslated region (UTR), 3'-UTR and gene body of *EGFP*[6,97,135,136] drastically diminished the gene expression by an unknown silencing mechanism. Removal of the 8-oxoG base by OGG1 and following strand incision by APE1 were essential to induce the transcriptional repression (Figure 4-6 A). BER-mediated gene silencing was also observed for 8-oxoG in promoters of human cytomegalovirus, the housekeeping gene beta actin and the tumour suppressor Ras association domain-containing protein 1[6], verifying that BER of 8-oxoG can regulate human promoters.

In contrast, OGG1 initiated BER of 8-oxoG in various other promoters was reported to enhance transcription of the affected gene in human cells[7,137,138]. For example, unproductive binding of OGG1 to 8-oxoG in NF-kB motives of the promoter of tumour necrosis factor (*TNF)* and C-C motif chemokine 20 *(CCL20)* genes enhances NF-kB binding to its target site and thereby activates transcription (Figure 4-6 B)[139].

8-oxoG in the promoter of DNA glycosylase Nth-endonuclease III-like protein 1 (NTHL1), histone deacetylase sirtuin-1 (SIRT1), protooncogene Kirsten rat sarcoma 2 viral oncogene homolog (KRAS) and Vascular Endothelial Growth Factor (VEGF) activates the gene expression by BER induced changes of the secondary DNA structure. More specifically, 8-oxoG repair within the negative calcium responsive elements of SIRT1 results in the unproductive binding of APE1 to the AP site intermediate, which facilitates the formation of a short hairpin (Figure 4-6 C)[137,138]. The AP site is presented by the DNA-hairpin structure and recruits X-ray repair cross-complementing proteins 5 and 6 (XRCC5/6) and RNAP2, eventually activating transcription.

In VEGF and NTHL1 promoters, incomplete BER of 8-oxoG at different positions within the GC-rich, GC box containing promoters indirectly enhances the gene expression (Figure 4-6 D)[7,140]. More specifically, OGG1 removes 8-oxoG from the DNA, generating an apurinic site intermediate. Unproductive APE1 binding to the AP site facilitates the formation of a complex G-quadruplex (G4)-fold from the initially unstructured potential quadruplex-forming sequence (PQS), thereby activating the gene expression[7,126,140,141]. Remarkably, G4-formation is essential to activate the VEGF promoter by 8-oxoG repair since transcriptional enhancement was diminished in G4-negative promoters, unable to fold into G4[7,140]. BER-dependent VEGF promoter activation is only observed for a PQS positioned in the non-transcribed strand (NTS), and is reverted when the PQS sequence is present in the transcribed strand (TS), where 8-oxoG induced transcriptional repression in a BER independent manner[140]. These results indicate, that not only the secondary structure formation, but also the DNA strand selection or orientation of the PQS element within the promoter can influence the regulatory function of 8-oxoG in G-quadruplex structures.

The KRAS promoter also contains a PQS but in contrast to VEGF and NTHL1, the genomic KRAS-PQS folds into a G4-structure in the absence of damage[138]. ROS-induced 8-oxoG in the KRAS promoter enhances the recruitment of the two transcription factors heterogeneous nuclear ribonucleoprotein A1 and MAZ (Myc-associated zinc finger protein), and destabilises the G4 structure, inducing the formation of a DNA-duplex (Figure 4-6 E)[138]. 8-oxoG is erased from the duplex DNA by OGG1 and BER, which restores the canonical DNA sequence to enable MAZ activated transcription.

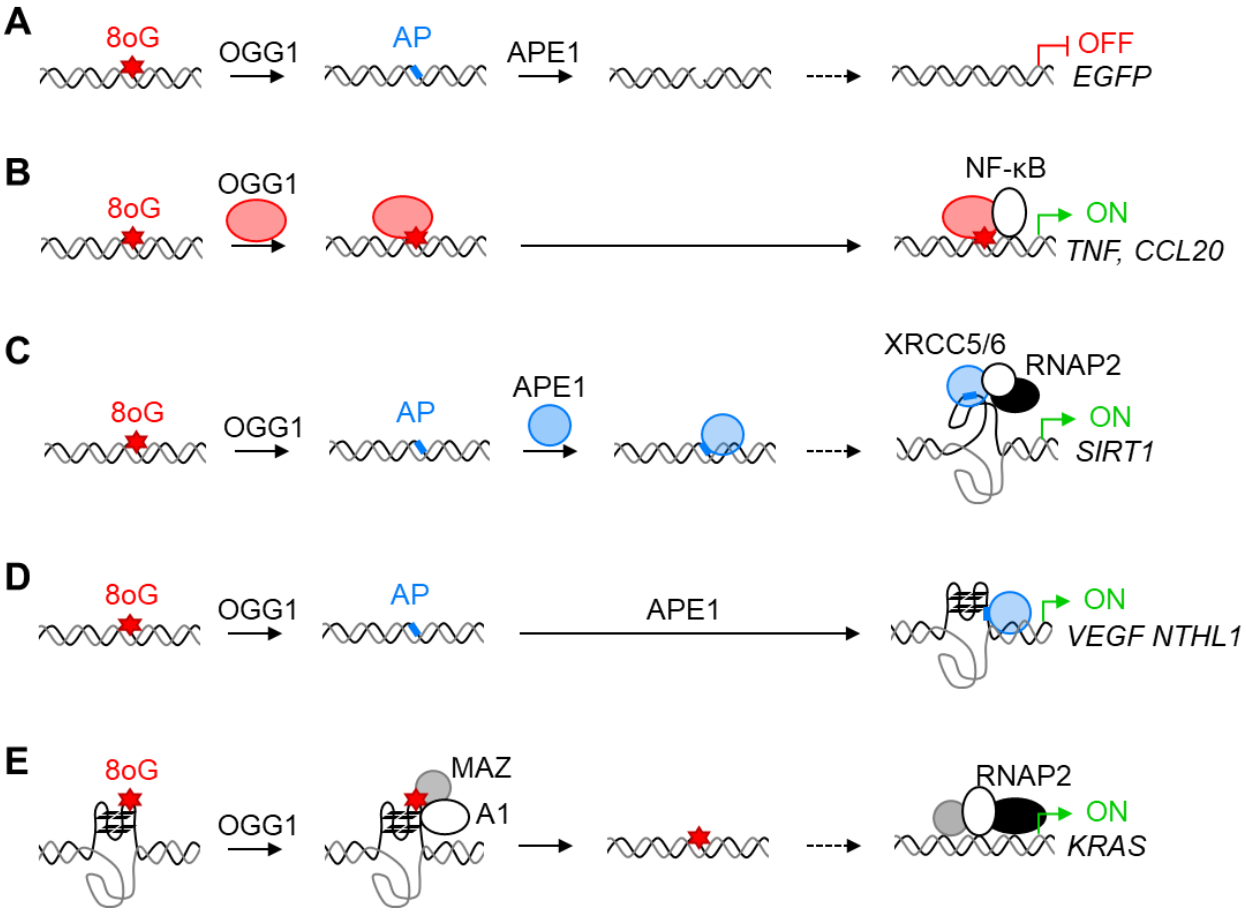


Figure 4-6: Functional consequences of 8-oxoguanine in promoter sequences for gene expression

Schematic representation of five mechanistically and functionally different consequences of 8-oxoG in human cells. (A) Base excision repair of 8-oxoG in the *EGFP* gene via OGG1-mediated base removal and APE1-mediated strand incision reduces the gene expression. (B) 8-oxoG within the *TNF* and *CCL20* gene promoters recruits OGG1. Unproductive binding of OGG1 enhances NF-kB binding and thereby activates transcription. (C) Base removal of 8-oxoG from the *SIRT1* promoter by OGG1 results in the unproductive binding of APE1 to the AP site intermediate. APE1 binding facilitates the formation of a short hairpin and recruits XRCC5/6 and RNAP2, thereby activating transcription. (D) Base removal of 8-oxoG from the *VEGF* and *NTHL1* promoters by OGG1 generates an AP site. Unproductive binding of APE1 to the AP site intermediate favours the formation of a G-quadruplex structure, activating the gene expression. (E) 8-oxoG formation within the *KRAS* promoter enhances the recruitment of the transcription factors A1 and MAZ, thereby destabilising the native G-quadruplex structure. OGG1 removes 8-oxoG from the duplex DNA and further BER processing enables MAZ activated transcription by RNAP2.

Although 8-oxoG has a specific distribution pattern in the human genome and can be generated by targeted enzymatic reactions, it remains unclear whether the DNA modification is a genuine regulatory mark with a specific regulatory function, due to the varying transcriptional consequences of 8-oxoG within different gene promoters (repressing in CRE versus activating in PQS).

The controversial effects of 8-oxoG repair on the gene expression can be reasoned by the different positions of 8-oxoG within the promoters, the different GC-content of the sequences, the promoters' varying ability to form non-canonical secondary structures and varying TF binding to the UREs within different promoters. To determine the basic functional consequences of 8-oxoG, it is therefore necessary to assess the

consequences of this base modification in a simple GC-rich promoter without a sophisticated secondary structure, thereby ensuring a minimal amount of transcription regulation complexity.

4.3 Removal of modified nucleobases from the human genome by base excision repair

Base excision repair was first described by Tomas Lindahl in 1974[142]. It removes a variety of small non-bulky nucleobase lesions from the human DNA[119,120,143,144], including 5-fC, 5-caC and 8-oxoG. BER is an evolutionary conserved DNA repair pathway, which plays a vital role in maintaining genome stability by preventing the accumulation of harmful DNA damage arising from oxidation, deamination and various other chemical reactions[9,120,143]. The BER process, which is depicted in Figure 4-7, can be divided into four steps, which enable lesion specific repair of the DNA damage. The first step is recognition and removal of the damaged base. The second step is strand incision at the generated abasic site, followed by replacement of the remaining sugar fragment for an intact nucleotide in the third step and nick sealing in the last step of BER.

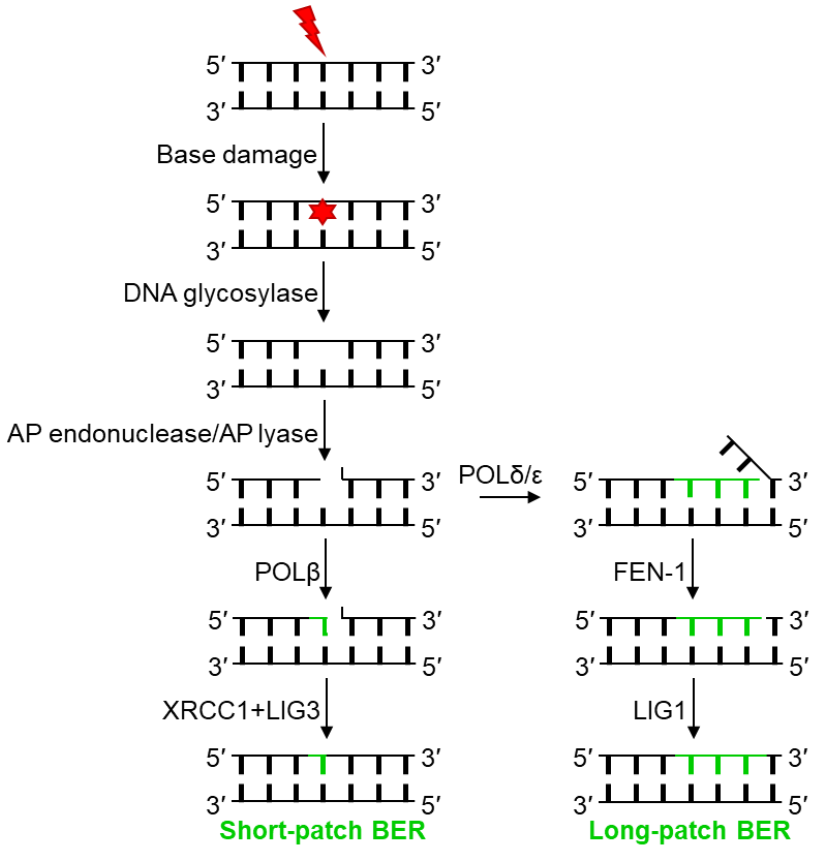


Figure 4-7: Overview of the base excision repair pathway in human cells

Schematic presentation of the short-patch (left panel) and long-patch base excision repair (right panel) of small, non-bulky DNA base modifications in human cells adapted from[120] showing the major participating enzymes and structural repair intermediates. DNA damage (red bolt) induced altered nucleobases (red star) are removed from the DNA by a DNA glycosylase creating an abasic site. An AP-lyase or AP-endonuclease incises the DNA strand at the AP site generating a 1 nt gap. If the remaining DNA ends possess a 3′-hydroxyl group and a 5′-dRp, BER progresses via short-patch BER (left panel), during which POLβ removes the 5′-dRp and fills the gap with a correct nucleotide. The remaining nick is sealed by a complex of XRCC1 and LIG3. If strand incision at the generated AP site results in non-conventional 5′-ends, long-patch BER processes the lesion (right panel). POLδ or POLε displace 2-12 deoxynucleotides of the original DNA sequence, including the non-conventional 5′-end for a newly synthesised DNA. FEN1 cleaves the 5′-overhang and DNA LIG1 seals the nick.

In the initial step of BER, the damaged base is recognised and removed from the DNA by a DNA glycosylase in a lesion-specific manner. The human genome encodes eleven DNA glycosylases, each of which recognises a different subset of damaged bases with partially overlapping substrate specificity[121,145,146] as set forth in Table 4-1. During base removal, the DNA glycosylase cleaves the N-glycosidic bond linking the damaged base and the 2-deoxyribose, generating an AP site. In the second step of BER, the AP site is targeted by an AP-lyase or AP-endonuclease, which incises the DNA 5′ from the AP site, generating a single-strand break (SSB). Depending on the inherent nature of the involved DNA glycosylase, strand incision is either performed by the DNA glycosylase itself (in case of bifunctional DNA glycosylases) or by additional repair enzymes (in case of monofunctional DNA glycosylases). In the latter case, APE1 hydrolyses the phosphate bond at position C5 of the generated nucleoside, inducing a

SSB with a priming 3′-hydroxyl group (-OH) and a 5′-dRP group[147,148]. In case of bifunctional DNA glycosylases, the intrinsic AP-lyase activity promotes DNA strand incision at the generated AP site resulting in a SSB. In the third step of BER, DNA polymerase β (POLβ) removes the residual 5′-dRP from the incised DNA and inserts an intact nucleotide into the gap, using the undamaged complementary DNA strand as a template[149–151]. The remaining nick is sealed by a complex of XRCC1 and LIG3a, thus completing the BER process and re-establishing the canonical DNA sequence[152]. If blocked 5′-ends are generated in the second step of BER instead of a 5′-dRP, BER switches from the classical short-patch BER described above (Figure 4-7, left panel) to the alternative long-patch BER pathway (Figure 4-7,right panel). During long-patch BER, the 3′-OH group acts as a starting point for DNA synthesis by POLδ/ε. The DNA polymerases displace 2-12 deoxynucleotides of the original DNA sequence, including the non-conventional 5′-end, for a newly synthesised DNA strand[153,154]. The generated 5′-flap of dispensed native ssDNA is removed from the DNA by flap endonuclease 1 (FEN1) and the remaining nick is sealed by LIG1[155,156].

Table 4-1: Human DNA glycosylases and their DNA substrates

DNA Glycosylase		Function	Major substrates	Ref.
MBD4	Methyl-binding domain DNA glycosylase4	Mono	U:G and T:G and 5-hmU in CpG islands in dsDNA	157–159
MPG	3-methyl-purine glycosylase	Mono	3-mA and 1-mG and 7-mG and ethenoA and hypoxanthine in ss- and dsDNA	160
MUTYH	MutY homolog DNA glycosylase	Mono	A opposite to 8-oxoG/C/G in dsDNA	161
NEIL1	Endonuclease VIII-like glycosylase 1	Bi	Oxidized pyrimidines (Tg, 5-hC, 5,6-dihydrouracil, Gh, Sp) and FapyG and FapyA in ss- and dsDNA	162
NEIL2	Endonuclease VIII-like glycosylase 2	Bi	Similar to NEIL1 in bubbles and loops	163
NEIL3	Endonuclease VIII-like glycosylase 3	Bi	Similar to NEIL1 and FapyG, FapyA, Sp and Gh in ssDNA	164
NTHL1	Endonuclease III-like1	Bi	Oxidised pyrimidines (Tg, 5-hC, 5-hU) and FapyG and FapyA in dsDNA	165,166
OGG1	8-oxoG DNA glycosylaseOGG1	Mono	Oxidised purines (8-oxoG:C and FapyG:C) in dsDNA	167
SMUG1	Single strand specific monofunctional uracil DNA glycosylase1	Mono	ssU, U:G, U:A, 5-hmU in ss- and dsDNA	168
TDG	Thymine DNA glycosylase	Mono	U:G, T:G, oxidised/deaminated 5-mC:G in dsDNA	169,170
UNG	Uracil-N glycosylase	Mono	U in ss- and dsDNA (UNG1-mitochondrial, UNG2-nuclear)	168,171

Abbreviations: 1-mG, 1-methylguanine; 3-mA, 3-methyladenine; 5-hC, 5-hydroxycytosine; 5-hU, 5-hydroxyuracil; 5-hmU, 5-hydroxymethyluracil; 5-mC, 5-methylcytosine; 7-mG, 7-methylguanine; 8-oxoG, 8-oxoguanine; A, Adenine; Bi: Bifunctional DNA glycosylase; C, Cytosine; FapyG, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; FapyA, 4,6-diamino-5-formamidopyrimidine; G, Guanine; Gh, 5-guanidinohydantoin; Mono, Monofunctional DNA glycosylase; T, Thymine; Tg, Thymine glycol; U, Uracil; Sp, Spiroiminodihydantoin;

4.3.1 Human base excision repair of 8-oxoG

Among other DNA modifications, BER also removes the oxidatively induced base modification 8-oxoG from the human genome as depicted in Figure 4-8[119]. Depending on the base pairing partner of 8-oxoG (C or A), adenine is replaced for C in miss-paired 8:oxoG:A pairs by MutY homologue DNA glycosylase (MUTYH)-initiated BER resulting in an 8-oxoG:C A)[118,161]. pair, from which 8-oxoG can be directly removed by OGG1-initiated BER[9,167,172]. During both BER processes, the responsible DNA glycosylases (OGG1 and MUTYH) act as monofunctional DNA-glycosylases in vivo[173–175]. If adenine has erroneously been inserted opposite to 8-oxoG during replication, MUTYH has to remove the miss-incorporated A from the non-canonical 8-oxoG:A base pair prior to

replication to prevent a potential fixation of the mutation in the DNA in a second replication round. After base removal, the generated AP site is further processed by APE1, inducing a SSB. POLλ in cooperation with proliferating-cell-nuclear-antigen (PCNA) and replication protein A fill the 1 nt gap and add one more nucleotide, while displacing the original DNA strand. FEN1 cleaves the 1 nt overhang and DNA LIG1 seals the remaining nick, generating an 8-oxoG:C base pair. The 8-oxoG:C pair is recognised by OGG1, which removes the oxidised guanine base from the DNA[176]. POLβ in combination with XRCC1 and LIG3 complete the BER process and re-establish the canonical G:C base pair.

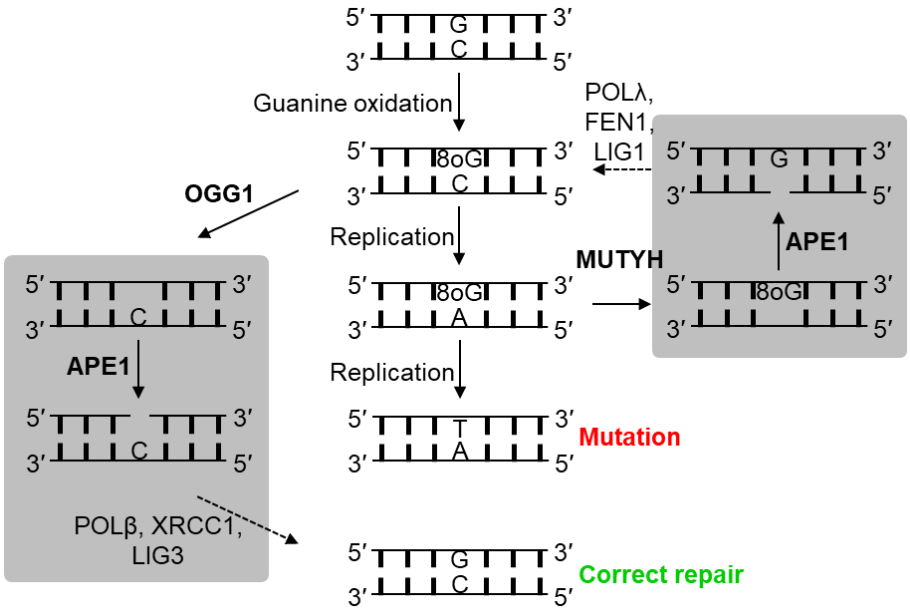


Figure 4-8: Repair of 8-oxoguanine by combined OGG1- and MUTYH-initiated BER in human cells

Schematic presentation of BER of 8-oxoG in human cells, showing the major participating enzymes, structural repair intermediates and functional outcomes of repair (adapted from[177]). 8-oxoG paired with C is removed from the DNA by the monofunctional action of OGG1, followed by classical BER which restores the correct DNA sequence. If replicative DNA polymerase incorporate A opposite of 8-oxoG during replication, the non-canonical base pair is recognised by MUTYH, removing A from the DNA. The generated AP site is further processed by APE1 and following BER steps. If the non-canonical 8-oxoG:A base pair persists till S-phase, DNA polymerases can introduce T opposite to A in a second replication round, fixing the mutation in the DNA.

Interestingly, the 8-oxoG repair capacity of a cell can be rapidly modified, which is achieved most effectively by regulating the enzymatic activities of OGG1, MUTYH and APE1. OGG1 activity can be regulated by post-translational protein modifications, including but not limited to acetylation, phosphorylation and oxidation. OGG1 undergoes acetylation by p300, which significantly reduces the activity of OGG1 towards the AP site product, thereby enhancing its turnover rate in vitro and in vivo[178,179]. OGG1 also interacts with HDACs, which are potentially responsible for its

deacetylation, resulting in a fraction of about 20% acetylated OGG1 in HeLa cells[178]. Phosphorylation of OGG1 by for example CDK4 can occur at serine and threonine residues, increasing OGG1's DNA glycosylase activity. In contrast, tyrosine phosphorylation of OGG1 by abelson murine leukaemia viral oncogene homolog 1 and serine/threonine phosphorylation by protein kinase C does not influence the repair capacity of OGG1[180,181]. The distinct functional outcomes of OGG1 phosphorylation by different protein kinases, suggests that OGG1's activity can be regulated based on the cells situation-dependent needs, by using different signal transduction pathways. Attachment of O-linked N-acetylglucosamine moieties, poly ADP-ribosyl moieties and protein oxidation have also been demonstrated for OGG1 and are associated with a decreased repair activity[182–184]. The most important protein interaction partners of OGG1 which modulate its repair activity are APE1 and XRCC1. APE1 interaction with OGG1 reduces the DNA glycosylases' affinity to the AP site product, enhancing the turnover rates[174,185]. XRCC1 interaction with OGG1 stimulates the formation of the OGG1-DNA-Schiff-base intermediate, thereby passing on the DNA intermediate from OGG1 to the endonuclease APE1 and accelerating the overall repair process[186,187]. Additionally, cut homeobox-1 and -2 proteins interact with OGG1 and stimulate its DNA glycosylase and AP lyase activities[188,189].

Similar to OGG1, APE1 is regulated by a multitude of post-translational modifications and protein interaction partners. Many post-translational protein modifications of APE1 have been controversially discussed in the field, because they were shown to differently affect APE1's repair activity, as shown for ubiquitination. On the one hand, APE1 ubiquitination drives proteasomal degradation of APE1, which downregulates BER[190]. On the other hand ubiquitination can enhance APE1's stability and its affinity to DNA, thereby enhancing BER[191]. Phosphorylation of APE1 by casein kinase 2 and CDK5 was shown to inhibit APE1's endonuclease activity in some studies[192,193], whereas others did not detect any effects of APE1 phosphorylation on the proteins' repair capacity[194–196]. The same holds true for acetylation of APE1, which can enhance the endonuclease activity[197] or only affect APE1's non-repair functions[196,198]. S-nitrosylation of APE1 by nitrosative stress is known to trigger the protein's exclusion from the nucleus[199], hence disabling it from participating in DNA repair processes. Protein-protein interactions also regulate APE1's activity[200], as shown for the DNA repair protein XRCC1. XRCC1 directly interacts with APE1 and stimulates its endonuclease activity[201]. LIG1 interaction with APE1 can both stimulate and inhibit

APE1's repair activity, depending on the type of moiety presented at the 5′ termini of the DNA break[202]. Protein kinase C and the growth arrest and DNA damage (GADD45) protein also directly modulate APE1's repair activity[195], whereas SIRT1 promotes its interaction with other BER proteins[203].

Modulation of 8-oxoG's repair by such an extensive regulation network could explain how the base modification and its BER can induce such great variety of transcriptional consequences ranging from promoter activation to total inhibition of the gene expression repression, marking BER an essential regulator of 8-oxoG outcomes.

4.3.2 Removal of 5-mC, 5-hmC, 5-fC and 5-caC from the human genome

BER is also involved in the removal of 5-mC from the human genome, however indirectly. In general, 5-mC is removed from the DNA by DNA demethylation processes which can occur either globally or locally. Global DNA demethylation is an essential driver of cell-differentiation and is observed during fertilisation and early embryonic development[204,205]. Locus-specific DNA demethylation at gene promoters and enhancers enables somatic cells to adapt to environmental stimuli and is important throughout all human life[206,207]. Global and local 5-mC removal in mammalian cells are accomplished by a combination of active and passive DNA demethylation processes[205,207], which return methylated DNA to its unmethylated state.

Passive DNA demethylation is accomplished by inhibiting maintenance and *de novo* DNA methylation during and after replication cycles[74,208–211], resulting in the dilution and ultimately elimination of 5-mC residues in the affected DNA sequence. In contrast, active DNA demethylation is based on the enzymatic removal of 5-mC from the DNA. Enzymes potentially initiating active DNA demethylation in human cells have been suggested plentifully and include but are not limited to activation-induced cytidine deaminase (AID)[212–214], apolipoprotein B editing complex (APOBEC)[212], MBD4[215,216], GADD45[206,217–219] and DNMTs [102,220,221], however experimental proof is fragmented and not always reproducible[222,223]. The active DNA demethylation pathway that is supported by most nowadays is TET-TDG-mediated DNA demethylation, as depicted in the left panel of Figure 4-9. The ancient mechanism of TET-TDG-mediated DNA demethylation is vital to mammalian development, as it can swiftly and site specifically remove the 5-mC silencing mark from the DNA[72,224,225]. During this process, 5-mC is iteratively oxidised by TETs to 5-hmC, 5-fC and 5-caC[59,62,226,227]. Whilst no direct repair mechanism for 5-hmC has been discovered, the modification can be further oxidised

to 5-fC and 5-caC, which are removed from the DNA by BER. BER of 5-fC and 5-caC is initiated by the monofunctional thymine DNA glycosylase followed by APE1-mediated strand incision, gap filling and nick sealing[61,170,227–231]. Following this notion, 5-hmC, 5-fC and 5-caC should be regarded not only as DNA lesions and potential regulatory marks, but also as active DNA demethylation intermediates.

Interestingly, TDG was found to be post-transcriptionally modified by addition of small ubiquitine-like modifier (SUMO) and acetyl groups in human cells. SUMOylation of lysine-330 by either SUMO1 or SUMO2 drastically reduces the DNA substrate and AP site binding affinity of TDG, thus enhancing the enzymatic turnover[232–234]. Acetylation of TDG does not directly affect the TDG activity but drastically reduces its interaction with APE1[235], which was shown to displace TDG from the generated AP sites[236]. Binding of the xeroderma pigmentosum, complementation group C protein enhances the slow AP site dissociation of TDG, thereby accelerating local and global active DNA demethylation kinetics[237,238]. Binding of the Nei-endonuclease VIII-like protein (NEIL) DNA glycosylases, as well as GADD45 also positively affects TDG-mediated base removal[217,239,240]. Many more protein-protein interactions of TDG have been found in human cells, indicating that they play an even bigger role in regulating TDG than the few post-translational protein modifications.

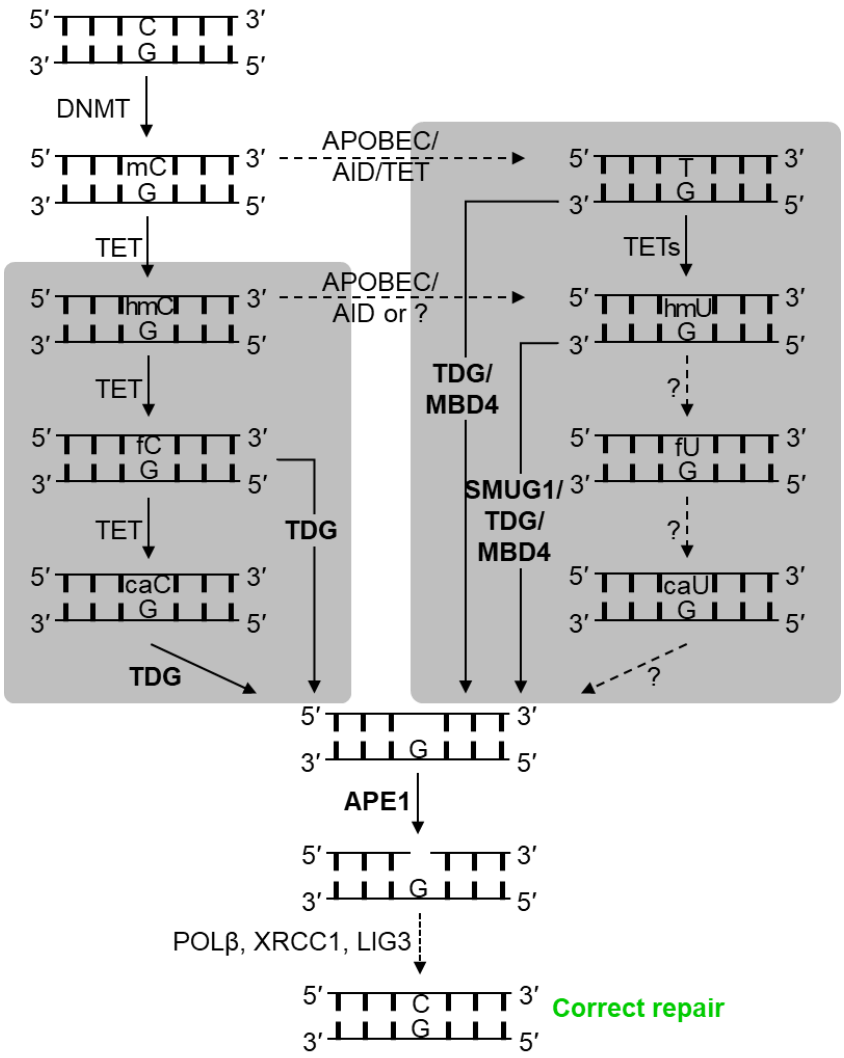


Figure 4-9: Potential mechanisms of active DNA demethylation by base excision repair in human cells

Schematic presentation of two potential mechanisms of active DNA demethylation by base excision repair in human cells, showing the major participating enzymes and structural repair intermediates. After DNMTs methylate C, the generated 5-mC can be actively removed from the genome by TET-TDG-mediated BER (left panel) and by a deamination induced BER-dependent pathway (right panel). TETs iteratively oxidise 5-mC to 5-hmC, 5-fC and 5-caC, of which 5-fC and 5-caC are removed from the DNA by TDG, initiating classical BER. Alternatively, 5-mC and 5-hmC can be deaminated by AID, APOBEC and TET, resulting in 5-mU:G (equals T:G) and 5-hmU:G base pairs respectively, however experimental proof is controversial (illustrated by dotted lines). 5-mU can be oxidised to 5-hmU by TETs and both DNA modifications can be removed from the DNA by TDG, MBD4 and SMUG1 initiated BER. It has been speculated, that 5-hmU is further oxidised to 5-fU and 5-caU, which could be further processed by BER, however experimental verification is lacking in human cells (dotted lines).

Another BER-mediated mechanism for active DNA demethylation that has been discussed over the last years, is the deamination of 5-mC and 5-hmC to T and 5-hmU respectively (Figure 4-9, right panel). Potential 5-mC and 5-hmC deaminating enzymes are AID[212,241], APOBEC[212,241] and TETs[241], all of which are known to be involved in transcription regulation in other instances[226,230,242–244]. T can then either be directly removed from the DNA by TDG or MBD4 [245,246], or be oxidised by TETs to 5-hmU[247], which can be removed from the DNA by SMUG1[248–250], TDG[245,251,252] or MBD4[252–254]. By analogy to 5-hmC, it is possible that 5-hmU is further oxidised to the potential BER

targets 5-formyluracil (5-fU) and 5-carboxyuracil (5-caU), however the contributing human enzymes have yet to be discovered. Base removal of T, 5-hmU, 5-fU and 5-caU results in the generation of an AP site, which is further processed by classical BER. The frequency of 5-hmU and 5-fU in the mammalian genome is estimated to vary between $1x10^{-7}$ and $1x10^{-6}$ modifications per nucleotide and although the importance of this alternative DNA demethylation pathway is yet to be determined, at least 7% of 5-hmU and 5-fU in mice are generated from 5-hmC[247,255].

Interestingly, BER of 5-hmU at important gene positions negatively influenced the expression of the affected gene[97]. Investigations in this lab showed, that at least 5-hmU opposite to A in a CRE containing promoter drastically reduced the gene expression in a SMUG1 and therefore BER-dependent manner[97]. 5-hmU can also be recognised by chromatin remodelling proteins and specific transcription factors in mammalian cells, supporting the hypothesis that genomic 5-hmU may have a regulatory function in human cells[247,255]. However, it is unclear if 5-hmU's function is limited to modifications opposite to A or if 5-hmU opposite to G also retains transcription regulation potential.

4.3.3 Base excision repair of the DNA lesions uracil, 5-hydroxymethyluracil and thymine glycol

In addition to 5-hmC, 5-fC, 5-caC and 8-oxoG, human DNA harbours several other DNA modifications processed by BER, including the well-characterised DNA lesion thymine glycol. With approximately 400 daily formed residues per cell[256], Tg is one of the most common oxidatively induced nucleobase lesions generated by ionising radiation and oxidative stress[257,258]. It mostly results from thymine oxidation generating Tg:A base pairs[259], but can also be generated by 5-mC oxidation and subsequent deamination of the unstable 5-methylcytosine glycol product generating Tg:G base pairs[34]. There is published evidence that, Tg constitutes a strong block for DNA replication[260], wherefore its removal is essential for genome stability and cellular survival. Tg repair in human cells is majorly accomplished by efficient BER initiated by the bifunctional NTHL1 DNA glycosylase[147,165,261] (depicted in Figure 4-10, left panel), with potential contribution of NEILs[262]. The large knowledge gained over the last decades about Tg generation, repair and functional consequences make the DNA lesion a perfect substrate when investigating BER.

A nucleobase lesion which is even more common in the human DNA than Tg is uracil[263], which is also repaired by BER (depicted in Figure 4-10, middle panel). U can either be generated by hydrolytic deamination of genomic cytosine, forming non-canonical U:G pairs, or by misincorporation of dUMP instead of dTMP during replication, creating a U:A pair [1,121,264]. Additionally, U can be enzymatically generated by AID-mediated deamination of cytosine[265]. AID induced U generation in human B-cells is essential for antibody diversification[73,243,264] and its misfunction can cause immune deficiencies like the Hyper-IgM-syndrome[16,244]. Although essential to establish immunity, U:G miss pairs in non-B-cells are especially dangerous for an organism, as they can give rise to C to T transition mutations, the most common base-mutation in cancer[266]. Mammalian cells counteract these deleterious effects by efficient BER removing U from the DNA. Uracil DNA glycosylase 2 (UNG2) is the main U repair enzyme in genomic DNA, efficiently removing uracil from U:G and U:A base pairs[16,168,267,268]. It has been reported, that the single-strand selective monofunctional uracil DNA glycosylase 1 (SMUG1), TDG und MBD4 support and back-up UNG2 in the repair of genomic U, however the significance of their contribution to U repair in vivo is still under debate. It was demonstrated that SMUG1 efficiently removes U from U:G and U:A base pairs in single- and double stranded DNA in vitro[168,269]. Furthermore, experiments in mice and mouse embryonic fibroblasts indicate that SMUG1 also contributes to U repair in vivo[270,271], although only a minimal contribution to U repair was detected in human cells[272,273]. TDG shows a preference towards U:G over U:C and U:T base pairs and is most active on bases in CpG contexts[169,274]. MBD4 removes U from U:G base pairs specifically in CpG dinucleotides, however due to its low enzymatic activity its significance in the cellular repair of genomic U is still unclear[158,159,264].

Comparable to Tg, the DNA modification 5-hmU can be generated from several nucleobases in the human genome including T, 5-mC and 5-hmC. TET, AID and APOBEC proteins can generate 5-hmU opposite to G by deamination of 5-mC or 5-hmC during alternative DNA demethylation[212,241,247]. Additionally, ROS induced thymine oxidation generates 5-hmU in 5-hmU:A base pairs, where the base modification is classically seen as a DNA lesion[275]. 5-hmU opposite to A and G is removed from the DNA by efficient BER, as depicted in the right panel of Figure 4-10. Removal of the 5-hmU opposite to A is initiated by SMUG1, whereas for 5-hmU:G pairs

the additional contribution of TDG and MBD4 to base removal has been demonstrated[248,251,252,254].

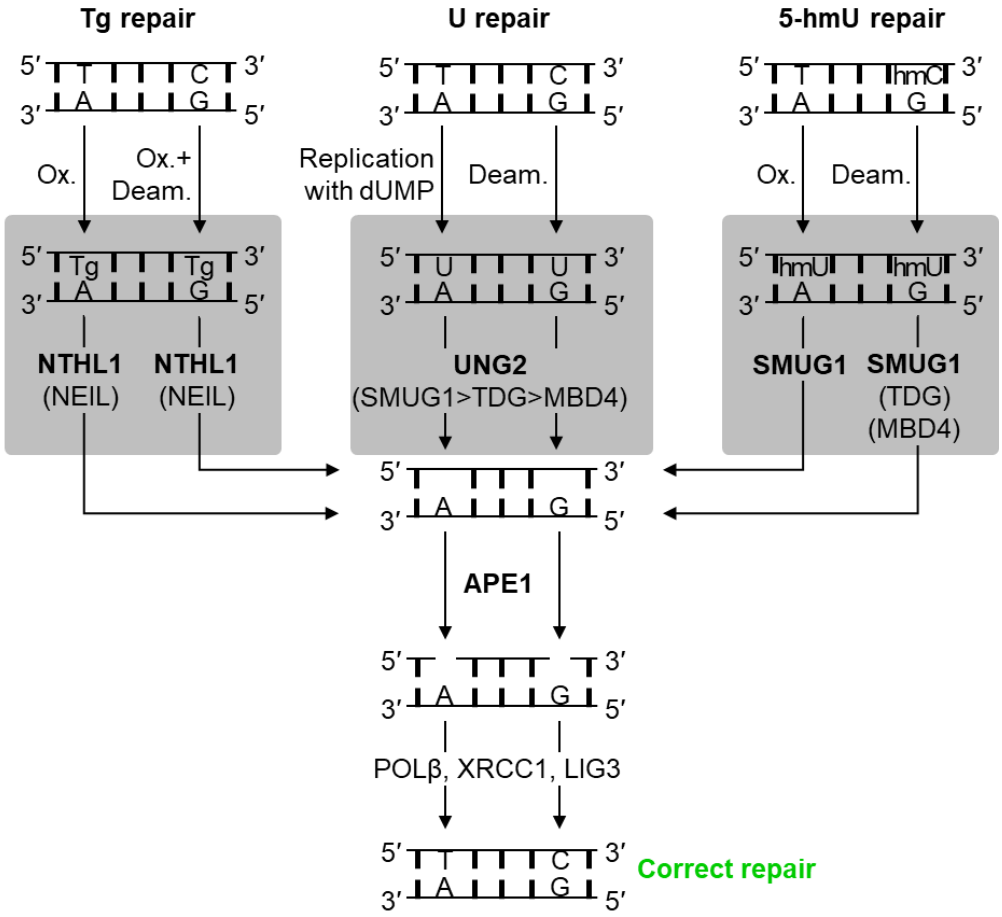


Figure 4-10: Base excision repair of thymine glycol, uracil and 5-hydroxymethyluracil in human cells

Schematic presentation of BER of Tg (left), U (middle) and 5-hmU (right) in human cells, showing the major participating enzymes and structural repair intermediates. BER targets Tg, U and 5-hmU are generated in the human genome by oxidation (Ox.) and deamination (Deam.) of T, C and 5-hmC. Additionally, dUMP can be incorporated opposite to A during DNA synthesis generating U:A base pairs. BER of Tg is initiated by NTHL1 with a potential back-up mechanism via NEIL initiated BER (left). BER of genomic U is majorly initiated by UNG2, although low U excision activity has also been demonstrated for SMUG1, TDG and MBD4 (middle). BER of 5-hmU opposite to A and G is initiated by SMUG1, whilst 5-hmU:G excision has also been demonstrated for TDG and MBD4 (right). As all these DNA glycosylases involved in Tg, U and 5-hmU repair are monofunctional, DNA strand incision at the generated AP sites is performed by APE1 and the 1 nt gap is filled with the canonical nucleotide by POLβ. Ligation of the remaining nick is performed by XRCC1/DNA LIG3, thereby restoring the correct DNA sequence.

5 Scope of this work

The aim of this investigation was to reveal the fundamental transcriptional consequences of the rare and poorly studied 5-mC oxidation products 5-fC and 5-caC and the controversially discussed major guanine oxidation product 8-oxoG. Effects of 5-fC and 5-caC on the gene expression were investigated in two common promoter motives. The GC-poor CRE upstream regulatory element was chosen because previous findings indicated that CRE activity is modulated by both DNA modifications[97] and the GC box was selected as a representative GC-rich URE. Since the basic transcriptional consequences of 8-oxoG have already been investigated in a simple GC-poor CRE promoter but not in a simple GC-rich promoter[7,276], this work focused on assessing the base modifications effects on gene expression in a minimal GC box promoter.

A plasmid-based reporter gene assay was used to investigate the functional consequences of 5-fC, 5-caC and 8-oxoG in human cells, thereby overcoming the hurdle of 5-fC's and 5-caC's rare frequency in the genome. 5-fC, 5-caC and 8-oxoG were site and sequence specifically introduce into the central CpG dinucleotide of CRE and/or GC box gene regulatory elements of EGFP reporters using the strand exchange protocol[277] as described in chapter 6.30. The transcriptional effects of the base modifications were then examined by reporter transfection into HeLa cells, followed by quantitative flow cytometric fluorescence analysis established previously[278]. Potential strand biases and influences of the promoter strength and URE choice on the transcriptional effects 5-fC and 5-caC were investigated using different EGFP reporters. Furthermore, the activity of promoters containing 8-oxoG at four different positions of the GC box was characterised, to determine how the DNA strand and the distance of 8-oxoG from the central CpG dinucleotide affected gene expression. As previous experiments indicated that 5-fC, 5-caC and 8-oxoG influence the gene expression by a BER-dependent mechanism[6,7,97], the transcriptional effects of the primary modifications needed to be differentiated from potential repair-induced effects. To achieve this task, excision of the modified base was inhibited using BER deficient DNA glycosylase knockdown-/knockout cells and chemically stabilised 2′-fluorinated DNA modifications. Furthermore, BER was abolished at the strand incision step using AP sites with a 5′-phosphorothioate, which protects the repair intermediated from APE1.

6 Materials and Methods

6.1 Instruments

Instrument	Manufacturer	Location
Centrifuge		
Heraeus FRESCO21 Centrifuge	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
VWR Mega Star 1.6 R Benchtop centrifuge	VWR INTERNATIONAL, BVBA	Leuven, Belgium
Cell counter		
Countess II FL Automated Cell Counter	LIFE TECHNOLOGIES CORPORATION	Carlsbad, CA, USA
Flow cytometer		
BD FACSCalibur™	BECTON DICKINSON GMBH,	Heidelberg, Germany
BD FACSAria™ III Cell Sorter	BECTON, DICKINSON AND COMPANY	Franklin Lakes, NJ, USA
Fluorescent microscope		
EVOS FLoid™ Cell Imaging Station	LIFE TECHNOLOGIES CORPORATION	Carlsbad, CA, USA
Gel documentation system		
Gel Doc™ EZ System	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
Gel electrophoresis system		
Wide Mini-Sub® Cell GT System	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
Mini-PROTEAN Tetra Cell	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
Photometer, Microtiter plate reader and infrared Imaging System		
Eppendorf BioPhotometer	EPPENDORF AG	Hamburg, Germany
LI-COR Odyssey 9120 infrared imaging system	LI-COR, INC.	Lincoln, NE, USA
NanoDrop™ 2000 Spectrophotometer	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
NanoPhotometer® N50	IMPLEN GMBH	Munich, Germany
TriStar^2 LB 942 Multimode Plate Reader	BERTHOLD TECHNOLOGIES GMBH & CO. KG	Bad Wildbad, Germany
Power supply		
PowerPac™ Basic Power Supply	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
Sonicator		
UP200Ht - Handheld Ultrasonic Homogenizer	HIELSCHER ULTRASONICS GMBH	Teltow, Germany
Thermal cycler		
MJ Mini™ Personal Thermal Cycler	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
T100™ Thermal Cycler	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA

6.2 Software

Software	Publisher	Location
DNA sequencing		
ClustalW2	EMBL-EBI	Cambridgeshire, UK
blastn	U.S. NATIONAL LIBRARY OF MEDICINE	Bethesda, MD, USA
Flow cytometry		
CellQuest™ Pro	BECTON DICKINSON GMBH	Heidelberg, Germany
Gel documentation		
Image Lab™	BIO-RAD LABORATORIES, INC.	Hercules, CA, USA
Primer design		
Primer Designer 4 V4.1	SCIENTIFIC & EDUCATIONAL SOFTWARE	Cary, NC, USA

6.3 Plasmid vectors

Construct name	Description	Origin
pCMV-1111-ZA-W	Stable EGFP under the control of a modified VS40 promoter containing several CRE, potential strand exchange of the NTS; vector also codes for kanamycin-resistance gene	J. Allgayer (Mainz, Germany)[136]
pCRE-UNO-ZA-W	Stable EGFP under the control of a modified VS40 promoter containing a single CRE, potential strand exchange of the NTS; vector also codes for kanamycin-resistance gene	J. Allgayer (Mainz, Germany)[5]
pCRE-UNO-ZA-C	Stable EGFP under the control of a modified VS40 promoter containing a single CRE, potential strand exchange of the TS; vector also codes for kanamycin-resistance gene	J. Allgayer (Mainz, Germany)[5]
pCRE-ZERO-ZA-W	Stable EGFP under the control of a modified VS40 promoter without any CRE, potential strand exchange of the NTS; vector also codes for kanamycin-resistance gene	J. Allgayer (Mainz, Germany)[5]
pDsRed Monomer N1 (pDsRed)	Vector coding for DsRed-Monomer N1 as transfection marker	CLONTECH LABORATORIES INC, Saint Germaine Laye, France
pX330-sgCAS9-HF	Vector coding for scrambled sgRNA, CAS9 protein and ampicillin resistance, potential exchange of the scrambled sgRNA sequence	ADDGENE, Watertown, MA, USA
pEGFP-mODC-ZAJ (pZAJ)	Modified pEGFP-mODC-ZA coding for stabilised EGFP with silent mutation in the kanamycin-resistance gene to eliminate an undesired restriction site	J. Allgayer (Mainz, Germany)[276]

6.4 Sequencing primer

Sequencing primer were purchased by EUROFINS GENOMICS GMBH (Ebersberg) as high purity salt free grade (HPSF).

Oligonucleotide	Template	Position	Sequence (5′-3′)
pZASS-PLUS-A1	pCRE-UNO-ZA-W and pCRE-UNO-ZA-C derived vectors	~220 bp upstream from CRE	TTCGCCACCTCTGACTTGA
U6 promoter forward	pX330-sgCAS9-HF derived vectors	~56 bp upstream from sgRNA sequence	CGTAACTTGAAAGTATTTC GATTTCTTGGC

6.5 Enzymes, inhibitors and marker

Product	Cat. Nr.	Producer	Location
Antarctic Phosphatase (5 U/µl)	M0289S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Aat II (20 U/ul)	R0117 S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
APE1 (10 U/µl)	M0282S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
FastDigest BpiI (BbsI)	FD1014	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
BSA (20 mg/ml)	B9000S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
BsrDI (5 U/µl)	R0574S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Endonuclease III (10 U/ul)	M0268S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Endonuclease IV (10 U/ul)	M0304S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Fpg (8 U/ul)	M0240S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Gel Loading Dye Purple (x6, no SDS)	B7025S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Gel Loading Dye Purple (x6, SDS)	B7024S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
GeneRuler Ultra Low Range DNA Ladder	SM1211	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
GeneRuler 1 kb DNA Ladder	SM0311	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
hOGG1 (1.6 U/µl)	M0241	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
Nb.BsrDI (10 U/µl)	R0648S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
PageRuler™ Plus Prestained Protein Ladder	26620	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
Proteinase K	A3830,0100	APPLICHEM GMBH	Darmstadt, Germany
PvuII-HF (20 U/µl)	R3151 S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany
S7 Fusion High-Fidelity DNA Polymerase (2 U/µl)	332530S	BIOZYM SCIENTIFIC GMBH	Hessisch Oldendorf, Germany
T4 DNA Ligase (30 U/µl)	EL0013	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
T4 Polynucleotide Kinase (10 U/µl)	EK0032	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
Taq DNA Polymerase (5 U/µl)	M0267S	NEW ENGLAND BIOLABS GMBH	Frankfurt, Germany

6.6 Kits

Kit	Cat. Nr.	Producer	Location
Amicon® Ultra-0.5 Centrifugal Filter Devices 30k	10256744	MERCK KGAA	Darmstadt, Germany
Effectene® Transfection Reagent	301427	QIAGEN GMBH	Hilden, Germany
GeneElut™ HP Plasmid Miniprep Kit	NA0150	SIGMA-ALDRICH GMBH	Seelze, Germany
Illustra™ GFX™ PCR DNA and Gel Band Purification Kit	10536295	GE HEALTHCARE	Chalfont St Giles, UK
QIAGEN Plasmid Mega Kit	12183	QIAGEN GMBH	Hilden, Germany

6.7 Antibodies

Target	Antibody	Cat. Nr.	Producer	Location
TDG	TDG Polyclonal Antibody from Rabbit (1µg/µl)	PA5-29140	THERMO FISHER SCIENTIFIC INC.	Waltham, MA, USA
HSP90	HSP90 Monoclonal Antibody from Mouse	AC88#ADI-SPA-830	ENZO LIFE SCIENCES (ELS) AG	Lausen, Switzerland
Mouse	IRDye® 680RD Donkey anti-Mouse IgG Secondary Antibody	926-68072	LI-COR, INC	Lincoln, NE, USA
Rabbit	IRDye® 800CW Donkey anti-Rabbit IgG Secondary Antibody	926-32213	LI-COR, INC	Lincoln, NE, USA

6.8 Antibiotics, buffer and media for cell culture

Chemical	Cat. Nr.	Producer	Location
Ampicillin	HP62.1	CARL ROTH GMBH	Karlsruhe, Germany
DMEM High Glucose (Dulbecco's Modified Eagle Medium, Gibco™)	41965-062	LIFE TECHNOLOGIES GMBH.	Darmstadt, Germany
FBS (Fetal Bovine Serum, Gibco™)	10270-106	LIFE TECHNOLOGIES GMBH	Darmstadt, Germany
LB-Medium (Lennox)	X964.1	CARL ROTH GMBH	Karlsruhe, Germany
LB-Medium (Luria/Miller)	X968.1	CARL ROTH GMBH	Karlsruhe, Germany
L-Glutamine solution (200 mM)	G7513-100ML	SIGMA-ALDRICH GMBH	Seelze, Germany
Penicillin-Streptomycin (10,000 U/mL)	15140-122	LIFE TECHNOLOGIES GMBH	Darmstadt, Germany
PBS (Gibco™)	882126-12	BIOZYM SCIENTIFIC GMBH	Hessisch Oldendorf, Germany
SOB medium	AE27.1	CARL ROTH GMBH	Karlsruhe, Germany
Tetracycline	T3258-5G	SIGMA-ALDRICH GMBH	Seelze, Germany
Trypsin-EDTA solution	T3924-100ML	SIGMA-ALDRICH GMBH	Seelze, Germany

6.9 Cell lines and bacteria

Cell line or bacteria	Description	Origin
E. coli SCS-8	*recA1 endA1 mcrA Δ(mcrBC-hsdRMS-mrr) Δ (argF-lac)U169 Φ80dlacZ ΔM15 Tn10 (Tetr)*	AGILENT TECHNOLOGIES, INC (Santa Clara, CA, USA)
HeLa	Human epithelial cervical cancer cells	R.J. Wiesner (Köln, Germany)
HeLa OGG1-GFP (Ser/Cys)	Human epithelial cervical cancer cells overexpressing OGG1-GFP fusion protein (Ser mutation variant)	J. Pablo Radicella (Fontenay aux Roses, France)[279]
HeLa OGG1-sh	Human epithelial cervical cancer cells stably transfected with OGG1 shRNA	J. Allgayer (Mainz, Germany)[276]
HeLa pEps	Human epithelial cervical cancer cells stably transfected with pENTR/pSUPER+-Vektor (pEpS+) containing non-targeting shRNA	B. Lühnsdorf (Mainz, Germany)[280]
HeLa SMUG-sh (Clone J22)	Human epithelial cervical cancer cells stably transfected with hSMUG1 shRNA	B. Lühnsdorf (Mainz, Germany)[280]
HeLa TDG-sh (Clone B6)	Human epithelial cervical cancer cells stably transfected with TDG shRNA	B. Lühnsdorf Mainz (Mainz, Germany)[280]
HeLa UNG-sh (Clone 12)	Human epithelial cervical cancer cells stably transfected with UNG1/2 shRNA	B. Lühnsdorf Mainz (Mainz, Germany)[280]

6.10 Composition of special buffers and solutions

Buffer or solution	Composition
Ampicillin solution	50 mg/ml Adjust in double distilled H_2O and store at -20 °C
BEH Buffer	10 mM Hepes, pH 7.5 200 mM NaCl 1 mM EDTA
Blocking solution for western blot	5% milk powder In 1x TBS-T
Blotting buffer for western blot	0.5x Lämmli (5x stock) 20% Ethanol Adjust to 1 l with double distilled H_2O
Bradford reagent	50 mg Serva Blue G 50 ml H_3PO_4 25 ml Ethanol Adjust to 500 ml with double distilled H_2O
Electrophoresis buffer for SDS PAGE	1x Lämmli buffer (5x stock) 10% SDS Adjust to 1 l with double distilled H_2O
Kanamycin solution	30 mg/ml Adjust in double distilled H_2O and store at -20 °C
Lämmli protein buffer (6x)	375 mM Tris-HCl pH 6.8 12% SDS 30% Glycerol 500 mM DTT 0.01% Bromphenol blue
Lämmli buffer (5x)	250 mM Tris 1.92 M Glycerol Adjust to 1 l with double distilled H_2O
Loading dye for SDS-PAGE	15,1 g Tris-HCl 94 g Glycerol 25 ml 20% SDS Adjust to 1 l with double distilled H_2O
LB-Medium 1x (2x)	5 (10) g Tryptone 2,5 (5) g Yeast extract 5 g NaCl Adjust to pH 7.5-8 ad 500 ml double distilled H_2O, autoclave
Lysis buffer for sonification	20 mM Tris-HCl, pH 8.0 1 mM EDTA, pH 8.0 5 M NaCl 1x protease inhibitor Adjust in double distilled H_2O, only directly prepare before usage and keep on ice at all times

NET-N	100 mM NaCl 10 mM Tris-HCl pH 8.0 10% Glycerol 1 mM EDTA 0.5% NP-40
PMSF	100 mM PMSF Adjust in isopropanol
PBS-PMSF (1x)	0.5 mM PMSF in PBS
Protease Inhibitor (7x)	1 tablet of cOmplete™ protease inhibitor cocktail (ROCHE DIAGNOSTICS GMBH, Mannheim, Germany) Dissolve in 1.5 ml of double distilled H_2O
SOB-Medium	0,5% Yeast extract 2% Tryptone 10 mM NaCl 2,5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ Adjust to 120 ml with double distilled H_2O Autoclave and store at room temperature or 4 °C
TB buffer	15 mM CaCl2 250 mM KCl Dissolve in autoclaved double distilled H_2O and adjust to pH 6.7 Add 10 mM PIPES 55 mM $MnCl_2$ Adjust to 100 ml with autoclaved double distilled H_2O, filter sterilise and store at 4 °C
TBS (10x)	200 mM Tris 1.4 M NaCl Adjust to pH 7.6 Adjust to 1 l with double distilled H_2O
TBS-T (x1, 0.05% Tween)	1x TBS 0.05% Tween 20 Adjust to 1 l with double distilled H_2O
TE buffer	10 mM Tris-HCl, pH 8.0 1 mM EDTA, pH 8.0 Adjust in double distilled H_2O
Tetracycline solution	20 mg/ml Adjust in methanol and store at -20 °C in absence of light
Transfer buffer WB	5.8 g Tris 2.9 g Glycerol 200 ml Ethanol Adjust to 1 l with double distilled H_2O

6.11 DNA quantification by spectrophotometry

DNA concentration of aqueous solutions was measured by spectrophotometry using the NanoDrop™ 2000 Spectrophotometer or the NanoPhotometer® N50. The DNA concentration was quantified and the absorption ratios A260/A280 and A260/A230 were used as an indicator for sample purity, with optimal values between 1.8-2.0 DNA solutions were stored at -20 °C and kept on ice after thawing at all times**.** If a DNA containing solution was used in the following reactions, it was added after the addition of water and any buffer solution.

6.12 Agarose gel electrophoresis

Standard agarose gel electrophoresis was performed with 0.8% agarose gels containing 0.625 µg/ml ethidium bromide (EthBr). The gels were prepared with UltraPure™ Agarose (THERMO FISHER SCIENTIFIC INC.) and 1xTAE. The gels were run at 80 V for 45 min in 1xTAE buffer followed by destaining in double distilled Water for 1-5 min using PowerPac™ Basic Power Supply. The EthBr-stained DNA was detected using Gel Doc™ EZ System (BIO-RAD LABORATORIES, INC.) and a picture was taken and analysed with Image Lab™ (BIO-RAD LABORATORIES, INC.) to calculate the amount of closed circle (cc), open circle (oc) and linear plasmid DNA in each sample. Since plasmids in their cc-form bind less DNA than in their oc- or linear-form, the band intensity of cc-plasmid DNA was multiplied by 2.4 when calculating the DNA content of the DNA bands to compensate for supercoiling[276].

Separation of DNA fragments shorter than 1000 bp such as polymerase chain reaction (PCR) products was performed in 1% agarose gels without EthBr. After the electrophoresis step, the gel was stained by incubation with a 100 µg/ml EthBr solution for 10-15 min, destained in a water bath for another 10-15 min and analysed for EthBr signals.

6.13 Cloning of EGFP reporters regulated by minimal GC box and TRE promoters

Reporters containing the EGFP gene under the control of a minimal GC box or 12-O-tetradecanoylphorbol 13-acetate-response element (TRE) promoter were generated by the following cloning procedure. In summary, the single CRE sequence of the parental pCRE-UNO-ZA-W and pCRE-UNO-ZA-C vectors was exchanged for a single GC box or TRE consensus sequence as depicted in Figure 6-1. Tandem restriction sites flanking CRE in the CRE-UNO constructs enabled the exchange of the

interspersed CRE-containing DNA for a GC box or TRE containing DNA sequence. Due to the inverted restriction sites in the parental vectors, the resulting expression constructs pGCbox-ZA-W/C and pTREC-ZA-W/C (for sequences see Appendix II) were suitable to replace the NTS or TS of the DNA based on the parental pCRE-UNO-ZA-W or pCRE-UNO-ZA-C vectors respectively.

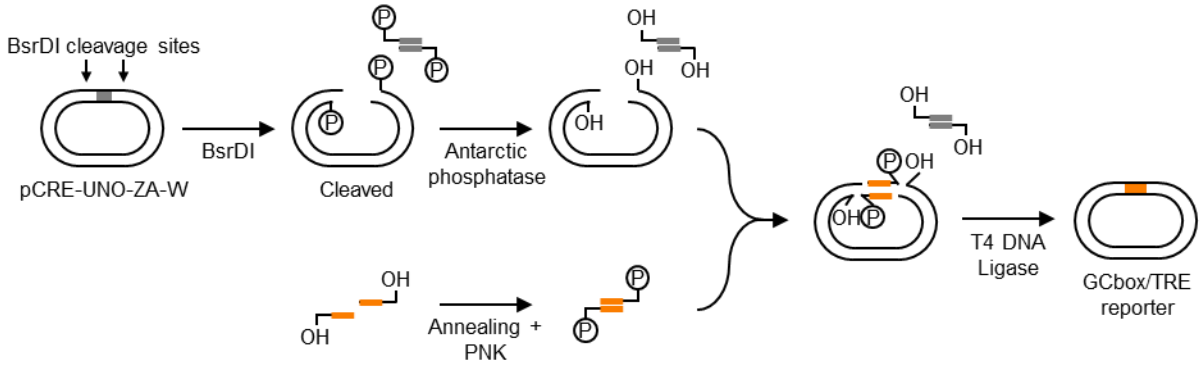


Figure 6-1: Overview of the cloning procedure for the generation of minimal GC box and TRE reporters

Schematic representation of the cloning procedure to generate EGFP reporters controled by minimal GC box or TRE promoters showing the major structural intermediates. GC box and TRE reporters were generated from the pCRE-UNO-ZA-W (depicted here) and pCRE-UNO-ZA-C vectors (not shown) by exchange of the single CRE site for a GC box or TRE consensus sequence. Parental CRE-UNO vectors were restricted with BsrDI and dephosphorylated by Antarctic Phosphatase as depicted for pCRE-UNO-ZA-W above. Then GC box and TRE coding inserts were generated by oligonucleotide annealing, phosphorylation by PNK and ligation with the restricted vector using T4 DNA Ligase.

During the first step of the plasmid cloning procedure, parental pCRE-UNO-ZA-W and pCRE-UNO-ZA-C vectors were restricted by the restriction enzyme BsrDI in a total reaction volume of 50 µl. BsrDI induces two double strand breaks at the designated restriction sites, resulting in the excision of a 18-nucleotide long DNA fragment including CRE. The reactions were performed in a thermocycler with reaction conditions and component concentrations as shown below.

Table 6-1: Reaction mix for the restriction of pCRE-UNO-ZA-W and pCRE-UNO-ZA-C by BsrDI

Reagent	Final conc./amount
Plasmid DNA	1.5 µg
NEBuffer 2.1 (x10)	1x
BsrDI (5U/µl)	15 U
Adjust to 75 µl with H_2O	

Table 6-2: Programme for the restriction of pCRE-UNO-ZA-W and pCRE-UNO-ZA-C by BsrDI

Step	Temperature	Time	Action
1	4 °C	∞	
2	65 °C	30 min	
3			Mix by pipetting up and down
4	65 °C	30 min	
5	80 °C	20 min	
6	4 °C	∞	

10 µl of the restricted CRE-UNO vector were kept as a ligation control for later experiments, whereas the remaining linearised vector was immediately dephosphorylated by incubation with antarctic phosphatase to prevent re-ligation of the linearised vector with the excised 18 nt DNA fragment.

Table 6-3: Reaction mix for the dephosphorylation of linearised pCRE-UNO-ZA-W and pCRE-UNO-ZA-C by antarctic phosphatase

Reagent	Final conc./amount
Restricted plasmid DNA (20 ng/µl)	1 µg
Antarctic Phosphatase buffer (x10)	1x
Antarctic Phosphatase (5 U/µl)	2 U
Adjust to 56 µl with H_2O	

Table 6-4: Programme for the dephosphorylation of linearised pCRE-UNO-ZA-W and pCRE-UNO-ZA-C by antarctic phosphatase

Step	Temperature	Time
1	4 °C	∞
2	37 °C	1 h
3	80 °C	10 min
4	4 °C	∞

In parallel to the linearisation of the CRE-UNO constructs, double stranded DNA inserts with BsrDI specific 5'-overhangs were generated, which code for the desired GC box or TRE sequences. The designated inserts were generated by annealing two complementary synthetic oligonucleotides (listed in Table 6-5) coding for the TF binding site of choice, which was separated from the BsrDI specific 5'_overhangs by a short linker sequence. Oligonucleotide annealing was performed in a thermocycler in the presence of T4 polynucleotide kinase (T4 PNK), which phosphorylates the 5'-end of the synthetic oligonucleotides to activate them for the following ligation step.

Synthetic oligonucleotides for the generation of GC box and TRE coding double stranded DNA inserts are listed in Table 6-5, which indicates the parental target vector in column 1, the gene regulatory element which is encoded on the insert in column 2 and the target strand and sequence of each oligonucleotide pair in columns 3-4.

Table 6-5: Oligonucleotides for cloning of GC box and TRE reporters

Oligonucleotides used to generate inserts for cloning of GC box and TRE reporters by annealing the NTS- and TS-sequences. Oligonucleotides were purchased by Eurofins Genomics GmbH (Ebersberg) as high purity liquid chromatography grade (HPLC).

Target vector	Insert	Strand	Sequence (5-3)
pCRE-UNO-W	GCbox-W	NTS	CATTGCATGGGCGGAGCG
		TS	CTCCGCCCATGCAATGAT
pCRE-UNO-C	GCbox-C	NTS	TGGGCGGAGCGCAATGTG
		TS	CATTGCGCTCCGCCCACG
pCRE-UNO-W	TREC-W	NTS	CATTGCATGAGTCAGCG
		TS	CTGACTCATGCAATGAT
pCRE-UNO-C	TREC-C	NTS	CTGACTCATGCAATGTG
		TS	CATTGCATGAGTCAGCG

Table 6-6: Reaction mix for the generation of GC box and TRE inserts by oligonucleotide annealing and phosphorylation using T4 PNK

Reagent	Final conc./amount
Oligonucleotide NTS	0.19 nmol (9.5µl of a 20 µmol/l solution)
Oligonucleotide TS	0.19 nmol (9.5µl of a 20 µmol/l solution)
T4 DNA Ligase buffer (x10)	1x
T4 PNK (10 U/µl)	16.7 U
Adjust to 25 µl with H_2O	

Table 6-7: Programme for the generation of GC box and TRE inserts by oligonucleotide annealing and phosphorylation using T4 PNK

Step	Temperature	Time
1	4 °C	∞
2	37 °C	30 min
3	95 °C	10 min
4	95→50°C	Ramp 0.1°C/s
5	50°C	1 min
6	4 °C	∞

The phosphorylated insert was ligated with the linearised dephosphorylated vector backbone by T4 DNA Ligase using plasmid:insert molarity ratios of 1:1 and 1:3. To determine the re-ligation frequency of restricted vector with residual 18 nt DNA fragment in the sample, a control without insert was prepared. The ligation efficiency

and potential secondary structure formation of restricted vector DNA were determined using a sample without ligase.

Table 6-8: Reaction mix for the ligation of GC box and TRE inserts with linearised pCRE-UNO-ZA-W/C using T4 DNA Ligase

Reagent	Final conc./amount
Insert mix	1:1 , 1:3 or 1:0 of plasmid:insert ratio (molarity)
Linearised plasmid DNA	100 ng
T4 DNA Ligase buffer (x10)	1x
T4 DNA Ligase (10 U/µl)	20 U
Adjust to 20 µl with H_2O	

Table 6-9: Programme for the ligation of GC box and TRE inserts with linearised pCRE-UNO-ZA-W/C using T4 DNA Ligase

Step	Temperature	Time
1	4 °C	∞
2	22°C	1 h
3	60°C	20 min
4	4 °C	∞

The generated GC box and TRE reporters were amplified in *E. coli* SCS-8 (chapter 6.15) and the plasmid DNA of 5 selected colonies per construct was extracted by plasmid mini preparation (chapter 6.16). Successful GC box and TRE construct generation was verified by analytical vector digestion of the extracted plasmid DNA by BsrDI/Nb.BsrDI (chapter 6.17) and sanger sequencing (chapter 6.18) using the pZASS-PLUS-A1 primer (chapter 6.4).

6.14 Generation of ultra-competent bacteria

E.coli SCS-8 were scraped from the -80 °C glycerol stock and transferred to 2 ml of lysogeny broth (LB) medium in a 50 ml falcon supplemented with 20 µg/ml of tetracycline. The bacteria were incubated at 37 °C on a rotary shaker with 250 rpm for approximately 8 hours. Afterwards, the cell suspension was transferred to a 500 ml Erlenmeyer flask containing 200 ml of LB medium supplemented with tetracycline. The cells were incubated at room temperature till reaching an optical density 600 (OD600) between 0.4 and 0.6 (~16 hours later). The suspension was cooled on ice for 10 min while shaking it gently and then centrifuged for 5 min at 4500 g and 4 °C. The medium was removed and the cells were resuspended in 40 ml of ice-cold Tris borate (TB)-buffer (chapter 6.10). After a second centrifugation step the bacteria were suspended in new 10 ml of ice-cold TB-buffer. 0.75 ml DMSO were added to obtain a final

concentration of 7% and the suspension was aliquoted into 2 ml screw cap tubes with approximately 500 µl bacteria suspension per aliquot and stored at -80 °C.

6.15 Transformation of ultra-competent bacteria

Ultra-competent *E.coli* SCS-8 cells (chapter 6.14) were defrozen on ice for approximately 30 min. 5 µl of a DNA solution of choice containing 2-25 ng of DNA (2 ng for supercoiled plasmid, 25 ng for newly cloned plasmids from ligation-reactions) and a control without DNA were incubated with 95 µl of ultra-competent *E.coli* SCS-8 for 10 min on ice. The bacteria were heat-shocked at 42 °C for exactly 90 s in the water bath, followed by incubation on ice for 2-10 min. The suspension was transferred to 2 ml of 2xLB medium without antibiotics and incubated at 37 °C for 30 min to 60 min on a rotary shaker at 250 rpm. LB agar plates containing plasmid specific antibiotics were used to plate approximately 50 µl of the cell suspension by glass bead assisted liquid distribution (1:1000 dilution of kanamycin stock solution for EGFP reporters and ampicilling stock solution for sgRNA-CAS9 expression vectors). The plated bacteria were cultivated at 37 °C for approximately 16 hours without shaking and the transformation and cultivation efficiency was analysed comparing the colony count. For cloned constructs, plate counts were compared from 1:1 and 1:3 vector:insert ratio samples and single colonies were selected only from plates with the highest number of spatially separated colonies.

6.16 Isolation of plasmid DNA from transformed bacteria

Plasmid DNA was isolated from transformed *E.coli* SCS-8 using mini- and mega-preparation. If smaller DNA amounts around 2 µg of plasmid were needed, mini-preparation was performed, whereas mega-preparation was performed to extract plasmid DNA on a larger scale with a target amount of approximately 3 mg of DNA.

For plasmid mini-preparation, transformed *E.coli* SCS-8 were initially collected either as a single colony from an agar plate or from a glycerol stock. The cells were incubated on a rotary shaker in 2 ml of 1xLB medium supplemented with the specific antibiotics for approximately 16 hours at 37 °C and 250 rpm. The plasmid DNA was extracted using the GeneElut™ HP Plasmid Miniprep Kit (SIGMA-ALDRICH GMBH) following suppliers' instructions using the spinning protocol and DNA was eluted in TE buffer (cf. chapter 6.10).

For plasmid mega-preparation, transformed *E.coli* SCS-8 from an agar plate or from a glycerol stock were incubated on a rotary shaker in 2 ml of 1xLB medium

supplemented with antibiotics for approximately 8 hours at 37 °C and 250 rpm. The suspension was transferred to 250 ml of 1xLB medium supplemented with specific antibiotics and the cells were expanded for another ~16 hours. The plasmid DNA was extracted after the plateau growth phase was reached, using the QIAGEN Plasmid Mega Kit (QIAGEN GMBH) following suppliers' instructions. The optional LyseBlue® reagent was used as mixing indicator and the solution was kept on ice at all times. Deviating from the protocol instructions, incubation with the neutralisation buffer was performed for only 10-15 min and the suspension was filtered through a double layer of pleated filter instead of centrifugation. Final plasmid-DNA pellets were suspended in 1.5-2.0 ml TE buffer.

The extracted plasmid DNA from mini- and mega preparations was analysed by spectrophotometry (chapter 6.11) and agarose gel electrophoresis (chapter 6.12) and stored at -20 °C.

6.17 Analytical digestion of plasmid DNA to verify successful reporter cloning or introduction of DNA modifications into EGFP reporters

Analytical digestion assays were used to qualitatively verify the presence and correct ligation of DNA inserts (chapter 6.13). Furthermore, analytical digestion assay were also used to verify the presence of a selected DNA modification in plasmid DNA, which was introduced into the reporter by the strand exchange method (chapter 6.30). In both cases, 100 ng of plasmid DNA was incubated with the sequence- or modification-specific enzyme (chapter 6.5) as summarised in Table 6-10 following the supplier's instruction. The reactions were performed in a total volume of 15 µl within the enzyme specific buffers and the treated DNA was analysed by agarose gel electrophoresis (cf. chapter 6.12) to verify the digestion outcome. Samples without enzyme were used as negative control. Digestion products which verify the successful cloning or nucleobase introduction are indicated in column 4 of Table 6-10. The table further lists the purpose of the digestion assay in column 1, the specific enzymes in column 2, their DNA target 3 in column as well as the amount of enzyme that is needed for the treatment of 100 ng of plasmid DNA in column 5. The incubation protocols of the different analytical digestions are summarised in Table 4-1

Table 6-10: Overview of the enzymes used for quantitative reporter digestion

Purpose	Enzyme	Target	Product	U/100 ng
Cloning success				
GC box and TRE inserts	BsrDI	GCAATG NN↓ CGTTAC↑NN	Linear DNA	1
	Nb.BsrDI	CGTTAC↑NN	Nicked DNA	0.1
Modification introduction				
Modified base at C within CpG of CRE	AatII	G ACGT↓C C↑TGCA G	Linear DNA	3
Tg in any sequence	EndoIII	Modified thymine	Nicked DNA	5
8-oxoG in any sequence	Fpg	Modified guanine	Nicked DNA	0.5
F and SF in any sequence	EndoIV	AP sites	Nicked DNA	4

Table 6-11: Programme for quantitative reporter digestion

	BsrDI & Nb.BsrDI		AatII		EndoIII & Fpg		EndoIV	
Step	**Temp.**	**Time**	**Temp.**	**Time**	**Temp.**	**Time**	**Temp.**	**Time**
1	4 °C	∞	4 °C	∞	4 °C	∞	4 °C	∞
2	65 °C	1 h	37 °C	1 h	37 °C	1 h	37 °C	1 h
3	80 °C	20 min	80 °C	10 min	60 °C	20 min	85 °C	20 min
4	4 °C	∞	4 °C	∞	4 °C	∞	4 °C	∞

6.18 Sanger sequencing of subcloned DNA fragments

The nucleotide composition of a selected DNA stretch was verified by Sanger sequencing using the service offered by STARSEQ® GMBH (Mainz) and EUROFINS GENOMICS GMBH (Ebersberg) (former GATC). Sample preparation is described below and the sequencing primers are described in chapter 6.4. Vectors derived from pCRE-UNO-ZA-W/C were sequenced using the primer pZASS-PLUS-A1 whereas vectors derived from pX330-sgCAS9-HF were sequenced using the U6 promoter forward primer.

Table 6-12: Reaction mix for the sequencing of a selected DNA stretch

	STARSEQ	GATC
Reagent	**Final conc./amount**	**Final conc./amount**
Plasmid DNA	450 ng (for 4.6-5.5 kb plasmids)	30-100 ng/µl
Primer	10 pmol	10 pmol
Adjust to 20 µl with H_2O		

After receiving the sequencing output, the sequencing performance was analysed using ClustalW (EMBL-EBI, Cambridgeshire, UK). To verify sequence correctness the nucleotide composition of the sample DNA was aligned with the computationally generated target sequence using the blastn online tool (U.S. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, Bethesda, MD, US).

6.19 Determination of the cell concentration in aqueous solutions using automated cell counting

The concentration of human cells within an aqueous solution was measured by automated cell counting using the Countess II FL Automated Cell Counter (LIFE TECHNOLOGIES) following suppliers' protocol. Living cells were distinguished from dead cells by standard trypan blue staining of the sample prior to analysis.

6.20 Design of single guide RNAs to introduce a CRISPR-CAS9-mediated gene knockout

Knock out of selected DNA glycosylases was established in HeLa cells by gene editing using the CRISPR-CAS9 (clustered regularly interspaced short palindrome repeats-CRISPR-associated protein 9) system. The active site coding nucleotides of the selected DNA glycosylase were removed from the genome by a plasmid-based gene editing approach following the "CRISPR Protocol for Genomic Deletions in Mammalian Cell Lines"[281] from ADDGENE (Watertown, USA). Removal of the enzyme's active site residue abolishes the base excision activity of the chosen DNA glycosylase, establishing a gene knockout.

To knock out a selected DNA glycosylase, a pair of CAS9 expression vectors needed to be generated per gene target. Each vector codes for the CAS9 protein and a single guide RNA (sgRNA), which was designed to guide the CAS9 protein to a selected site within the target gene. The sgRNA pairs were designed in a way that sgRNA-I targets a sequence upstream from the active site coding nucleotides of the target gene, whereas sgRNA-II targets a downstream DNA sequence (Figure 6-2). Thus, sgRNA pairs induced DNA cleavage at both sides of the active site codon, thereby enabling the deletion of the enclosed DNA stretch.

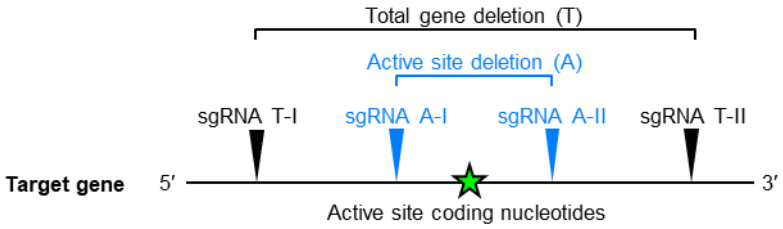


Figure 6-2: Position of single guide RNAs designed to induce total gene deletion or active site deletion of the target gene by CRISPR-CAS9-mediated gene editing

Schematic representation of the sgRNA positioning to ensure the deletion of the total gene sequence (T) or active site coding sequence (A) of the target gene by simultaneous incision at both sgRNA sites by CRISPR-CAS9-mediated gene editing. Scheme of the target gene: gene coding region (black line), active site coding nucleotides (green star), sgRNA targets for total gene deletion (black triangle), sgRNA for active site deletion (blue triangle), sgRNA target site upstream from active site coding nucleotides (sgRNA-I), sgRNA target site downstream from active site coding nucleotides (sgRNA-II).

Two strategies were tested to ensure efficient gene knockout: A) deletion of the exon that contains the catalytic residue by targeting this very exon or the neighbouring ones (active site knockout; A) deletion of the whole protein-coding sequence by targeting the first and the last exon whenever possible (total gene knockout; T) (Figure 6-2).

To design two pairs of sgRNAs (sgRNA-I and sgRNA-II for both strategies), the target gene sequence was searched for potential single guide RNAs, using the CHOPCHOP v2 online tool[282,283]. The gene name was added into the search engine as search criterion; homo sapiens was selected as target species and the nucleotide triade "NGG" was selected as protospacer adjacent motif (PAM) to introduce the gene knockout. From the output list two pairs of sgRNAs were selected. Exon positioned sgRNAs positioned with high editing efficiency and low off-target effect scores were selected preferentially. After selecting two sgRNA pairs, each sgRNA sequence was computationally modified according to the ADDGENE protocol[281] as depicted in (Figure 6-3) to enable the usage with the CAS9 expression vector pX330-sgCAS9-HF. The "NGG" PAM was deleted from the output sequence and a double stranded DNA sequence was generated from the single stranded guide sequence. To make the insert compatible with BpiI-mediated restriction cloning of pX330-sgCAS9-HF the ends of the DNA sequence needed to be trimmed. Thus, BpiI specific sticky ends "CACC(G)" (for the NTS, with G only if the first sgRNA output nucleotide is no G) and "AAAC" (for TS) were added to the 5′-ends of the modified guide sequence resulting in a 24-25 bp long double stranded sgRNA coding insert. If a G was added to the 5′-end of the NTS, the complementary TS was modified to contain an additional canonical C at the 3′-end. Designed sgRNA pairs were ordered as complementary HPLC grade oligonucleotides from EUROFINS GENOMICS GMBH (Ebersberg) as depicted in Table 6-15.

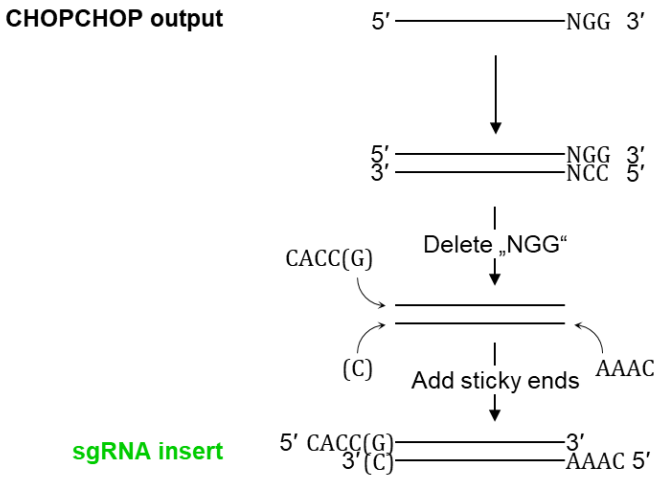


Figure 6-3: Generation of pX330-SgCas9-HF compatible sgRNA inserts from CHOPCHOP output sequences

Schematic representation of the double stranded sgRNA insert design for CRISPR-CAS9-mediated gene silencing based on the "CRISPR Protocol for Genomic Deletions in Mammalian Cell Lines"[281], showing the major structural intermediates. The selected CHOPCHOP sgRNA output sequence was used as a template to generate the reverse complementary strand and the ends were trimmed to compatible with the sticky ends of the pX330-SpCAS9-HF vector after restriction with BbiI enzyme. Therefore the "NGG" PAM sequence from CHOPCHOP output was removed and the CACC(G) and AAAC sticky ends are added to the 5′-end of the NTS and the 3′-end of TS respectively. (G) was only added if the CHOPCHOP output does not start with a G in the NTS, requiring the addition of a canonical C base pairing partner on the 3′-end of the TS.

6.21 Subcloning of sgRNA inserts into the pX330-sgCas9-HF expression vector to generate knockout constructs

To knock out selected DNA glycosylases in human HeLa cells by CRISPR-CAS9-mediated gene editing, knockout constructs coding for the CAS9 protein as well as a selected sgRNA were generated. The CAS9 coding pX330-sgCAS9-HF vector from ADDGENE was used as backbone, in which the scrambled sgRNA sequence was exchanged for sgRNA sequences of choice designed as described in chapter 6.20. The cloning procedure, as depicted in Figure 6-3, followed the "CRISPR Protocol for Genomic Deletions in Mammalian Cell Lines"[281], with modifications as described in the text below.

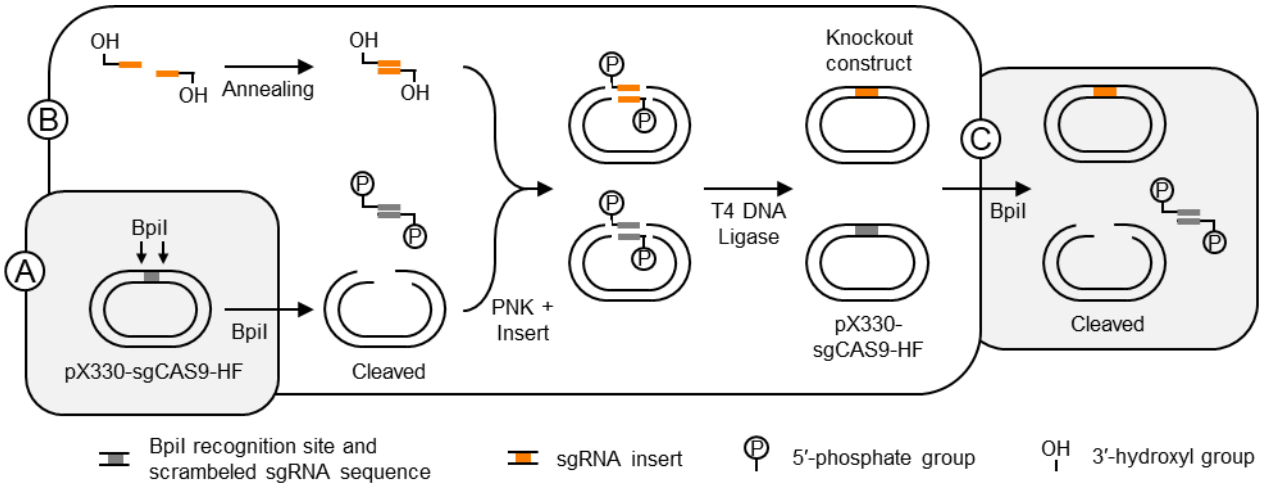


Figure 6-4: Overview of the cloning procedure for the generation of gene specific sgRNA-CAS9 expression vectors

Schematic representation of the cloning procedure to generate knockout constructs for CRISPR-CAS9-mediated gene editing showing the major structural intermediates. Knockout constructs were generated from pX330-sgCAS9-HF vector from ADDGENE in three sequential reactions, as indicated by square boxes A-C. (A) Parental pX330-sgCAS9-HF vector was restricted with BpiI. (B) sgRNA inserts targeting the gene of choice were generated by oligonucleotide annealing, phosphorylation and ligation with the restricted vector. (C) The reaction mixture was incubated with BpiI to cleave residual pX330-sgCAS9-HF vector. In contrast to parental pX330-sgCAS9-HF, knockout constructs carrying the sgRNA sequence are not restricted by BpiI because they lack the BpiI recognition site encoded on the excised DNA sequence.

In the first cloning step, pX330-sgCAS9-HF was digested with the restriction enzyme BpiI-HF, whose recognition sites are situated at the periphery of the scrambled sgRNA sequence within the vector. BpiI treatment results in the release of a 22 bp long DNA fragment coding for the scrambled sgRNA sequence and the BpiI recognition site. Digestion conditions and reaction components are listed in the tables below.

Table 6-13: Reaction mix for the digestion of the pX330-sgCAS9-HF vector with BpiI

Reagent	Final conc./amount
Plasmid DNA	3 µg
Fast digest buffer (x10)	1x
BpiI-HF	3 µl
Adjust to 150 µl with H_2O	

Table 6-14: Programme for the digestion of the pX330-sgCAS9-HF vector with BpiI

Step	Temperature	Time
1	4 °C	∞
2	37 °C	10 min
3	65 °C	20 min
4	4 °C	∞

Afterwards, the released 22 bp scrambled sgRNA sequence was removed from the sample by Amicon® Ultra-0.5 Centrifugal Filter Devices 30k (MERCK KGAA, DARMSTADT) following manufacturer's protocol. The purified linear pX330-sgCAS9-HF was adjusted to 25 µl with TE and the DNA concentration was determined by spectrophotometry (chapter 6.11).

In the second step of the cloning procedure, double stranded sgRNA inserts were generated and ligated with the linearised pX330-sgCAS9-HF vector. Complementary synthetic oligonucleotides coding for the sgRNA of choice, as designed in section 6.20 and listed in Table 6-15 were annealed to form the double stranded sgRNA insert. Complementary oligonucleotides are listed pairwise in Table 6-15, in which the target gene is indicated in column 2, the sgRNA position in columns 3-4 and the NTS an TS sequence of the insert (as complementary oligonucleotides) in column 6 (e.g. oligonucleotides TDG sgRNA T-I NTS and TDG sgRNA T-I TS). The oligonucleotides were phosphorylated by T4 PNK to enabled ligation with the linearised pX330-sgCAS9-HF vector by T4 DNA Ligase. All steps were performed in one reaction. To find optimal ligation conditions, vector:insert molarity ratios of 1:1 and 1:3 were used for ligation. A sample without oligonucleotide was used as a control to verify efficient plasmid restriction and purification.

Table 6-15: Oligonucleotides for the cloning of sgRNA CAS9 expression vectors

Oligonucleotides used to generate sgRNA inserts for cloning of sgRNA-CAS9 expression vectors by annealing the NTS- and TS-sequences. Oligonucleotides were purchased by Eurofins Genomics GmbH (Ebersberg) as high purity liquid chromatography grade (HPLC).

Target vector	Insert		Distance from TSS	Strand	Sequence (5-3)
pX330-sgCAS9-HF	TDG-sgRNA	T-I	93	NTS	CACCGAACGCGGGCAGGTAATACCG
				TS	AAACCGGTATTACCTGCCCGCGTTC
		T-II	18908	NTS	CACCGACGAAATATGGACGTTCAAG
				TS	AAACCTTGAACGTCCATATTTCGTC
		A-I	11012	NTS	CACCGATGGCTGAAGCTCCTAATA
				TS	AAACTATTAGGAGCTTCAGCCATC
		A-II	16905	NTS	CACCGCTACCAGGGAAGTATGGTAT
				TS	AAACATACCATACTTCCCTGGTAGC
	MBD4-sgRNA	T-I	-339	NTS	CACCGCCGAGCGCGCATGTCCGAAA
				TS	AAACTTTCGGACATGCGCGCTCGGC
		T-II	8652	NTS	CACCGTGGGCCCCTAGCTTTAGCA
				TS	AAACTGCTAAAGCTAGGGGCCCAC
		A-I	7651	NTS	CACCGTACACCACTACAGGAAAGC
				TS	AAACGCTTTCCTGTAGTGGTGTAC
		A-II	8479	NTS	CACCGCAGTAGCAAAGATCCACCAT
				TS	AAACATGGTGGATCTTTGCTACTGC
	SMUG1-sgRNA	T-I	-742	NTS	CACCGCATCTAAGGCAAGATGGCGT
				TS	AAACACGCCATCTTGCCTTAGATGC
		T-II	1953	NTS	CACCGAATACGTTTCCCAGCGACC
				TS	AAACGGTCGCTGGGAAACGTATTC
		A-I	153	NTS	CACCGGGCATCATCTACAATCCCG
				TS	AAACCGGGATTGTAGATGATGCCC
		A-II	1546	NTS	CACCGGAGTAAGGTTGCGCCCGCT
				TS	AAACAGCGGGCGCAACCTTACTCC

Table 6-16: Reaction mix for the generation of sgRNA inserts and their ligation to the linearised pX330-sgCAS9-HF vector

Reagent	Final conc./amount
Linearised plasmid DNA	175 ng
Oligonucleotide NTS	1:1 or 1:3 of plasmid:insert ratio
Oligonucleotide TS	1:1 or 1:3 of plasmid:insert ratio
T4 DNA Ligase buffer (x10)	1x
T4 PNK (10 U/µl)	0.5 U
Adjust to 16 µl with H_2O	
T4 DNA Ligase (30 U/µl)	0.07 U

Table 6-17: Programme for the generation of sgRNA inserts and their ligation to the linearised pX330-sgCAS9-HF vector

Step	Temperature	Time	Action
1	4 °C	∞	
2	37 °C	30 min	
3	95 °C	5 min	
4	95→20 °C	Ramp 0.1 °C/s	
5	4 °C	∞	
6			Add ligase
7	20 °C	1 h	
8	65 °C	20 min	
9	4 °C	∞	

To eliminate any rests of original pX330-sgCAS9-HF vector from the ligation mixture an additional BpiI digestion was performed in the third step of the cloning procedure. Due to plasmid design, the BpiI recognition site within the scrambled sgRNA sequence of the original pX330-sgCAS9-HF vector was destroyed in knockout constructs when replacing it with the sgRNA of choice. Hence, BpiI solely linearises unaltered parental constructs, marking them for degradation after *E.coli* transformation.

Table 6-18: Reaction mix for the digestion of remaining parental pX330-sgCAS9-HF vector within the ligation mix with BpiI

Reagent	Final conc./amount
Plasmid DNA	175 ng (complete ligation mix)
Fast digest buffer (x10)	1x
BpiI-HF (1:10)	1.3 µl

Table 6-19: Programme for the digestion of remaining parental pX330-sgCAS9-HF vector within the ligation mix with BpiI

Step	Temperature	Time
1	4 °C	∞
2	37 °C	10 min
3	65 °C	20 min
4	4 °C	∞

In the fourth step of the cloning procedure, the generated sgRNA-CAS9 expression vectors were amplified in transformed *E.coli* SCS-8*.* To verify *E.coli* SCS-8 competence, the bacteria were transformed in parallel with 5 ng of original pX330-sgCAS9-HF. Both parental pX330-SgCas9-HF vector and the derived knockout constructs encode for ampicillin-resistance and enable bacteria growth on the ampicillin-containing agar plates. 25 ng of linearised purified pX330-sgCAS9-HF from step A was used to verify BpiI digestion and efficient removal of the scrambled RNA

sequence. Three *E.coli* clones were selected per knockout construct for further amplification and plasmid DNA was extracted using the GeneElute™ HP Plasmid Miniprep Kit (SIGMA-ALDRICH GMBH) (chapter 6.16) eluting in 50 µl of elution solution. The DNA concentration was measured (chapter 6.11) and sgRNA presence in the knockout constructs was verified by analytical BsrDI/Nb.BsrDI plasmid digestion (chapter 6.17) and Sanger sequencing (chapter 6.18) using the U6 promoter forward primer (chapter 6.4). The sequencing data was analysed using ClustalW software (EMBL-EBI, Cambridgeshire, UK) and the sequence output was aligned with computationally generated target sequences using the blastn online tool (U.S. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, Bethesda, MD, US).

6.22 HeLa transfection with a pair of knockout constructs to induce gene editing

To introduce a DNA glycosylase knockout in the human genome, HeLa cells were transfected with a pair of sgRNA-CAS9 expression vectors generated in chapter 6.21. Simultaneous transfection with both vectors enabled target gene incision at both sites of the active site coding nucleotides by CRISPR-CAS9-mediated gene editing. The generated non-homologous ends could be fused by non-homologous end joining under exclusion of the DNA stretch between the sgRNA sites. Since this DNA stretch contains the active site coding nucleotide, cells with such gene editing events were expected to be depleted of the specific DNA glycosylase activity. Single cell sorting was applied to the transfected cells to select clones with efficient CRISPR-CAS9-mediated gene editing. The selected clones were examined for the desired gene editing events on the genome and protein level and clones with efficient gene editing were expanded to establish the desired DNA glycosylase knockout cell line[281].

Before generating the final knockout cell line, a PCR approach was set to up to detect the desired gene knockout in transfected cells and the editing conditions were optimised in HeLa. To identify the desired gene deletion in the HeLa genome, gene specific PCR primers were designed (chapter 6.26) and PCR conditions were optimised as described in chapter 6.27 to amplify non-rearranged target gene sequences.

Afterwards, HeLa cells were co-transfected with sgRNA-CAS9 expression vector pairs (summarised in Table 6-20) and the editing efficiency was examined by PCR to establish optimal knockout conditions. Different combinations of knockout constructs

were tested for optimal gene editing, starting with the sgRNA combination inducing total gene deletion and active site deletion (Figure 6-2). If these combinations of sgRNA-CAS9 expression constructs did not efficiently induce the desired gene knock out, inter-pair combinations were tested for their gene editing efficiency, with sgRNA pairs targeting a sequence that encloses the active site coding nucleotides.

Table 6-20: Knockout constructs for the induction of CRISPR-CAS9 mediated gene editing events in human cells

Knockout constructs encoding for a gene specific sgRNA and the CAS9 protein used to induce CRISPR-CAS9 mediated gene editing events by combined transfection of human cells.

sgRNA-CAS9 expression vector	Target	Deletion approach	sgRNA	Target site position	Target deletion (bp)
pX330-SpCAS9-HF1_ΔTDG+93	TDG	T	T-I	93	18815
pX330-SpCAS9-HF1_ΔTDG+18908	TDG	T	T-II	18908	
pX330-SpCAS9-HF1_ΔTDG+11012	TDG	A	A-I	11012	5893
pX330-SpCAS9-HF1_ΔTDG+16905	TDG	A	A-II	16905	
pX330-SpCAS9-HF1_ΔMBD4-339	MBD4	T	T-I	-339	8991
pX330-SpCAS9-HF1_ΔMBD4+8652	MBD4	T	T-II	8652	
pX330-SpCAS9-HF1_ΔMBD4+7651	MBD4	A	A-I	7651	828
pX330-SpCAS9-HF1_ΔMBD4+8479	MBD4	A	A-II	8479	
pX330-SpCAS9-HF1_ΔSMUG1-742	SMUG1	T	T-I	-742	2695
pX330-SpCAS9-HF1_ΔSMUG1+1953	SMUG1	T	T-II	1953	
pX330-SpCAS9-HF1_ΔSMUG1+153	SMUG1	A	A-I	153	1393
pX330-SpCAS9-HF1_ΔSMUG1+1546	SMUG1	A	A-II	1546	

HeLa cells were plated to reach an amount of 200,000 (chapter 6.19) exponentially growing cells within one well of a 6-well plate at the day of transfection. The cells were cultivated at 37 °C with a 5% carbon dioxide concentration in the air and 99% humidity in 6-well plates in 2.5 ml of Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% v/v FBS, 2 mM L-glutamine, 110 U/µl penicillin and 110 µg/ml streptomycin (from now on called medium). At the day of transfection, the medium was replaced with 1.5 ml of warm medium and cells were co-transfected with a pair of knockout constructs and EGFP coding pZAJ transfection marker (cf. chapter 6.3), using 3.2 µl enhancer and 5 µl of the effectene transfection reagent from QIAGEN, following manufacturer's instructions. 300 ng of each knockout construct and 50 ng of transfection marker were used to induce the target gene knockout in HeLa cells, whereas untransfected cells and cells transfected with 50 ng of pZAJ only (mock transfected) were used as controls. The cell viability, transfection success and gene editing efficiency was determined at 8-, 24-, 48- and 72-hours after transfection. The cell viability was assessed under the light- and fluorescent microscope by examining

the shape, size and distribution of the transfected HeLa, as well as the number of detached cells. Transfection success was documented as the presence of green fluorescent cells using the EVOS FLoid™ Cell Imaging Station (LIFE TECHNOLOGIES CORPORATION). To analyse the gene editing efficiency, genomic DNA was isolated from transfected and control cells as described in chapter 6.24 and used as templated in a dual PCR approach (chapter 6.27). Non-rearranged and rearranged target genes were amplified using gene specific PCR primers (chapter 6.26) and editing efficiencies were determined as successful amplification of rearranged genes. The time point with optimal gene editing conditions was determined by qualitatively comparing the amplification of edited genes in genomic DNA obtained from cells at different time points after transfection. If no optimal gene editing conditions were found, the transfection conditions can be changed according to producer's instructions, different combinations of knockout constructs can be tested for their gene editing efficiency and as a last resort new targeting sgRNAs can be designed.

After determining the optimal editing conditions, DNA glycosylase knockout cell lines were generated by HeLa transfection with the selected pair of sgRNA-CAS9 expression vectors using the optimal editing conditions. Single cell sorting (chapter 6.23) was applied to the transfected cells at the determined optimal time point after transfection to select cells with efficient gene editing and the gene knockout was verified on the gene and protein level[281].

200,000 exponentially growing HeLa cells in a 6-well plate (chapter 6.19) were transfected with the optimal pair of knockout constructs as described above and mock transfected as well as untransfected HeLa cells were prepared in parallel. 300 ng of each knockout construct and 50 ng of transfection marker were used for transfection, whereas mock transfected cells were transfected with 50 ng of pZAJ only. To enable the sorting of 192-288 single cell clones from the pool of transfected cells, HeLa cells were plated in a way that at the time of sorting, three million cells were available for the control samples. At least seven million transfected cells are needed per gene knockout approach, thus the number of transfected cells was calculated accordingly. Standardly, HeLa cells were sorted one day after transfection, wherefore six wells per gene knockout and three wells per control containing 200,000 cells each needed to be transfected. If the cells were sorted later, the transfected cells were split 24 hours after transfection in a way that ensured exponential growth till the sorting point. In case

transfected cells should be split before sorting, less wells needed to be transfected in the beginning, calculated according to the splitting ratio.

6.23 Single cell sorting and processing of HeLa cells transfected with knockout constructs

Hela cells transfected with sgRNA-CAS9 expression vectors were sorted as single cells and cultivated to validate the desired gene knockout on the gene and protein level. First, transfected cells were washed with PBS, detached via trypsination and resuspended in 500 µl DMEM and samples containing the same constructs were merged. Afterwards, the cells were washed again twice with PBS and the liquid was removed by centrifugation at 7000 g for 1 min. HeLa cells were resuspended in PBS substituted with 0.5 mM EDTA and 1% FBS. The cell concentration (chapter 6.19) was adjusted to five million cells per millilitre and the cell suspension was transferred to a 15 ml tube. The concentrated cells were sorted as single cells according to their EGFP fluorescence using the BD FACSAria™ III Cell Sorter (BECTON DICKINSON AND COMPANY, Franklin Lakes, NJ, US) in the Flow Cytometry Core Facility of the Institute for Molecular Biology Mainz under supervision of scientific staff. Live cell gating and gating of transfected cells by EGFP expression was performed using untransfected and pZAJ transfected cells as calibration samples. For each construct 192-288 living transfected single cell clones with top 5% and top 20% of EGFP expression were sorted into 96-well plates containing 200 µl of medium. Additionally, the remaining cells of the selected population were bulk sorted into tubes containing 2 ml medium and transferred to 25 mm^2 flasks. The bulk sorted cells were expanded till reaching approximately 70% confluency and then analysed for transfection, sorting and gene editing efficiency by PCR (chapter 6.27). Single cell clones were initially expanded in the 96-well plates (only move after 24 hours to allow cell attachment) and were analysed for transfection marker expression as well as cell division speed and colony size one, four and eleven days after sorting and fluorescence images were taken for each knockout construct. If after eleven days, less than 30 single cell clones developed into sufficiently big colonies, the medium of the insufficiently grown clones was replaced for 200 µl of new medium and the cells were expanded further. After developing into colonies of several thousand cells, at least 30 single clones were transferred to 25 mm^2 flasks after cell detachment with 20 µl of trypsin and were cultivated with medium refreshment at intervals of approximately seven days. If the

cells were growing in separate patches, they were transferred to a new 25 mm^2 flask to allow even cell-spreading over their flasks surface. When reaching freezing density (70%-90% of confluency) approximately 14 days after sorting, each single cell clone population was removed from their flask. For each clone the cell suspension was halved and one aliquot was stored in medium supplemented with 10% dimethyl sulfoxide (DMSO) in liquid nitrogen following the standard freezing protocol. The second aliquot was used to prepare “quick” cell lysates (chapter 6.25), to verify the DNA glycosylase knockout on the gene level by PCR.

6.24 Isolation of genomic DNA from HeLa cells for the detection of CRISPR-CAS9-mediated gene editing events

Genomic DNA was isolated from HeLa and HeLa derived knockout cells in order to detect the desired CRISPR-CAS9-mediated gene editing events by PCR. Cells were plated in 6-well plates in a way that they reached 70-80% confluency after an incubation time of 16-24 hours at 37 °C. To isolate the genomic DNA, the cells were washed twice with ice cold PBS and incubated with RNAse containing buffer P1 from the QIAGEN Plasmid Mega Kit (chapter 6.6) for 5 min at room temperature. Afterwards, 40 µl of a 10% SDS solution were added and the plate was tilted gently to ensure an even liquid coverage of the surface. The viscous DNA containing solution was absorbed from the plate using a 1000 µl pipette after cutting the tip ~5 mm from the top. To ensure maximal DNA uptake, the pipette was moved in circles over the plate during absorption, before transferring the absorbed liquid to a 2 ml tube. 20 µl of a 10 mg/ml Proteinase K solution were added to the solution and incubated for at least one hour at 50 °C in a water bath. 1 ml of a Phenol:Chloroform:Isoamyl alcohol solution 25:24:1 was mix vigorously with the DNA containing liquid by vortexing. After centrifuging for 1 min at 13,000 g at room temperature, the mixture separated into two phases and the upper, DNA containing phase was carefully transferred to a new 2 ml tube without disrupting the interphase. 40 µl of a three molar sodium acetate solution (pH 5.2) and 1 ml of ice cold 100% ethanol were added to the solution, mixed vigorously and centrifuged for 1 hour at 13,000 g and room temperature. The pellet was suspended in 900 µl of ice cold 70% ethanol, mixed vigorously and centrifuged for another 20 min. Afterwards, the pellet was air dried for 5 min, resuspended in 40 µl of Tris-HCl (pH 8.0) and the DNA concentration was quantified as described in chapter 6.11 before verifying the DNA glycosylase knockout by PCR (cf. chapter 6.27).

6.25 Preparation of “quick” cell lysates from HeLa cells for the detection of CRISPR-CAS9-mediated gene editing events

“Quick” cell lysates were prepared from HeLa derived single cell clones and the respective untransfected control cells to detect CRISPR-CAS9-mediated gene editing events by PCR. Aliquoted cells (c.f. chapter 6.23) were washed twice in PBS and resuspended in 20 µl of double distilled water. The samples were incubated at 95 °C for 30 min on the heating block in 1.5 ml tubes and centrifuged for 1 min at 1000 g. The supernatant was transferred into a new 1.5 ml tube and used as a template for PCR reactions. The preparation of cell lysates as PCR templates is advantageous over the classical extraction of genomic DNA, as the described “quick” cell lysate procedure spares time and resources when analysing big sample amounts. In addition to the cell lysates of single cell clones, genomic DNA (chapter 6.24) and cell lysates of transfected unsorted cells, transfected bulk sorted cells and untransfected cells were used as PCR controls.

6.26 Design of PCR primers for the detection of CRISPR-CAS9-mediated gene editing events

CRISPR-CAS9-mediated gene editing of selected DNA glycosylases in human HeLa cells was verified on the gene level by PCR. Following the previously described PCR approach[281], a PCR primer pair was designed for each sgRNA aiming to amplify the sgRNA target site in the gene of choice (e.g. primer pair TDG T-I forward and TDG T-I reverse, for amplification of the sgRNA T-I target site; Table 6-21) as depicted in Figure 6-5. Importantly, the primers were design to also detect the desired gene editing events using mixed primer pairs. Therefore, the forward primer for the amplification of sgRNA-I and the reverse primer for the amplification of sgRNA-II were designed to be compatible (e.g. primers TDG T-I forward and TDG T-II reverse; Table 6-21). In unedited target genes, the distance between the primers of such combined pairs would most of the times be too long to efficiently yield a PCR product. If, however, the target gene was simultaneously cut at both sgRNA sites and non-homologous end-joining took place, the resulting deletion would bring the primers in much closer proximity, enabling PCR amplification. The primers used in this study were designed to yield PCR products sized between 150 and 1100 base pairs, as predicted based on the edited genomic DNA sequences.

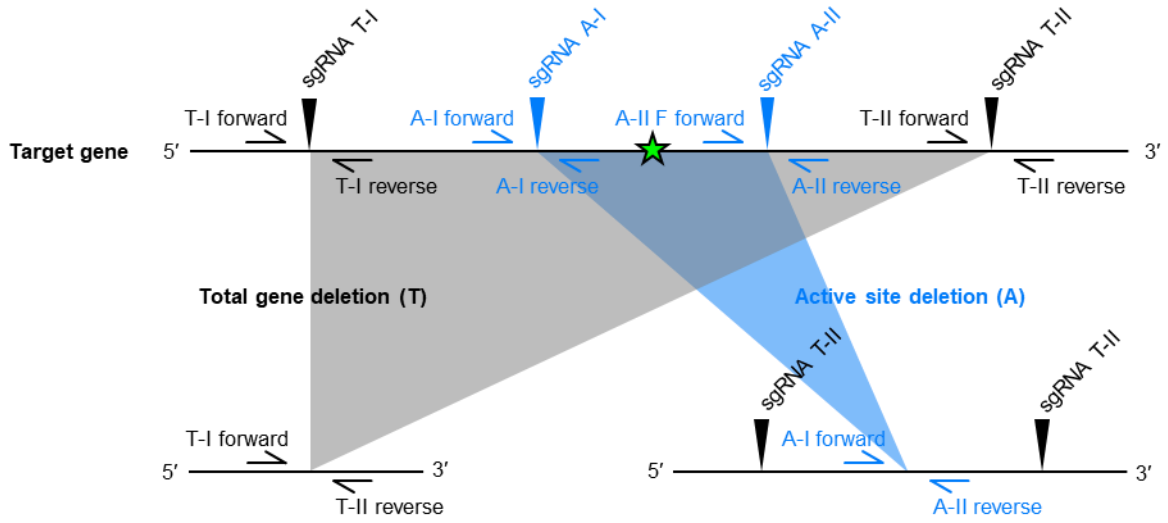


Figure 6-5: Position of PCR primers used to detect non-rearranged target genes as well as total gene deletions and active site deletions by PCR

Schematic representation of the PCR primers used to detect gene editing events of the target gene by CRISPR-CAS9 mediated total gene deletion and active site deletion. Scheme of the non-rearranged target gene (upper panel) and rearranged target gene after total gene deletion (lower left panel) or active site deletion (lower right panel): gene coding region (black line), active site coding nucleotides (green star), sgRNA targets for total gene deletion (black triangle), sgRNA for active site deletion (blue triangle), PCR primer for total gene deletion (black half arrows), PCR primer for active site deletion (black half arrows).

PCR primer pairs were designed using Primer Designer 4 V4.1 (Scientific & Educational Software, Cary NC, USA) and selected based on their GC-content, annealing temperature, repetition- and dimerisation-probability. If no optimal PCR primer pairs was found using default software settings, the parameters for primer design were altered by increasing the melting temperature range, miss-pairing rate and primer length. After verifying the inter-pair compatibility of the primers according to the upper parameters (forward of sgRNA-I and reverse of sgRNA-II), the primers were purchased from EUROFINS GENOMICS GMBH (Ebersberg) as PCR primers with HPSF grade. Table 6-21 lists the DNA glycosylase specific primers, showing the purchased oligonucleotides in pairs according to sgRNA site amplification. The target gene is depicted in column 1, whilst the sgRNA site which is amplified by the primer pair is indicated in column 2 with distance from the gene's TSS indicated in column 3. Optimal annealing temperatures are listed in column 6 and the expected product length of the sgRNA specific pairs is calculated in column 7.

Table 6-21: PCR primer for the detection of the detection of CRISPR-CAS9 mediated gene editing events

PCR primer were purchased by Eurofins Genomics GmbH (Ebersberg) as HPSF grade.

Target (sgRNA)		Distance from TSS	Primer	Sequence (5-3)	T_{Anneal}	Product length
TDG	T-I	-29	Forward	TCTTACCGCAGTGAGTACCA	60°C	423 bp
		414	Reverse	TGGCATCCAGAAAGACACAT	60°C	
	T-II	18598	Forward	TCACTTCCTGACTTGGTAAT	60°C	797 bp
		19395	Reverse	TGCAGAAGTGCCAGAGTAGA	60°C	
	A-I	10818	Forward	TCCTCTGTAATCCACTCTAA	60°C	698 bp
		11516	Reverse	ATGTCCCTACTCTGATCTTT	60°C	
	A-II	16227	Forward	TAGGGCAACTGATAGTAATG	50°C	849 bp
		17076	Reverse	AGCTCAGCTTGAACTAGATA	50°C	
MBD4	T-I	-409	Forward	CACTTTGGCTACCTGCGTTA	60°C	520 bp
		111	Reverse	AACTTACCGGAGGTCATTCG	60°C	
	T-II	8479	Forward	CAATGGTGGATCTTTGCTAC	65 °C	450 bp
		8929	Reverse	AGTGTGGAGCTGTCAACAAT	65 °C	
	A-I	7172	Forward	TTGGGAGGGTGTCTTTAGAA	65 °C	855 bp
		8027	Reverse	GTGGCGATAACATGAGTCAA	65 °C	
	A-II	7991	Forward	CCACCCCACATTCTAAGTCA	65 °C	740 bp
		8731	Reverse	CTCAGCCTTCCGAGATTACA	65 °C	
SMUG1	T-I	-865	Forward	GGTTGTGTAGCTCGTAAGAT	55°C	287 bp
		-579	Reverse	GTATTACTGCTCCCCTGTTA	55°C	
	T-II	1872	Forward	ATTCAAGACCTCGAAGTCAT	60°C	371 bp
		2243	Reverse	GTATCCTGGCAAGATATTTC	60°C	
	A-I	-158	Forward	GGTGGTTGGTAGATGACTGA	65 °C	740 bp
		582	Reverse	ACAGGGGAGATCCAGTAAAG	65 °C	
	A-II	1424	Forward	CCCACAGTCAGAAGTGAGTG	65 °C	811 bp
		2243	Reverse	GTATCCTGGCAAGATATTTC	65 °C	

6.27 Validation of CRISPR-CAS9-mediated gene editing events in HeLa cells by polymerase chain reaction

The polymerase chain reaction was used to qualitatively verify the presence of non-rearranged and rearranged CRISPR-CAS9 target genes in extracted genomic DNA (chapter 6.24) or “quick” cell lysates (chapter 6.25) of HeLa cells transfected with a pair of sgRNA-CAS9 expression vectors or parental control cells. The PCR reactions were performed in a thermal cycler with the primers designed in chapter 6.26. As schematically depicted in Figure 6-5, primer pairs closely flanking an sgRNA site (e.g. T-I forward and T-I reverse) were used to detect non-rearranged copies of the target gene. Rearranged target gene copies were detected using the forward primer for sgRNA-I combined with the reverse primer for sgRNA-II (e.g. T-I forward and T-II reverse). The DNA template was added to the lid of the PCR tube and only mixed with

other solutions, which were applied into the tube, by centrifugation immediately before starting the reaction. The S7 Fusion Polymerase from BIOZYME was standardly used for all PCR reactions. As an exception, PCR reactions performed to establish the optimal annealing temperature of PCR primers were performed with the cheaper *Taq* DNA Polymerase from NEW ENGLAND BIOLABS. Usage of *Taq* DNA Polymerase for any PCR is indicated in the figure legends of the corresponding experiments.

For both enzymes, a table with reaction mixtures and PCR programmes is depicted below and the optimal annealing temperatures (T_{Anneal}) of PCR primers are listed in Table 6-21, page 69. If new primer combinations should be used, a pilot PCR was performed to establish find optimal product amplification conditions. Several sample preparations were incubated at different annealing temperatures in the same reaction using an annealing temperature gradient between 55-65 °C.

Table 6-22: Reaction mix for the amplification of a selected DNA sequence by PCR using S7 Fusion polymerase

Reagent	Final conc./amount
DNA template	10 ng (genomic DNA) 3 µl (cell lysate for PCR)
HF buffer (x5)	1x
dNTPs (10 mM)	200 µM
Forward primer (10 mM)	0.5 µM
Reverse primer (10 mM)	0.5 µM
S7 Fusion polymerase (2 U/µl)	0.4 U
Adjust to 20 µl with H_2O	

Table 6-23: Programme for the amplification of a selected DNA sequence by PCR using S7 Fusion polymerase

Step	Temperature	Time	Action
1	4 °C	∞	
2	98 °C	2 min	
3	98 °C	20 s	Repeat steps 3 to 5 for 34 times
4	T_{Anneal}	30 s	
5	72 °C	30 s	
6	72 °C	5 min	
7	4 °C	∞	

Table 6-24: Reaction mix for the amplification of a selected DNA sequence by PCR using Taq DNA polymerase

Reagent	Final conc./amount
DNA template	10 ng (genomic DNA) 0.2 ng (plasmid DNA)
Thermopol. reaction buffer (x10)	1x
dNTPs (10 mM)	200 µM
Forward primer (10 mM)	0.2 µM
Reverse primer (10 mM)	0.2 µM
Taq DNA polymerase (5 U/µl)	2 U
Adjust to 25 µl with H_2O	

Table 6-25: Programme for the amplification of a selected DNA sequence by PCR using *Taq* DNA polymerase

Step	Temperature	Time	Action
1	4 °C	∞	
2	95 °C	5 min	
3	95 °C	30 s	Repeat steps 3 to 5 for 34 times
4	T_{Anneal}	30 s	
5	72 °C	1 min	
6	68 °C	2 min	
7	4 °C	∞	

PCR products were analysed by agarose gel electrophoresis (chapter 6.12) and single cell clones containing rearranged genes yet absent non-rearranged genes according to PCR results were analysed for target protein presence by classical western blot (WB) analysis (chapter 6.28).

6.28 Immunodetection of proteins in extracts of HeLa cells by western blot analysis

Western blot analysis was used to validate CRISPR-CAS9-mediated gene editing in single cell clones with potential gene knockout as indicated by PCR (chapter 6.27). Thus, the presence of the target protein within whole cell extracts of the single cell clones was verified by western blot analysis. Preparation of western blot buffers is described in chapter 6.10. The parental cell line was used as a control as well as a knockdown and/or overexpression cell line of the desired gene if available.

To generate whole cell extracts, the selected HeLa and HeLa derived knockdown, knockout or overexpression cells were cultivated in a 75 mm^2 flask till reaching ~70% confluency. The cells were washed twice with 20 ml of ice-cold PBS and incubated with 10 ml of PBSCMF for 1-3 minutes. Afterwards, the cells were scratched from the

surface, suspended in PBS supplemented with 0.5 mM PMSF (chapter 6.10) and centrifuged for 5 min at 4000 g and 4 °C. The supernatant was thoroughly removed and the pellet was either directly used to prepare cell extracts or frozen in liquid nitrogen and stored at -80 °C for later usage. One fifth of the pellet was incubated with ~100 µl of NET-N buffer for 30 min on ice in a 1.5 ml tube. The suspension was sonicated with the UP200Ht - Handheld Ultrasonic Homogenizer (HIELSCHER ULTRASONICS GMBH) for 10 seconds with an amplitude of 40, 10 pulses and C 10% followed by centrifugation at 13,000 rpm at 4 °C for 25 min. The supernatant was transferred to a fresh tube and either stored at -80 °C or directly used to determine the protein concentration using classical Bradford assay.

A fraction of the cell lysate was diluted by a factor of 10, whilst a dilution row of the bovine serum albumin (BSA) standard was generated ranging from 25 µg/ml to 2.5 µg/ml. 10 µl of each BSA dilution and cell extract sample was plated in duplicates in a 96-well plate, 200 ml of Bradford reagent were added and the absorption of the samples at 595 nm was measured using TriStar² LB 942 Multimode Plate Reader (BERTHOLD TECHNOLOGIES). After generating the BSA concentration standard curve, the protein concentration of each cell lysate sample was calculated from the OD_{595} values using standard curve.

The proteins within the cell extracts were separated by size using an SDS-Polyacrylamide gel electrophoresis (SDS-PAGE). All cell extracts that should be compared for the presence of the selected protein were adjusted to the highest possible common concentration. 40 µg of each sample were incubated with the lämmli protein buffer (6x) for 5 min at 95 °C. The samples and the PageRuler™ Plus Prestained Protein Ladder (THERMO FISHER SCIENTIFIC INC) were loaded on a 10% sodium-dodecyl-sulphate-polyacrylamide gel as described in Table 6-26 and electrophoresis was performed for 15 min at 95 V in the stocking gel and 90 min at 125 V in the separation gel in lämmli buffer.

Table 6-26: Reaction mix for the preparation of sodium dodecyl sulphate polyacrylamide gels for SDS-PAGE

Reagent	Separation gel	Collection gel
Double distilled water	5.7 ml	4.4 ml
1.5 M Tris pH 8.8	3 ml	-
1.5 M Tris pH 6.8	-	760 µl
10% SDS	120 µl	60 µl
Acrylamide	3 ml	760 µl
10% APS	60 µl	60 µl
TEMED	6 µl	6 µl

The size separated proteins in the polyacrylamide gel were then transferred to a nitrocellulose membrane by western blot to enable selective protein detection. The nitrocellulose membrane was equilibrated prior to running in double distilled water, whilst six papers of blotting paper and the synthetic sponge were equilibrated in blotting buffer. The blotting chamber was assembled by stacking the synthetic sponge, followed by three blotting papers, the acrylamide gel, the nitrocellulose membrane and three blotting papers finishing with a synthetic sponge. The proteins were transferred from the polyacrylamide gel to the nitrocellulose membrane by applying 300 mA for 2 hours in the blotting chamber filled with blotting buffer.

To detect the protein of choice on the nitrocellulose membrane harbouring the size separated proteins of the cell extracts, the membrane was washed in TBS-T (chapter 6.10) for 5 min at room temperature. Afterwards, the membrane was blocked applying 5%TM/TBS-T for 1 hour and washed again trice with TBS-T for 5 min each. To detect the protein of choice, the membrane was incubated with a protein specific primary antibody diluted in TBS-T supplied with 5% of milk and 5% BSA in a 50 ml tube for ~18 hours at 4 °C under constant rolling. The membrane was washed trice with TBS-T and incubated with the secondary antibody diluted in TBS-T for 1.5 hours at room temperature in an optically opaque 50 ml tube under constant rolling. The membrane was washed trice again under exclusion of light and the fluorescent signal of the secondary antibody and the protein ladder were detected according to manufacturer's instructions in separate channels of LI-COR Odyssey 9120 infrared imaging system (LI-COR INC.). If necessary, the membrane was used again to detect a different protein e.g. loading control by repetition of protein detection using another primary antibody from a different organismic origin after a washing step of three minutes with TBS-T. The membrane was dried and stored at room temperature.

Based on PCR and WB results, two single cell clones per gene knockout approach were selected, expanded and used for future experiments.

6.29 Preparation of whole cell extracts for DNA cleavage assays

Cell extracts were prepared from HeLa and HeLa derived cell lines to quantify the enzyme activity of selected proteins by plasmid cleavage assays. The cell extracts were prepared from exponentially growing HeLa cells, cultivated in a 75 mm^2 flask at 37 °C till reaching ~70% confluency. From now on, the cells were kept on ice at all times and the work was performed on the bench. Preparation of the buffers used in

this procedure are described in chapter 6.10. After washing the cells twice with 20 ml of ice-cold PBS, they were incubated with 10 ml of PBS supplemented with 0.5 mM PMSF for 1-3 minutes. Next, the cells were scratched from the surface, suspended in PBS supplemented with 0.5 mM PMSF and cells from the same origin were merged in a 50 ml tube. The samples were centrifuged for 5 min at 4,000 g at 4 °C, the pellet was suspended in 500 µl of lysis buffer and the sample was transferred to a 1.5 ml screw cap tube. The suspension was sonicated on ice slurry with the UP200Ht - Handheld Ultrasonic Homogenizer (HIELSCHER ULTRASONICS) with two pulses for 40 seconds with a 60 s interval at 10% power settings and an amplitude of 20, and C 40%. Afterwards, the suspension was centrifuged twice at 21,000 g and 4 °C for 25 min in new 2 ml tubes and the supernatant was split into 1.5 ml tubes each containing 50 µl of cell extract and stored at -80 °C. Either directly or after freezing, one sample was used to determine the protein concentration of the cell extract based on the A280 absorbance determined by spectrophotometry (cf. chapter 6.11) using the lysis buffer as blank.

6.30 Site and sequence specific incorporation of single DNA modifications into EGFP reporters

EGFP expression constructs carrying a single DNA modification within the CRE or GC box gene regulatory element were generated by the strand exchange protocol [277] as depicted in Figure 6-6. In summary, the CRE and GC box reporters were site specifically nicked by Nb.BsrDI nicking endonuclease, which due to the design of the enzyme only incises one DNA strand. The vectors contain two Nb.BsrDI specific nicking sites in the promoter sequence, which flank the CRE or GC box gene regulatory element selected to contain the DNA modification of choice. The Nb.BsrDI sites have same the orientation and are 18 nts apart from each other, thus Nb.BsrDI treatment results in the excision of a single stranded DNA fragment of 18 nts. The excised single stranded DNA fragment was exchanged for a complementary synthetic oligonucleotide containing the modified base of choice. Analytical digestion was used to verify the presence of the selected DNA modification, thus confirming the successful generation of a reporters containing a single DNA modification of choice.

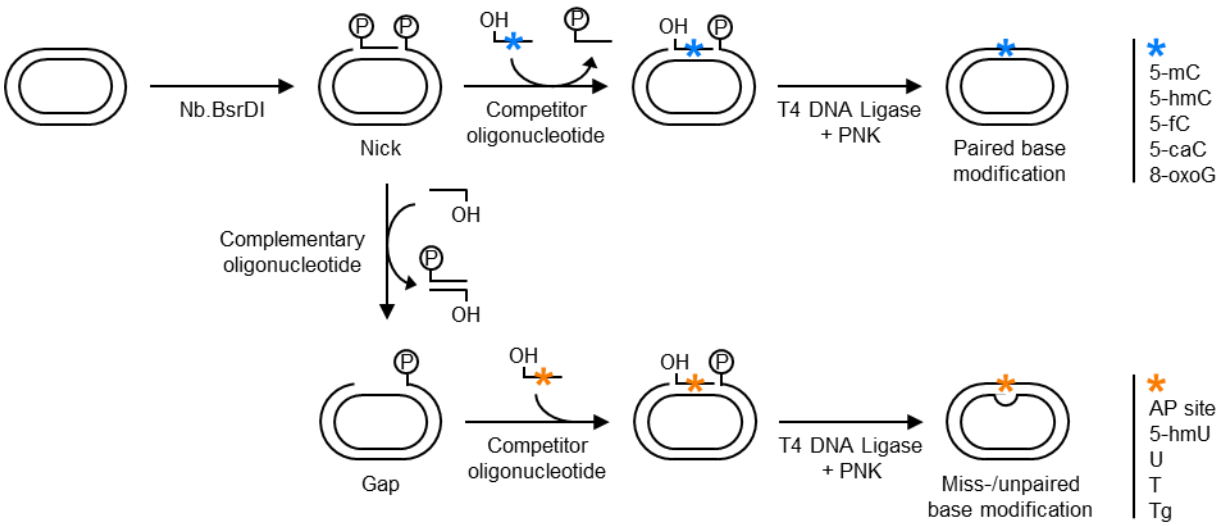


Figure 6-6: Generation of reporters containing a single DNA modification of choice

Schematic representation of the site and sequence specific introduction of a single DNA modification into an EGFP reporter using the strand exchange method[277], showing the participating enzymes and intermediate structures. The parental vector is treated with the nicking endonuclease Nb.BsrDI generating double nicked vectors. If a pairing DNA modification should be introduced, the excised DNA fragment is directly exchanged for a synthetic competitor oligonucleotide containing the modification of choice, which is phosphorylated by PNK and ligated to the vector by T4 DNA Ligase (upper panel). If a miss- or unpaired DNA modification should be introduced, the nicked vector is incubated with complementary oligonucleotides, which bind to the excised DNA fragment, enabling its removal by chromatography before annealing and ligating the competitor oligonucleotide (lower panel). Base modifications that were introduced into reporter DNA by the plasmid nicking or gapping procedure are indicated to the right.

First, the parental vectors targeted to introduce the DNA modification of choice (pCMV-III-ZA-W, pCRE-UNO-ZA-W, pCRE-UNO-ZA-C, pGCbox-ZA-W and pGCbox-ZA-C) were nicked by the nicking endonuclease Nb.BsrDI. Nb.BsrDI introduced a nick at the two recognition sites within the promoter sequence of the vector DNA, thus excising a 18 nt long single stranded DNA fragment. Based on the vector design the nicking took place either in the NTS (-W vectors) or the TS (-C vectors) of the DNA. The protocol and reaction components are presented in the tables below.

Table 6-27: Reaction mix for the nicking of CMV-1111, CRE-UNO and GC box reporters by Nb.BsrDI

Reagent	Final conc./amount
Plasmid DNA	100 µg
Cut smart buffer (x10)	1x
Nb.BsrDI (10 U/µl)	150 U
Adjust to 500 µl with H_2O, split into 100 µl	

Table 6-28: Programme for the nicking of CMV-1111, CRE-UNO and GC box reporters by Nb.BsrDI

Step	Temperature	Time
1	4 °C	∞
2	65 °C	2 h
5	80 °C	20 min
6	4 °C	∞

The nicking efficiency by Nb.BsrDI and the ligation efficiency of the synthetic oligonucleotides with the nicked vector was analysed by analytical ligation. 200 ng of nicked vector DNA was incubated with synthetic oligonucleotides, which were complementary to the single stranded DNA sequence of the reporter and contain the modification of choice. Synthetic oligonucleotides used for the introduction of DNA modifications into CMV-1111, CRE and GC box reporters are listed in Table 6-29. Target reporters are indicated in column 1, the specific DNA modification in column 2, their distance from TSS in column 3, the targeted DNA strand for the oligonucleotide incorporation in column 4, the DNA sequence in column 5 and the purchaser in column 5. Presence of the desired base modification was verified by the purchaser via mass spectrometry using matrix-assisted laser desorption/ionization (MALDI) and time-of-flight (TOF) mass analyser and the analysis output was purchased with material
Simultaneous to strand annealing, the synthetic oligonucleotides were phosphorylated at the 5' end by T4 PNK activating the ends for the subsequent ligation by T4 DNA Ligase. The reactions were performed in a thermocycler with reaction components and protocol shown in the Table 6-30 and Table 6-31.

Table 6-29: Oligonucleotides for the introduction of single DNA modifications

Oligonucleotides containing a single DNA modification and their unmodified counterparts (controls) were used for the site and sequence specific introduction of a single DNA modification into EGFP reporter vectors. The oligonucleotides were purchased as HPLC purified grade by Eurofins Genomics GmbH (Ebersberg), BIOSPRING GMBH (Frankfurt am Main), KANEKA EUROGENTEC S.A. (Seraing, Belgium) or GENE LINK INC. (Orlando, FL, US) or were kindly prepared by our collaboration partners in the lab of Thomas Carell in Munich.

Target vector	Modification	Distance from CpG	Strand	Sequence (5-3)	Provider
pCRE UNO-ZA-W	-	-	NTS	CATTGCGTGACGTCAGCG	Euro
pCRE UNO-ZA-W	5-fC	-1	NTS	CATTGCGTGA5-fCGTCAGCG	Carell lab
pCRE UNO-ZA-W	F-fC	-1	NTS	CATTGCGTGAF-fCGTCAGCG	Carell lab
pCRE UNO-ZA-W	5-caC	-1	NTS	CATTGCGTGA5-caCGTCAGCG	Carell lab
pCRE UNO-ZA-W	F-caC	-1	NTS	CATTGCGTGAF-caCGTCAGCG	Carell lab
pCRE UNO-ZA-W	5-hmU	-1	NTS	CATTGCGTGA5-hmUGTCAGCG	E-tec
pCRE UNO-ZA-W	T	-1	NTS	CATTGCGTGATGTCAGCG	Euro
pCRE UNO-ZA-W	Tg	-1	NTS	CATTGCGTGATgGTCAGCG	Gene
pCRE UNO-ZA-W	U	-1	NTS	CATTGCGTGAUGTCAGCG	Euro
pCRE UNO-ZA-W	F	-1	NTS	CATTGCGTGAFGTCAGCG	Bio
pCRE UNO-ZA-W	SF	-1	NTS	CATTGCGTGASFGTCAGCG	Bio
pGCbox-ZA-W	-	-	NTS	CATTGCATGGGCGGAGCG	Euro
pGCbox-ZA-W	5-mC	-1	NTS	CATTGCATGGG5-mCGGAGCG	Euro
pGCbox-ZA-W	5-hmC	-1	NTS	CATTGCATGGG5-hmCGGAGCG	Carell lab
pGCbox-ZA-W	5-fC	-1	NTS	CATTGCATGGG5-fCGGAGCG	Carell lab
pGCbox-ZA-W	F-fC	-1	NTS	CATTGCATGGGF-fCGGAGCG	Carell lab
pGCbox-ZA-W	5-caC	-1	NTS	CATTGCATGGG5-caCGGAGCG	Carell lab
pGCbox-ZA-W	F-caC	-1	NTS	CATTGCATGGGF-caCGGAGCG	Carell lab
pGCbox-ZA-W	8oG	-3	NTS	CATTGCATG8oGGCGGAGCG	Bio
pGCbox-ZA-W	8oG	-2	NTS	CATTGCATGG8oGCGGAGCG	Bio
pGCbox-ZA-W	8oG	+1	NTS	CATTGCATGGGC8oGGAGCG	Bio
pGCbox-ZA-W	F	-1	NTS	CATTGCATGGGFGGAGCG	Bio
pGCbox-ZA-W	SF	-1	NTS	CATTGCATGGGSFGGAGCG	Bio
pGCbox-ZA-W	F	+1	NTS	CATTGCATGGGCFGAGCG	Bio
pGCbox-ZA-W	SF	+1	NTS	CATTGCATGGGCSFGAGCG	Bio
pGCbox-ZA-C	-	-	NTS	CGTGGGCGGAGCGCAATG	Euro
pGCbox-ZA-C	5-mC	+1	TS	CATTGCGCTC5-mCGCCCACG	Euro
pGCbox-ZA-C	5-hmC	+1	TS	CATTGCGCTC5-mCGCCCACG	Carell lab
pGCbox-ZA-C	5-fC	+1	TS	CATTGCGCTC5-fCGCCCACG	Carell lab
pGCbox-ZA-C	F-fC	+1	TS	CATTGCGCTCF-fCGCCCACG	Carell lab
pGCbox-ZA-C	5-caC	+1	TS	CATTGCGCTC5-caCGCCCACG	Carell lab
pGCbox-ZA-C	F-caC	+1	TS	CATTGCGCTCF-caCGCCCACG	Carell lab
pGCbox-ZA-C	8oG	-1	TS	CATTGCGCTCC8oGCCCACG	Bio
pGCbox-ZA-C	F	-1	TS	CATTGCGCTCCFCCCACG	Bio
pGCbox-ZA-C	SF	-1	TS	CATTGCGCTCCSFCCCACG	Bio
pGCbox-ZA-C	F	+1	TS	CATTGCGCTCFGCCCACG	Bio
pGCbox-ZA-C	SF	+1	TS	CATTGCGCTCSFGCCCACG	Bio

Eurofins Genomics GmbH (Euro), BioSpring GmbH (Bio), Kaneka Eurogentec S.A. (E-tec), Gene Link Inc. (Gene), Lab of Thomas Carell (Carell lab)

Table 6-30: Reaction mix for the analytical ligation of synthetic oligonucleotides with nicked CMV-1111, CRE-UNO or GC box reporters by T4 DNA Ligase

Reagent	Final conc./amount			
	- Ligase	**+Ligase**	**+Oligo**	**+Oligo +PNK**
Nicked plasmid DNA (200 ng/µl)	200 ng	200 ng	200 ng	200 ng
T4 DNA Ligase buffer (x10)	1x	1x	1x	1x
Oligonucleotide (2 µM)	-	-	1.5 µM	1.5 µM
T4 PNK (10 U/µl)	-	-	-	5 U
Adjust to 15 µl with H_2O → annealing + phosphorylation				
T4 DNA Ligase (30 U/µl)	-	2 U	2 U	2 U
	(2 U in 5 µl of 1x buffer, added to each 15 µl reaction)			

Table 6-31: Programme for the analytical ligation of synthetic oligonucleotides with nicked CMV-1111, CRE-UNO or GC box reporters by T4 DNA Ligase

Step	Temperature	Time	Action
1	4 °C	∞	
2	37 °C	30 min	
3	80 °C	10 min	
4	4 °C	∞	
			Add ligase
5	37 °C	1 h	
6	65 °C	15 min	
6	4 °C	∞	

The products of the analytical ligation were analysed by agarose gel electrophoresis (cf. chapter 6.12). Afterwards, the nicked vectors were either directly used for large scale ligation (preparative ligation) or were used to generate gapped vectors.

The additional step of vector gapping was performed in case a miss- or unpaired base pair should be generated in reporters with DNA modifications such as AP site, 5-hmU, U, T or Tg. In these cases, the nicked vector DNA was incubated with a 46-times excess of complementary oligonucleotide which due to sequence complementarity anneals with the excised 18 nt DNA fragment, outcompeting vector binding and thereby enables the removal of the excised native DNA from the reaction mix. Complementary oligonucleotides for the generation of gapped pCRE-UNO-ZA-W, pGCbox-ZA-W and pGCbox-ZA-C plasmid DNA are listed in Table 6-32, with the target vector indicated in column 1, the strand the oligonucleotide is derived from in column 2 and the excised DNA strand the oligonucleotide is complementary to indicated in column 4. The reactions were performed in a thermocycler with reaction components and protocol shown in Table 6-33 and Table 6-34.

Table 6-32: Complementary oligonucleotides for the generation of reporters containing a miss- or unpaired base pair

Complementary oligonucleotides were used to perform vector gapping by trapping the excised DNA fragment from nicked CRE- and GC box reporters by annealing due to sequence complementarity. The complementary oligonucleotides were generated as HPLC purified grade and were purchased by Eurofins Genomics GmbH (Ebersberg).

Target vector	Strand	Sequence (5-3)	Complementary to
pCRE-UNO-ZA-W	TS	CGCTGACGTCACGCAATG	NTS
pGCbox-ZA-W	TS	CGCTCCGCCCATGCAATG	TS
pGCbox-ZA-C	NTS	CGTGGGCGGAGCGCAATG	TS

Table 6-33: Reaction mix for the gapping of nicked CRE-UNO or GC box reporters

Reagent	Final conc./amount
Nicked plasmid DNA	10 µg
T4 DNA Ligase buffer (x10)	1x
Competitor oligonucleotide (200 µM)	5.57 µM
Adjust to 500 µl with H_2O, split into 100 µl	

Table 6-34: Programme for the gapping of nicked CRE-UNO or GC box reporters

Step	Temperature	Time	Action
1	4 °C	∞	
2	37 °C	30 min	
3	80 °C	10 min	
6	4 °C	∞	

Scavenging of excised DNA fragments by complementary oligonucleotides was removed from the solution using Amicon® Ultra-0.5 Centrifugal Filter Devices 30k (MERCK KGAA, Darmstadt) rendering gapped plasmid DNA. A second analytical ligation was performed (as described above) to verify efficient vector gapping before performing preparative ligation reactions.

Nicked or gapped vector DNA was used for preparative ligation reactions and the same control samples (directly removed from ligation master mix) were included as described for analytical ligation reactions. A lower concentration of synthetic oligonucleotide was used for gapped vector DNA since the ligation efficiency of this DNA was higher than for nicked vector DNA. Preparative ligation reactions were performed in a thermocycler with reaction components shown in the tables below and the incubation protocol as shown for analytical ligations (cf. Table 6-30).

Table 6-35: Reaction mix for the preparative ligation of synthetic oligonucleotides with nicked CRE-UNO or GC box reporters by T4 DNA Ligase

Reagent	Final conc./amount	
	Nicked DNA	**Gapped DNA**
Plasmid DNA	8 µg	8 µg
T4 DNA Ligase buffer (x10)	1x	1x
Oligonucleotide (200 µM)	1.5 µM (0.055 pmol/ng)	0.39 µM (0.0144 pmol/ng)
T4 PNK (10 U/µl)	100 U	100 U
Adjust to 300 µl with H_2O, split into 75 µl → annealing + phosphorylation		
T4 DNA Ligase	2 U (10 U in 25 µl of 1x buffer, added to each 75 µl reaction)	

Samples of preparative ligation reactions with the same content were pooled in a 1.5 ml Eppendorf tube and a portion of each sample was analysed for nicking, phosphorylation and ligation efficiency by agarose gel electrophoresis (cf. chapter 6.12). Afterwards, the ligated samples were purified using Amicon® Ultra-0.5 Centrifugal Filter Devices 30k from MERCK KGAA (Darmstadt) and the presence of the desired DNA modification was analysed by analytical digestion (cf. chapter 6.17).

6.31 Quantification of the excision efficiency of plasmid DNA containing a single 8-oxoG or apurinic site

Quantitative cleavage assays were performed to analyse the enzymatic activity of pure human OGG1, APE1 or whole cell extracts (cf. chapter 6.29) towards reporters containing a single 8-oxoG or apurinic site respectively at different positions within the GC box gene regulatory element. The method was based on the conversion of circular closed plasmid DNA into its open circular form by DNA strand incision at AP sites, which were either directly introduced into the DNA (F) or generated as a repair intermediate following removal of the 8-oxoG base.

In addition to the analysis of the cell extract cleavage activity, 8-oxoG containing plasmids were incubated with pure hOGG1 in the presence or absence of APE1, whereas F containing plasmids were only incubated with APE1. Reactions containing cell extract were performed in magnesium-free BEH buffer (chapter 6.10) supplemented with 0.1 mg/mL BSA (BEH-BSA), whereas reactions containing pure enzymes were standardly performed in NEBuffer4. Cleavage assays with APE1 only were also performed in BEH-BSA. Whenever predilutions of the enzymes were necessary, hOGG1 was diluted in NEBuffer2 and APE1 was diluted in NEBuffer4. The cleavage samples were treated stepwise so that incubation times for each reaction were exactly the same and the reaction was stopped by the addition of SDS containing

loading dye and heat-inactivation. The reaction components and incubation times are shown in the tables below.

Table 6-36: Reaction mix for the quantitative cleavage of modified reporters by pure OGG1, APE1 or cell extracts

	8-oxoG		AP site		8-oxoG + AP site
Reagent	**hOGG1**	**hOGG1+APE1**	**NEBuffer4**	**BEH BSA**	**Cell extract**
Plasmid DNA	150 ng	150 ng	150 ng	150 ng	100 ng
NEBuffer4 (x10)	1x	1x	1x	-	-
BEH-BSA	-	-	-	1x	1x
hOGG1 (1.6 U/µl)	various	various	-	-	-
APE1 (10 U/µl)	-	0.01 U/ng	Various	Various	-
Cell extract	-	-	-	-	Various
Adjust to 15 µl with H_2O					

Table 6-37: Programme for the quantitative cleavage of modified reporters by pure OGG1, APE1 or cell extracts

hOGG1		
Step	**Temp.**	**Time**
1	37 °C	50 min
2	Add 4 µl loading dye	
3	65 °C	20 min
4	4 °C	∞

APE		
Step	**Temp.**	**Time**
1	37 °C	1 h
2	Add 4 µl loading dye	
3	65 °C	20 min
4	4 °C	∞

Cell extract		
Step	**Temp.**	**Time**
1	37 °C	1 h
2	Add 4 µl loading dye	
3	4 °C	∞

The amount of open and closed circular plasmid DNA in each sample was analysed by agarose gel electrophoresis (cf. chapter 6.12) and quantified with Image Lab™. The difference of oc-converted plasmid in control versus modified constructs indicates the glycosylase activity towards the modifications of choice in reporter DNA.

6.32 Host cell reactivation assay

The expression of a plasmid-borne EGFP gene containing a selected DNA modification or the respective control in CMV-1111, CRE or GC box gene regulatory element was quantified in transiently transfected HeLa cells by host cell reactivation assay (HCR).

In the first step, HeLa cells were transfected with covalently closed plasmid DNA using Effectene® transfection reagent from QIAGEN. One day prior to transfection 150 000-450,000 cells - as determined by automated cell counting (chapter 6.19) - were plated in a 6-well plate in 2.5 ml of medium and were incubated for 16-24 hours as described in chapter 6.22 paragraph 4 till reaching 70-80% confluency. The medium was replaced with 1.5 ml of warm new medium and the cells were co-transfected with

400 ng of the respective EGFP reporter (CMV-1111, CRE, GC box or TRE-reporters) and 400 ng of transfection marker pDsRed Monomer N1 as described previously by this lab[278]. 3 µl of enhancer and 4 µl of effectene were used for standard transfection of all human cell lines. Transfected cells were cultivated further till harvested at either one or several time points. If gene expression should be determined at several time points, the cells were split 6 hours after transfection in a way that they grow exponentially till the time of fixation.

After the desired amount of time, the cells of each well were washed with 2 ml of PBS, detached by incubation with 500 µl of trypsin for 5 min at 37 °C and resuspended in 1 ml of medium. Collected cells were washed twice with PBS and fixed in PBS supplemented with 1% of formaldehyde as described previously by this lab[278]. The fluorescence of fixed cells was quantified by flow cytometric fluorescence analysed using BD FACSCalibur from BECTON DICKINSON and the data was analysed with CellQuest Pro from BECTON DICKINSON. BD FACSCalibur was calibrated using untransfected cells, as well as cells transfected exclusively with pDsRed Monomer N1. Afterwards, the green fluorescence of the EGFP protein encoded on the CMV-1111, CRE-, GC box- and TRE-vectors in transfected cells was determined in fluorescence channel 1 (FL-1, 515-545 nm) in parallel to the red fluorescence of the DsRed-Monomer in fluorescence channel 2 (FL-2, 564-606 nm). The fluorescence was given in a unitless, logarithmic scale output, used to generate dot-blot diagrams showing the green and red fluorescence of each living cell. To quantify the effects of selected base modifications on the EGFP expression, the EGFP fluorescence of transfected cells needed to be examined separately from untransfected cells. In order to exclude untransfected cells from the quantification, only the EGFP expression of cells with DsRed monomer fluorescence (transfection indicator) above the threshold value of 30 was included in the calculation. Afterwards, the cell number was plotted against EGFP expression in a fluorescence distribution plot and the median EGFP expression was determined. Normalised median EGFP expression levels were calculated for better inter-experimental comparison as the ration of the median EGFP expression of cells transfected with modified versus unmodified EGFP reporters.

7 Results

7.1 Consequences of 5-formylcytosine and 5-carboxycytosine in CRE and GC box gene regulatory elements

7.1.1 5-fC and 5-caC in a standalone CRE upstream regulatory element repress the gene expression in a BER-dependent manner

Within the scope of this work the functional consequences of 5-fC and 5-caC in CRE promoters were investigated in human cells using a plasmid-based reporter gene assay. The research was based on findings of Julia Allgayer (AK Khobta), showing that both base modifications influenced the CRE activity not only by direct TF binding inhibition, but also by a TDG-dependent mechanism[97], implying a causal connection between the repair of 5-fC and 5-caC and transcription regulation (cf. chapter 4.2.1 last section). Since previous experiments were insufficient to verify the link between BER and the transcriptional repression caused by 5-fC and 5-caC in CRE, excision of the modified bases was completely inhibited in the following experiments, to distinguish the effects of the primary base modifications from repair-induced effects on gene expression. BER inhibition was achieved using BER deficient DNA glycosylase knockdown- or knockout cells and/or chemically stabilised 2′-fluorinated BER-resistant DNA modifications.

The impacts of BER-sensitive and BER-resistant 5-fC and 5-caC were analysed in the minimal CRE-UNO promoter of the EGFP reporter pCRE-UNO-ZA-W. The strand exchange protocol[277] (chapter 6.30) was used to site and sequence specifically introduce a single 5-fC or 5-caC residue at the central CpG dinucleotide in the NTS of the CRE consensus sequence 5′-TGA<u>C</u>GTCA (position of modification underlined) of pCRE-UNO-ZA-W (Figure 7-1 A) generating hemi-modified reporters. Therefore, the reporter was double-nicked by Nb.BsrDI and the excised 18 nucleotides of the CRE containing NTS were exchanged for a synthetic oligonucleotide. The synthetic oligonucleotide either contained 5-fC and 5-caC or its BER-resistant 2′-fluorinated derivatives F-fC and F-caC, which were chemically stabilised by fluorination at the second carbon atom (C2) to protect the modified nucleotides against base removal by DNA glycosylases and thus BER[284]. Poly nucleotide kinase (PNK) and T4 DNA Ligase were sequentially added to the reaction, enabling the seamless ligation of

oligonucleotides and nicked reporter as described in chapter 6.30. Agarose gel electrophoresis was applied as described in chapter 6.12 and the ligation efficiency was quantified as the fraction of covalently closed (cc) plasmid DNA.
Nicked expression constructs were expected to occupy an open circular form (oc), thereby running slower than untreated constructs which occupy cc form. When ligase was added to the nicked expression constructs the excised fragment religated to the plasmid backbone and formed cc-plasmid, thus verifying that sufficient amounts of enzyme were used to enable DNA-ligation. When further supplementing the reaction with an excess of synthetic oligonucleotides, these oligonucleotides outcompeted the excised fragment for binding to the single stranded DNA stretch in nicked constructs. Since no PNK was added, the synthetic oligonucleotides could not be ligated with the plasmid backbone, because they possess a 5'-hydroxyl group instead of the 5'-phosphate required for ligation, resulting in oc-form plasmid only. Efficient oligonucleotide phosphorylation and ligation with nicked plasmid DNA was verified by the formation of cc-plasmid DNA in samples containing synthetic oligonucleotides, PNK and ligase.

The successful nicking and ligation of CRE-UNO reporters with synthetic oligonucleotides was verified by agarose gel electrophoresis, demonstrating complete construct nicking (Figure 7-1 B, lane 1, 100% oc-form plasmid) and efficient religation of the excised fragment in the absence of synthetic oligonucleotide (Figure 7-1 B, lane 2, 88% cc-form plasmid). Synthetic oligonucleotides completely outcompeted the excised fragment for binding to the single stranded DNA stretch in the nicked vector, as verified by absent cc-form plasmid DNA in PNK, ligase and oligonucleotide containing samples (Figure 7-1 B, lanes 3, 5, 7, 9 and 11). Synthetic oligonucleotides were efficiently ligated with nicked pCRE-UNO-ZA-W reporters in the presence of PNK, with samples showing on average 88% of ligated, covalently closed plasmid fraction (Figure 7-1 B, lanes 4, 6, 8, 10 and 12). The cc-plasmid fractions were similar for modified synthetic oligonucleotides and the C control, proving that single 5-fC, 5-caC, F-fC and F-caC residues did not detectably influence the ligation efficiency.

Analytical AatII digestion was used to verify the presence of 5-fC, 5-caC, F-fC and F caC in the generated reporters as described in chapter 6.17. The restriction enzyme AatII specifically recognises and cleaves 5′-GACGTC-3′ sequences and is inhibited by modifications within its recognition site. As the 4408 bp long pCRE-UNO-ZA-W reporter contains only one AatII site, which is located within the CRE sequence, AatII

cleavage linearises control reporters but not reporters containing a base modification in CRE. As expected, agarose gel electrophoresis of the AatII treated reporters showed complete cleavage of the control reporters whereas 5-fC, 5-caC, F-fC and F-caC containing constructs remained uncleaved (Figure 7-1 C). Absence of reporter cleavage verified the presence of modified nucleobases in the AatII cleavage site. In combination with the purchased MALDI-TOF mass spectrometry which verified the presence of desired DNA modifications, AatII cleavage assessment demonstrated the successful generation of 5-fC and 5-caC containing reporters.

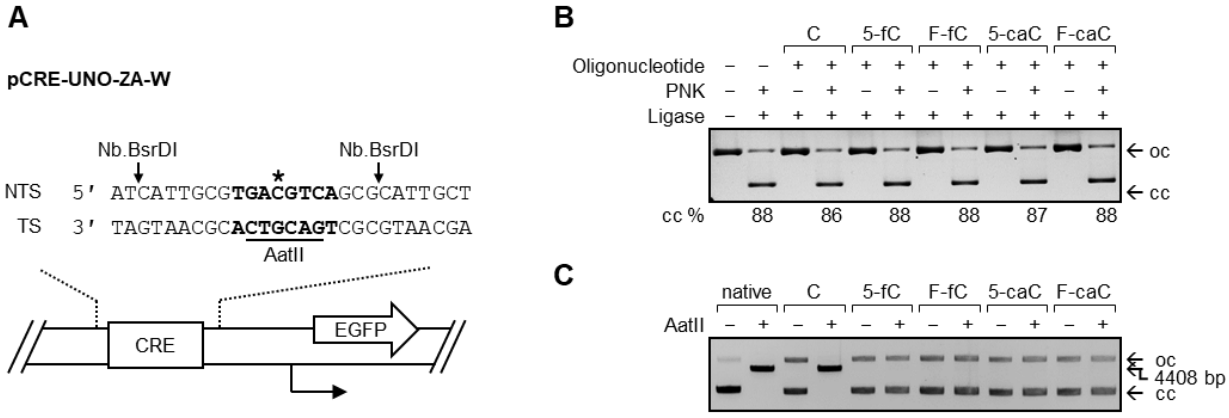


Figure 7-1: Construction of CRE-UNO reporters containing single BER-resistant or BER-sensitive 5-fC or 5-caC at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element

(A) Scheme of the standalone CRE gene regulatory element within the pCRE-UNO-ZA-W reporter: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking sites (black arrows) and position of 5-fC, 5-caC, F-fC and F-caC in the incorporated oligonucleotides (asterisk). (B) Ligation of Nb.BsrDI-nicked CRE-UNO reporters with synthetic oligonucleotides containing the specified DNA modifications or the C control in the presence and absence of PNK. The migration positions of open circular- (oc) and closed circular (cc)-form plasmid DNA are indicted by arrows and the cc-fraction of plasmid DNA is quantified (underneath). (C) Incubation of the generated constructs with the restriction enzyme AatII to verify the presence of the modifications based on the inhibition of the plasmid cleavage. Arrows indicate migration positions of the linearised vector (4408 bp) and of the cc- and oc-form vector.

The effects of BER-sensitive and BER-resistant 5-fC, 5-caC on the gene expression were investigated by reporter transfection into human cells, followed by quantitative fluorescence analysis via flow cytometry[135], as described in chapter 6.32 and depicted in Figure 7-2 A. The easy-to-handle human HeLa cell line was chosen for the expression analysis due to its high transfection efficiency and potent BER, which are essential features to obtain high-quality results from flow cytometric analysis of reporters containing a single base modification. To quantify the EGFP expression in cells, HeLa were co-transfected with equal amounts of a reporter coding for a red-fluorescent protein (DsRed-Monomer N1) as transfection indicator (chapter 6.32). Cells were fixed at the selected time after transfection and examined for green and red fluorescence via flow cytometry on single cell level (chapter 6.32). Red fluorescence was used to distinguish transfected from untransfected cells and the median EGFP

fluorescence of the transfected population was calculated. EGFP expression of reporters containing a base modification was normalised against control constructs and used to compare reporter expression in different experiments and cell lines.

The EGFP expression analysis of transfected cells showed, that BER-resistant F-fC and F-caC had a slight negative impact on the reporter activity, resulting in an approximate residual EGFP expression of 81% and 89% at 24 hours respectively (Figure 7-2 B-D). The reduction of the EGFP fluorescence was already detectable 6 hours after transfection and expression remained stable over time. BER-sensitive 5-fC and 5-caC also negatively impacted the gene expression at the 6-hour time point, yet the effect intensified over 24 hours. The EGFP expression levels strongly decreased from the initial ~92% to 30% (5-fC) and 27% (5-caC). Since BER-sensitive 5-fC and 5-caC, but not BER-resistant F-fC and F-caC, caused a time-dependent decline of the EGFP repression, these data indicate that both modifications caused transcriptional repression by a BER-dependent mechanism. The utilised HeLa cells contained normal amounts of TDG, yet transcriptional repression was only observed for BER-sensitive 5-fC and 5-caC. The findings thus rule out that the mere presence of TDG protein was sufficient to induce gene silencing by 5-fC and 5-caC. Combined with previous experiments showing that repression of the gene expression by 5-fC and 5-caC correlated with the TDG amount in cells[97], my experimental data support the hypothesis that gene silencing caused by 5-fC and 5-caC was induced by a mechanism which depends on TDG-initiated BER.

In summary, the EGFP expression analysis showed that, 5-fC and 5-caC in a minimal CRE promoter had a small but statistically significant negative effect on the gene expression in absence of BER. Furthermore, BER of 5-fC and 5-caC drastically reduced the EGFP expression in agreement with earlier results obtained in this lab[97]. The impact of 5-fC and 5-caC was analysed in the NTS only, as previous results in this lab showed that the outcomes of BER were similar in both DNA strands[97].

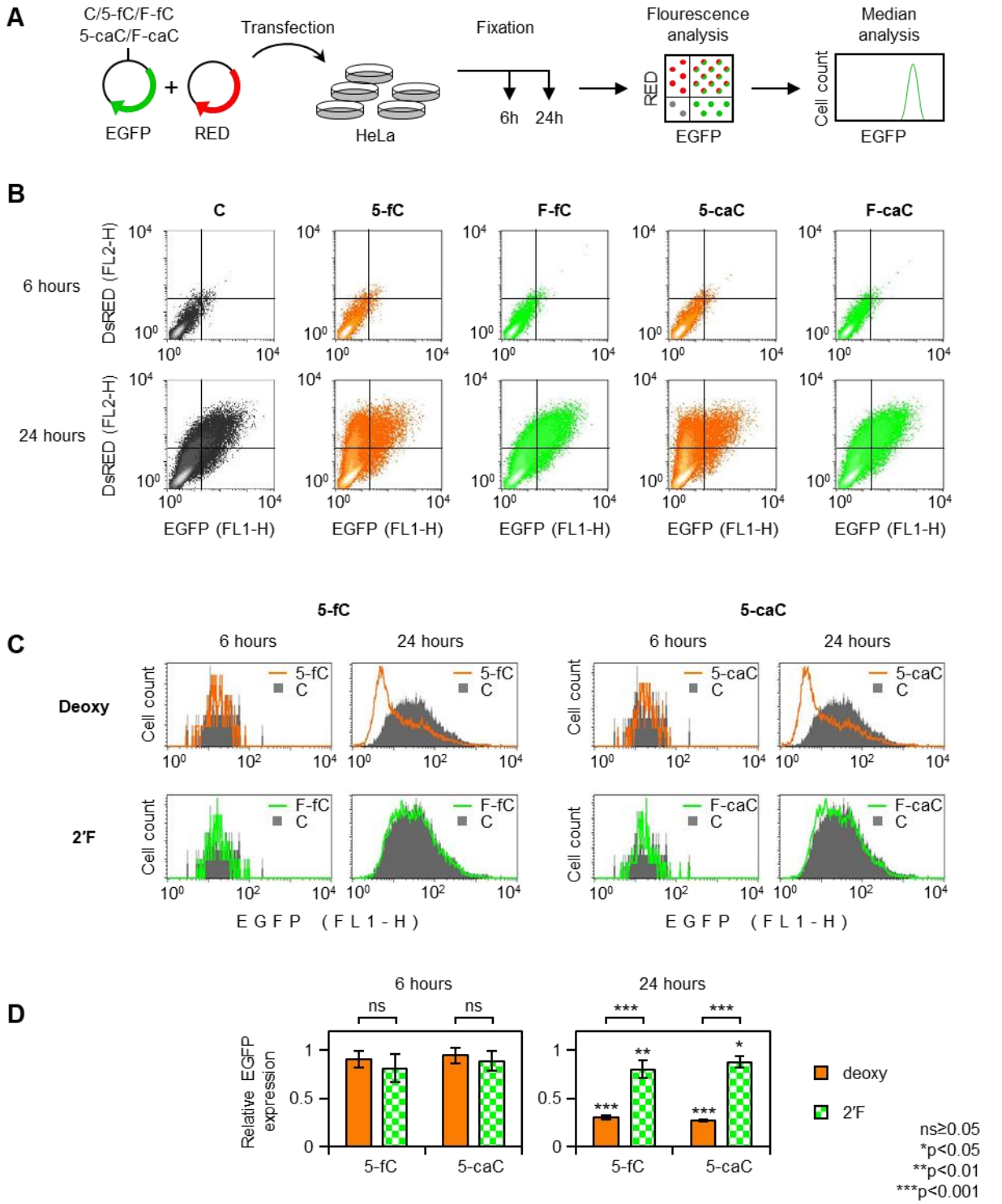


Figure 7-2: Expression of CRE-UNO reporters containing single BER-resistant or BER-sensitive 5-fC or 5-caC at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pCRE-UNO-ZA-W reporters containing single BER-sensitive 5-fC and 5-caC or their BER-resistant 2′-fluorinated derivatives F-fC and F-caC at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE upstream regulatory element. (A) Schematic representation of experimental set-up to quantify the EGFP expression of reporters containing a single base modification. HeLa cells were co-transfected with the EGFP reporter and a reporter coding for the DsRed-Monomer N1 as transfection marker, incubated and samples were taken for formaldehyde fixation 6-, 12-, 24-, and 48-hours after transfection. The fixed cells were subjected to quantitative fluorescence analysis by flow cytometry. Transfected cells were selected by red fluorescence and their EGFP fluorescence was quantified (B) Representative scatter plots and (C) fluorescent distribution plots of HeLa cells 6- and 24-hours after transfection. (D) Quantification of the relative EGFP expression values of transfected HeLa cells in five independent experiments (mean ± SD).

7.1.2 APE1 mediated incision of the DNA at apyrimidinic sites is essential to induce CRE silencing

Given that 5-fC and 5-caC repressed transcription in a BER-dependent manner, the nature of the repair intermediate provoking the observed transcription reduction was investigated. To model an apyrimidinic site intermediate, CRE-UNO constructs carrying either a BER-sensitive AP site analogue tetrahydrofuran (F), its BER-resistant derivative phosphorothioate tetrahydrofuran (SF) or the respective C control at the position of 5-fC and 5-caC (Figure 7-3 A) were generated by the strand exchange procedure (chapter 6.30). At SF residues, incision of the DNA by AP endonucleases such as APE1 is inhibited by the presence of the 5′-phosphorothioate, which renders SF resistant to BER. In contrast, F harbours a 5′-phosphate, posing as a good substrate for APE1. Agarose gel electrophoresis (cf. chapter 6.12) of the CRE-UNO constructs containing the specified types of AP sites showed efficient plasmid gapping and ligation with the synthetic oligonucleotides, resulting in approximately 86% of cc-form plasmid (Figure 7-3 B). Analytical digestion with the major AP endonuclease EndoIV as described in chapter 6.17 showed the expected full cleavage of constructs containing F, which verifies the presence of an AP site within the plasmid DNA (Figure 7-3C). SF containing constructs were cleaved significantly less frequent compared to F containing constructs, verifying that the 5′-phosphorothioate rendered the AP site at least partially resistant to AP-endonucleases.

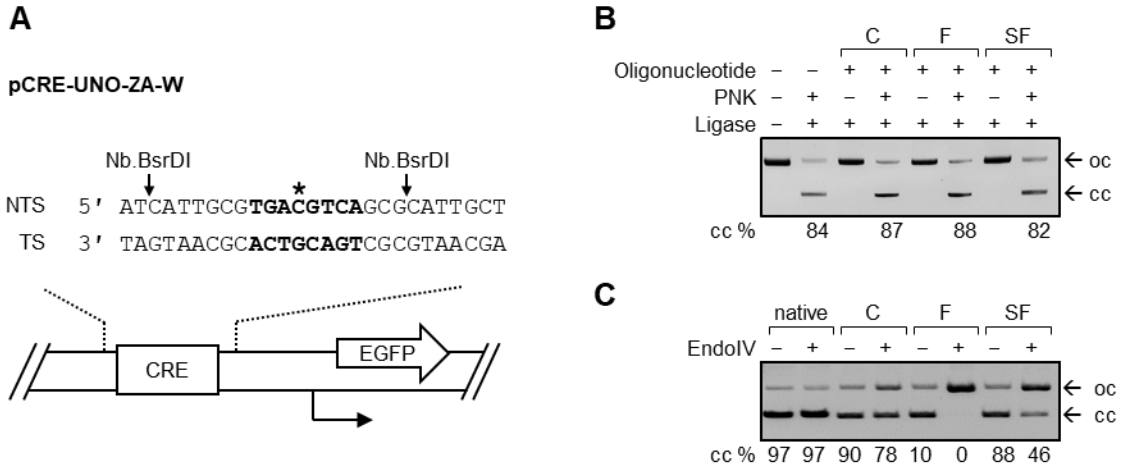


Figure 7-3: Construction of CRE-UNO reporters containing a single BER-resistant or BER-sensitive apyrimidinic site at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element

(A) Scheme of the standalone CRE gene regulatory element within the pCRE-UNO-ZA-W reporter: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), Nb.BsrDI nicking sites (black arrows) and position of F and SF in the incorporated oligonucleotides (asterisk). (B) Ligation of Nb.BsrDI-nicked and gapped pCRE-UNO constructs with synthetic oligonucleotides containing a single F, SF or the respective C control in the presence and absence of PNK. (C) Incubation of the generated constructs with EndoIV to verify the presence of F by EndoIV nicking and the AP endonuclease protection of SF. The fraction of cc-form plasmid is quantified (underneath).

Expression analysis of AP site-containing CRE reporters in HeLa cells (chapter 6.32) showed that a single BER-resistant AP site mildly reduced the gene expression, resulting in 99% and 86% of residual EGFP expression after 6- and 24 hours respectively (Figure 7-4 A-C). BER-sensitive AP sites caused a strong reduction of the gene expression over 24 hours, down to 20% of residual EGFP expression. The repressive effect was not yet detectable 6 hours after transfection. TDG as an upstream BER enzyme had no influence on the transcriptional repression by F in CRE, as the results in HeLa cells were exactly reproduced in the isogenic TDG deficient cell line (Figure 7-4 D).

Comparing the effects of BER-resistant F and BER-sensitive SF within the CRE gene regulatory element showed that a single AP site per se did not influence the promoter activity. Only in the presence of BER did an AP site cause a drastic reduction of the gene expression, indicating that APE1-mediated DNA strand incision was essential to initiate transcriptional repression. Since BER-resistant AP sites did not induce transcriptional repression, it is inferred that the DNA strand incision step also mediated the transcriptional repression observed for 5-fC and 5-caC in CRE.

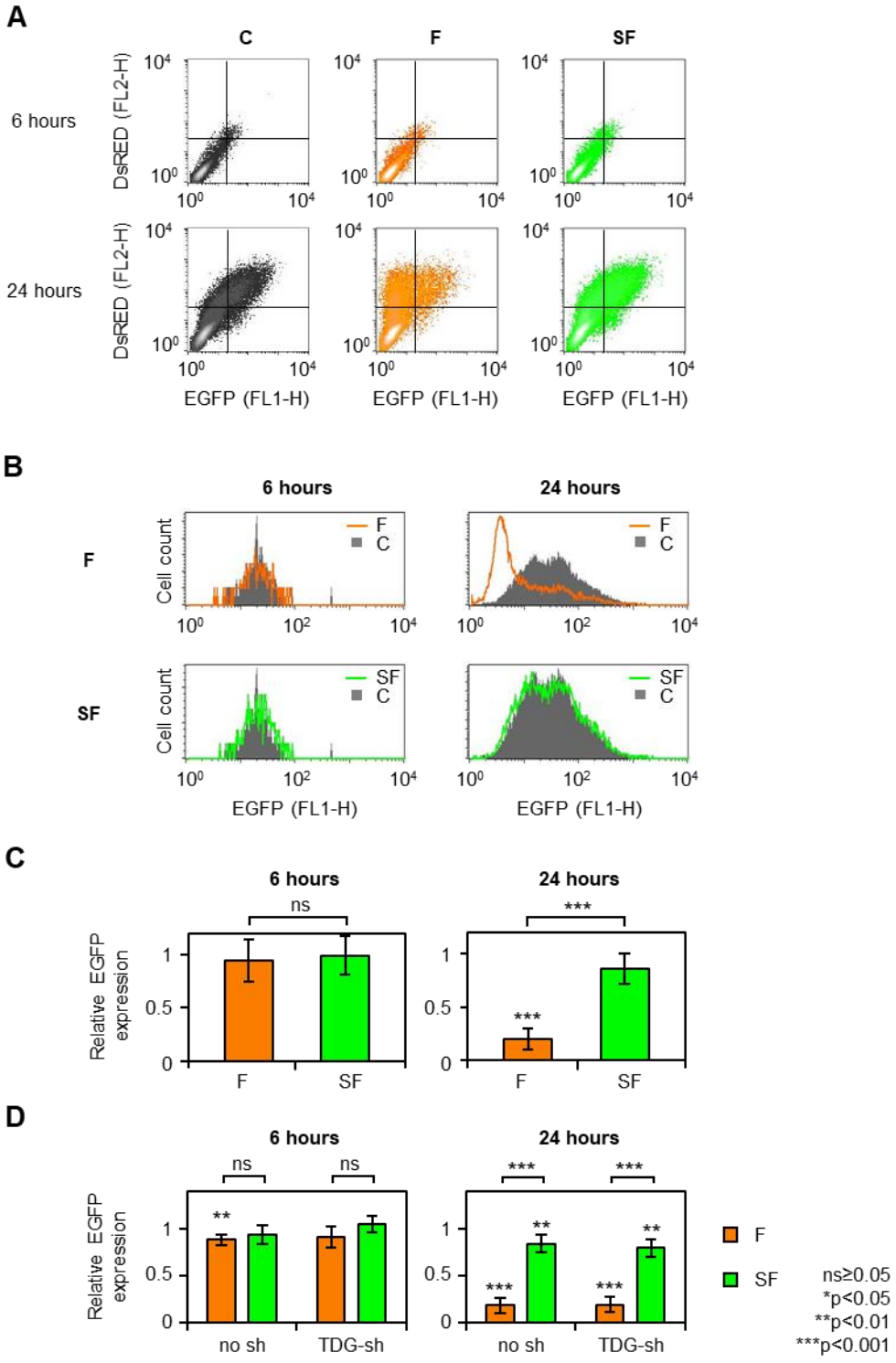


Figure 7-4: Expression of CRE-UNO reporters containing a single BER-resistant or BER-sensitive apyrimidinic site at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa and HeLa derived TDG knockdown cells transfected with pCRE-UNO-ZA-W reporters containing either a single apyrimidinic site analogue F, its BER protected derivative SF or the C control at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE upstream regulatory element. Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6- and 24 hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in four independent experiments (mean ± SD). (D) Quantification of the relative EGFP expression of transfected TDG proficient (no sh) and isogenic TDG deficient HeLa cells (TDG-sh) in four independent experiments (mean ± SD).

7.1.3 Transcriptional repression by BER of 5-fC and 5-caC in CRE is independent from promoter strength

Next, it was investigated whether the impact of 5-fC and 5-caC on gene expression was influenced by the promoter strength, a component of transcription regulation with yet unknown effects on the base modifications' functional consequences. For this reason, the transcriptional effects of 5-fC and 5-caC were analysed in the strong CMV derived CMV-1111 promoter. The CMV-1111 promoter exhibits an enhanced transcriptional activity due to the presence of four CRE elements in contrast to one CREB site in the minimal CRE-UNO promoter. pCMV-1111-ZA-W reporters[97,136] were used to introduce 5-fC, 5-caC, F-fC, F-caC or C in the central CpG dinucleotide of the third CRE consensus sequence of the CMV-1111 promoter (Figure 7-5 A) by the strand exchange procedure (cf. chapter 6.30). Agarose gel electrophoresis (cf. chapter 6.12) showed efficient pCMV-1111-ZA-W nicking and further ligation with 5-fC, 5-caC, F-fC, F-caC or C containing synthetic oligonucleotides on average resulting in 89% of cc-form plasmid (Figure 7-5 B). Analytical AatII digestion, as described in chapter 6.17, was used to verify the presence of the DNA modifications by restriction pattern alterations compared to control constructs. AatII-induced DNA cleavage at CRE sites is inhibited by 5-fC, 5-caC, F-fC, F-caC, thereby changing the restriction of the 4872 bp long pCMV-1111-ZA-W vector from four fragments with 53 bp, 83 bp, 177 bp and 4559 bp (control construct) to 53 bp, 260 bp and 4559 bp (construct containing a base modification). As expected, altered restriction patterns were observed for all reporters containing single 5-fC, 5-caC, F-fC or F-caC (Figure 7-5 C), which together with MALDI-TOF mass spectrometry analysis of the synthetic oligonucleotides verified the successful generation of CMV-1111 reporters containing a single DNA at the central CpG dinucleotide of the target CRE site. Due to its small length, the signal of the 53 bp fragment is undetectable in the EthBr stained gel.

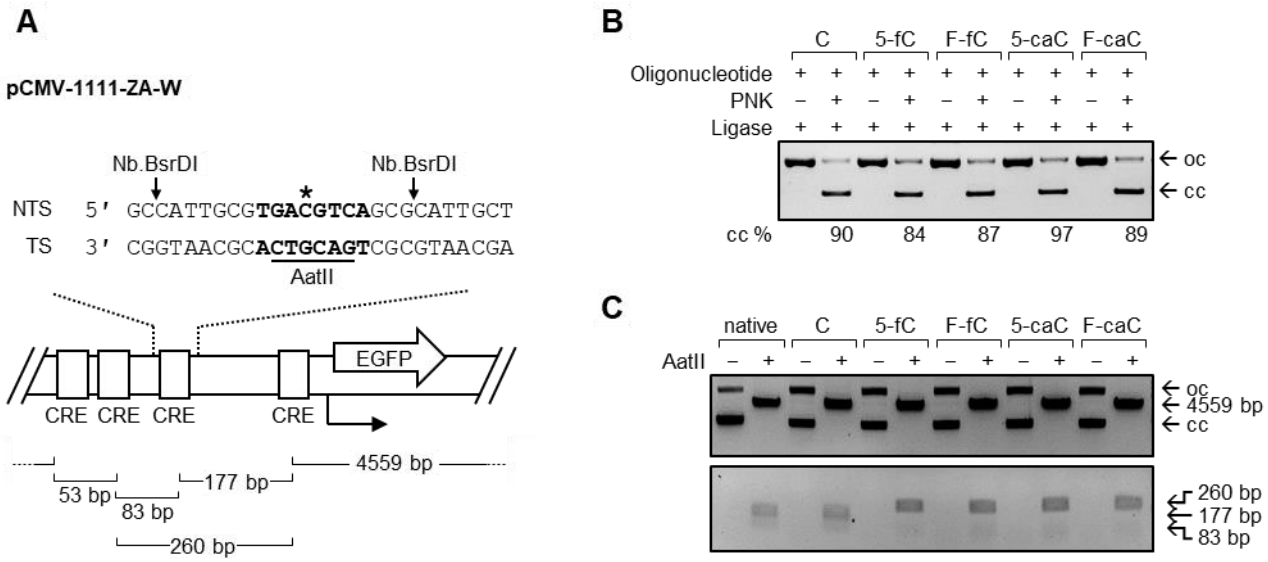


Figure 7-5: Construction of CMV-1111 reporters containing single BER-resistant or BER-sensitive 5-fC and 5-caC at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter

(A) Scheme of the four canonical CRE sites within the pCMV-1111-ZA-W reporter, with the distances between the CRE indicated below and focus on the third CRE gene regulatory element: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking sites (black arrows) and position of BER-sensitive 5-fC, 5-caC and BER-resistant F-fC and F-caC in the incorporated oligonucleotides (asterisk). (B) Ligation of Nb.BsrDI-nicked CMV-1111 constructs with synthetic oligonucleotides containing the specified cytosine modification or the C control in the presence and absence of PNK. (C) Incubation of the generated constructs with the restriction enzyme AatII to verify the presence of the modifications based on the inhibition of the plasmid digestion at the modified CRE. Arrows indicate migration positions of the DNA fragments of cleaved control reporter (4559 bp, 177 bp, 83 bp and 53 bp fragments) and fragments CMV-1111 reporters containing a base modification (4559 bp, 260 bp and 53 bp) and of the cc- and oc-form vector.

EGFP expression analysis of transfected HeLa cells (chapter 6.32) showed that also in the strong CMV-1111 promoter a single 5-fC and 5-caC residue negatively affected the gene expression in a BER-dependent manner (Figure 7-6). Thus, gene expression was not significantly affected by 5-fC and 5-caC at 6 hours after transfection. It declined to 22% (5-fC) and 17% (5-caC) of residual EGFP expression at 24 hours, with a higher repression magnitude for 5-caC compared to 5-fC. Inhibition of BER using 2′-fluorinated 5-fC and 5-caC completely abolished the time-dependent transcriptional repression by both modifications (Figure 7-6). These results indicate that transcriptional silencing of CMV-1111 was mediated by a similarly BER-mediated mechanism as observed for 5-fC and 5-caC in CRE-UNO. Remarkably, the repressive effect of BER-sensitive 5-fC and 5-caC in CMV-1111 was only marginally stronger than in the weak CRE-UNO promoter (cf. chapter 7.1.1), indicating that promoter strength does not affect the BER-mediated transcriptional repression by 5-fC and 5-caC in CRE. These findings further demonstrate that transcriptional repression induced by BER of 5-fC or 5-caC spreads beyond the local modified CRE, covering the whole promoter.

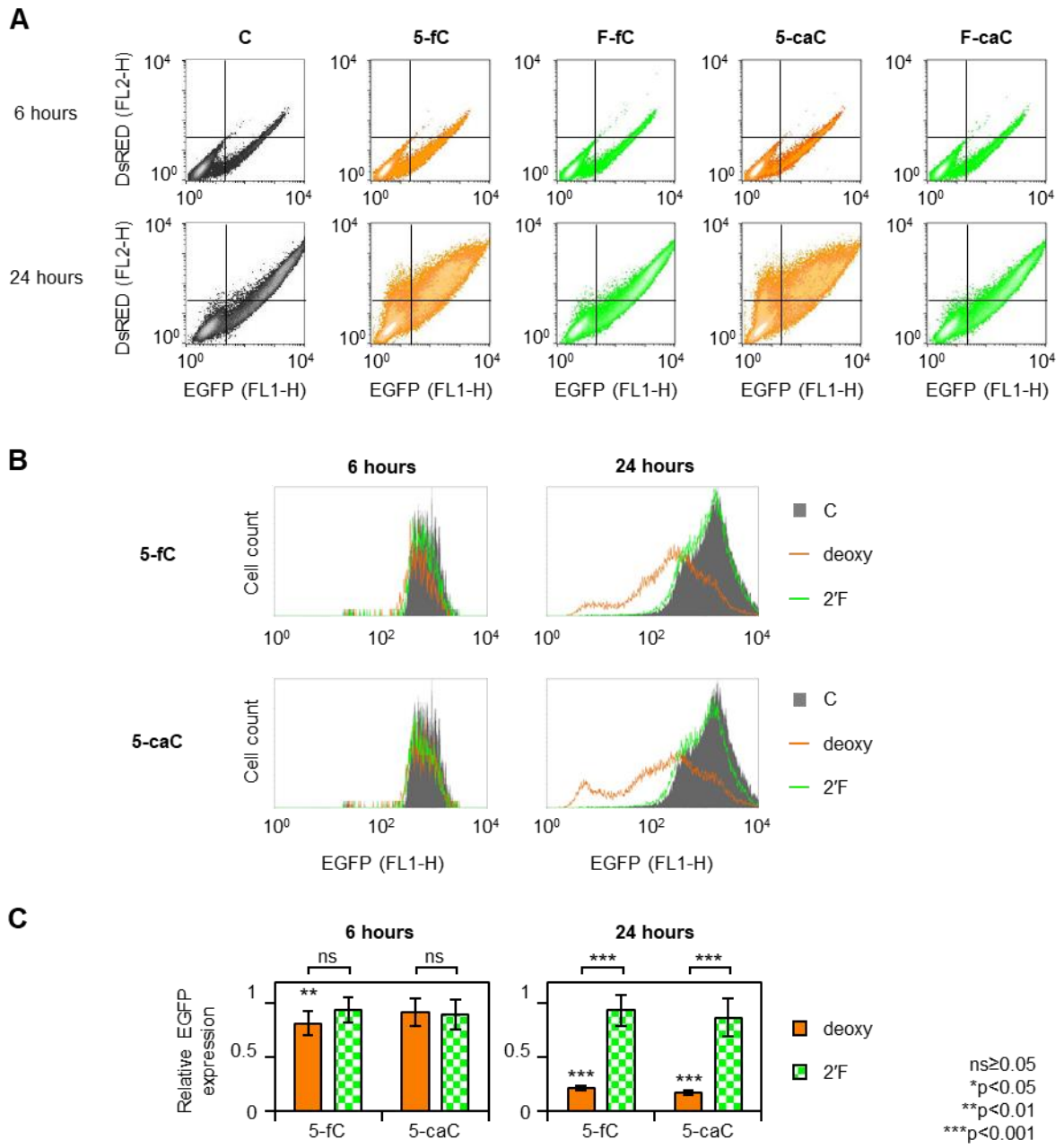


Figure 7-6: Expression of CMV-1111 reporters containing BER-resistant or BER-sensitive 5-fC and 5-caC at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pCMV-1111-ZA-W reporters containing single BER-sensitive 5-fC and 5-caC or their BER-resistant 2′-fluorinated derivatives at the central CpG dinucleotide within the non-transcribed strand of the third CRE gene regulatory element of the CMV-1111 promoter. Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6- and 24-hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in five independent experiments (mean ± SD).

7.1.4 TDG knockdown partially rescues the transcriptional repression induced by BER of 5-fC and 5-caC in CRE

Previous experiments indicate that not only BER-proficiency (cf. chapter 7.1.1), but also physiological amounts of cellular TDG are essential to induce transcriptional repression by 5-fC and 5-caC in the CRE-UNO promoter (cf. chapter 4.2.1 last section)[97]. It is thus assumed, that 5-fC and 5-caC in CRE-UNO promoters induce

transcriptional repression by a mechanism that requires TDG initiated BER of both base modifications. Since transcriptional repression by BER of 5-fC and 5-caC was also observed in the CMV-1111 promoter, TDG involvement was assessed using TDG deficient cells. The functional consequences of 5-fC and 5-caC in CMV-1111 were analysed in HeLa derived cell lines with stable knockdown of one of the three human uracil DNA glycosylases TDG, SMUG or UNG (cf. above). The cells were transfected with CMV-1111 reporters containing 5-fC, 5-caC or the respective C control in the selected CRE site and the transcription was analysed at 6-, 12-, 24- and 48 hours after transfection (chapter 6.32). In the DNA glycosylase proficient cells, 5-fC and 5-caC induced the characteristic time-dependent decline of the EGFP expression (as observed previously in chapter 7.1.3), which intensified between the 24- and 48-hour time points (Figure 7-7). Comparison between the isogenic parental (no sh) cell line with TDG knockdown cells, showed that TDG deficiency significantly minimised the transcriptional repression induced by 5-fC and 5-caC. More specifically, a 26% (5-fC) and 29% (5-caC) recovery of the expression was observed comparing the EGFP fluorescence of TDG deficient- with TDG proficient cells (TDG-sh versus no sh) at 24 hours after transfection. These data demonstrate that the transcriptional repression by 5-fC and 5-caC in CMV-1111 was at least partially TDG-dependent. Knockdown of SMUG1 and UNG DNA glycosylases did not affect the transcriptional repression induced by 5-fC and 5-caC.

In summary, 5-caC and 5-fC in the weak CRE-UNO and the strong CMV-1111 promoter dynamically decreased the gene expression, which was attenuated by TDG knockdown. Since the complete inhibition of base removal fully restored the transcription of 5-fC and 5-caC containing reporters (cf. chapter 7.1.1), it was hypothesised that the residual TDG activity in TDG-sh cells caused the incomplete recovery of the transcription under TDG knockdown conditions. Nevertheless, it was still possible that another DNA glycosylase contributed to BER of 5-fC and 5-caC in vivo, wherefore functional consequences of 5-fC and 5-caC were next analysed in the complete absence of TDG.

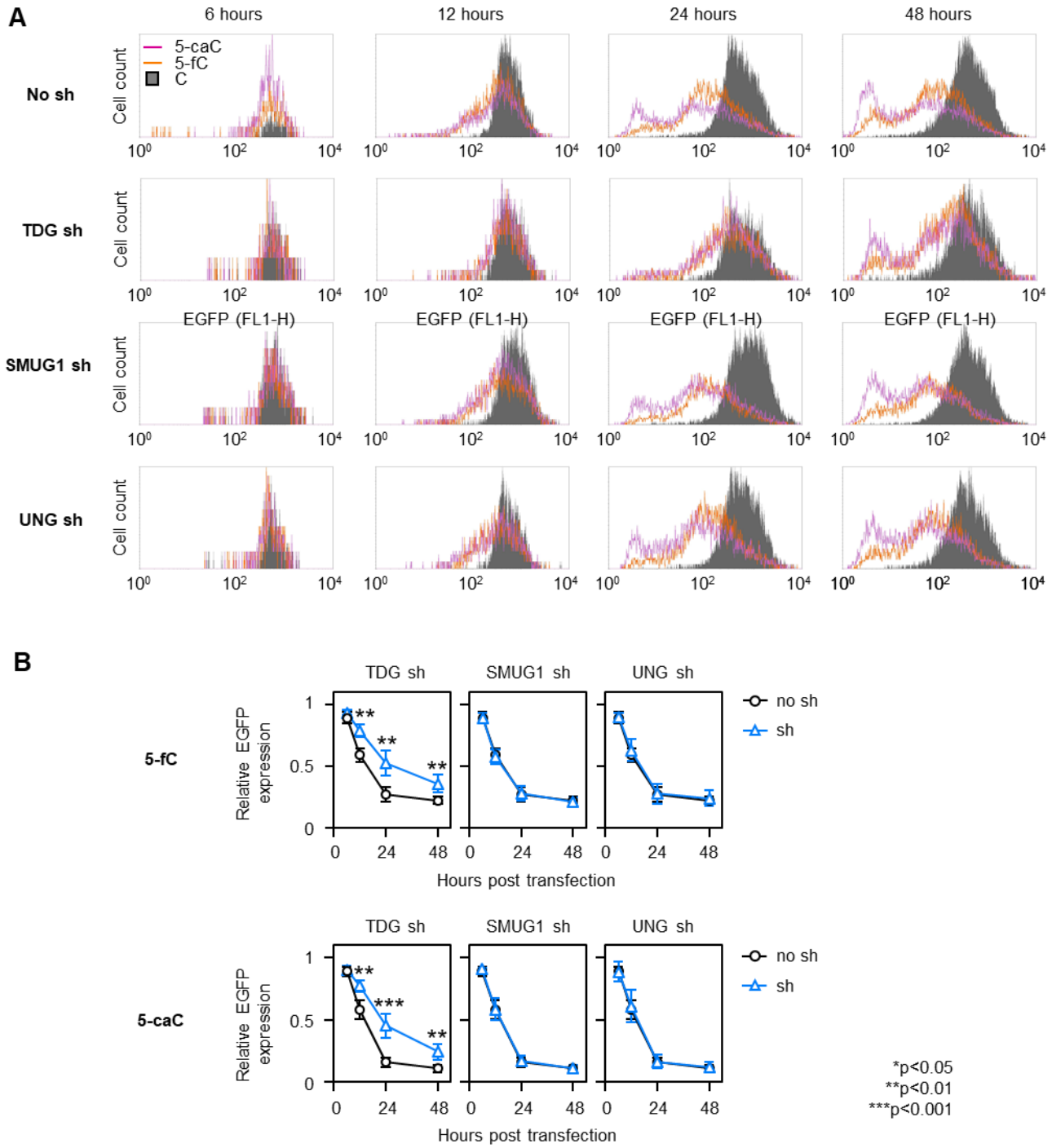


Figure 7-7: Expression of CMV-1111 reporters containing 5-fC and 5-caC at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter in isogenic cell lines with stable knockdown of the uracil DNA glycosylases TDG, SMUG1 and UNG

Quantitative EGFP expression analysis of HeLa cells deficient in the TDG (TDG-sh), SMUG1 (SMUG1-sh) and UNG (UNG-sh) DNA glycosylase in comparison to the respective control cell line (no sh) transfected with pCMV-1111-ZA-W reporters containing single 5-fC and 5-caC in the central CpG dinucleotide within the non-transcribed strand of the third CRE gene regulatory element of the CMV-1111 promoter. (A) Representative fluorescent distribution plots of HeLa cells 6-, 12-, 24- and 48 hours after transfection. (B) Quantification of the relative EGFP expression of transfected HeLa and HeLa derived glycosylase knockdown cells in five independent experiments (mean ± SD).

7.1.5 Optimisation of CRISPR-CAS9-mediated TDG knockout in HeLa cells

To determine if TDG was the only DNA glycosylase to remove 5-fC and 5-caC bases from CMV-1111 promoters and induce transcriptional repression, TDG activity should be completely abolished in target cells. To completely abolish TDG activity in human cells, a stable TDG knockout HeLa cell line was generated by CRISPR-CAS9 mediated gene editing. The TDG knockout was established following the "CRISPR: Protocol for Genomic Deletions in Mammalian Cell Lines"[281] from ADDGENE. The protocol was modified as described in chapters 6.20-6.28, in order to adapt the procedure to the HeLa cell line. With this technique, the CRISPR-RNA and CAS9 protein are introduced into the cells, encoded on vector DNA (knockout constructs), that encodes both the CAS9 protein and a *TDG* targeting sgRNA.

To ensure complete loss of the TDG activity, the *TDG* gene was simultaneously targeted at two different positions by two sgRNAs, enclosing the nucleotides coding for the R140 active site amino acid as schematically depicted in Figure 7-8 A. Two approaches were tested to introduce complete loss of TDG activity. On the one hand, the whole protein-coding sequence should be deleted by targeting sites within the *TDG* gene which ly as much up- and downstream from the active site coding nuclides as possible (total gene deletion-T). On the other hand, gene knockout can be accomplished by minimal impact gene editing, selectively deleting the exon containing the catalytic residue by targeting the closest possible adjacent sites (active site deletion-A). Thus, sgRNA set T was designed as described in chapter 6.20 to target positions +253 and +18908 upstream from the GTG start codon, whereas sgRNA set A was designed to target positions +11012 and +16905 within the *TDG* gene as schematically depicted in Figure 7-8 A (sgRNA sequence details: Table 6-15, p. 60). The corresponding knockout constructs (e.g. pX330-spCAS9-HF_TDG+253 for sgRNA +253) were cloned based on the pX330-sgCAS9-HF vector (ADDGENE) as described in chapter 6.21 (for sequences cf. Appendix III) and verified by analytical BsrDI digestion (chapter 6.17) and Sanger sequencing (chapter 6.18) (data not shown).

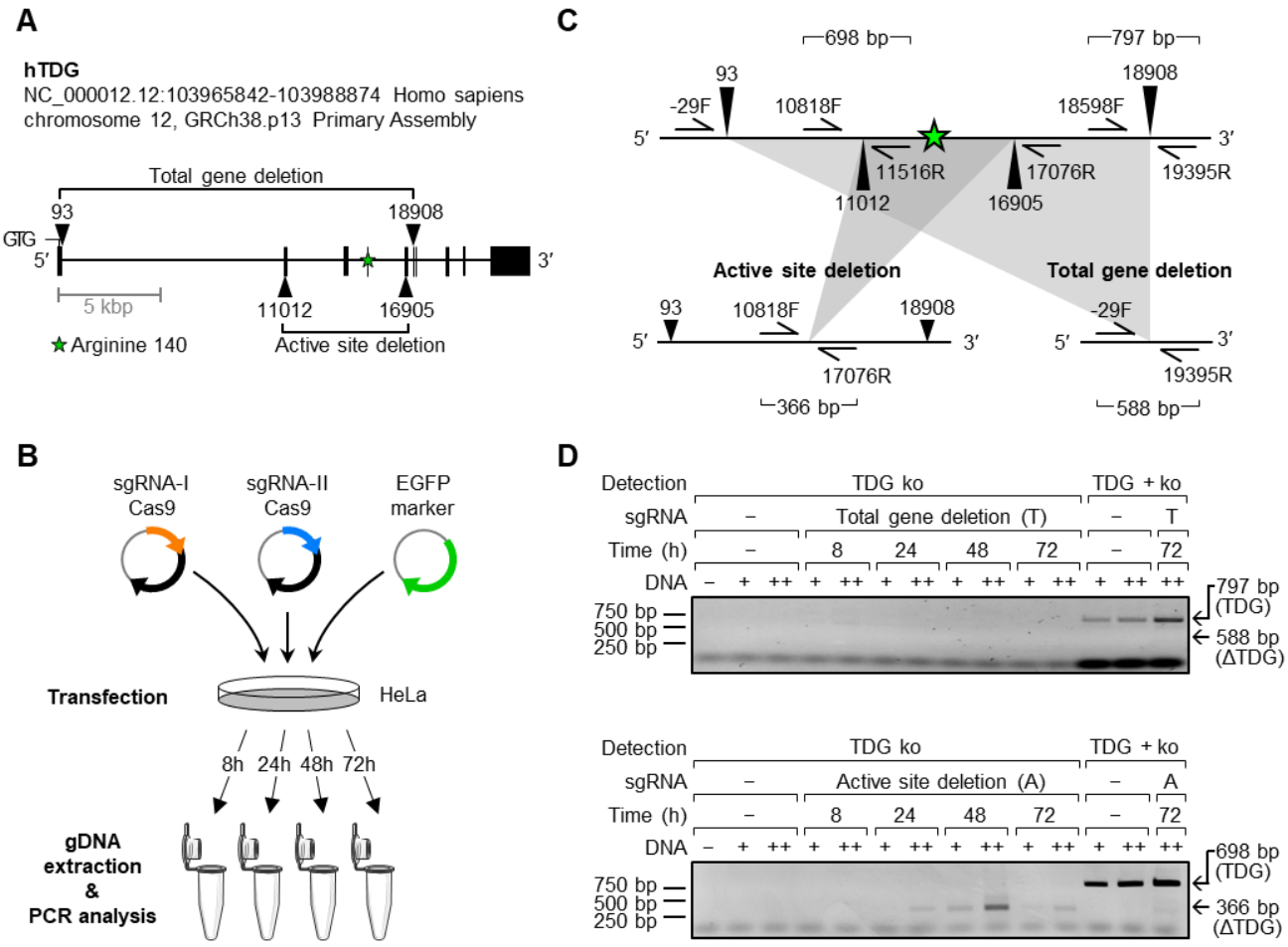


Figure 7-8: CRISPR-CAS9-mediated TDG knockout in HeLa cells by simultaneous targeting of two sgRNA sites within the TDG gene

Establishment of a TDG knockout within HeLa cells by either deleting the total TDG gene (T) or specifically deleting the active site coding exon (A) by CRISPR-CAS9 mediated gene editing using the sgRNA combinations sgRNA+93/+18908 or sgRNA+11012/+16905 respectively. (A) True to scale scheme of the human genomic TDG sequence: gene coding region (black line), position of GTG start codon (broken line), exons (black boxes), arginine 140 active site coding nucleotides (green star), position of sgRNA site (triangle, with distance from the TSS indicated above). (B) Schematic representation of the experimental setup to determine the highest TDG editing efficiency in HeLa cells transfected with a pair of knockout constructs coding for a targeting sgRNA and the CAS9 protein and an EGFP transfection marker: knockout constructs (circles with orange/blue arrows indicating the sgRNA sequence and black arrows indicating the CAS9 sequence), EGFP transfection marker (circle with green arrow indicating EGFP sequence). (C) Schematic representation of the non-rearranged TDG gene (upper panel) and rearranged TDG gene after active site deletion (lower left panel) or total gene deletion (lower right panel): active site coding nucleotides (green star), position of sgRNA sites (triangle) and PCR primers (half arrow with distance from the TSS indicated above) with the according PCR product length. (D) Detection of the TDG knockout in HeLa cells by PCR screening. gDNA from cells transfected with different sets of knockout constructs was extracted 8-, 24-, 48-, and 72 hours post transfection and either 10 ng (+) or 30 ng (++) of DNA were used as a PCR template with gDNA of untransfected cells as a control. In pX330-spCAS9-HF_sgRNA+93/+18908 transfected cells the total gene deletion (ko) was detected using primers TDG_-29F/19395R whilst primer TDG_18598F was added to detect non-rearranged TDG (TDG). In pX330-spCAS9-HF_sgRNA+11012/+16905 transfected cells the active site deletion was detected using primers TDG_10818F/17076R whilst primer TDG_11516R was added to detect non-rearranged TDG.

In order to find the sgRNA combination with highest editing efficiency, HeLa cells were transfected with a pair of knockout constructs coding for sgRNA-I and sgRNA-II in parallel (chapter 6.22) and the genomic DNA was extracted (chapter 6.24) at different time points as depicted in Figure 7-8 B. It was expected, that CAS9-mediated incision of the TDG gene at both sgRNA target sites results in the ligation of the cleaved DNA

by non-homologous end-joining under exclusion of the excised active site-containing DNA stretch, thus inducing the desired gene knockout. Cells were transfected either with pX330-spCAS9-HF_TDG+253 and pX330-spCAS9-HF _TDG+18908 to induce total gene knockout, or with pX330-spCAS9-HF_TDG+11012 and pX330-spCAS9-HF _TDG+16905 to induce active site knockout. Genomic DNA extraction was performed 8-, 24-, 48- and 72 hours after transfection as described in chapter 6.23 to find optimal editing conditions and the *TDG* editing efficiency was analysed by PCR.

Two separate PCR reactions were performed as described in chapter 6.27 to assess the presence of non-rearranged or rearranged genomic *TDG* in HeLa cells. For each sgRNA site, primer pairs were designed (chapter 6.26) to closely flanked the target site and thereby specifically amplifying an sgRNA-containing product. Owing to their peculiar design, the same primers can be used to detect the desired total- or active site TDG deletion in the human genome by mixed primers of different pairs. Thus, the forward primer for sgRNA-I and the reverse primer for sgRNA-II can only amplify a PCR product from *TDG* when gene editing took place in the desired fashion. First, a separate PCR was performed as describe in chapter 6.27 to determine the primers optimal annealing temperature using genomic DNA of untransfected cells as template (Appendix I 1) and primer pairs TDG_18598F/19395R and TDG_10818F/11516R were selected to detect non-rearranged TDG in HeLa cells in further experiments. The position of the primers is depicted in Figure 7-8 C.

After establishing optimal annealing conditions, PCR was used to assess gDNA (chapter 6.24) of transfected cells for *TDG* editing using10 ng or 30 ng as template. DNA from untransfected HeLa cells was used as a negative control. Amplification of non-rearranged *TDG* was used to verify the template quality and PCR product amplification efficiency by addition of a third primer in a multiplexing PCR. For cells transfected with sgRNA set T, the desired total *TDG* deletion remained undetected using the primer pair TDG_-29F/19395R (Figure 7-8 D, upper gel, lane 4-11, absence of PCR product with 588 bp). Multiplexing by addition of primer TDG_18598F effectively yielded the expected PCR product from non-rearranged *TDG* in samples containing DNA of untransfected and transfected HeLa cells (Figure 7-8 D, upper gel, lane 12-14, 797 bp). This excluded an unsuccessful PCR reaction as the reason for the absence of rearranged *TDG* product, indicating that sgRNA set T was not effective to induce CRISPR-CAS9-mediated gene editing of *TDG* in HeLa. For cells transfected with sgRNA set A, the desired *TDG* deletion of ~16,000 bp was detected 24-, 48- and

72 hours after transfection, with optimal product amplification at the 48-hour time point using the primer pair TDG_10818F/17076R (Figure 7-8 D, lower gel, lane 4-11, PCR product with 366 bp). Non-rearranged (untransfected HeLa) and to a lesser extend also rearranged *TDG* (untransfected and transfected HeLa) was detected by multiplexing with primer TDG_16227F, verifying effective DNA amplification (Figure 7-8 D, lower gel, lane 12-14, PCR product with 797 bp). The PCR results show that transfection of HeLa cells with knockout constructs coding for sgRNA set A (active site deletion) successfully induced the desired TDG knockout in HeLa cells, with highest editing efficiency at 48 hours after transfection.

7.1.6 Generation and validation of HeLa derived single cell clones with TDG knockout

The optimised transfection and gene editing conditions were used to generated HeLa derived single cell clones harbouring the desired TDG knockout as depicted in Figure 7-9 A. HeLa cells were co-transfected with pX330-spCAS9-HF_TDG+11012, pX330-spCAS9-HF_TDG+16905 and the EGFP transfection parker pZAJ as described in chapter 6.22 and used for single cell sorting 39 hours after transfection (chapter 6.23). Single cell sorting was performed in the Flow Cytometry Core Facility of the Institute for Molecular Biology Mainz under supervision of scientific staff. Based on the assumption, that plasmid uptake directly correlates with the protein expression, and thus gene editing efficiency in transfected cells, HeLa cells with high EGFP-transfection marker expression were selected. Therefore, HeLa cells were sorted according to top 5% of EGFP fluorescence (Figure 7-9 A+B) in order to maximise the gene editing probability. A second population was sorted according to top 20% of EGFP fluorescence for comparison purposes, to verify if the induced amounts of CRISPR-CAS9 activity were critically high and impaired genome stability.

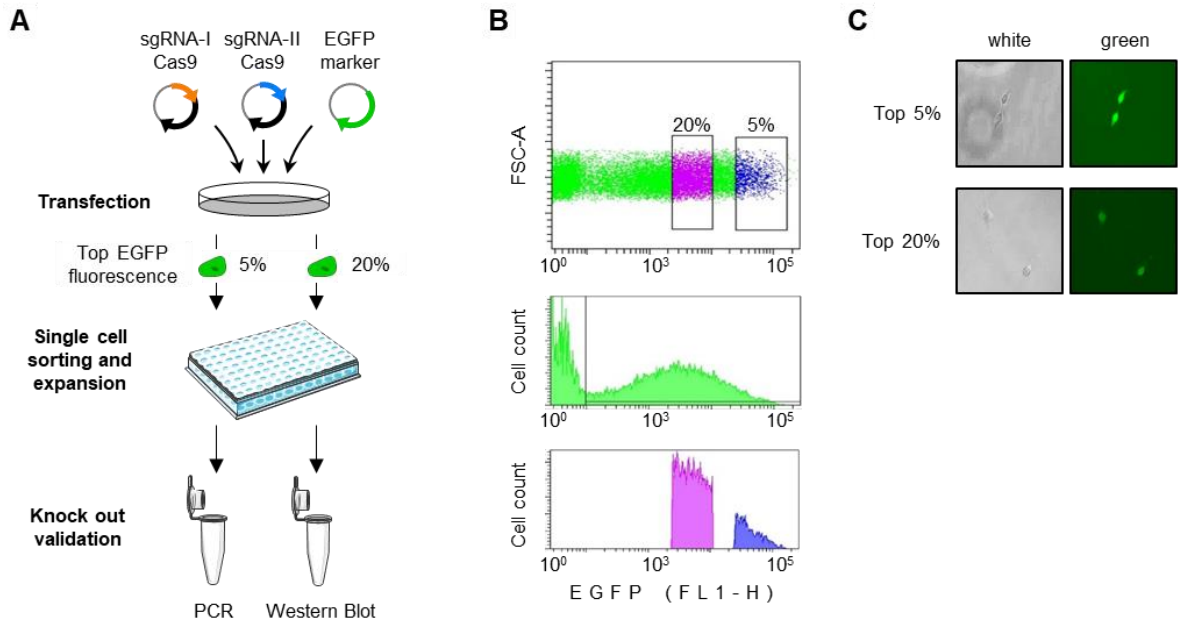


Figure 7-9: Single cell sorting of HeLa cells with potential CRISPR-CAS9-mediated TDG knockout

Single cell sorting of HeLa cells with potential TDG knockout 39 hours after transfection with knockout constructs pX330-SpCAS9-HF1_TDG+11012 and pX330-SpCAS9-HF1_TDG +16905 and the pZAJ EGFP transfection marker. (A) Schematic representation of the sorting approach to select HeLa cells with high TDG knockout potential. Based on the assumption that the knockout efficiency correlates with the transfection efficiency the cells were sorted according to high transfection marker expression, selecting clones with top 5% and top 20% of EGFP fluorescence. (B) Quantitative EGFP expression analysis of transfected HeLa during the sorting procedure, selecting cells with top 5% and top 20% of EGFP fluorescence: representative scatter plot (upper panel), fluorescent distribution plot of all living cells (middle panel) and fluorescent distribution of the selected populations with top 5% and top 20% EGFP fluorescence (lower panel). (C) Representative fluorescence microscope images of transfected HeLa clones one day after sorting.

192 single cells from the pool of living, transfected HeLa cells with top 5% or top 20% of EGFP fluorescence were sorted into 96-well plates and analysed for growth and EGFP expression over eleven days. As expected, top 5% EGFP sorted cells showed higher fluorescent signals compared to top 20% EGFP sorted cells, which was verified by more intense EGFP signals observed by fluorescence microscopy 24 hours after sorting (Figure 7-9 C). Clones from both sorting types were able to multiply and form colonies as exemplary shown in Figure 7-9 C and summarised in Table 7-1. Colony growth verified that the transfection and sorting conditions were sufficiently mild, hence, negligibly affected the HeLa cells. Clonal survival was marginally higher for top 20% EGFP sorted clones compared to top 5% EGFP sorted clones, of which 49% and 58% were able to divide. Of the dividing clones, 72% of the top 5% EGFP sorted clones and 89% of the top 20% EGFP sorted clones expanded into colonies with several thousand cells, indicating that the increased amount of transfected DNA was mildly toxic for HeLa cells. Approximately 4% of the wells with surviving cells contained multiple colonies and were therefore excluded from the screening test.

Table 7-1: Overview of the HeLa derived single cell clones potentially harbouring the desired TDG knockout with indications on the clonal growth and TDG verification outcome

Summary of the sorted HeLa derived single cell clones potentially harbouring the desired TDG knockout. HeLa cells were transfected with the knockout constructs pX330-SpCAS9-HF1_TDG+11012/+16905 and an EGFP reporter with the aim to delete the active site coding exon within the TDG gene by CRISPR-CAS9-mediated gene editing. Single cells were sorted into 96-well plates according to top 5% and top 20% of EGFP fluorescence (two plates per sorting condition) and the clonal growth was analysed eleven days after transfection. Afterwards, selected clones were screened for the desired TDG knockout on the genome and protein level. The absence of the TDG gene was validated by PCR and the absence of the full length TDG protein was verified by western blot analysis.

	Top 5% EGFP fluorescence		**Top 20% EGFP fluorescence**	
Sorted clones	**192**		**192**	
Empty wells	97	(51%)	80	(42%)
Wells with dividing cells	95	(49%)	112	(58%)
Wells with colonies	**95**		**112**	
Dead	23	(24%)	9	(7%)
Multiple	6	(3%)	8	(4%)
Single	66	(69%)	95	(85%)
Clones selected for validation	48		18	
Clones screened on genome level	**48**		**18**	
Clones without non-rearranged TDG	8	(17%)	0	(0%)
Clones screened on protein level	**8**			
Clones without TDG protein	6	(75%)		
Clones selected for HCR	2-F3 (and 2-C11)			

From the surviving single cell clones, 48 clones with top 5% of EGFP expression and 18 clones with top 20% of EGFP expression were expanded and analysed for TDG knockout on the gene level by PCR. To increase time and resource efficiency when testing high amounts of samples, a PCR assay was established which uses “quick “cell lysates of boiled HeLa cells as DNA template (depicted in Appendix I 2). “Quick “cell lysates were prepared from the 66 selected single cell clones as described in chapter 6.25 and *TDG* editing was assessed by PCR.

The primers TDG_10818F/11516R were used to detect the desired ~16,000 bp deletion within the *TDG* gene. PCR screening of the selected 66 single cell clones revealed that 19 clones contained rearranged *TDG* genes (Figure 7-10, lower panel, bold labelled clones). The expected 366 bp PCR product from the active site TDG deletion was more common in clones sorted for top 5% EGFP fluorescence compared to clones sorted for top 20% EGFP fluorescence, with 31% and 17% of positive clones respectively (Figure 7-10 and Table 7-1). It is therefore assumed, that the transfection efficiency and knockout efficiency correlate positively with each other. Genomic DNA (chapter 6.24) and “quick” cell lysate (chapter 6.25) from transfected, unsorted HeLa cells were used to validate the PCR performance and efficiently amplified the 366 bp PCR product (Figure 7-10, lower panel). As expected, the control sample without DNA

template, as well as samples containing gDNA or cell lysate of untransfected HeLa, did not amplify the *TDG* deletion product.

The absence of non-rearranged *TDG* alleles was verified in a second PCR using primers TDG_10818F/17076R. 64 out of the tested 66 single cell clones showed efficient PCR product amplification (Figure 7-10, upper gel). Interestingly, the fragment size did not always correspond to the expected size of 698 bp, indicating that for the outliers, *TDG* gene editing occurred in an unexpected way. The positive control (untransfected HeLa) showed efficient and selective product amplification from non-rearranged *TDG*, validating the primer specificity. Comparing the results from both PCR reactions, it was observed that six clones efficiently amplified rearranged *TDG* products but did not amplify DNA with the size expected for non-rearranged *TDG* products. It was thus assumed that clones 2-B4, 2-C11, 2-D6, 2-D11, 2-F3 and 2-F11 of top 5% EGFP fluorescent sorted cells potentially harbour the desired *TDG* deletion in all four alleles within their genome (Figure 7-10, labelled with asterisks). The *TDG* status of clones 2-C5 and 2-G12 (top 5% EGFP) was unclear, as no PCR product was detected in either reaction. None of the top 20% EGFP sorted clones demonstrated disruption of all wild-type *TDG* alleles based on the PCR results, indicating that the lower transfection efficiency was not sufficient to establish a complete TDG knockout in HeLa cells. Based on these results, only top 5% EGFP fluorescence is used as a selection criterion to generate HeLa knockout cell lines in the future.

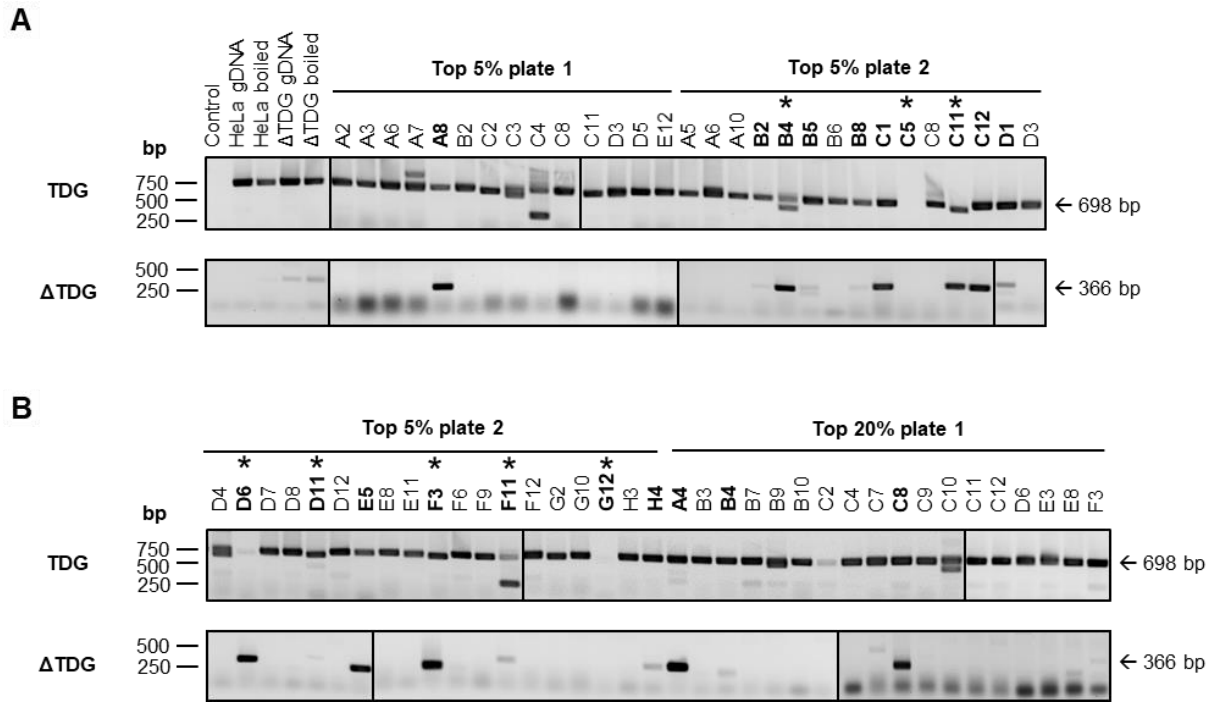


Figure 7-10: Validation of the TDG knockout in selected HeLa single cell clones on the genome level via PCR

(A+B) Screening of the selected HeLa derived single cell clones for the desired TDG knockout by PCR. 66 of the sorted single cell clones with potential TDG knockout were screened via PCR to detect rearranged- (lower panel) and non-rearranged TDG genes (upper panel). "Quick" cell lysates were prepared from the clones and used as a PCR templates to verify the deletion of the active site coding sequence within the TDG gene (366bp PCR fragment) with primers TDG_10818F/17076R (lower panels). A separate PCR was performed to detect non-rearranged TDG (698bp PCR fragment) using primers TDG_16227F/17076R (upper panels). The migration positions of the respective PCR products within the agarose gel are indicted by arrows. Clones amplifying rearranged TDG are indicated in bold and clones with absent non-rearranged PCR product amplification and clones which didn't amplify any product are marked with asterisks.

In the following step, the *TDG* status of the six single cell clones 2-B4, 2-C11, 2-D6, 2-D11, 2-F3 and 2-F11 (top 5% EGFP) harbouring potentially disruptions of all wild-type TDG alleles based on the PCR results and of clones C5, G12 (top 5% EGFP) was analysed on the protein level. Clones 1-A6 and 2-D8 were used as single cell clone controls without the desired gene knockout. Whole cell lysates from the eight selected clones, the parental HeLa cell line and an isogenic TDG knockdown cell line were used for western blot analysis as described under section 6.28. Protein detection showed, that no TDG detection between 60 kilodaltons (kDa) and 86 kDa for clones 2-B4, 2-C5, 2-C11, 2-F3, 2-F11 and 2-G12 (Figure 7-11, lane 6, 7, 8, 11, 12 and 13). Clones 2-D6 and 2-D11 showed TDG signals with smaller molecular weight than the full-length protein, indicating the presence of a truncated TDG variant (Figure 7-11, lane 9 and 10). As expected, TDG knockdown in the isogenic HeLa-derived cell line decreased the TDG signal detectably (Figure 7-11, lane 3), verifying the specific binding of the selected antibody to the TDG protein. Interestingly, clone 2-D8 - which did not show a characteristic TDG-knockout pattern in the PCR assay - contained truncated TDG protein only, indicating that gene editing took place although not at the expected

positions. Clone 1-A6 showed decreased amounts of full length TDG compared to TDG proficient HeLa cells. Equal signal intensities of the housekeeping protein heat shock protein 90 (HSP90) for all samples verifying that similar protein amounts were loaded for each sample.

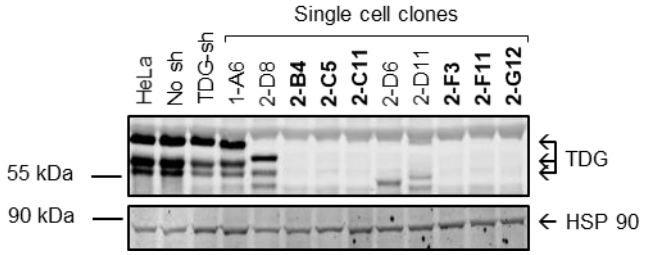


Figure 7-11: Validation of the TDG knockout in selected HeLa single cell clones by western blot analysis

Eight HeLa derived single cell clones with potential *TDG* knockout were screened for TDG protein absence by western blot analysis. Clones 2-B4, 2-C5, 2-C11, 2-D6, 2-D11, 2-F3, 2-F11 and 2-G12 with positive results in the PCR screen were selected for *TDG* knockout validation on the protein level, where the TDG signal was detected between 60-86 kDa in cell extracts. Extracts of parental HeLa cells, as well as a *TDG* knock down cell line (TDG sh) and the respective control (no sh) were used to examine the antibody specificity, whilst Clones 1-A6 and 2-D8 were used to verify potential clonal diversification. The housekeeping protein HSP90 with 90kDa is used as a loading control. The migration of the HSP90 and the TDG proteins within the acrylamide gel are indicted by arrows. Clones with absent TDG protein are indicated in bold.

Western blot analysis verified the successful generation of six HeLa derived single cell clones containing the desired TDG knockout. It was shown that the active site deletion within the *TDG* gene (~16,000 bp) completely abrogated the expression of the TDG protein, generating HeLa cells which were expected to have no residual TDG activity. Based on the presented data, the established transfection and sorting parameters proved to be successful in generating and selecting viable HeLa derived single cell clones with the desired gene knockout, a summary of which is shown in Table 7-1 (clonal growth and TDG knockout validation).

7.1.7 Transcriptional repression by BER of 5-fC and 5-caC in CRE is completely TDG-dependent

The HeLa derived TDG knockout clone 2-F3 was selected to assess the effects of 5-fC and 5-caC in CMV-1111 in order to determine the impact of TDG on BER-dependent transcriptional repression. TDG deficient and TDG proficient parental HeLa cells were transfected with pCMV-1111-ZA-W constructs containing 5-fC, 5-caC or the respective C control and the EGFP expression was monitored at 6- and 24 hours (chapter 6.32). As expected, 5-fC reduced the gene expression down to 25.3% of residual EGFP expression in TDG proficient HeLa cells (Figure 7-12 A). TDG knockout in the clonal cell line 2-F3 completely reverted the 5-fC induced transcriptional repression, showing

EGFP levels of 100% and 97.3% at 6- and 24 hours respectively. Analogously, the time-dependent transcriptional repression characteristically induced by 5-caC in the CMV promoter was absent in TDG deficient cells (Figure 7-12 B). EGFP levels of TDG knockout cells transfected with 5-caC containing reporters were comparable at the 6- and 24-hour time point, whereas the EGFP fluorescence in HeLa cell showed a strong time dependent reduction of 68.1%. These results were exactly reproduced in the TDG knockout clones 2-B4, 2-C5, 2-C11, 2-F3, 2-F11 and 2-G12 ruling out that clonal diversity influenced TDG-dependent transcriptional repression by 5-caC (Appendix I 3).

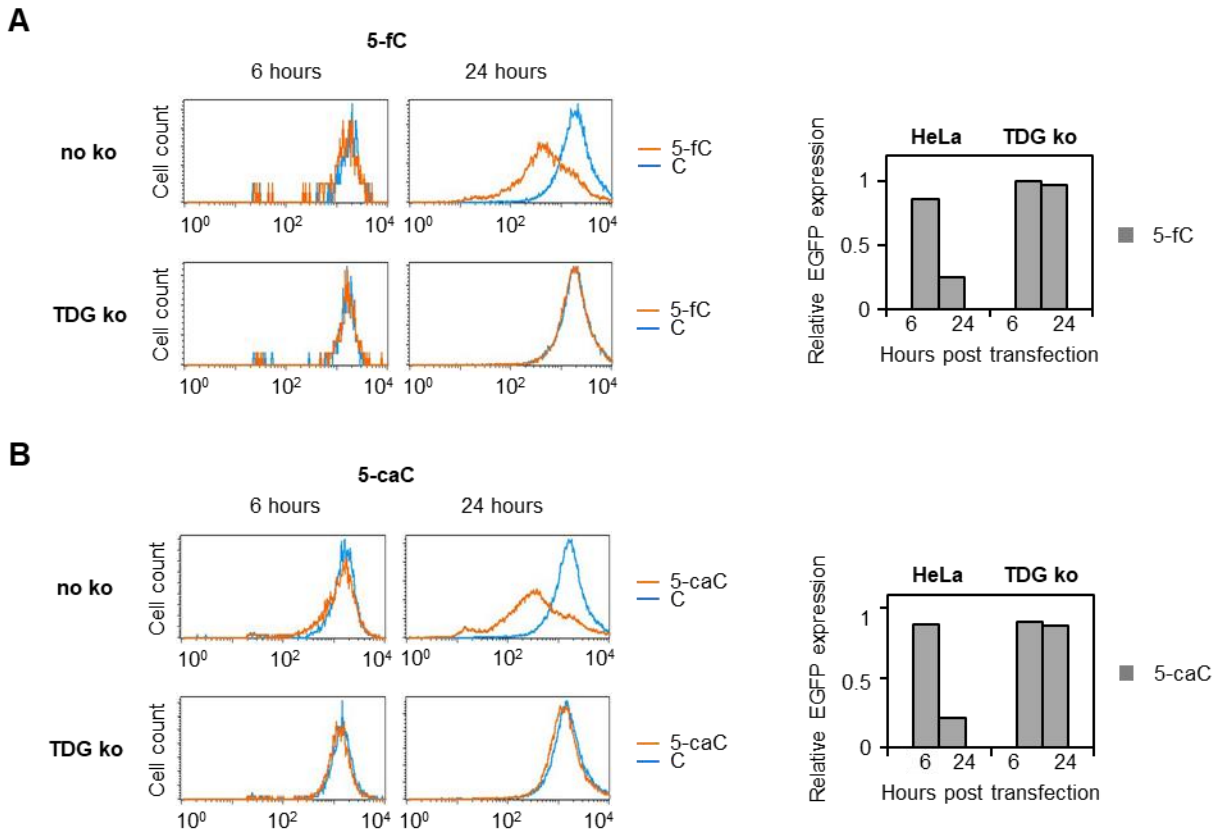


Figure 7-12: Expression of CMV-1111 reporters containing a single 5-fC or 5-caC at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter in *TDG* knockout cells

Quantitative EGFP expression analysis of the HeLa derived clonal *TDG* knockout cell line 2-F3 (TDG ko) and the isogenic TDG proficient HeLa cell line (no ko) transfected with pCMV-1111-ZA-W reporters containing single 5-fC (A) or 5-caC residue (B) at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter. (Left panel) Fluorescent distribution plots and (right panel) relative EGFP expression of transfected HeLa and HeLa derived TDG knockout cells 6- and 24 hours after transfection (n=1).

The EGFP expression analysis indicated, that transcriptional repression by a single 5-fC and 5-caC residue within the strong CMV promoter completely depended on TDG-mediated BER, since TDG knockout reverted the repressive phenotype. These results further verified, that incomplete recovery of the 5-fC and 5-caC induced

transcriptional repression under TDG knockdown conditions was caused by residual TDG activity in the TDG-sh cells (cf. chapter 7.1.4).

In summary, the presented experiments provide important insights into the interactions of the 5-mC oxidation products 5-fC and 5-caC with transcription in human cells. TDG plays an essential role in modulating the transcriptional consequences of 5-fC and 5-caC in CRE gene regulatory elements, changing from slight promoter inhibition by the primary base modifications to drastic transcriptional repression by BER. The repair intermediate which caused the transcriptional repression lied downstream from AP sites. Transcriptional repression was not only observed in the weak CRE-UNO-promoter but also in the strong CMV-1111 promoter, demonstrating that the repressive signal spread beyond the local modified CRE over the whole promoter, indicating the involvement of an epigenetic gene silencing mechanism.

7.1.8 Generation of minimal GC box and TRE reporters suitable to introduce DNA modifications into a standalone GC box or TRE upstream regulatory element

Quantitative expression analysis of CRE reporters showed that already a single 5-fC and 5-caC residue has an extremely negative effect on the gene expression controlled by GC-poor CRE-UNO and CMV-1111. To understand whether transcriptional repression is a CRE-specific consequence of 5-fC and 5-caC or a general trade of these DNA modifications, their effect on the gene expression was investigated in an additional gene regulatory element. The GC box was chosen to explore the impacts of 5-fC and 5-caC, because like CRE, its consensus sequence contains a central CpG dinucleotide which is methylation sensitive[87,90,91] and therefore a target of active DNA demethylation. Yet the GC box binds a different set of transcription factors than CRE and has a higher GC-content, making it an optimal representative GC-rich URE for the analysis of the functional consequences of 5-fC and 5-caC. The GC box gene regulatory element is also optimal to analyse the basic functional consequences of 8-oxoG in a GC-rich URE, because 8-oxoG in this sequence context has been reported to influence the activity of several complex PQS promoters (cf. chapter 4.2.2). Furthermore, the tetradecanoylphorbol-13-acetate responsive element (TRE) was considered as a URE in which to investigate the effects of 8-oxoG on gene expression. TRE was chosen because it binds a different set of transcription factors than CRE and GC box UREs and occurs in one-third of all human gene promoters[285].

To analyse the functional consequences of 5-fC, 5-caC and later on 8-oxoG in the GC box gene regulatory element, EGFP reporters needed to be designed first, in which the reporter expression is controlled by a minimal GC box promoter. The promoter was designed to consists of a standalone GC box gene regulatory element coding for the second most common GC box consensus sequence 5′-GGGCGGAG[286]. Its transcription activating effect is equivalent to the most prominent GC box consensus sequence 5′-GGGCGGGG, yet it is less prone to form uncontrollable non-canonical structures. As the GC box sequence is commonly found in both directions within human promoters[287] the orientation of the asymmetric GC box sequence within the reporters mimics those of DHFR gene, in which the activating function of GC box has been comprehensively characterised[288]. As set forth in materials and methods (cf. chapter 6.13) and exemplified in chapter 7.1.8, the CRE sequence of the pCRE-UNO-ZA-W and pCRE-UNO-ZA-C reporters was replaced for a GC box sequence by BsrDI mediated vector cloning (as depicted in Figure 7-13 A), generating pGCbox-ZA-W and pGCbox-ZA-C respectively. Due to the parental vector design the resulting GC box constructs contain Nb.BsrDI nicking sites with opposite orientations, allowing the selective replacement of the purine-rich (pGCbox-ZA-W) or the pyrimidine-rich strand of the GC box (pGCbox-ZA-C) respectively, verified by analytical digestion (Appendix I 4) and sequencing (data not shown). The newly generated pGCbox-ZA-W and pGCbox-ZA-C constructs (for sequences cf. Appendix II) were analysed for GC box functionality by quantitative reporter gene expression analysis in HeLa cells 24 hours after transfection (chapter 6.32). EGFP fluorescence analysis showed, that regardless of the orientation of the nicking sites, the GC box enhanced the gene expression by a factor of >2 with respect to the background expression level (pGCbox versus pCRE-Zero), thus confirming its function as an activating upstream element in both reporters (Figure 7-13 B, C, D).

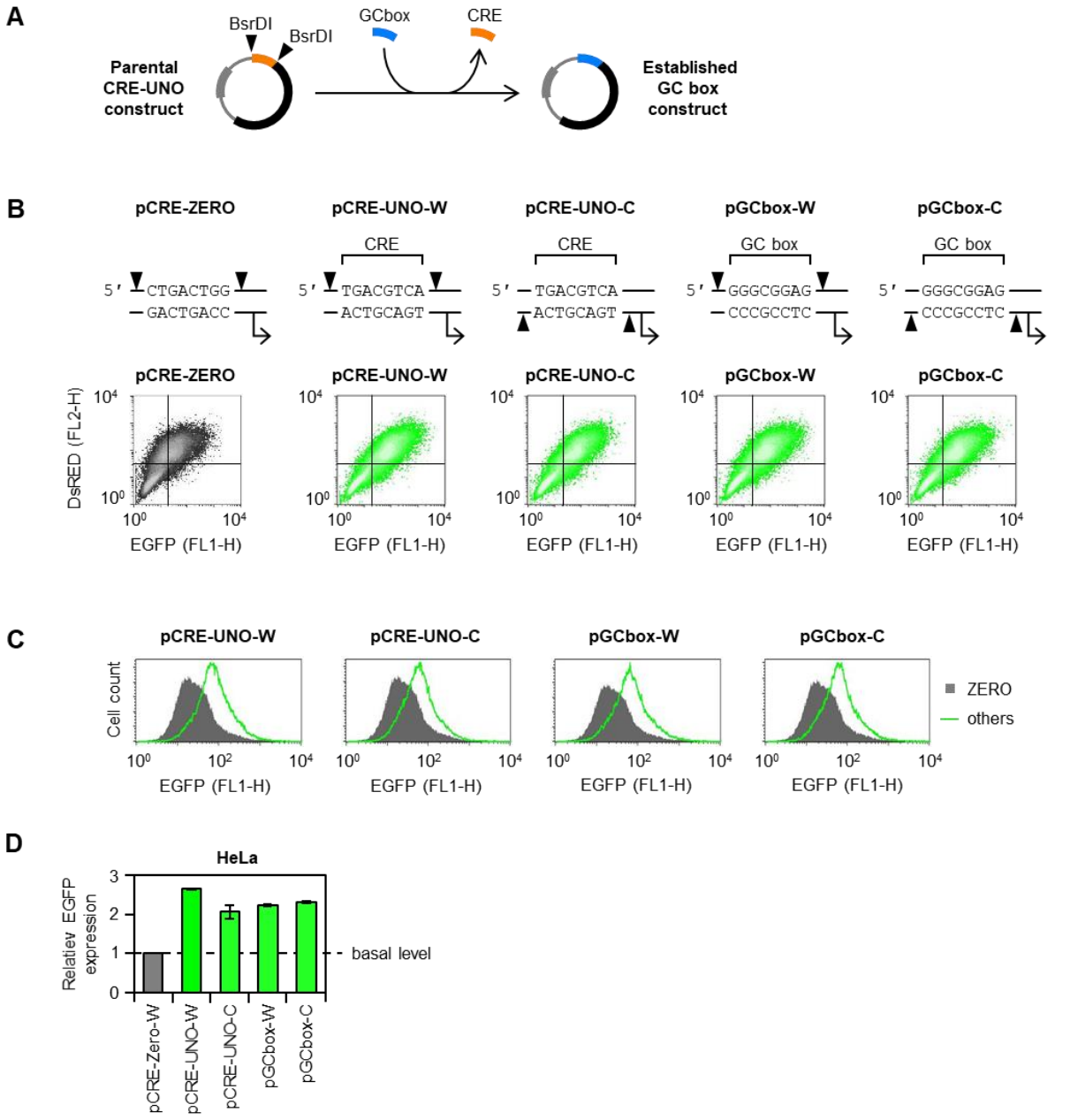


Figure 7-13: Construction and expression analysis of EGFP reporters that enable the site- and sequence-specific incorporation of DNA modifications in a minimal promoter consisting of a single GC box gene regulatory element

(A) Schematic representation of the experimental BsrDI cloning setup to generate a pair of GC box reporters in which the EGFP expression is under the control of a minimal GC box promoter from parental pCRE-UNO-ZA-W and pCRE-UNO-ZA-C constructs: EGFP gene (black half circle), CRE containing sequence (orange box), GC box containing insert (blue box), BsrDI restriction sites (black arrowheads). (B) Scheme of the reporter's URE, showing the DNA sequence, TSS (broken arrow) and Nb.BsrDI nicking sites (black arrowheads) above representative scatter plots of HeLa cells 24 hours after transfection. (C) Representative fluorescent distribution plots of HeLa cells 24 hours after transfection. (D) Quantification of the relative EGFP expression of transfected HeLa cells in three independent experiments (mean ± SD). Dotted line indicates the basal expression level from reporters without any activating URE (pCRE-Zero-W).

In parallel to the GC box reporters, a second set of EGFP coding reporters was generated, in which the reporter expression is controlled by a minimal TRE gene regulatory element. The minimal TRE promoter consists of a standalone TRE consensus sequence 5′-TGACTCA flanked a by short linker sequence. TRE is known to attract the prominent transcription factor activated by activating protein-1[289–293],

which consisting of homo- and heterodimers of basic region-leucine zipper proteins belonging to the JUN- and FOS-family. TRE is present in human genes in the forward and reverse direction by equal portions[285], wherefore a pair of reporter was designed which contained TRE once in reverse (pTRE-W with 5′-TGACTCA in transcribed strand) and once forward direction (pTRE-C with 5′-TGACTCA in the non-transcribed strand). The minimal TRE reporters were generated from CRE-UNO vectors by exchanging the standalone CRE site for a TRE gene regulatory element as depicted in Figure 7-14 A. Reporter generation followed the same cloning procedure as for the GC box (cf. chapter 6.13) and resulted in the generation of pTRE-ZA-W and pTRE-ZA-C (for sequences cf. Appendix II), verified by analytical digestion (Appendix I 4) and sequencing (data not shown). The vectors allow the selective replacement of TRE's purine-rich DNA strand for synthetic oligonucleotides through the strand exchange method described in chapter 6.30. The reporters can therefore be used to assess the functional consequences of a selected DNA modification contained in the synthetic oligonucleotide by reporter expression analysis in human cells.

EGFP fluorescence analysis of the newly cloned TRE reporters (chapter 6.32) showed that TRE activated the gene expression, with an enhancement-factor of 1.7 and 1.4 for pTRE-ZA-W and pTRE-ZA-C respectively (pTRE versus pCRE-Zero), thus confirming its function as an activating upstream element in both reporter vectors (Figure 7-14 B, C). Since the dynamic range of TRE promoters was considerably smaller than those of GC box promoters (enhancement-factor of 1.7 and 1.4 versus >2 for TRE and GC box reporters respectively), the GC box were selected to investigate the effects of 5-fC and 5-caC and at a later point also 8-oxoG on the gene expression controlled by non-CRE gene regulatory elements.

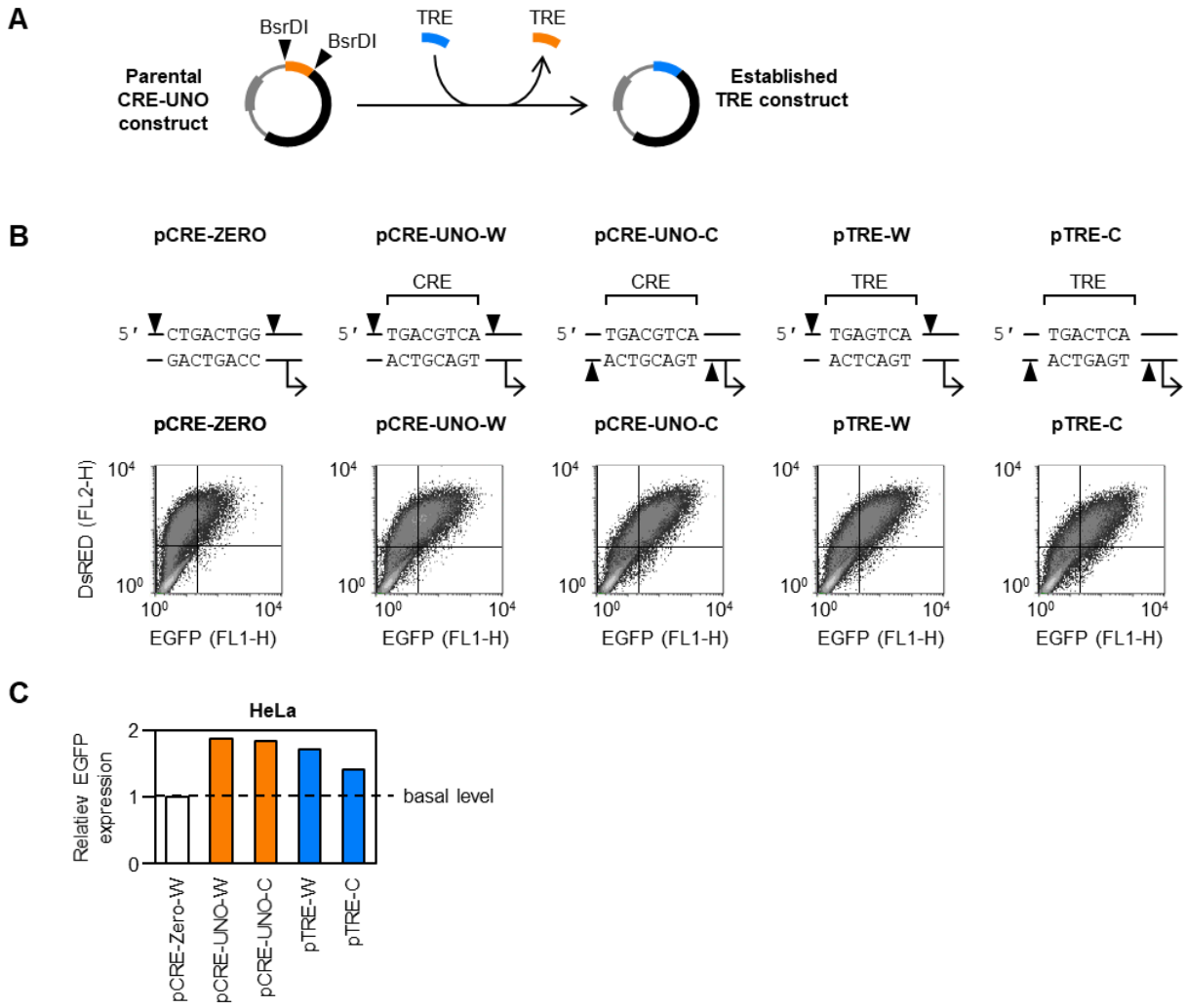


Figure 7-14: Construction and expression analysis of EGFP reporters that enable the site- and sequence-specific incorporation of DNA modifications in a minimal TRE promoter consisting of a single TRE gene regulatory element

A) Schematic representation of the experimental BsrDI cloning setup to generate a pair of TRE reporters in which the EGFP expression is under the control of a minimal TRE promoter from pCRE-UNO-ZA-W and pCRE-UNO-ZA-C constructs: EGFP gene (black half circle), CRE containing sequence (orange box), TRE containing insert (blue box), BsrDI restriction sites (black arrowheads). (B) Scheme of the reporter's URE, showing the DNA sequence, TSS (broken arrow) and Nb.BsrDI nicking sites (black arrowheads) above representative scatter plots of HeLa cells 24 hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells (n=1). Dotted line indicates the basal expression level from reporters without any activating URE (pCRE-Zero-W).

7.1.9 5-fC and 5-caC but not 5-mC and 5-hmC in the purine-rich strand of the GC box cause a time-dependent reduction of the gene expression

pGCbox-ZA-W and pGCbox-ZA-C plasmids were used to introduce 5-fC, 5-caC and their precursor DNA modifications 5-mC and 5-hmC into the central CpG nucleotide of the purine-rich and the pyrimidine-rich strand of the GC box consensus sequence respectively (Figure 7-15 A). Afterwards, the impact 5-mC, 5-hmC, 5-fC or 5-caC on the GC box activity was quantified by reporter transfection into HeLa cells and subsequent expression analysis.

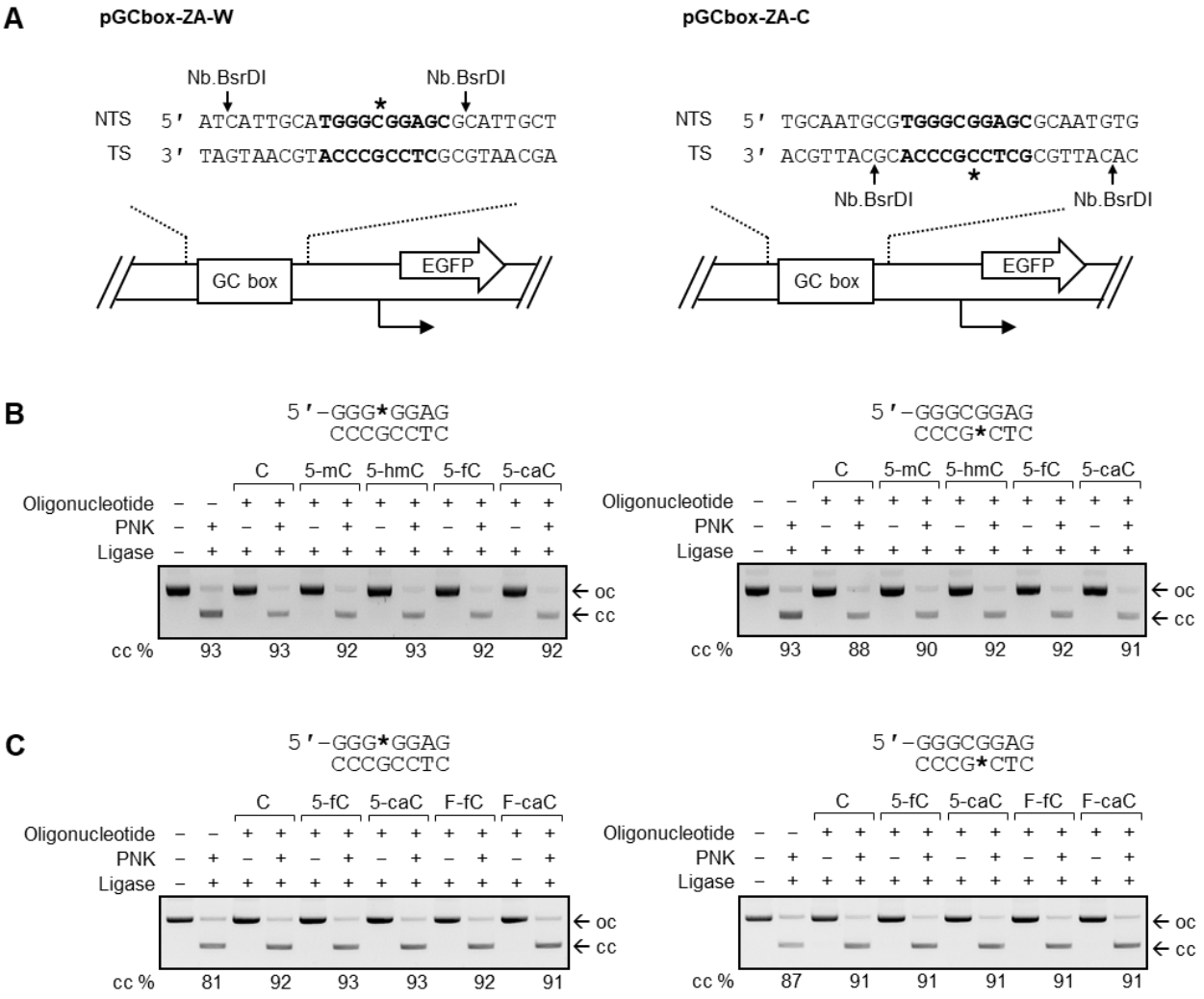


Figure 7-15: Construction of GC box reporters containing a single 5-mC, 5-hmC, 5-fC, 5-caC, F-fC or F-caC at the central CpG dinucleotide within either strand of the standalone GC box gene regulatory element

(A) Scheme of the standalone GC box gene regulatory element within pGCbox-ZA-W (left panel) and pGCbox-ZA-C reporters (right panel) used for the introduction of cytosine modifications at the central CpG dinucleotide of the purine- (pGCbox-ZA-W) and pyrimidine-rich strand (pGCbox-ZA-C) of the GC box: EGFP coding sequence (white arrow), TSS (broken arrow), GC box sequence (bold), Nb.BsrDI nicking sites (black arrows) and position of 5-mC, 5-hmC, 5-fC, 5-caC, F-fC and F-caC in the incorporated oligonucleotides (asterisk). (B+C) Ligation of the Nb.BsrDI-nicked GC box constructs with synthetic oligonucleotides containing the specified BER-sensitive (B) and BER-resistant (C) cytosine modification or the C control in the presence and absence of PNK.

Modified GC box reporters were generated following the strand exchange protocol as described under section 6.30, using synthetic oligonucleotides listed in Table 6-29 that contained a single 5-mC, 5-hmC, 5-fC, 5-caC or the respective C control residue at the central CpG dinucleotide of the GC box sequence (Figure 7-15 A, B). In addition to these reporters, expression constructs containing single BER-resistant F-fC and F-caC were generated, in order to investigate the transcriptional impact of BER (Figure 7-15 A, C). Analytical agarose gel electrophoresis (cf. chapter 6.12) of ligated reporters showed that a fraction of approximately 90% of all plasmids was present in their cc-form, verifying that both pGCbox-ZA-W and pGCbox-ZA-C were efficiently nicked and ligated with the synthetic oligonucleotides (Figure 7-15 B, C). There was no detectable difference comparing the ligation efficiency of modified synthetic oligonucleotides versus the C control, proving that single 5-mC, 5-hmC, 5-fC, 5-caC, F-fC and F-caC did not detectably influence the ligation efficiency. Since no GC box or base modification specific enzyme was available for analytic digestion assays, MALDI-TOF mass spectrometry analysis confirming the presence of desired DNA modifications in the synthetic oligonucleotides together with efficient outcompetition of excised DNA fragments were used as indicators for the successful generation of 5-mC, 5-hmC, 5-fC, 5-caC, F-fC and F-caC containing GC box reporters.

The modified pGCbox-ZA-W constructs were used to investigate the impact of BER-sensitive 5-mC, 5-hmC, 5-fC and 5-caC in the purine-rich strand of the GC box on the reporter activity. Quantitative EGFP expression analysis of transfected HeLa cells (chapter 6.32) showed that 5-fC and 5-caC had a strong negative effect on the reporter activity over 48 hours (Figure 7-16). The decline of the EGFP expression was time-dependent for both modifications, with 23% and 26% of residual EGFP fluorescence after 48 hours for 5-fC and 5-caC containing reporters respectively. It should be noted, that the slope of the transcriptional repression induced by 5-fC was steeper than that of 5-caC. Interestingly, both modifications inhibited the reporter activity already 6 hours after transfection, with a more pronounced effect for 5-caC compared to 5-fC (reducing the EGFP fluorescence to 87% and 75% respectively). In contrast, single 5-mC and 5-hmC did not affect the GC box activity over 48 hours.

The EGFP expression analysis verified, that single 5-fC and 5-caC residues in the purine-rich strand of the GC box had a negative effect on the gene expression which intensified with time. 5-mC and 5-hmC did not influence the gene expression.

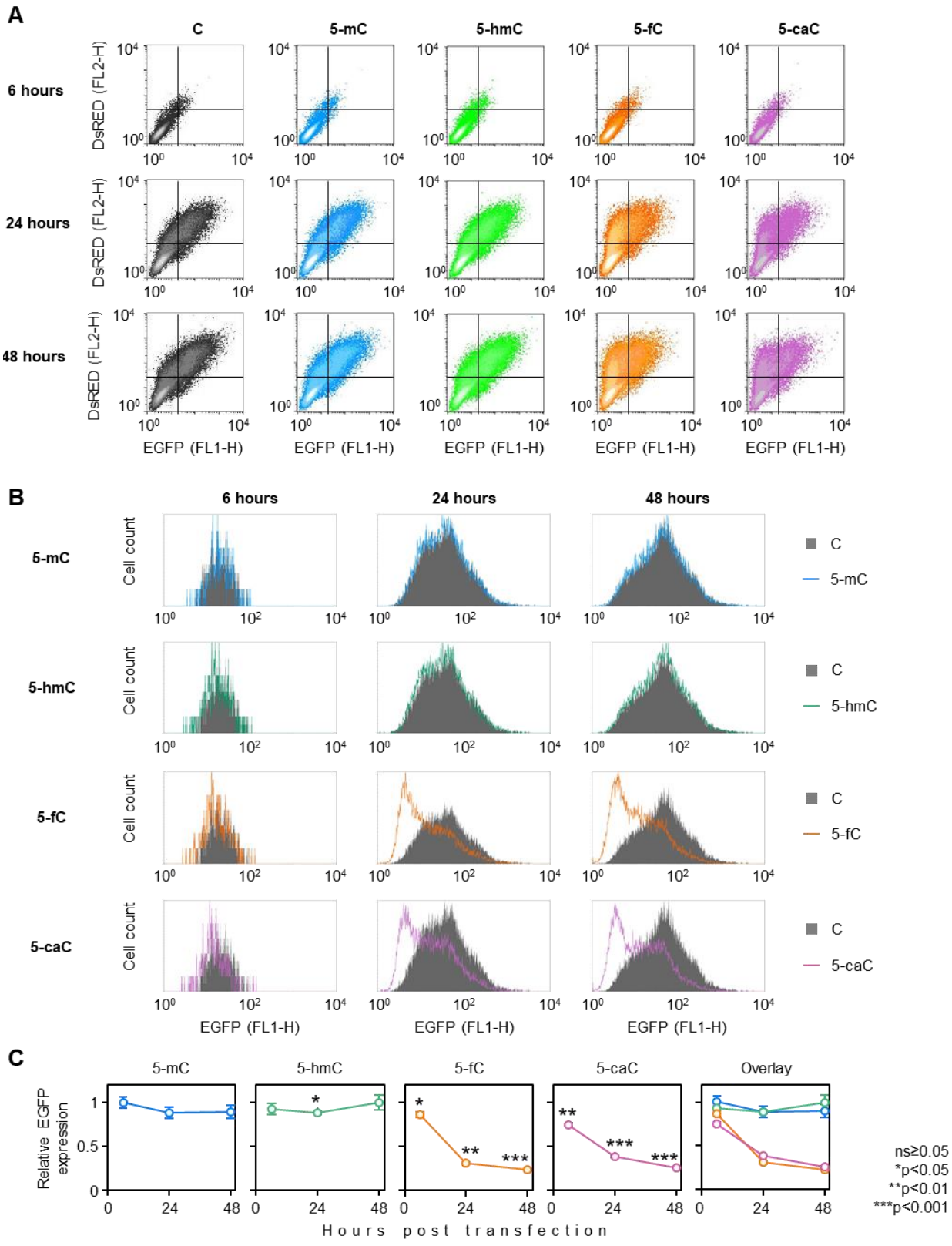


Figure 7-16: Expression of GC box reporters containing a single 5-mC, 5-hmC, 5-fC or 5-caC residue at the central CpG dinucleotide within the purine-rich strand of the standalone GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W reporters containing either a single 5-mC, 5-hmC, 5-fC, 5-caC or the C control at the central CpG dinucleotide within the purine-rich strand of the GC box upstream regulatory element. Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6-, 24- and 48-hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in three independent experiments (mean ± SD).

7.1.10 Transcriptional repression by 5-fC and 5-caC in the GC box is independent from the DNA strand

In the next step it was investigated, whether the impact of the four cytosine modifications on the gene expression was DNA strand biased. To answer this question, Hela cells were transfected with pGCbox-ZA-C reporters containing 5-mC, 5-hmC, 5-fC or 5-caC in the pyrimidine-rich strand of the GC box and the EGFP expression was monitored over 48 hours (chapter 6.32). Quantitative EGFP expression analysis showed, that the consequences of the four cytosine modifications in the pyrimidine-rich strand of the GC box exactly recapitulated their consequences in the purine-rich DNA strand. Thus, 5-fC and 5-caC in the pyrimidine-rich strand of the GC box negatively affected the gene expression in a time-dependent manner (Figure 7-16), whilst 5-mC and 5-hmC had no detectable effect.

In summary, the EGFP expression analysis verified, that 5-fC and 5-caC progressively repressed the GC box driven reporter gene expression in a strand independent manner. 5-mC and 5-hmC did not affect the GC box activity and showed similar expression profiles in purine-rich and pyrimidine-rich strand.

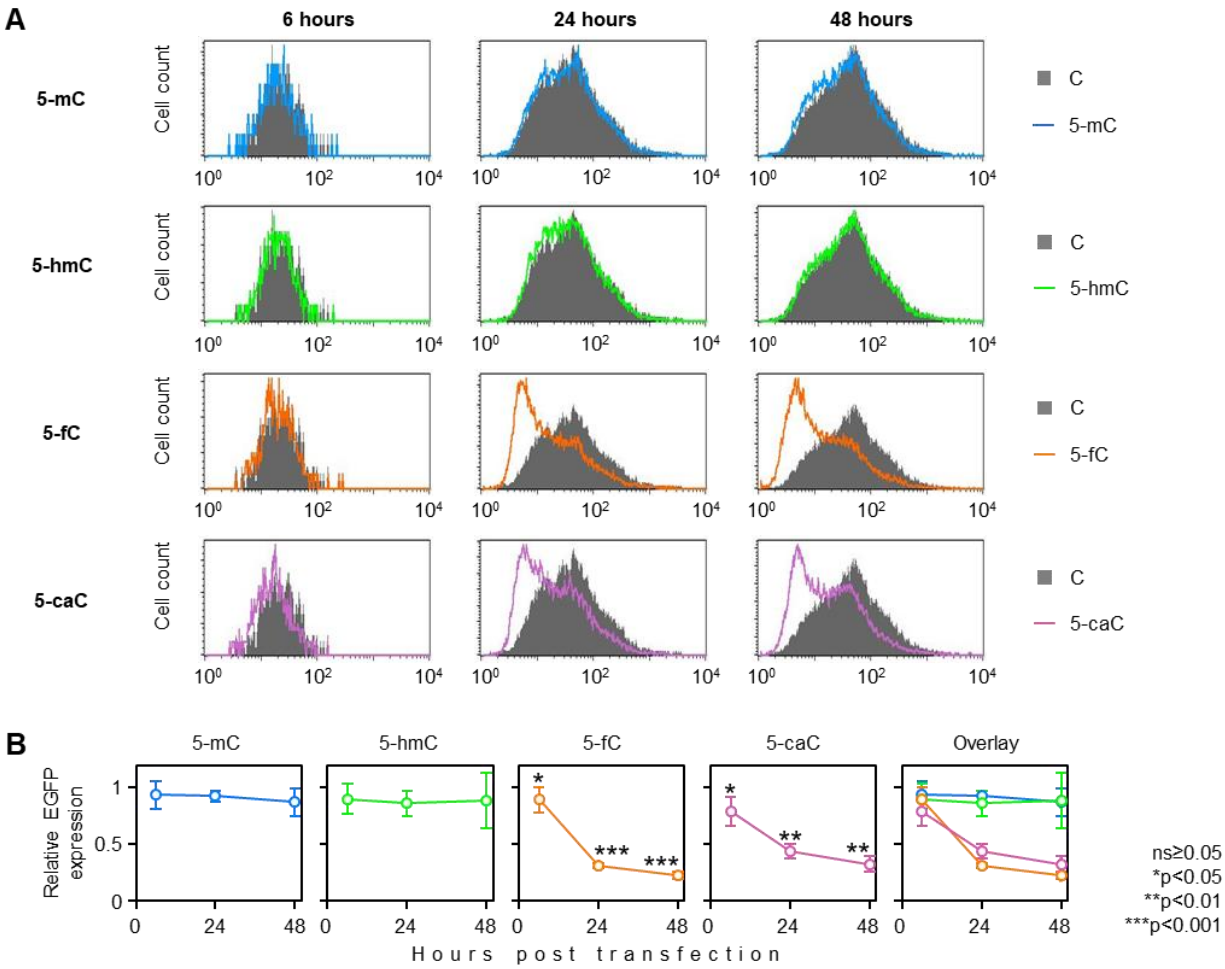


Figure 7-17: Expression of GC box reporters containing a single 5-mC, 5-hmC, 5-fC or 5-caC residue at the central CpG dinucleotide within the pyrimidine-rich strand of the standalone GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-C reporters containing either a single 5-mC, 5-hmC, 5-fC, 5-caC or the C control at the central CpG dinucleotide within the pyrimidine-rich strand of the GC box upstream regulatory element. (A) Representative fluorescent distribution plots of HeLa cells 6-, 24- and 48 hours after transfection. (B) Quantification of the relative EGFP expression of transfected HeLa cells in three independent experiments (mean ± SD).

7.1.11 Inhibition of BER abolishes the transcriptional repression induced by 5-fC and 5-caC and reveals a direct negative effect of 5-caC on the GC box activity

To assess whether BER contributed to the time-dependent reduction of the gene expression by 5-fC and 5-caC in the GC box as previously shown for CRE, 2′-fluorinated derivatives were analysed for their functional consequences in the GC box. Reporters containing single 5-fC, 5-caC, their BER-resistant 2′-fluorinated derivatives F-fC, F-caC or the respective C control in either strand of the GC box (generation described in chapter 7.1.8) were transfected into HeLa cells followed by quantitative EGFP expression analysis (chapter 6.32).

Expression analysis of pGCbox-ZA-W showed that BER-resistant F-fC in the purine-rich strand of the GC box had a small negative effect on the gene expression resulting

in a stably decreased expression of 83% at 24 hours after transfection(Figure 7-18 A, C left panel). A similar level of EGFP expression was observed for BER-sensitive 5-fC at 6 hours after transfection. As seen before, the EGFP expression of pGCbox-ZA-W constructs containing BER-sensitive 5-fC decreased drastically at 24 hours, down to 31% of residual EGFP fluorescence. BER-sensitive 5-caC also reduced the gene expression from 80% at 6 hours down to 37% at 24 hours after transfection. Strikingly, elimination of BER not only reverted transcriptional repression by 5-caC but also revealed a direct negative effect of 5-caC on the gene expression, resulting in approximately 40% of residual EGFP fluorescence. The expression reduction by F-caC was already detectable 6 hours after transfection, indicating that the modified base directly inhibited the GC box activity. Unexpectedly, EGFP from GC box reporters containing 5-caC was expressed twice as much as EGFP from F-caC containing reporters at the 6 hour time point. Since BER-inhibition by 2′-fluorination was designed to selectively inhibit base removal while leaving other biological processes unaffected[294], the different functional outcomes of F-caC and 5-caC were solely attributed to differential BER-processing of the base modifications. It is therefore anticipated that promoter inhibition by F-caC is an effect of the primary base modification, whilst 5-caC consequences are a combination of modification- and repair-dependent effects. Consequently, also the initially higher expression in 5-caC containing GC box reporters compared to F-caC was caused by BER, reverting the inhibitory effect of the primary base modification and reactivating the promoter activity.

Quantitative EGFP expression analysis of pGCbox-ZA-C transfected cells showed, that BER-resistant F-caC stably reduced the gene expression down to ~56% of residual EGFP expression (Figure 7-18 B, C right panel). The repression observed for F-caC in the pyrimidine-rich strand was significantly weaker than in the purine-rich strand. Interestingly, BER-resistant F-fC in the pyrimidine-rich strand of the GC box did not significantly alter the gene expression, which differs from results in the opposite strand. The effects of BER-sensitive 5-fC and 5-caC on the gene expression were exactly reproduced in the purine-rich strand showing progressive transcriptional repression by both modifications. Again, EGFP expression of 5-caC containing reporters was significantly higher than the respective BER-resistant counterpart, with a difference of 26% (83% EGFP expression by 5-caC versus 57% EGFP expression by F-caC), indicating that repair of 5-caC initially enhanced the gene expression.

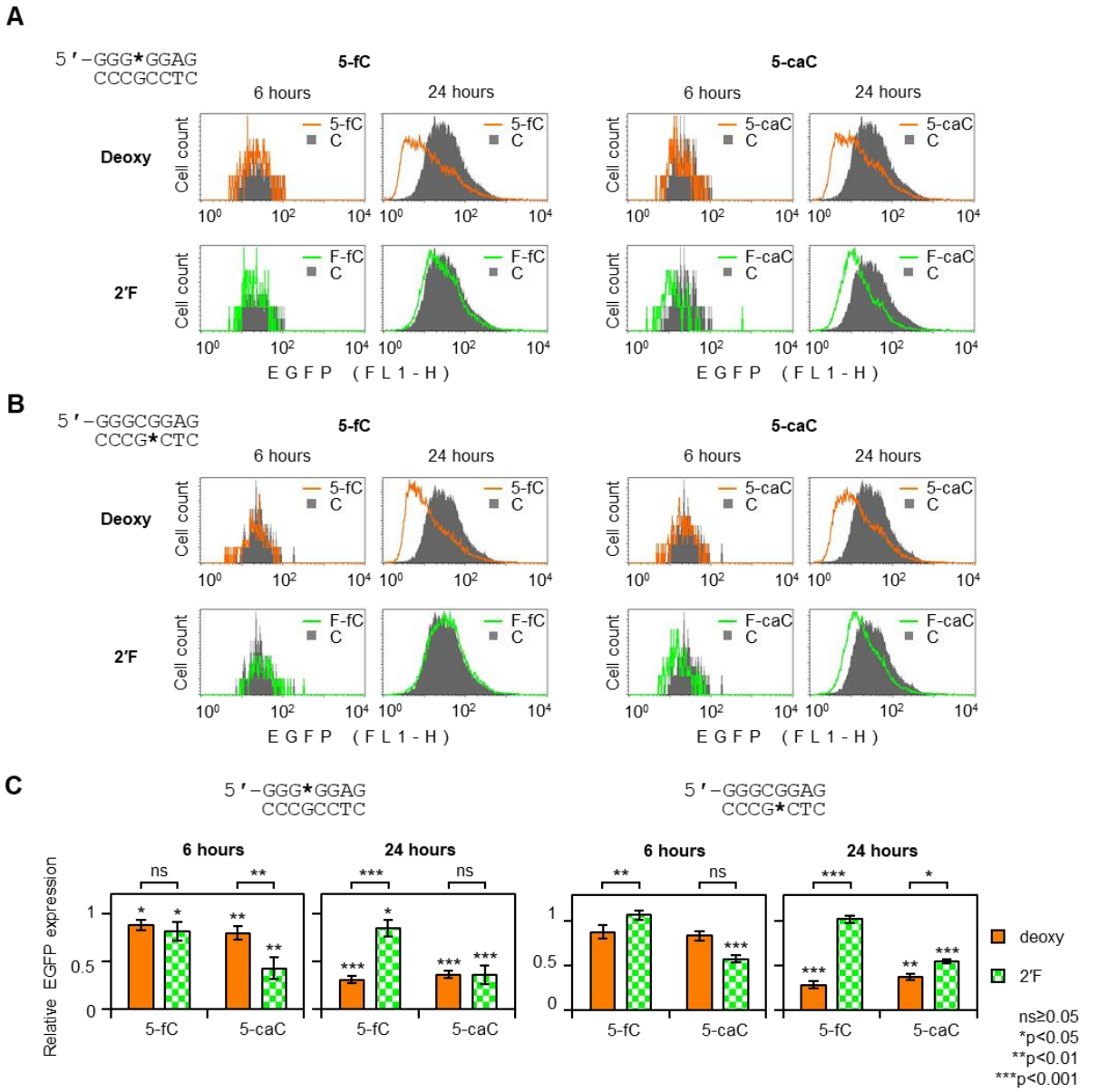


Figure 7-18: Expression of GC box reporters containing BER-resistant or BER-sensitive 5-fC and 5-caC in either strand of the central CpG dinucleotide of the GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W and pGCbox-ZA-C reporters containing single BER-sensitive 5-fC and 5-caC or their BER-resistant 2′F derivatives at the central CpG dinucleotide within the purine-rich strand (A+C left panel) or pyrimidine-rich strand (B+C right panel) of the standalone GC box upstream regulatory element: position of the modified base (asterisk). (A+B) Representative fluorescent distribution plots and (C) quantification of the relative EGFP expression of transfected HeLa cells 6- and 24 hours after transfection in four independent experiments (mean ± SD).

In summary, the EGFP expression analysis showed that in a minimal GC box promoter 5-fC per se had a small negative effect on the gene expression, which was only detectable in the purine-rich strand of the GC box. The negative effect of 5-caC as primary base modification was much stronger than the promoter inhibition by 5-fC and was detectable in both DNA strands with a significant bias towards the purine-rich DNA strand. Strikingly, BER of 5-caC reactivated the GC box activity, indicating that repair of 5-caC initially enhances transcription of the affected gene. Reactivation of 5-caC

containing promoters was superseded later on by time-dependent transcriptional silencing. Transcriptional repression was also observed for BER-sensitive but not BER-resistant 5-fC, indicating that the silencing mechanism depends on BER.

7.1.12 TDG knockdown does not detectably rescue the transcriptional repression induced by BER of 5-fC and 5-caC in the GC box

The BER-dependent transcriptional repression observed for 5-fC and 5-caC in the GC box strongly resembled their negative functional consequences in CRE, which were induced by TDG-initiated BER. Due to outcome analogy, it was investigated whether TDG was equally essential for BER-mediated transcriptional repression by 5-fC and 5-caC in the GC box. For this reason, the consequences of GC box positioned 5-mC, 5-hmC, 5-fC and 5-caC were investigated in TDG proficient HeLa cells (TDG knockdown cells). HeLa cells harbouring a stable SMUG and UNG knockdown were used to verify the impact of other uracil DNA glycosylases on 5-fC and 5-caC repair.
Quantitative expression analysis of reporters (chapter 6.32) containing a single cytosine modification in the purine-rich strand of the GC box showed a 78% and 73% decrease of the EGFP expression by 5-fC and 5-caC in control no sh, as well as SMUG sh and UNG sh cell lines (Figure 7-19). TDG knockdown did not detectably affect the transcriptional repression induced by 5-fC and 5-caC, with an insignificant increase of the EGFP fluorescence of 13% and 11% 24 hours after transfection (TDG sh versus no sh). Similar results were obtained in the pyrimidine-rich strand of the GC box, where 5-fC and 5-caC repressed the transcription by approximately 70% and 60% in all cell lines. Knockdown of TDG, SMUG1 and UNG did not significantly affect the expression of 5-mC and 5-hmC containing GC box reporters.

The discrepancy in the influence of TDG knockdown on the effects of 5-fC and 5-caC in CRE[97] versus GC box promoters (significant weakening of the transcriptional repression versus no significant effect of TDG-sh) can be explained by the following hypotheses.
Hypothesis-I: TDG is essential for BER-dependent reduction of the gene expression by 5-fC and 5-caC in both CRE and GC box promoters, however the TDG-dependency in the GC box was not detectable in the given assay. A possible reason for this detection inability is that the dynamic range of EGFP expression by GC box constructs was not big enough to determine the effect of TDG knock down on 5-fC and 5-caC

induced transcriptional repression. Weaker activation of the gene expression by the GC box activating upstream element compared to CRE supports this hypothesis (cf. chapter 7.1.8). Another potential reason is that the remaining TDG activity in TDG-sh cells was sufficient to successfully initiate BER of 5-fC and 5-caC and cause transcriptional repression, thus no detectable difference was observed in TDG-deficient and TDG proficient cells (TDG-sh versus no sh). To abolish any residual TDG activity, the previously generated TDG knockout cells can be used for HCR. Alternatively, the dynamic range of GC box promoters can be enhanced by substituting one CRE site in the strong CMV-1111 promoter for a GC box. 5-fC and 5-caC could be introduced into such CMV-1111 derived GC box reporter by the strand exchange method (chapter 6.30) and the generated expression constructs can be used for transfection of TDG knockdown cells.

Hypothesis-II: The inability to detect an effect of TDG knockdown on the transcriptional repression by BER of 5-fC and 5-caC could be for the simple reason that TDG is not essential for the transcription inhibition mechanism. In that case, base removal of 5-fC and 5-caC from the GC box during BER is accomplished by an alternative DNA glycosylase which acts on the modified bases either in addition, or instead of TDG. In this case, the effects of 5-fC and 5-caC on gene expression need to be assessed in the complete absence of TDG to distinguish if TDG was irrelevant for 5-fC and 5-caC base removal or if there was a backup-DNA glycosylase for TDG. If no impact of TDG absence on the transcriptional consequences of 5-fC and 5-caC is determined, the activity of the other 10 human DNA glycosylases needs to be abolished in turns and analysed for their effects on transcriptional repression by 5-fC and 5-caC in the GC box.

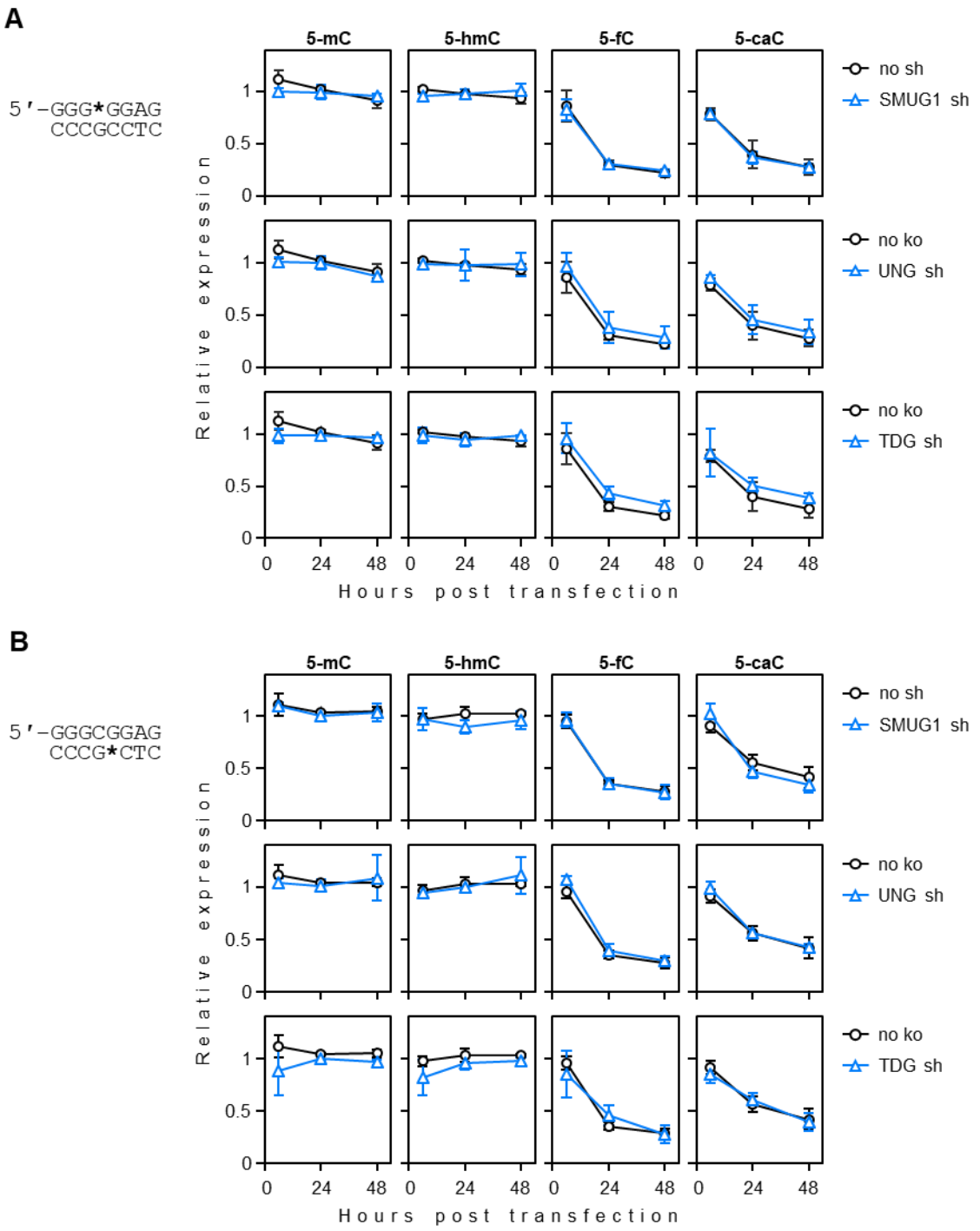


Figure 7-19: Expression of GC box reporters containing 5-mC, 5-hmC, 5-fC or 5-caC in either strand of the central CpG dinucleotide of the GC box gene regulatory element in the TDG-, SMUG1- and UNG knockdown cells

Quantitative EGFP expression analysis of HeLa cells with unaltered (no sh) and decreased TDG-, SMUG- or UNG-levels (TDG sh, SMUG sh and UNG sh respectively) transfected with pGCbox-ZA-W (A) or pGCbox-ZA-C reporters (B) containing single 5-mC, 5-hmC, 5-fC or 5-caC at the central CpG dinucleotide within the purine-rich strand (A) or pyrimidine-rich strand (B) of the standalone GC box upstream regulatory element: position of the modified base (asterisk). Quantitative analysis of the relative EGFP expression values of transfected cells 6-, 24- and 48 hours after transfection in three independent experiments (mean ± SD).

7.1.13 Transcriptional repression by BER of 5-fC and 5-caC in GC box is completely TDG-dependent

To examine more deeply whether TDG was essential for the transcriptional repression by 5-fC and 5-caC in the GC box, the sensitivity of the host cell reactivation assay towards determining TDG's effects was improved. This can be achieved by either using stronger GC box promoters or TDG knockout- instead of TDG knockdown cells. Since TDG knockout cells were already at hand, they were used to examine the functional consequences of 5-fC and 5-caC in the GC box by reporter transfection and subsequent EGFP expression analysis. The two TDG knockout cell lines ΔTDG-2-C11 and ΔTDG-2-F3 and two isogenic TDG proficient cell lines (parental HeLa and clonal HeLa derived ΔNTHL1) were transfected with GC box reporters containing 5-mC, 5-hmC, 5-fC and 5-caC in either DNA strand and the EGFP expression was monitored over 48 hours (chapter 6.32). The NTHL1 DNA glycosylase is uninvolved in 5-fC and 5-caC repair, wherefore NTHL1 knockout can be used to control that any effects on 5-fC and 5-caC induced GC box silencing observed in TDG knockout cells is specific for TDG absence. Knockout of NTHL1 was performed following the optimised version of the "CRISPR: Protocol for Genomic Deletions in Mammalian Cell Lines"[281] as described for TDG (cf. chapters 7.1.5 and 7.1.6), with assistance by master student Aalaa Farag (for knockout details see[295]). PCR and western blot analysis verified the efficient NTHL1 knockout in three HeLa clones[295] and clone 1-F5 was selected for the following expression analysis.

Quantitative expression analysis of pGCbox-ZA-W transfected cells showed, that the relative EGFP fluorescence of TDG deficient cells transfected with 5-fC and 5-caC containing constructs remained stable over 48 hours with a variation of < 9%. (Figure 7-20). At the same time, the EGFP expression values of TDG proficient cells declined rapidly over 48 hours with an approximate magnitude of ~55% (5-fC) and ~50% (5-caC). The presence of transcriptional repression caused by 5-fC and 5-caC in TDG proficient but not TDG deficient cells demonstrates, that TDG is essential for establishing the transcriptional silencing. These results thus prove that the residual TDG activity of TDG-sh cells was the reason why TDG effects on the transcription reduction by 5-fC and 5-caC could not be verified in the HCR assay with TDG knockdown cells earlier (cf. chapter 7.1.12). The expression analysis of GC box reporters in TDG knockout cells also showed, that BER inhibition by TDG knockout did

not completely revert the repressive effects of 5-caC on the gene expression, since it resulted in constantly decreased EGFP fluorescence of approximately 59%. The reduced EGFP levels in TDG deficient cells demonstrate that 5-caC per se has a negative effect on the reporter activity as observed earlier (cf. chapter 7.1.11). The expression of 5-caC containing reporters in TDG proficient cells 6 hours after transfection was ~20% higher than the expression in TDG deficient cells, (EGFP expression of 84% and 80% for HeLa and ΔNTHL1-1-F5 versus 67% and 57% for ΔTDG-2-C11 and ΔTDG-2-F3 respectively). This significant expression enhancement of 5-caC containing GC box reporters in TDG proficient cells compared to TDG deficient cells, indicates that removal of the 5-caC base by TDG initially reactivated the GC box promoter. A similar reactivation was observed earlier when comparing the expression of GC box reporters containing a BER-resistant versus BER-sensitive 5-caC derivative (cf. chapter 7.1.11). 5-fC in TDG deficient cells inhibited the reporter activity to a much smaller extend than 5-caC, with an approximate residual EGFP expression of 85%. As expected, single 5-mC and 5-hmC residues in the purine-rich strand of the GC box did not excessively alter the gene expression in any cell line. No significant transcriptional differences were observed comparing HeLa with NTHL1-deficient cells and ΔTDG-2-C11 with ΔTDG-2-F3, indicating that clonal diversification did not affect reporter activity.
The effects of 5-mC, 5-hmC, 5-fC and 5-caC in the purine-rich strand of the GC box (cf. above) were exactly reproduced in the pyrimidine-rich strand of the GC box, again verifying the absence of any DNA strand bias.

In summary, expression analysis of GC box reporters in TDG knockout cells showed that removal of the 5-fC and 5-caC base by TDG is essential to establish transcriptional repression. TDG is equally important to initially reactivate the GC box activity by BER of 5-caC. Furthermore, the primary base modifications inhibit the GC box activity in the absence of repair, an effect which is stronger for 5-caC than 5-fC. This dual mechanism of transcription regulation by 5-fC and 5-caC in the GC box was not only observed by BER-inhibition using TDG knockout cells, but also using chemically stabilised BER-resistant derivatives of 5-fC and 5-caC.

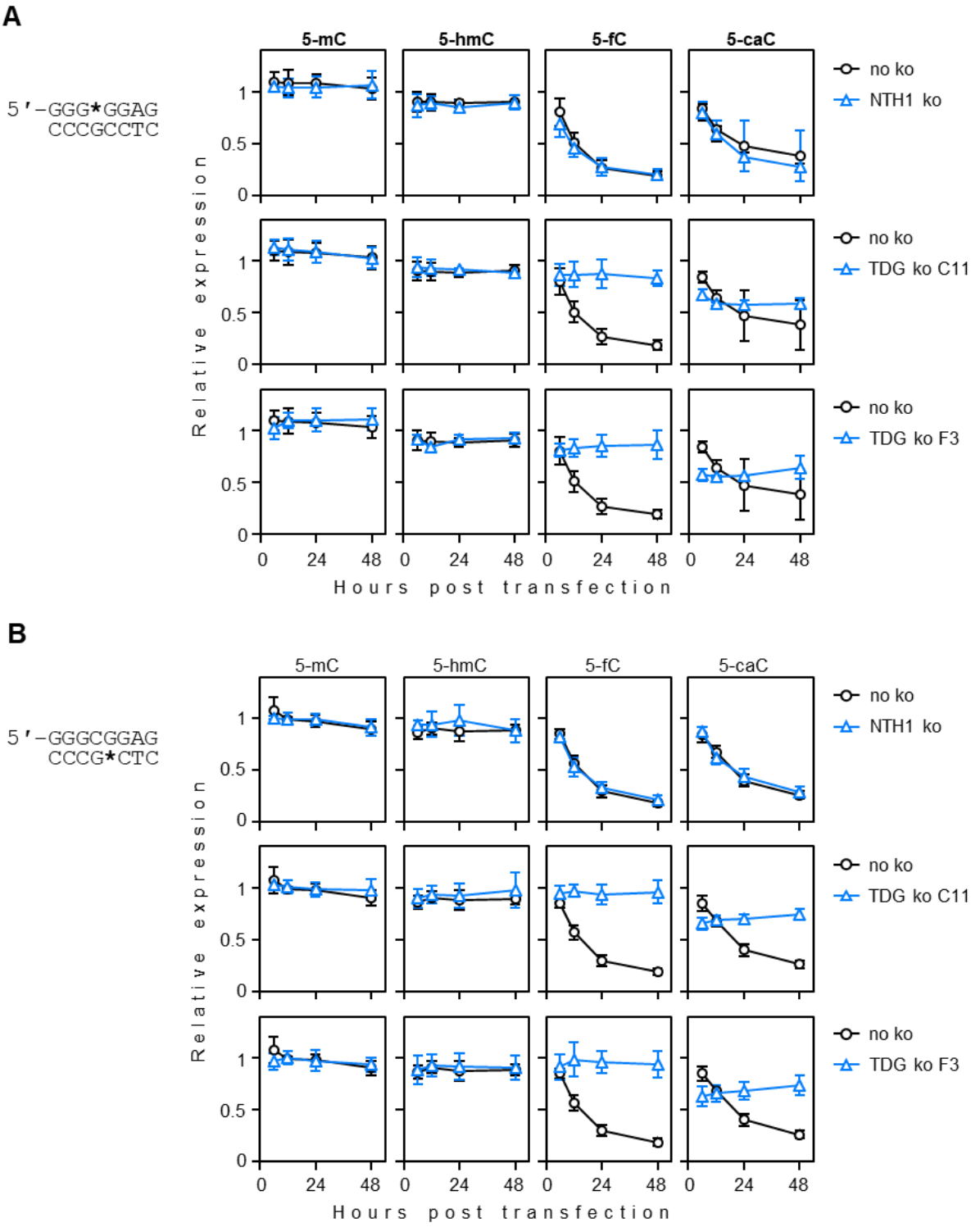


Figure 7-20: Expression of GC box reporters containing 5-mC, 5-hmC, 5-fC or 5-caC in either strand of the central CpG dinucleotide of the GC box gene regulatory element in the TDG knockout cells

Quantitative EGFP expression analysis of TDG proficient (no ko and NTH1 ko) and TDG deficient HeLa cells (TDG ko clones 2-C11 and 2-F3) transfected with pGCbox-ZA-W (A) or pGCbox-ZA-C reporters (B) containing single 5-mC, 5-hmC, 5-fC or 5-caC at the central CpG dinucleotide within the purine-rich strand (A) or pyrimidine-rich strand (B) of the standalone GC box upstream regulatory element: position of the modified base (asterisk). Quantification of the relative EGFP expression of transfected cells 6-, 24- and 48 hours after transfection in three independent experiments (mean ± SD).

7.1.14 TDG is the only DNA glycosylase to initiate BER of 5-fC and 5-caC in the GC box and induce transcriptional repression

To assess if TDG is the only DNA glycosylase which removes 5-fC and 5-caC from the GC box and thereby causing transcriptional repression, the functional consequences of BER-sensitive and BER-resistant base modifications were analysed in TDG deficient cells. The experimental setup is based on the assumption that in TDG-ko cells removal of 2′-deoxy 5-fC and 5-caC is only possible if a DNA glycosylase other than TDG (which is absent) processes the DNA modifications. Since 2′-fluorinated 5-fC and 5-caC are resistant to any DNA glycosylase and cannot be repaired by BER at all, they can be used to distinguish consequences of the primary base modification from repair induced effects. Therefore, expression differences between GC box reporters containing 2′-deoxy- and 2′-fluorinated 5-fC and 5-caC in TDG knockout cells are an indication for any backup DNA glycosylase of TDG.

TDG knockout cells (ΔTDG-2-F3 clonal cell line) were transfected with GC box reporters containing 5-fC, 5-caC or their BER-resistant 2′-fluorinated derivatives F-fC and F-caC and analysed for expression differences over 24 hours (chapter 6.32). Quantitative expression analysis of the transfected TDG knockout cells showed that neither BER-sensitive- nor BER-resistant 5-fC and 5-caC in the GC box induced time-dependent transcriptional repression (Figure 7-21). Here, 5-caC and F-caC in both DNA strands inhibited the promoter activity to the same extend, resulting in an average residual EGFP expression of 50% for the purine-rich strand and 58% for the pyrimidine-rich strand. Transcriptional repression was also absent for 5-fC and F-fC, which did not strongly affected transcription in the purine- and pyrimidine-rich strand of the GC box. Although transcriptional repression was undetectable for both base modifications, it should be noted that the gene expression was significantly higher for F-fC compared to 5-fC in the pyrimidine-rich strand of the GC box, which had not been statistically significant in the earlier experiments (cf. chapter 7.1.11).
Since transcriptional silencing was absent for BER-resistant and BER-sensitive 5-fC and 5-caC in TDG knockout cells, these data indicate that TDG is the only DNA glycosylase to remove 5-fC and 5-caC from the GC box consensus sequence to induce transcriptional repression. Similar effects of BER-sensitive and BER-resistant 5-fC and 5-caC on the gene expression in TDG knockout cells further verified that TDG knockout

and 2′-fluorination of the nucleotides can be equally used to abolish BER of 5-fC and 5-caC in human cells.

Based on the presented results, it is assumed that the low dynamic range of the GC box (16% less active than CRE cf. chapter 7.1.8) combined with the residual TDG activity in TDG knockdown cells masked the TDG dependency of 5-fC and 5-caC induced GC box silencing in earlier experiments (cf. HeLa versus TDG-sh; chapter 7.1.12).

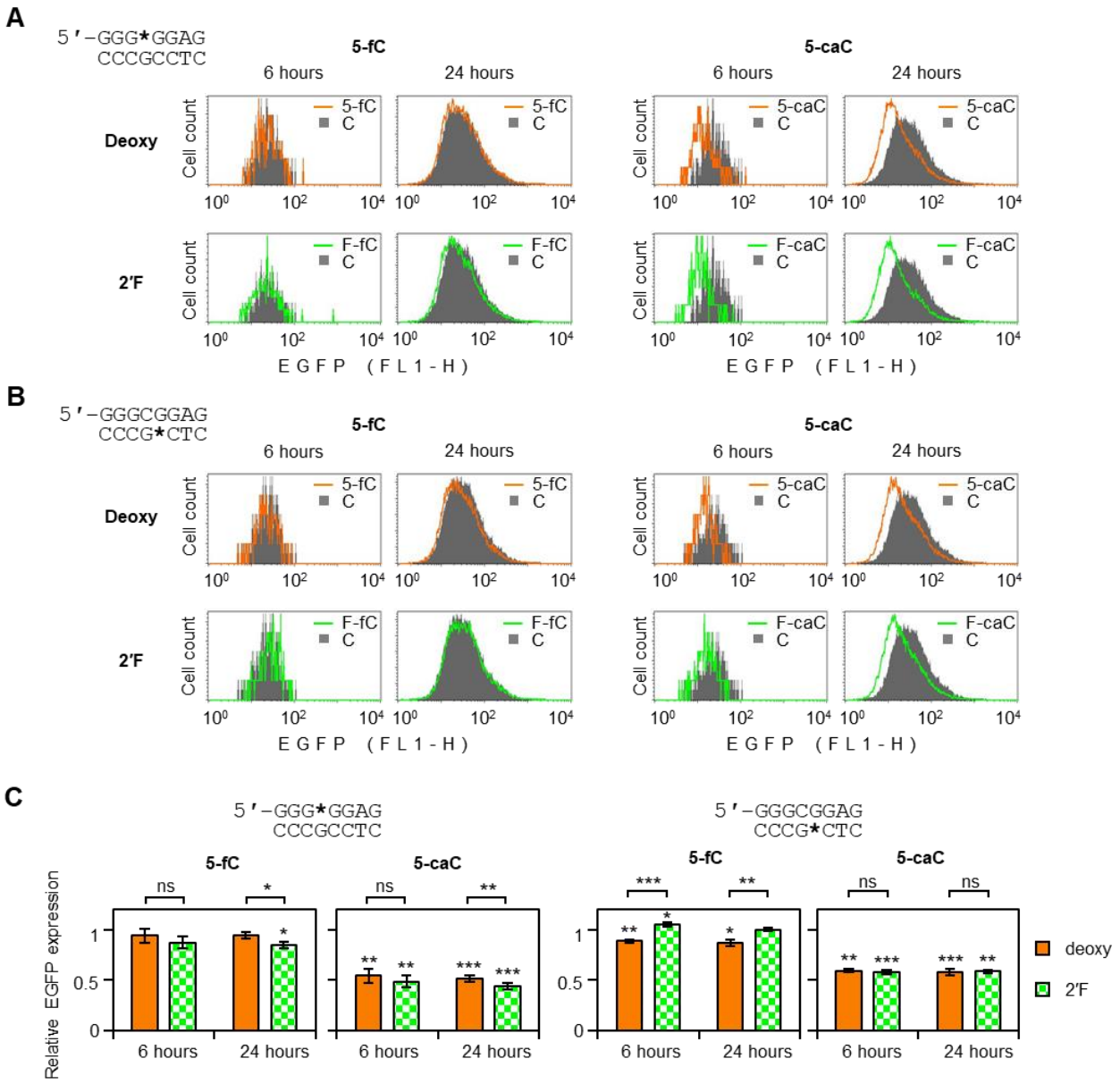


Figure 7-21: Expression of GC box reporters containing BER-resistant or BER-sensitive 5-fC and 5-caC in either strand of the central CpG dinucleotide of the GC box gene regulatory element in a TDG knockout cells

Quantitative EGFP expression analysis of HeLa derived clonal TDG knockout cells (ΔTDG-2-F3) transfected with pGCbox-ZA-W or pGCbox-ZA-C reporters containing single BER-sensitive 5-fC and 5-caC or their BER-resistant 2′-fluorinated derivatives at the central CpG dinucleotide within the purine-rich strand (A+C left panel) or pyrimidine-rich strand (B+C right panel) of the standalone GC box upstream regulatory element: position of the modified base (bold). (A+B) Representative fluorescent distribution plots and (C) quantification of the relative EGFP expression of TDG knockout HeLa cells 6- and 24 hours after transfection in three independent experiments (mean ± SD). Students' T-test was performed with p-values: ns≥0.05, *p<0.05, **p<0.01, ***p<0.001.

In summary, the expression analysis of GC box reporters in TDG knockout cells showed that the transcriptional repression by BER of 5-fC and 5-caC in either strand of the GC box is completely TDG-dependent. Since TDG knockout equalised the functional consequences of BER-sensitive and BER-resistant 5-fC and 5-caC, indicating that TDG is the only DNA glycosylase to excise the modified based from the GC box in vivo, thereby initiating BER.

7.1.15 DNA strand incision at apyrimidinic sites is essential to induce GC box silencing

To determine which repair intermediate induced the transcriptional repression during BER of 5-fC and 5-caC in the GC box the consequences of a single BER-resistant and BER-sensitive apyrimidinic site on the GC box activity were analysed in HeLa cells.
Modified GC box constructs were generated by the strand exchange method (cf. chapter 6.30) to carry a single AP site at the central CpG dinucleotide of the purine- and pyrimidine-rich strand in the GC box (same position as cytosine modifications). Synthetic oligonucleotides containing either a BER-sensitive tetrahydrofuran AP site (F), its BER-resistant derivative SF or the respective C control (Table 6-29) were used for construct generation. Agarose gel electrophoresis (cf. chapter 6.12) of the ligated reporters showed efficient GC box reporter gapping and ligation with all synthetic oligonucleotides (~85% cc plasmid fraction) (Figure 7-22 A+B upper panels). EndoIV digestion as described in chapter 6.17, showed efficient plasmid nicking of F but not C and SF containing reporters, verifying the presence of BER-sensitive AP sites within the desired constructs (Figure 7-22 A+B lower panels).

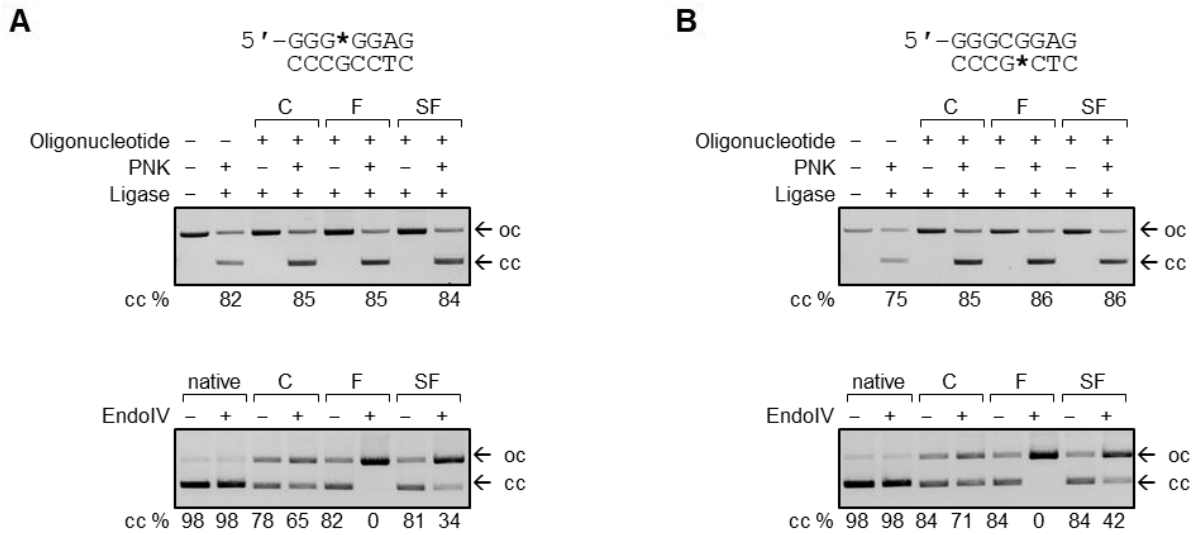


Figure 7-22: Construction of GC box reporters containing a single BER-resistant or BER-sensitive apyrimidinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element

Representative agarose gels of pGCbox-ZA-W (A) and pGCbox-ZA-C (B) constructs containing F, SF or the respective dC control at the central CpG dinucleotide of the purine- (A) or pyrimidine-rich strand (B) of the GC box gene regulatory element: position of the modification (asterisk). Ligation of Nb.BsrDI-nicked GC box constructs with synthetic oligonucleotides containing the indicated modifications in the presence and absence of PNK (upper panel). Incubation of the generated constructs with EndoIV to verify the presence of F by EndoIV nicking and the APE1 protection of SF by absent reporter nicking (lower panel).

Quantitative EGFP expression analysis of HeLa cells transfected with constructs containing a single AP site in either strand of the GC box (chapter 6.32) showed that BER-resistant AP sites only mildly affect the gene expression over 24 hours (Figure 7-23). SF stably decreased the EGFP expression to an average value of 88% (purine-rich strand) and 82% (pyrimidine-rich strand). In contrast, BER-sensitive AP sites drastically decreased the gene expression over time. The transcriptional repression was already detectable six hours after transfection resulting in 81% of residual EGFP expression for F in both DNA strands. The increase of transcriptional repression was significantly stronger by F in the pyrimidine-rich strand of the GC box with a repression magnitude of 60% compared to 52% by F in the purine-rich strand of the GC box. All effects were reproduced in the isogenic TDG knockout cell line ΔTDG-2-F3 (Appendix I 5), verifying that TDG knockout did not influence BER efficiency at steps downstream from DNA base removal.

The EGFP expression analysis demonstrated, that a BER-resistant AP site at the central CpG dinucleotide of the GC box induced a slight reduction of the EGFP expression, independent from the DNA strand and TDG status of the transfected cell line. In the presence of BER, single AP sites progressively reduced the gene expression in a time-dependent manner, with more drastic effects in the pyrimidine-rich strand of the GC box. The findings therefore indicate, that single-strand break

induction and downstream BER processes at AP sites within the GC box were essential for the induction of transcriptional repression.

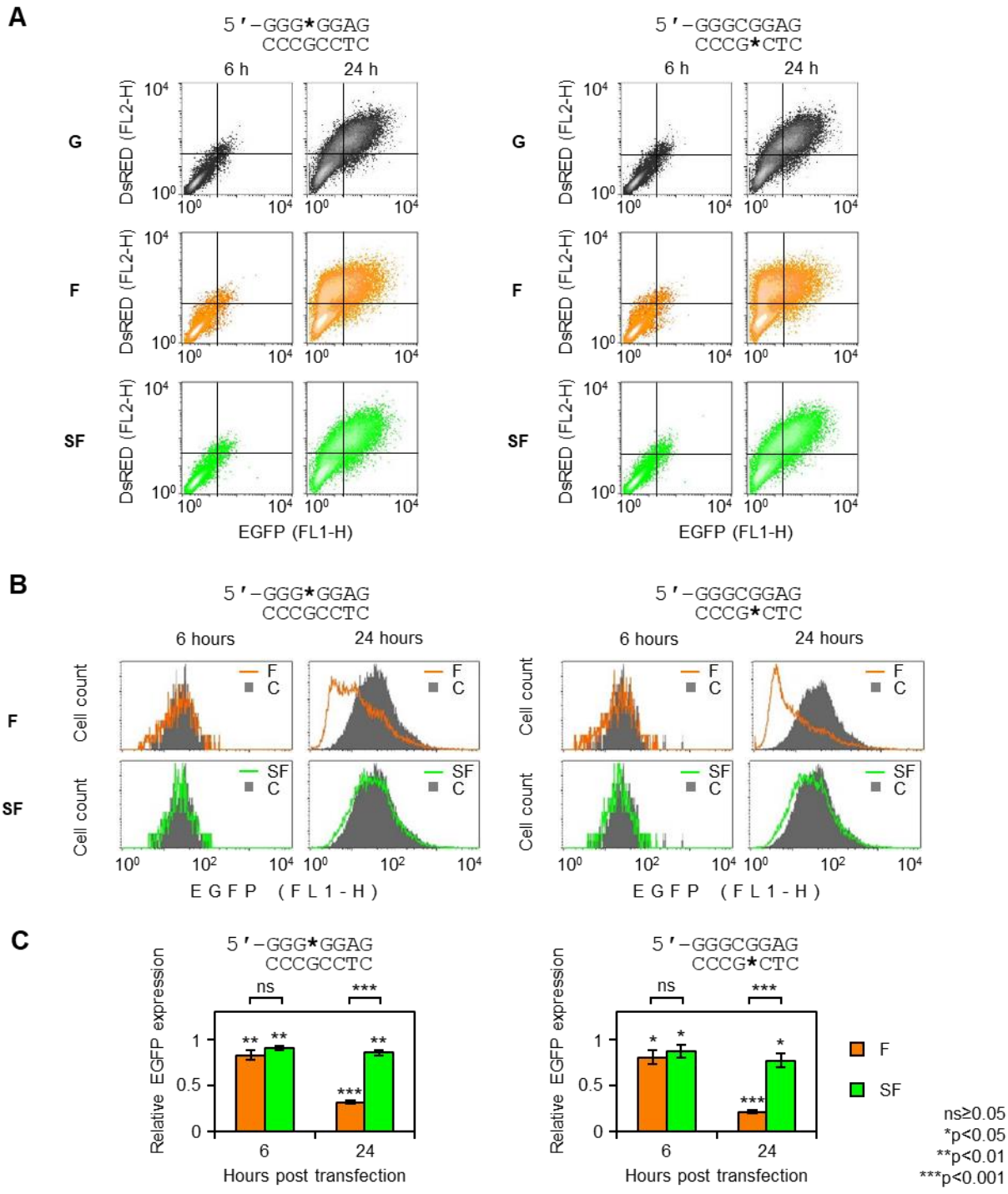


Figure 7-23: Expression of GC box reporters containing a single BER-resistant or BER-sensitive apyrimidinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W or pGCbox-ZA-W reporters containing either a single apyrimidinic site analogue F, its endonuclease protected derivative SF or the C control at the central CpG dinucleotide in the purine- (A) or pyrimidine-rich strand (B) of the GC box upstream regulatory element: position of the modified base (asterisk). Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6- and 24 hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in four independent experiments (mean ± SD).

Comparing the functional consequences of 5-fC and 5-caC in CRE and GC Box reporters, it was demonstrated that UREs significantly affect the consequences of both primary base modifications for promoter activation. Thus, 5-fC and 5-caC per se inhibit the CRE activity to similar extend, whilst promoter inhibition was much stronger for 5-caC in the GC box. Expression analysis in GC box reporters revealed that TDG is of great importance for the transcriptional regulation by 5-fC and 5-caC, indicating that TDG initiated removal of the 5-caC base reactivates promoter activity. Furthermore, the work supplies evidence that the critical role of strand incision for transcriptional silencing by BER of 5-fC and 5-caC may be common for various UREs, since it was verified in both CRE and GC box reporters.

7.2 Consequences of 8-oxoguanine at four selected positions within a standalone GC box gene regulatory element

7.2.1 8-oxoG at the central CpG dinucleotide of a GC box gene regulatory element negatively affects the gene expression in a strand-dependent manner

To better understand the function of 8-oxoG in transcription regulation of GC box containing GC-rich promoters, the consequences of this base modification were investigated in a minimal GC box promoter consisting of a standalone GC box gene regulatory element. The designed minimal GC box promoter does not form complex secondary structures, yet still contains an SP1 transcription factor binding site, in contrast to previously investigated PQS containing promoters (cf. chapter 4.2.2). That way, the number of variables was reduced to a minimum to distinguish 8-oxoG's modification effects from structure specific effects.

The previously generated GC box reporters (chapter 7.1.8) were modified to contain a single 8-oxoG residue at the central CpG dinucleotide of the purine-rich (pGCbox-ZA-W) or pyrimidine-rich strand of the GC box (pGCbox-ZA-C) (depicted in Figure 7-24 A and B, left panels) using the strand exchange method (cf. chapter 6.30) and synthetic oligonucleotides listed in Table 6-29. Agarose gel electrophoresis (cf. chapter 6.12) verified, that both reporters were efficiently nicked, gapped and ligated with synthetic oligonucleotides containing a single 8-oxoG residue or the respective G control (Figure 7-24 A and B, right top panels). Ligation of gapped GC box reporters resulted in an average amount of 93% cc-form plasmid irrespective of the synthetic oligonucleotide, proving that a single 8-oxoG did not detectably influenced the ligation efficiency. Analytical reporter digestion with formamidopyrimidine DNA glycosylase (Fpg) as described in chapter 6.17 resulted in the conversion of 100% of modified reporters to their oc-form, which together with MALDI-TOF analysis of the synthetic oligonucleotides confirmed the presence of 8-oxoG in opposite DNA strands of the GC box (Figure 7-24 A and B, right bottom panels). Fpg digestion only minimally increased the fraction of unmodified constructs in their oc-form (increase of ~10% in G samples) verifying the Fpg specificity.

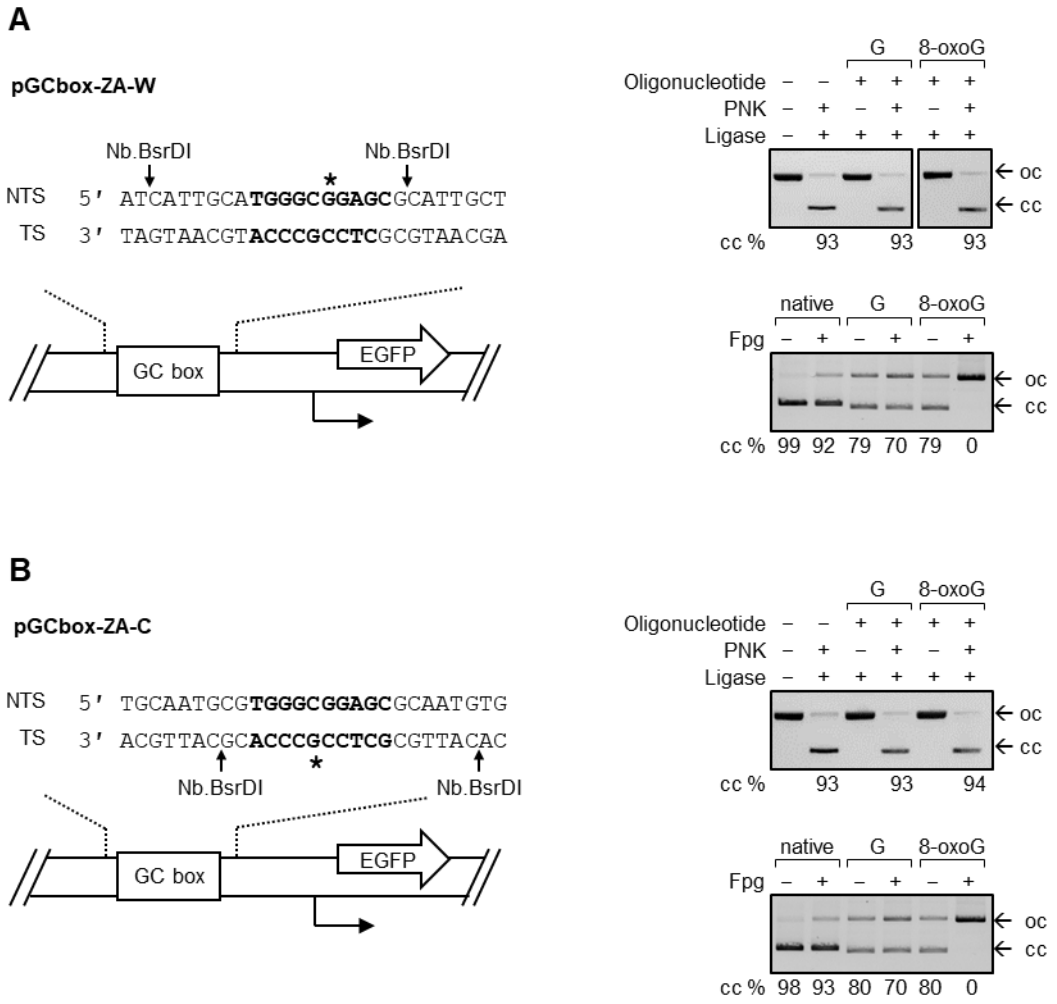


Figure 7-24: Construction of GC box reporters containing single 8-oxoG in either strand of the central CpG dinucleotide of the GC box gene regulatory element

Generation of pGCbox-ZA-W or pGCbox-ZA-C reporters containing a single 8-oxoG residue in the purine- (A) and pyrimidine-rich strand (B) of the central CpG dinucleotide of the standalone GC box upstream regulatory element. (Left panel) Scheme of the standalone GC box gene regulatory element within pGCbox-ZA-W (A) and pGCbox-ZA-C reporters (B) used for the introduction of 8-oxoG in the GC box: EGFP coding sequence (white arrow), TSS (broken arrow), GC box sequence (bold), Nb.BsrDI nicking sites (black arrows) and position of 8-oxoG in the incorporated oligonucleotides (asterisk). (Upper right panel) Ligation of the Nb.BsrDI-nicked GC box constructs with synthetic oligonucleotides containing 8-oxoG or the G control in the presence and absence of PNK. (Lower right panel) Incubation of the generated constructs with Fpg to verify the presence of 8-oxoG by plasmid incision.

The modified GC box constructs were used to transfect HeLa cells and perform quantitative expression analysis (chapter 6.32). Fluorescence analysis of transfected cells by flow cytometry showed that 8-oxoG negatively affected the gene expression over 48 hours (Figure 7-25). 8-oxoG in the purine-rich strand of the GC box caused a 31% decrease of the EGFP level at 48 hours after transfection. This decrease was observed at all time points taken between 6 and 48 hours post transfection and was significantly different from the control G construct ($p<0.01$). In contrast, 8-oxoG in the pyrimidine-rich strand of the GC box repressed the EGFP expression in a gradual time-dependent manner. The EGFP signal was not yet affected at 6 hours after transfection,

however was significantly decreased by 39% and 65% at the 24- and 48-hours' time points respectively.

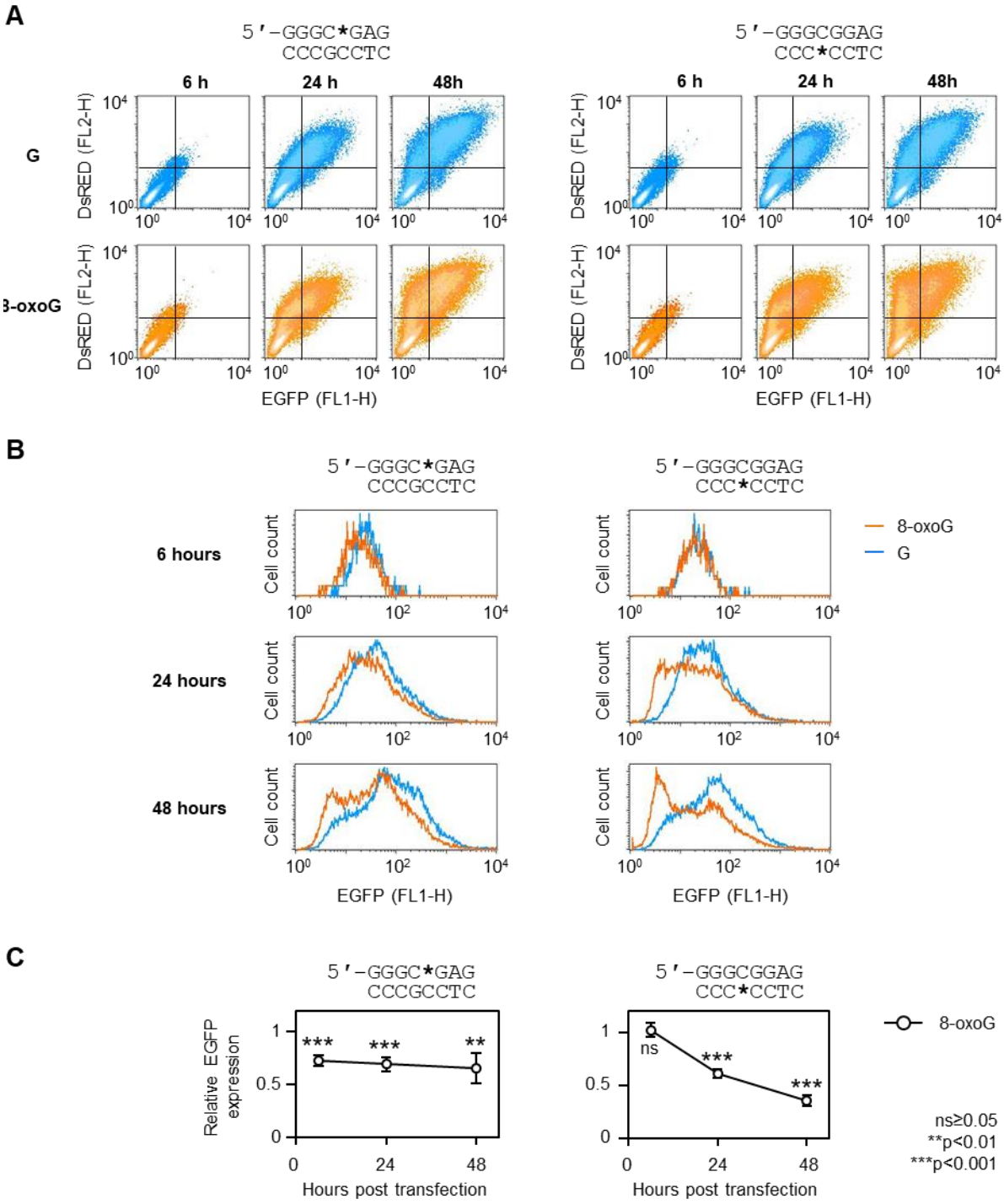


Figure 7-25: Expression of GC box reporters containing single 8-oxoG in either strand of the central CpG dinucleotide of the GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W (left panel) or pGCbox-ZA-C reporters (right panel) containing either a single 8-oxoG or the G control at the central CpG dinucleotide in the purine- (left) and pyrimidine-rich strand (right) of the GC box upstream regulatory element: position of the modified base (asterisk). Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6-, 24- and 48 hours after transfection. C) Quantification of the relative EGFP expression of transfected HeLa cells in five independent experiments (mean ± SD).

The quantitative EGFP expression analysis demonstrated, that a single 8-oxoG residue in either strand of a minimal GC box promoter decreased the promoter activity. Interestingly, the magnitude and progression of the GC box inhibition was very different for 8-oxoG situated in different DNA strands, with a stable decrease for 8-oxoG in the purine- rich strand and a drastic, time-dependent reduction for 8-oxoG in the pyrimidine-rich DNA strand.

7.2.2 The magnitude of GC box inhibition induced by 8-oxoG in the pyrimidine-rich strand but not by 8-oxoG in the purine-rich strand of the GC box correlates with the amount of OGG1 in cells

In the next step, the functional consequences of 8-oxoG on the gene expression were examined in BER deficient cells to distinguish the DNA modification specific from the repair induced transcriptional effects of 8-oxoG. The stable OGG1 knockdown HeLa cell line (OGG1-sh)[276] and the isogenic OGG1 proficient HeLa cell line (no sh) were transfected with GC box reporters containing G or 8-oxoG in either strand of the GC box and the EGFP expression was monitored over 48 hours (chapter 6.32). Strikingly, OGG1 knockdown had very different influences on the inhibition of the gene expression by 8-oxoG situated in different DNA strands (Figure 7-26). No notable difference of EGFP expression levels was observed for reporters containing 8-oxoG in the purine-rich strand of the GC box when comparing cell lines with different OGG1 status. In the parental and the OGG1 knockdown cell line, 8-oxoG in the purine-rich strand of the GC box decreased the promoter activity, resulting in approximately 70% of residual EGFP expression at all time points taken between 6- and 48 hours post transfection. In contrast, the outcome of 8-oxoG in the pyrimidine-rich strand of the GC box differed in OGG1-deficient and OGG1-proficient cells. As observed before, 8-oxoG in the pyrimidine-rich strand of the GC box induced a gradually decline of the EGFP expression down to 41% in OGG1 proficient cells (no sh). OGG1 knockdown significantly reverted the transcriptional repression by 8-oxoG in the pyrimidine-rich strand of the GC box resulting in an EGFP fluorescence of 80% after 48 hours.

Overall, the quantitative expression analysis indicated that 8-oxoG in the pyrimidine-rich strand of the GC box reduced the gene expression in an OGG1-dependent manner, which was not the case if the modification is present in the purine-rich DNA strand.

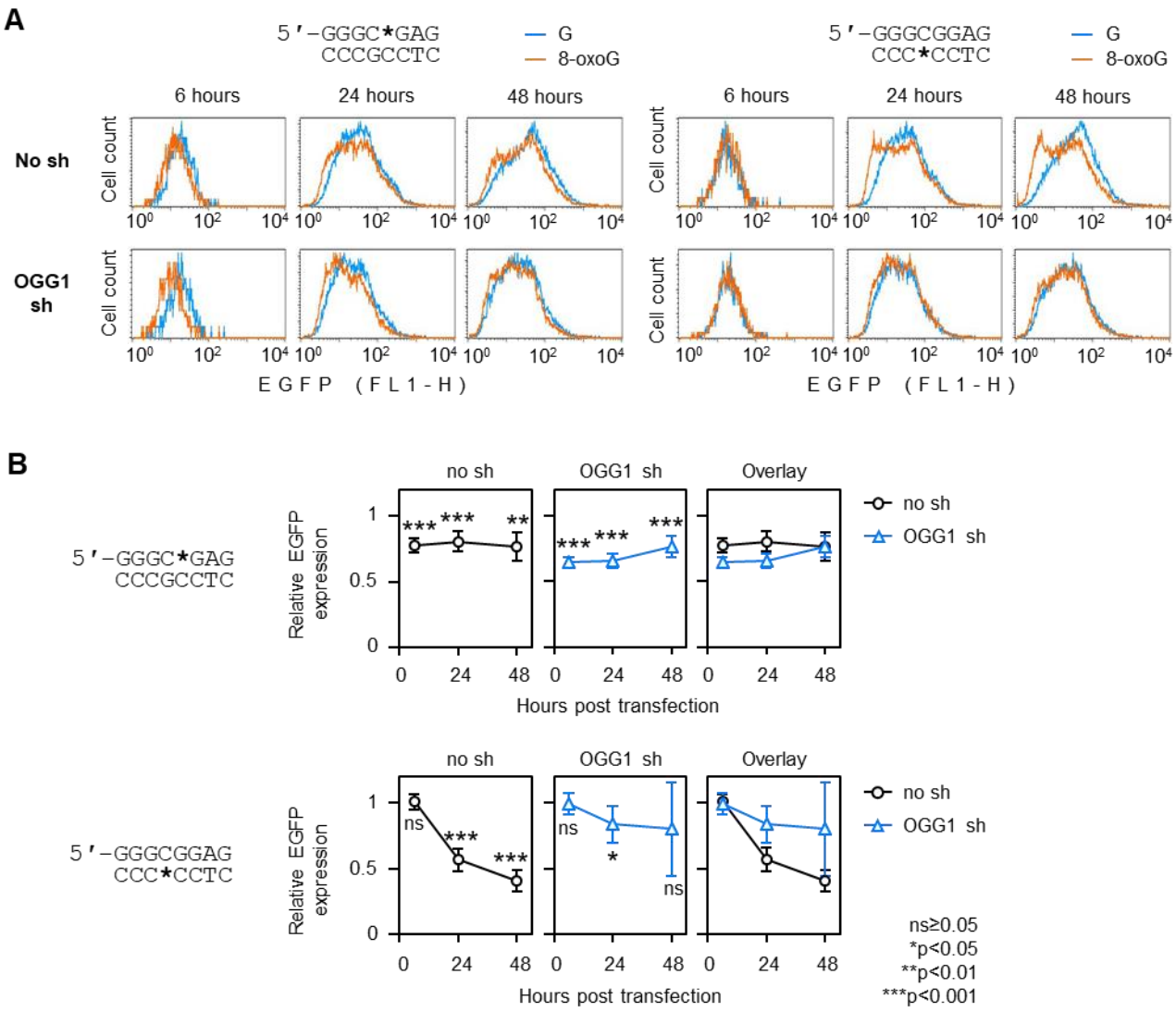


Figure 7-26: Expression of GC box reporters containing a single 8-oxoG residue in either strand of the central CpG dinucleotide of the GC box gene regulatory element in OGG1 knockdown cells

Quantitative EGFP expression analysis of HeLa derived OGG1 knockdown cells and the parental OGG1 proficient HeLa cells transfected with pGCbox-ZA-W or pGCbox-ZA-C reporters containing either a single 8-oxoG or the G control at the central CpG dinucleotide in the purine- and pyrimidine-rich strand of the GC box upstream regulatory element: position of the modified base (asterisk). (A) Representative fluorescent distribution plots of HeLa and the isogenic OGG1 knockdown cells 6-, 24- and 48 hours after transfection. B) Quantification of the relative EGFP expression of transfected HeLa cells in seven independent experiments (mean ± SD).

From on the results of the presented expression analysis it was deduced that 8-oxoG residues situated in different stands of the GC box sequence affected the gene expression by two distinct mechanisms. In the purine-rich strand, 8-oxoG directly inhibited the promoter activity, although to a relatively small extend. In pyrimidine-rich strand, 8-oxoG itself seemed to be almost neutral for the promoter activity, whilst in the presence of OGG1 it induced transcriptional repression in a time-dependent manner. The OGG1-dependent transcriptional repression by 8-oxoG in the pyrimidine-rich strand of the GC box correlates with previous reports, showing that a single 8-oxoG in

the gene body, 5'-UTR and 3'-UTR of an EGFP reporter progressively repressed the gene expression[6,135,136,276]. This repression was found to be induced by OGG1-initiated BER of 8-oxoG and not by the modification itself[6,135,136]. Based on comparable effects, it is assumed that the inhibition of the gene expression by 8-oxoG in the pyrimidine-rich strand but not the purine-rich strand of the GC box was also induced by OGG1 initiated BER.

7.2.3 Strand biases of the transcriptional repression induced by 8-oxoG in the GC box correlate with incision preference of OGG1

Since 8-oxoG at the central CpG dinucleotide of the GC box negatively impacted the gene expression in an OGG1-dependent and OGG1-independent manner in the pyrimidine-rich and purine-rich strand respectively it was questioned whether the DNA glycosylase preferred a particular DNA strand. To address this question, in vitro incision assays with human OGG1 (cf. chapter 6.31) were performed, comparing the cleavage of constructs containing 8-oxoG (or the respective G control) in the pyrimidine-rich or purine-rich strand of the GC box. After identifying a concentration range in which pure OGG1 measurably but incompletely incised 8-oxoG containing constructs (data not shown), its incision activity towards the different DNA strands was analysed as the fraction of incised plasmid DNA generated by equivalent enzyme concentrations. Due to the fact that OGG1 has a low processivity at the post-excision step[174,185], APE1 endonuclease was added to the reaction to increase the rate of strand cleavage at 8-oxoG sites, thereby ensuring that base removal was rate limiting.

In the absence of APE1, pure human OGG1 was much more active on constructs containing single 8-oxoG in the pyrimidine-rich strand of the GC box (Figure 7-27 A). Incubation of 100 ng of plasmid DNA with 0.02 U of hOGG1 resulted in 32.3±1.0% and 52.1±0.8% and of oc plasmid DNA for 8-oxoG placed in the purine-rich and pyrimidine-rich DNA strand respectively. Importantly, control constructs containing G were only cleaved to a very minor extend (<7% increase of oc-form in addition to 11% already present in the absence of OGG1), indicating that the vector DNA contained negligible amounts of OGG1-sensitive base modifications at non-specific sites. Thus, OGG1 induced plasmid incision was solely attributed to the removal of 8-oxoG at the indicated positions within the GC box sequence. As expected, addition of APE1 drastically increased the amount of incised 8-oxoG containing GC box reporters (oc-form 74.8% in purine-rich strand and 90.6% in pyrimidine-rich strand). Still, plasmid cleavage was

more complete when 8-oxoG was present in the pyrimidine-rich strand of the GC box sequence, in agreement with preferential activity of pure human OGG1 towards 8-oxoG in this strand. Saturating APE1 conditions were used as demonstrated by full cleavage of constructs containing a single synthetic AP site analogue F instead of 8-oxoG (Figure 7-27 A last lane). It should be noted that considerable background cleavage was observed after incubation of control constructs with OGG1 in the presence of APE1 (additional 12% of oc plasmid in pGCbox-ZA-W/C comparing hOGG1 + APE1 versus untreated). This was not unexpected, since all processed plasmid DNA inherently contains a small amount of AP sites generated by spontaneous depurination during the preparation process. 8-oxoG containing reporters displayed a similar cleavage rate at these unspecific sites by APE1 (Figure 7-27 B, cc-form increase of 9.2% and 8.0% for pGCbox-ZA-W and pGCbox-ZA-C respectively), which is why the differential activities towards the 8-oxoG substrates were accounted entirely to OGG1.

The cleavage assay was repeated with five times lower hOGG1 concentration, because cleavage rates by 0.02 U hOGG1 in the presence of APE1 were too high for quantitative comparison. Cleavage assays with 0.004 U of hOGG1 in the presence of APE1 showed a clear cleavage preference for 8-oxoG in the pyrimidine-rich strand of the GC box with an increase in oc plasmid DNA by 66.4±0.4% (Figure 7-27 B). For the purine-rich strand, the corresponding increase only amounted 34.4±0.3%. Again, APE1 had no effect on the strand preferences of OGG1 (0.004 U versus 0.002 U hOGG1), indicating that the observed differences in plasmid cleavage were attributed to the DNA glycosylase. APE1 alone only marginally cleaved the constructs containing G or 8-oxoG, suggesting that the vector DNA contained negligible amounts of APE1-sensitive base modifications at non-specific sites (<11% increase of oc-form in addition to 11% already present in the absence of OGG1). Comparison between the two different OGG1 concentrations in the presence or absence of APE1, showed that APE1 enhances the cleavage rates of OGG1 towards 8-oxoG containing constructs by a factor greater than five. It was concluded, that pure human OGG1 preferentially removes 8-oxoG from the pyrimidine-rich strand of the GC box, independent from the subsequent strand incision activity.

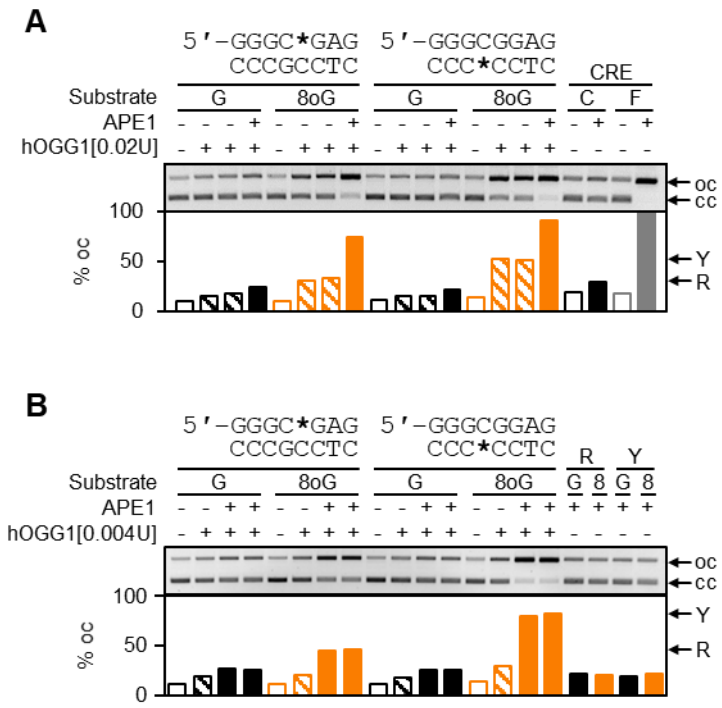


Figure 7-27: Incision of GC box reporters containing a single 8-oxoG residue in either strand of the central CpG dinucleotide of the GC box gene regulatory element by pure human OGG1

Quantitative incision analysis of pGCbox-ZA-W or pGCbox-ZA-W reporters containing a single 8-oxoG residue in either strand of the central CpG dinucleotide of the GC box gene regulatory element by pure human OGG1 in the presence and absence of 1-unit APE1 per 100 ng of DNA (position of the modified base: asterisk). Incubation of covalently closed circular DNA substrates with either 0.02 units (A) or 0.004 units (B) of human OGG1. The quantification of oc-form plasmid fraction after treatment is shown underneath the respective agarose gel images. Incision efficiencies were measured as percentages of conversion into the open circular-form. CRE-UNO reporters containing a single AP lesion were used as a control to verify the APE1 activity (compared to C control, A, last four lanes).

Next, cleavage assays were performed with HeLa cell extracts (cf. chapter 6.31) instead of pure OGG1 and APE1 to evaluate incision efficiencies towards 8-oxoG containing substrates under a physiological stoichiometry of BER components. Protein concentrations were chosen which result in a significant cleavage of AP site containing GC box constructs, corresponding to 66.2% and 94.7% oc DNA for constructs containing F in the pyrimidine-rich and purine-rich DNA strand, respectively (Figure 7-28 A). However, the cleavage of 8-oxoG containing constructs was barely detectable under these conditions. The relatively low OGG1 glycosylase activity detected is in line with previous reports of low OGG1 levels in human cells and low OGG1 activity in human cell extracts[296]. Since higher amounts of protein extracts caused significant band shifting which interfered with the signal quantification (data not shown), it was deduced that the comparison of the incision activity of cellular OGG1 towards 8-oxoG placed in the purine-rich versus the pyrimidine-rich strand is not possible under the given conditions.

Next, protein extracts from OGG1 overexpressing HeLa cells (HeLa OGG1-GFP) were generated as described in chapter 6.29 and used to enhance the glycosylase activity towards 8-oxoG. HeLa OGG1-GFP extracts were reported to exhibit an OGG1 activity high enough to incise plasmid DNA, that contained 8-oxoG in a AGC and CGG context (position underlined)[97,136], indicating that their OGG1 activity is sufficient to excise 8-oxoG from the GC box.

In this work, the OGG1 overexpressing HeLa cell extracts clearly showed a strand bias towards 8-oxoG in the pyrimidine-rich strand of the GC box when analysing the incision of modified reporters (Figure 7-28 B), comparable to the preferences of pure human OGG1 (shown above). More precisely, increasing amounts of cell extract formed 8.8%, 36.5% and 67.5% of additional oc plasmid in comparison to the basal level (7.7%) when 8-oxoG was placed in the pyrimidine-rich strand. For constructs containing 8-oxoG in the purine-rich strand of the GC box, the corresponding oc plasmid increase was only 4.8%, 14.1% and 28.0% respectively. As expected, the extract concentration positively correlated with the amount of incised 8-oxoG but not G control constructs (Figure 7-28 B, lanes 1-16), verifying the lesion specific glycosylase activity in the extract. It should be noted, that constructs containing an AP site at the central CpG dinucleotide of the GC box were cleaved more efficiently than the respective 8-oxoG containing constructs with equivalent cell extract concentrations. The increased strand incision at AP sites compared to 8-oxoG indicates that base removal was rate-limiting under the given conditions. It is thus assumed, that the observed differences in plasmid cleavage was solely attributed to the OGG1 DNA glycosylase.

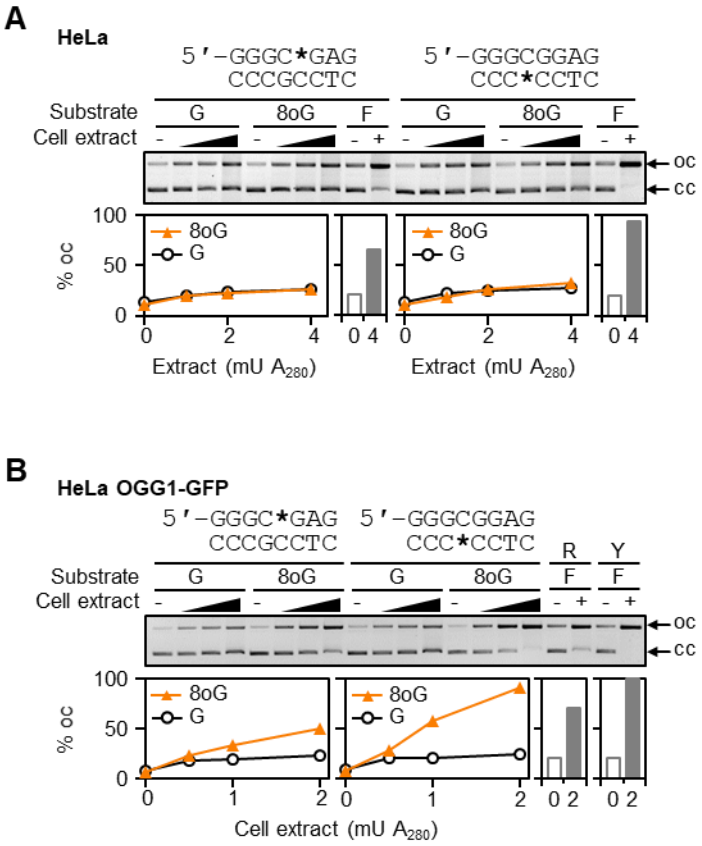


Figure 7-28: Incision of GC box reporters containing a single 8-oxoG residue in either strand of the central CpG dinucleotide of the GC box gene regulatory element by extracts of HeLa and derived OGG1 overexpressing cells

Quantitative incision analysis of pGCbox-ZA-W or pGCbox-ZA-C reporters containing a single 8-oxoG residue in either strand of the central CpG dinucleotide of the GC box upstream regulatory element using extracts from HeLa cells and a derived cell line overexpressing OGG1-GFP fusion protein (position of the modified base: asterisk). Incubation of covalently closed circular DNA substrates with increasing amounts of extract from HeLa cells (A) and OGG1-GFP overexpressing HeLa cells (B). GC box reporters containing a tetrahydrofuran AP lesion at the same position as 8-oxoG were used as a control to determine the APE1 activity.

The combined analysis of 8-oxoG containing reporter cleavage by pure enzymes and cell extracts showed, that OGG1 as pure protein and in cell extracts prefers 8-oxoG in the pyrimidine-rich strand of the GC box in vitro, which correlates with 8-oxoG induced gene silencing at this position in vivo. These data further indicate, that the OGG1 initiated repair dynamics of 8-oxoG in the GC box upstream regulatory element strongly depend on the DNA strand choice.

7.2.4 APE1 prefers apurinic sites in the pyrimidine-rich strand of the GC box under physiological conditions

Interestingly, single AP sites within the GC box were also incised with different efficiencies depending on the DNA strand (cf. chapter 7.2.3). It is therefore assumed, that APE1-mediated strand incision (following base removal by OGG1) contributed to the different repair dynamics of 8-oxoG in the pyrimidine-rich compared to the purine-rich strand of the GC box. For this reason, reporters containing a single AP site at the position of 8-oxoG were generated from pGCbox-ZA-W and pGCbox-ZA-C (Figure 7-29 A) using the strand exchange protocol (cf. chapter 6.30). Gapped parental vectors were ligated with synthetic oligonucleotides containing a single BER-sensitive AP site analogue F, its BER-resistant derivative SF or the respective G control in either strand of the GC box (for sequences see Table 6-29, p. 77). Agarose gel electrophoresis (cf. chapter 6.12) verified that all synthetic oligonucleotides were efficiently ligated to the gapped plasmid DNA resulting in ~87% of cc-form plasmid (Figure 7-29 B). Analytical incision with EndoIV (chapter 6.17) confirmed the presence of F in opposite DNA strands of the GC box, converting 100% of F-containing reporters to their oc-form, whilst control constructs were only minimally incised (oc-form increased by ~14% by EndoIV addition in G sample) (Figure 7-29 C). As expected, constructs containing SF were protected from strand cleavage, resulting in a drastically reduced incision frequency of 39% (cc-fraction treated versus untreated).

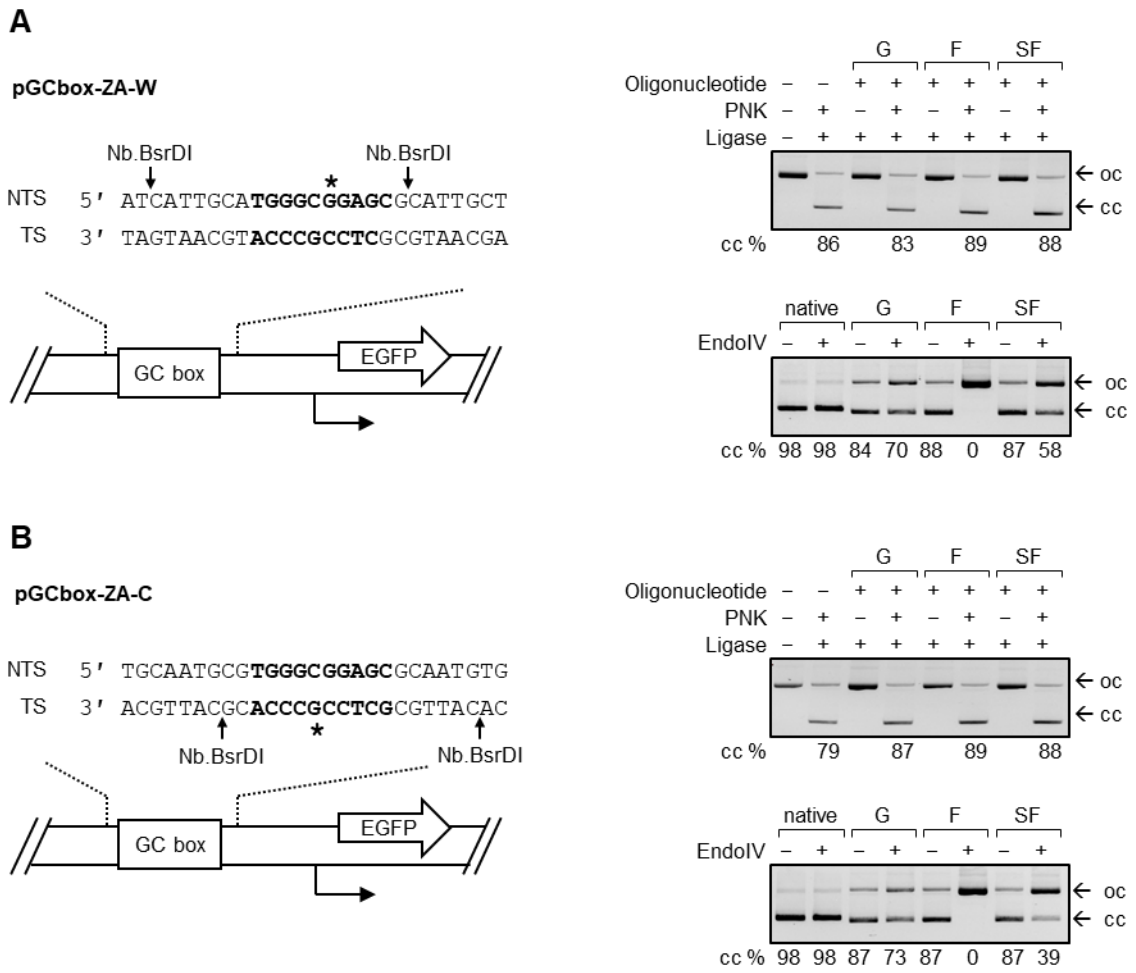


Figure 7-29: Construction of GC box reporters containing a single BER-resistant or BER-sensitive apurinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element

Generation of pGCbox-ZA-W or pGCbox-ZA-C reporters containing a single BER-resistant or BER-sensitive apurinic site in the purine- (A) and pyrimidine-rich strand (B) of the central CpG dinucleotide of the standalone GC box upstream regulatory element. (Left panel) Scheme of the standalone GC box gene regulatory element within pGCbox-ZA-W (A) and pGCbox-ZA-C reporters (B) used for the introduction of BER-resistant SF and BER-sensitive F apurinic sites into the GC box: EGFP coding sequence (white arrow), TSS (broken arrow), GC box sequence (bold), Nb.BsrDI nicking sites (black arrows) and position of apurinic sites in the incorporated oligonucleotides (asterisk). (Upper right panel) Ligation of the Nb.BsrDI-nicked and gapped GC box constructs with synthetic oligonucleotides containing F, SF or the respective G control in the presence and absence of PNK. (Lower right panel) Incubation of the generated constructs with EndoIV to verify the presence of F by EndoIV nicking and the APE1 protection of SF (lower panel).

Cleavage rates of AP sites in the different DNA strands were quantified under the limited incision conditions established above, with serial dilutions of HeLa and OGG1 overexpressing HeLa cell extracts as described in chapter 6.31. Agarose gel electrophoresis showed, that the cleavage activity of HeLa cell extracts towards constructs containing F in the pyrimidine-rich strand of the GC box was at least three times higher compared to constructs containing F in the purine-rich strand (Figure 7-30 A). Already 0.5 mU_{A280} of HeLa cell extract sufficed to form additional 8.5% oc-form plasmid DNA when F was placed in the pyrimidine-rich strand of the GC box. For the purine-rich strand, a similar increase (8.3%) was only achieved with 2 mU_{A280} of cell extract. Comparable results were obtained with OGG1 overexpressing HeLa cell

extract, where 0.5 mU_{A280} and 1 mU_{A280} were needed to incise similar amounts of plasmid DNA when F was placed in the pyrimidine-rich and purine-rich strand respectively (Figure 7-30 B). The extract concentration correlated positively with the amount of incised F constructs for both cell lines. It should be noted, that the incision of SF and G control constructs was higher than expected, with an average incision frequency of 23%. Such non-specific DNA nicking was likely caused by nucleases present in the cell extracts despite the presence of EDTA in the reaction buffer. As the chelating agent EDTA can negatively influence the APE1 activity[297,298], it was unclear whether the observed incision preferences are of biological relevance.

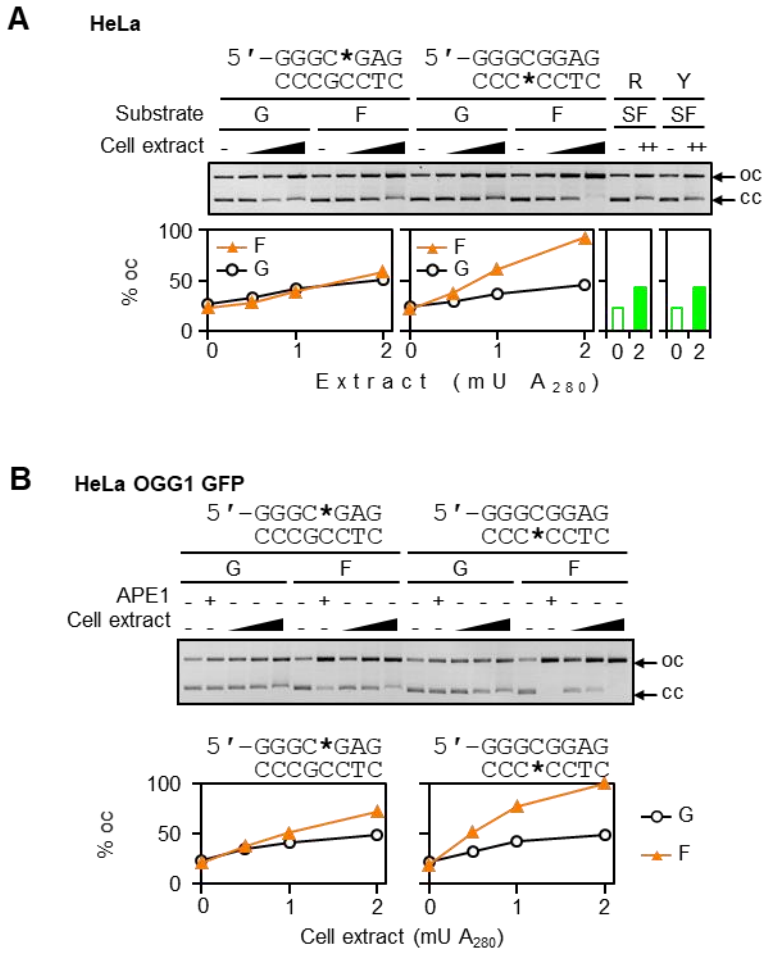


Figure 7-30: Incision of GC box reporters containing a single apurinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element by extracts of HeLa and derived OGG1 overexpressing cells

Quantitative incision analysis of pGCbox-ZA-W and pGCbox-ZA-C reporters containing a single synthetic tetrahydrofuran AP site (F) in either strand of the central CpG dinucleotide of the GC box upstream regulatory element by extracts from HeLa cells and a derived cell line overexpressing OGG1-GFP fusion protein (position of the modified base: asterisk). Incubation of covalently closed circular DNA substrates with increasing amounts of extract from HeLa cells (A) and OGG1-GFP overexpressing HeLa cells (B). Nicking activities of 2 mU A_{280} of HeLa extract towards GC box reporters containing a BER-resistant F analogue (SF) are shown for comparison (green bars).

To examine the potential of EDTA to influence reporter incision biases, the incision preferences towards AP sites in different strands of the GC box were analysed with purified human APE1 both in the presence of EDTA (BEH-BSA buffer as with cell extracts) and under the optimal magnesium concentration as recommended by the producer (NEBuffer4). In vitro incision assays were performed under conditions were APE1 measurably but incompletely incised the DNA in both buffer systems. In the presence of EDTA (BEH-BSA buffer), incubation with 1 U of APE1 resulted in 82.3% and 100% of oc plasmid DNA for F placed in the purine-rich and pyrimidine-rich DNA strand respectively (Figure 7-31 A). APE1 preferences could not be established with low APE1 concentrations (0.005 U) because F specific plasmid incision was undetectable. SF was in great parts protected from strand incision resulting in 36% of oc-form plasmid DNA for both reporters with 1 U of APE1.

Under optimal magnesium concentrations, APE1-mediated nicking of F containing constructs was strongly stimulated (approximately 200-fold, Figure 7-31 A versus B) in agreement with previous reports[297,298]. Interestingly, the preferential cleavage of constructs containing F in the pyrimidine-rich strand of the GC box observed in BEH-BSA buffer (chapter 6.8) was no longer detectable under optimal APE1 cleavage conditions (Figure 7-31 B). Incubation of 100 ng of plasmid DNA with 0.005 U of APE1 showed similar incision rates for F placed in the purine-rich and pyrimidine-rich DNA strand of the GC box resulting in 91.6% and 81.7% of oc plasmid DNA respectively, which was absent for control constructs treated with equivalent enzyme amounts. BER-resistant SF was largely protected from strand cleavage, yet showing a significantly higher amount of plasmid incision at SF in the purine-rich strand of the GC box by 1 U APE1.

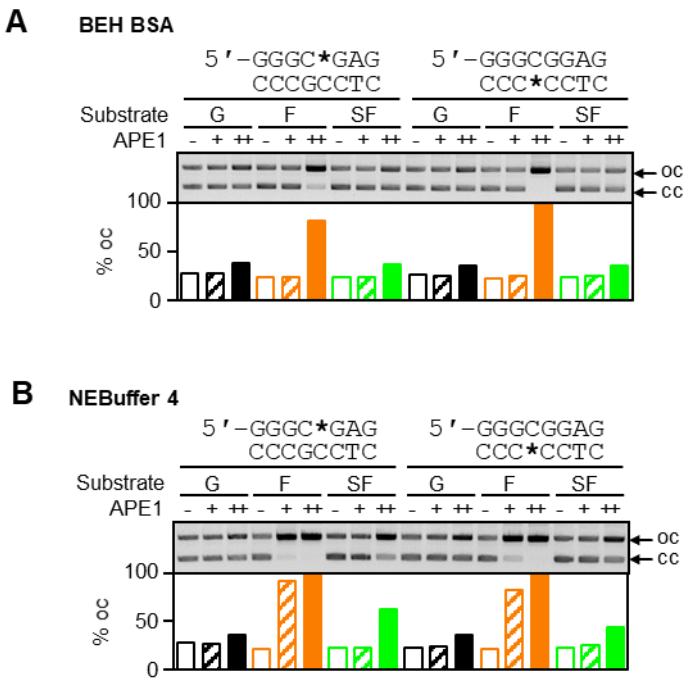


Figure 7-31: Incision of GC box reporters containing a single BER-resistant or BER-sensitive apurinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element by APE1 in two different buffer systems

Quantitative incision analysis of pGCbox-ZA-W and pGCbox-ZA-C reporters containing a single BER-resistant or BER-sensitive apurinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element by 0.005 units (+) and 1 unit (++) of pure APE1 in two different buffer systems (position of the modified base: asterisk). Incubation of covalently closed circular DNA substrates with purified APE1 either in the EDTA-containing, magnesium-free BEH-BSA buffer (A) or the EDTA-free, magnesium-containing NEBuffer 4 (B). Reporters containing APE1-resistant SF were incubated in parallel.

Quantitative plasmid incision assays with purified APE1 and cell extracts indicate that APE1 incision preferences towards AP sites in either strand of the GC box greatly depends on the buffer conditions. In the presence of EDTA, cell extracts as well as pure APE1 had a strand bias towards AP sites in the pyrimidine-rich strand of the GC box, which was abolished in the presence of magnesium.

7.2.5 BER of apurinic sites in the GC box causes gene silencing in a strand independent manner

The functional consequences of AP sites in the GC box were further investigated with respect to their promoter regulation potential. Quantitative EGFP expression analysis (chapter 6.32) was performed in HeLa cells transfected with reporters containing a single BER-resistant or BER-sensitive AP site in either strand of the GC box. BER-resistant SF in the purine-rich strand of the GC box sequence mildly inhibited the promoter activity over time. Nonetheless the reduction was significant at 6- and 24 hours post transfection, resulting in a residual EGFP expression of 87% and 75% respectively (Figure 7-32). Strikingly, SF in the pyrimidine-rich strand of the GC box

significantly enhanced the gene expression at both time points (116% at 6 hours and 118% at 24 hours), indicating that AP sites per se only mildly altered transcription, yet with different effects for the different strands. Although small in size, the consequences of AP sites in both DNA strand were regarded as potentially biologically meaningful, due to the standalone GC box promoter offering a rather narrow (an approximately two-fold) dynamic range of transcriptional activation in relationship to the basal level of transcription (cf. chapter 7.1.8). In contrast to SF, BER-sensitive AP sites in both strands of the GC box drastically reduced the gene expression in a time-dependent manner. Transcriptional repression was already detectable 6 hours after transfection (~84% of residual EGFP expression for both F containing reporters) and increased over 24 hours down to ~14% of residual EGFP fluorescence, with a similar repression magnitude for modifications in the purine-rich and the pyrimidine-rich strand (Figure 7-32).

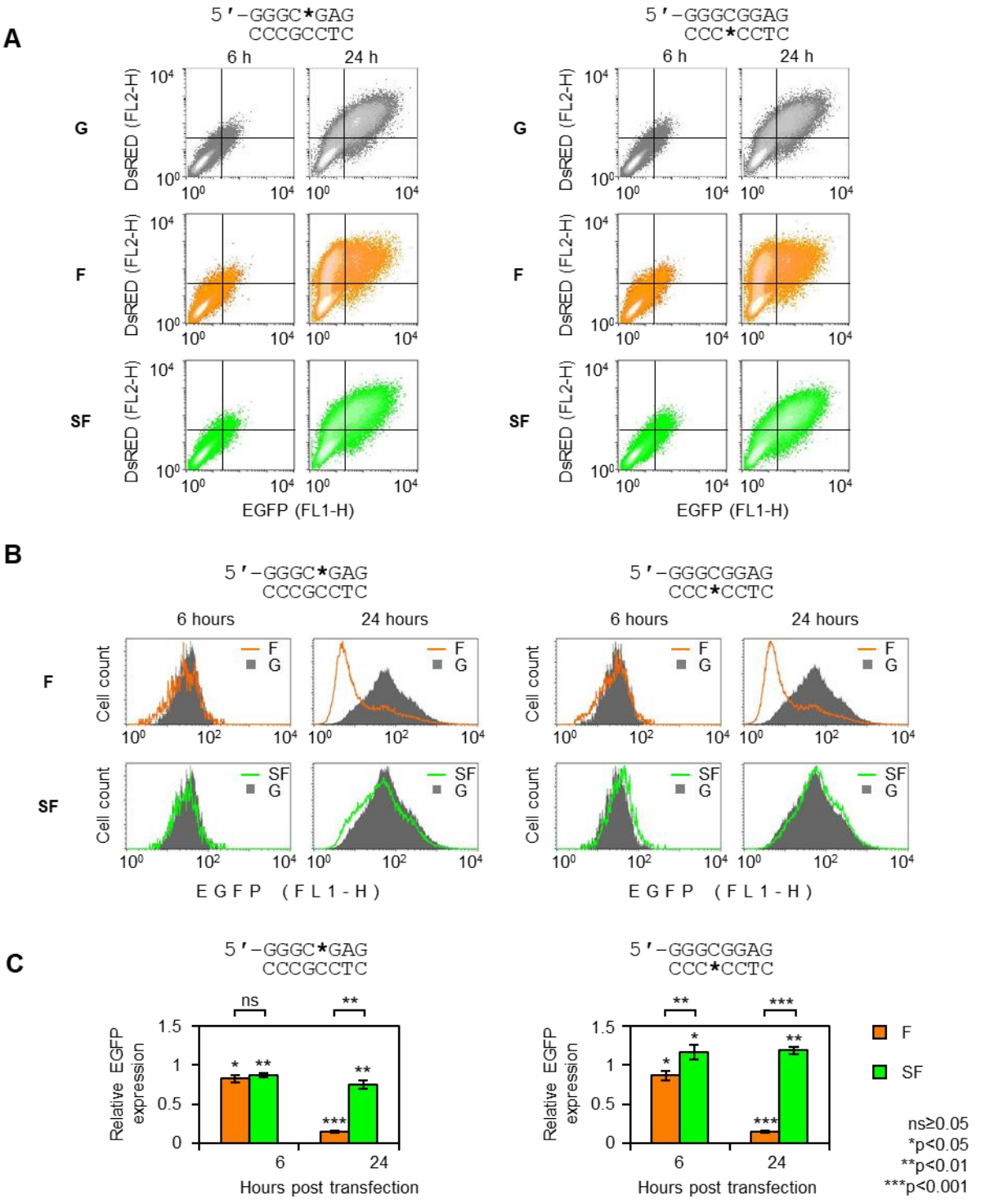


Figure 7-32: Expression of GC box reporters containing a single apurinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W (left panel) and pGCbox-ZA-C (right panel) reporters containing either a single apurinic-site analogue F, its endonuclease protected derivative SF or the G control in either strand of the central CpG dinucleotide of the GC box upstream regulatory element (position of the modified base: asterisk). Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa cells 6- and 24 hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in five independent experiments (mean ± SD).

The EGFP expression analysis showed that AP sites in the GC box per se slightly influenced the promoter activity in a DNA-strand specific manner. In the presence of BER, incision of AP lesions by APE1 and downstream BER processes triggered

transcriptional repression. This is in line with previous findings, showing that transcriptional repression by AP sites in gene coding regions dependet on strand incision by APE1[136]. BER-sensitive AP lesions in different GC box strands inflicted similar transcriptional silencing responses, indicating that APE1 efficiently incised both AP site containing DNA strands in vivo. Since transcriptional repression by 8-oxoG was strand-dependent, (cf. chapter 7.2.2), these strand biases can now be attributed to OGG1 preferences (not APE1), in coherence with OGG1's strand incision selectivity observed in vivo (cf. chapter 7.2.3).

7.2.6 8-oxoG in the purine-rich DNA strand negatively impacts the GC box activity in a position-dependent manner

Considering the completely different outcomes of 8-oxoG residues in the two strands of the GC box, it was questioned whether these diverse consequences depend on the strand or the position of 8-oxoG within the GC box sequence. To answer this question, the impacts of 8-oxoG located at three different positions within the purine-rich ("R") strand of the GC box were investigated. In addition to 8-oxoG at position R+1 analysed above (with G of the central CpG-dinucleotide indexed as R+1), 8-oxoG at positions R-2 and R-3 was examined for its functional consequences (Figure 7-33 A). These positions were selected because they are at varying distances from the central CpG dinucleotide and are conserved in the mammalian GC box consensus sequence[287], indicating an important role in URE function. Reporters containing a single 8-oxoG residue at the positions R-3, R-2 and R+1 were generated by the strand exchange method (cf. chapter 6.30), similar to previous GC box constructs. Efficient ligation of the synthetic oligonucleotides carrying 8-oxoG at the desired positions (for oligonucleotides sequences see Table 6-29, p. 77) with gapped GC box reporters was verified by agarose gel electrophoresis (cf. chapter 6.12) demonstrating an approximate cc-form plasmid fraction of 90% for all constructs (Figure 7-33 B). Analytical digestion of the generated constructs with Fpg (chapter 6.17) demonstrated that 100% of the constructs with synthetic 8-oxoG containing oligonucleotides were converted into their oc-form, whilst the G control was only marginally incised (oc-form increased by 16% in addition to the basal level in G sample), verifying the presence of 8-oxoG (Figure 7-33 C).

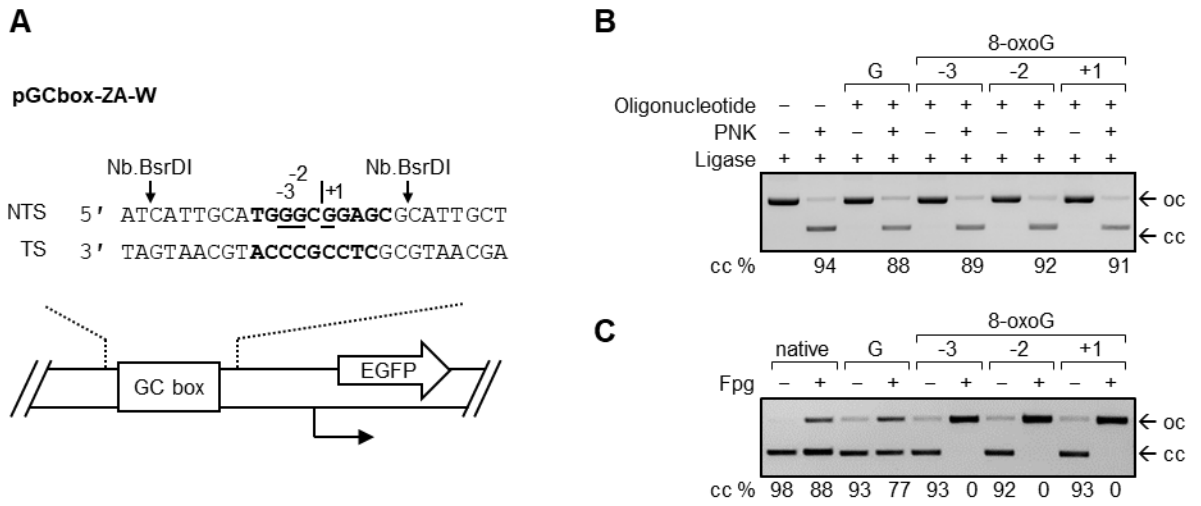


Figure 7-33: Construction of GC box reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box gene regulatory element

Generation of pGCbox-ZA-W or pGCbox-ZA-C reporters containing a single 8-oxoG at three selected positions in the purine-rich strand of the GC box upstream regulatory element. (A) Scheme of the standalone GC box gene regulatory element within the pGCbox-ZA-W reporter used for the introduction of 8-oxoG into the GC box: EGFP coding sequence (white arrow), TSS (broken arrow), GC box sequence (bold), Nb.BsrDI nicking sites (black arrows) and position of 8-oxoG in the incorporated oligonucleotides (underlined, with distance from central CpG dinucleotide indicated above). (B) Ligation of the Nb.BsrDI-nicked GC box construct with synthetic oligonucleotides containing a single 8-oxoG residue or the G control at positions -3, -2 and +1 in the purine-rich strand of the GC box in the presence and absence of PNK. (C) Incubation of the generated constructs with Fpg to verify the presence of 8-oxoG by plasmid incision.

Quantitative expression analysis in HeLa cells (chapter 6.32) showed that 8-oxoG at any of the three selected positions in the purine-rich strand of the GC box caused a significant decline of the EGFP expression 48 hours after transfection (Figure 7-34). For 8-oxoG at positions R-3, the inhibition of the promoter activity was stable in magnitude throughout the 48 hours' time course, which matched the outcome of 8-oxoG at position R+1. More specifically, 8-oxoG at R-3 and R+1 decreased the EGFP expression to 73-66%, which equals the degree of promoter inhibition by 8-oxoG at position R+1 observed earlier (Figure 7-34 versus Figure 7-25). In contrast, 8-oxoG at position R-2 repressed the gene expression in a time-dependent manner. The transcriptional repression was already observed 6 hours after transfection (87% of residual EGFP expression) and intensified over time, with 63% and 52% of residual EGFP expression at the 24- and 48-hours' time point respectively (Figure 7-34). Although modest in its magnitude, this effect was highly reproducible and significant and showed similarities to the gradual loss of EGFP expression by 8-oxoG at position -1 in the pyrimidine-rich strand of the GC box (Figure 7-34, R-2 versus Figure 7-25, Y-1).

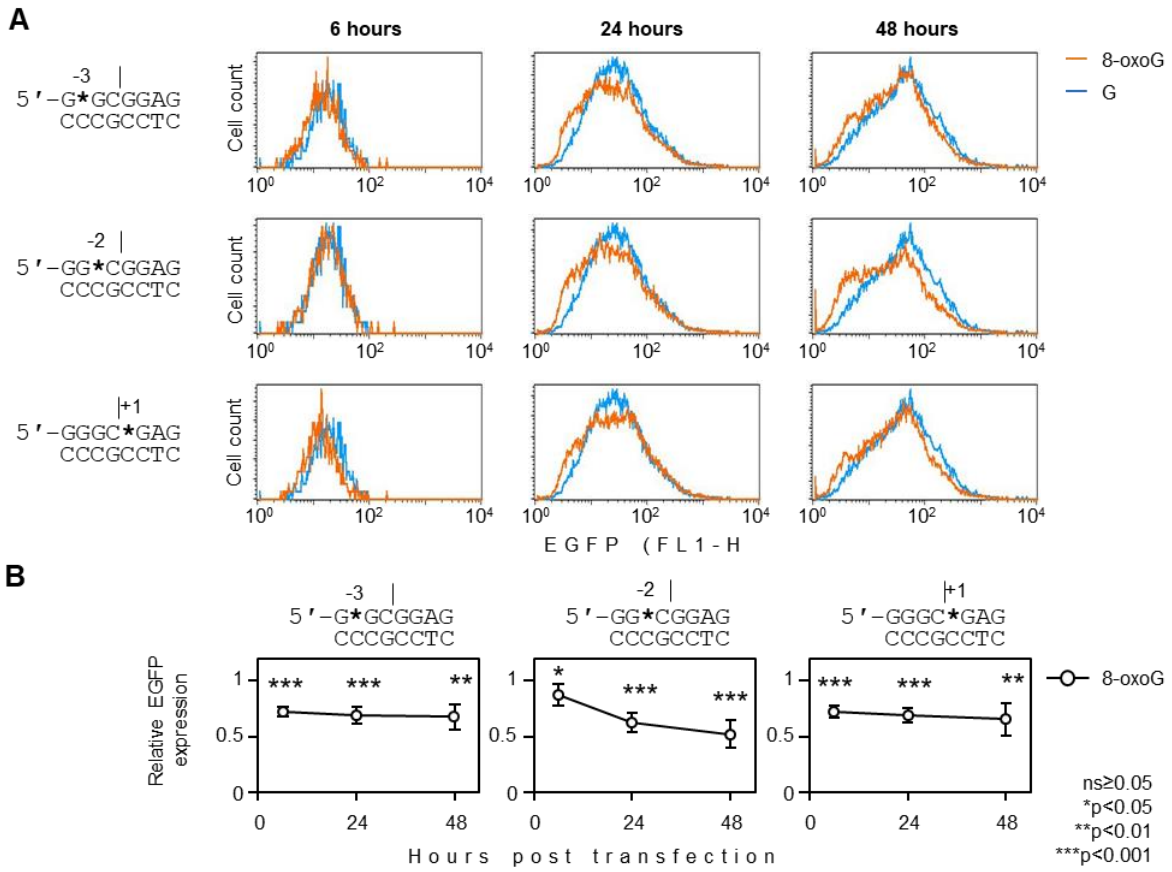


Figure 7-34: Expression of GC box reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W reporters containing either a single 8-oxoG or the G control at position -3, -2 and +1 of the purine-rich strand of the GC box upstream regulatory element (position of the modification: asterisk). (A) Representative fluorescent distribution plots of HeLa cells 6-, 24- and 48 hours after transfection. (B) Quantification of the relative EGFP expression of transfected HeLa cells in five independent experiments (mean ± SD).

Comparing the transcriptional effects of 8-oxoG at the four analysed positions of the GC box (R-3, R-2 and R+1 in the purine-rich strand and Y-1 in the pyrimidine-rich strand), it is assumed that the base modification 8 negatively affects the gene expression in a position- rather than strand-dependent manner.

7.2.7 OGG1 knockdown only rescues the transcriptional repression induced by 8-oxoG at position R-2 of the GC box

Based on the observation that the transcriptional repression by 8-oxoG at position R-2 and Y-1 of the GC box was similarly time-dependent, it was questioned whether these similarities also extended to the mode of repression. Since the transcriptional repression by 8-oxoG at position Y-1 was found to be OGG1-dependent (cf. chapter 7.2.2), quantitative expression analysis (chapter 6.32) of GC box reporters containing 8-oxoG at positions R-3, R-2 and R+1 was performed in OGG1-depleted cells (Figure 7-35). As expected, 8-oxoG at position R-2 but not R-3 and R+1 induced a progressive

41% decline of the EGFP fluorescence in OGG1 proficient cells over the 48 hours' time course. OGG1 knockdown completely reverted the time-dependent transcriptional repression by 8-oxoG at position R-2, verifying that OGG1 was essential for this process. It should be noted, that 8-oxoG at position R-2 still reduced the GC box activity in the absence of repair, resulting in a stably reduced EGFP expression of approximate 80%. For 8-oxoG at positions R-3 and R+1, the outcomes in OGG1-deficient and OGG1-proficient cells were similar, suggesting that 8-oxoG at these positions inhibited the promoter activity by a BER independent mechanism.

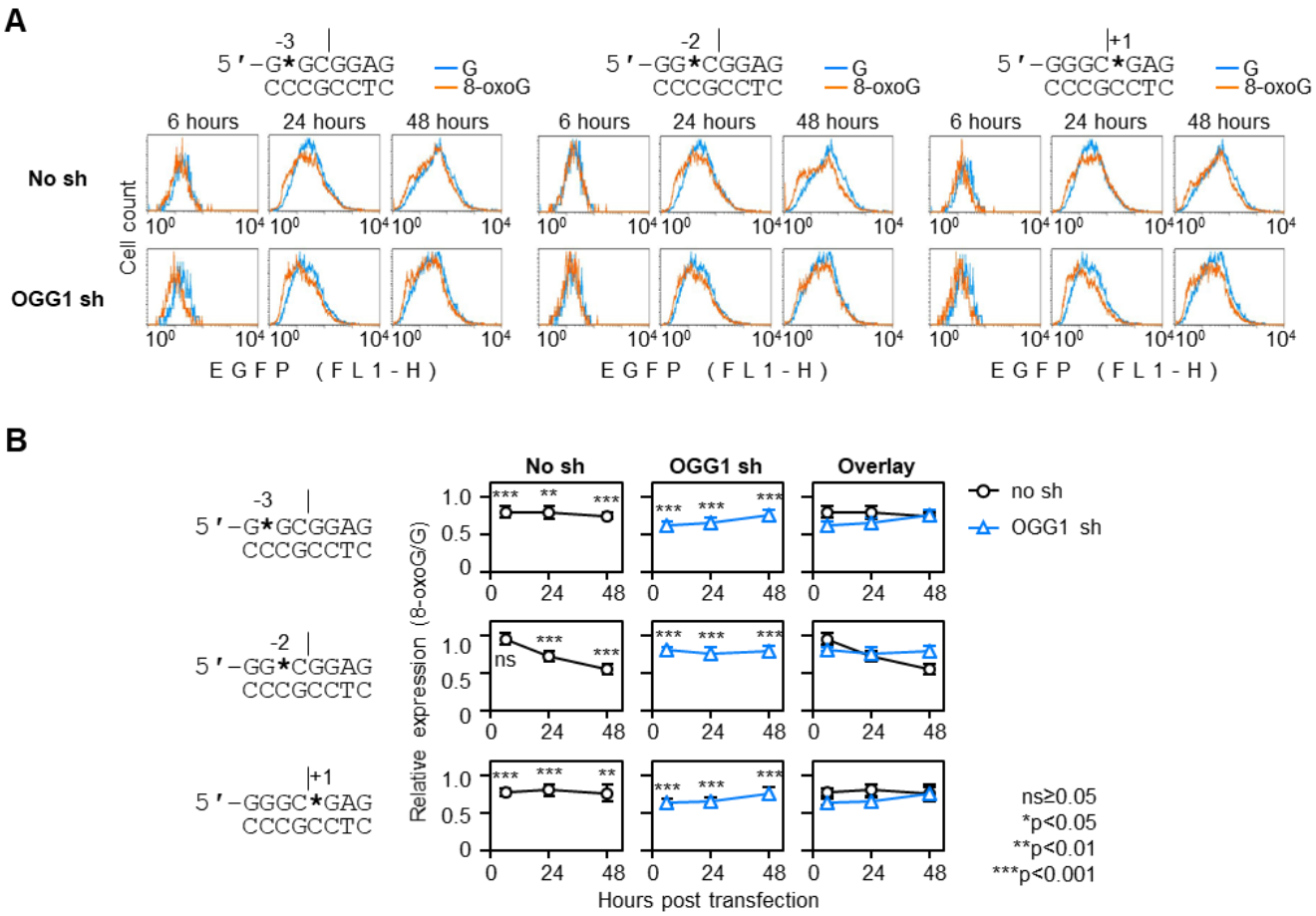


Figure 7-35: Expression of GC box reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box gene regulatory element in OGG1 knockdown cells

Quantitative EGFP expression analysis of HeLa derived OGG1 knockdown cells and the parental OGG1 proficient HeLa cells transfected with pGCbox-ZA-W reporters containing either a single 8-oxoG or the G control at three selected positions in the purine-rich strand of the GC box upstream regulatory element (position of the modified base: asterisk). (A) Representative fluorescent distribution plots of HeLa and the isogenic OGG1 knockdown cells 6-, 24- and 48 hours after transfection. (B) Quantification of the relative EGFP expression of transfected cells in seven independent experiments (mean ± SD).

Comparing the EGFP expression in BER deficient and BER proficient cells, it was demonstrated that 8-oxoG at three different positions within the purine-rich strand of the GC box (R-3, R-2 and R+1) per se reduced the promoter activity in a position dependent manner. Furthermore, 8-oxoG at position R-2 induced time-dependent transcriptional repression by OGG1 initiated BER, comparable to 8-oxoG in the pyrimidine-rich strand of the GC box.

7.2.8 Transcriptional repression by 8-oxoG at position R-2 of the GC box correlates with high plasmid incision efficiency by pure human OGG1 and cell extracts

To find out if the different functional consequences of 8-oxoG in the purine-rich strand of the GC box were mediated by varying OGG1 preference as shown for the central CpG dinucleotide (cf. chapter 7.2.3), incision efficiencies at different positions by cell extracts and pure human OGG1 were analysed as described in chapter 6.31. In vitro incision assays with extracts from HeLa cells showed 41% cleavage for constructs containing a single tetrahydrofuran AP site at position R+1, yet the cleavage of 8-oxoG containing constructs was barely detectable (Figure 7-36 A). Therefore, the OGG1 activity of the reaction was enhanced using extracts from OGG1 overexpressing cells in the next incision assay. Here, a clear preference for constructs containing 8-oxoG at position R-2 (Figure 7-36 B) was observed with a two-fold higher incision efficiency compared to positions R-3 and R+1. 2 mU_{A280} were needed to incise 29.7% and 22.9% of constructs containing 8-oxoG at position R-3 and R+1, whereas 1mU_{A280} was sufficient to incise 26.5% of constructs containing 8-oxoG at position R-2 (compare oc-form differences 8-oxoG versus G). The incision differences were specifically attributed to 8-oxoG, since the corresponding G control construct was only incised at minor rates. It should be noted, that constructs containing a single AP site at position R+1 were cleaved more efficiently than the respective 8-oxoG containing constructs, verifying that base removal remained the rate limiting step of the overall incision reaction under OGG1 overexpression conditions.

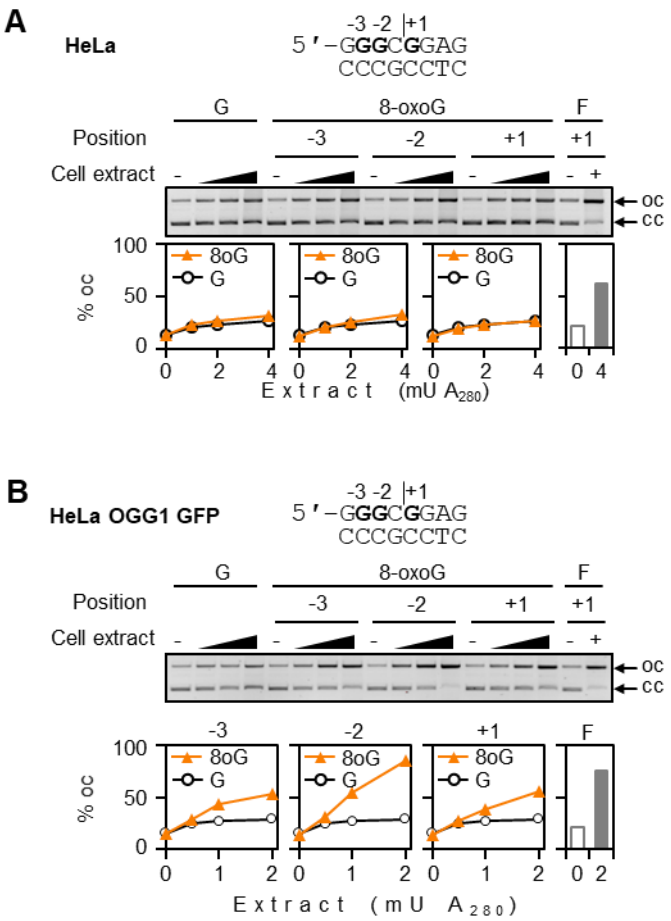


Figure 7-36: Incision of GC box reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box gene regulatory element by extracts of HeLa and derived OGG1 overexpressing cells

Quantitative incision analysis of pGCbox-ZA-W reporters containing a single 8-oxoG at three selected positions in the purine-rich strand the GC box upstream regulatory element by extracts of HeLa cells and a derived cell line overexpressing OGG1-GFP fusion protein (position of the modified base: bold). Incubation of covalently closed circular DNA substrates with increasing amounts of extract from HeLa cells (A) and OGG1-GFP overexpressing HeLa cells (B). GC box reporters containing a tetrahydrofuran AP lesion (F) at the same position as 8-oxoG were used as a control to determine the APE1 activity.

To verify that preferential 8-oxoG removal at position R-2 of the GC box by OGG1 biases was indeed the cause for varying plasmid incision frequencies, in vitro incision assays with purified human OGG1 were performed. APE1 was added to the reaction to enhance the strand incision step and thereby ensured that base removal was rate limiting. The plasmid incision analysis showed that 8-oxoG at R-2 was again cleaved with the highest efficiency of the three selected positions in the GC box (Figure 7-37). Thus, incubation of 100 ng of plasmid DNA with 0.04 U of hOGG1 and 1U of APE1 resulted in 71.0±1.15%, 82.3±1.1% and 50.8±1.75% of oc plasmid DNA for 8-oxoG placed at position R-3, R-2 and R+1, respectively. Minor cleavage of control G constructs (<15% increase of oc-form in addition to 15% already present in the absence of OGG1) with equivalent enzyme concentration verified that the differential activities towards the 8-oxoG substrates should be accounted entirely to OGG1. Fpg

digestion of the reporters (chapter 6.17) again verified the presence of 8-oxoG within the GC box by selective plasmid incision.

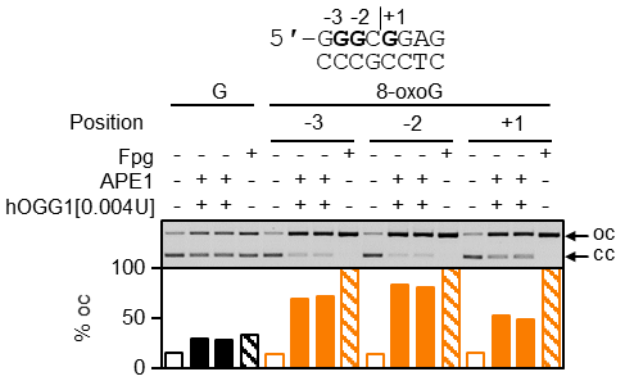


Figure 7-37: Incision of GC box reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box gene regulatory element by pure human OGG1

Quantitative incision analysis of pGCbox-ZA-W reporters containing a single 8-oxoG residue at three selected positions in the purine-rich strand of the GC box upstream regulatory element by pure human OGG1 in the presence of APE1 (position of the modified base: asterisk). Incubation of covalently closed circular DNA substrates with 0.004 units of hOGG1 and 1 unit of APE1 per 100 ng of DNA. Parallel incubation with 2 Units of Fpg was performed to verify the presence of 8-oxoG by plasmid incision.

Based on the cleavage assays performed with pure human OGG1 and HeLa cell extracts, it is concluded that OGG1 preferred 8-oxoG at position R-2 over positions R-3 and R+1 under cell free conditions. The high incision rate of 8-oxoG at position R-2 correlated with the OGG1-dependent downregulation of the gene expression observed for this position. No OGG1-dependent repression was observed for 8-oxoG at positions R-3 or R+1 with lower incision rates. Since a similar correlation between strand specific transcriptional repression by BER of 8-oxoG and incision preferences of OGG1 has also been established for the central CpG dinucleotide of the GC box previously (cf. chapters7.2.2 and 7.2.3), it is tempting to suggest a causal connection between BER efficiency and the outcome of 8-oxoG for promoter activity.

7.3 Functional consequences of various unrelated BER targets in a CRE upstream regulatory element

7.3.1 DNA lesions 5-hmU, U and Tg but not T opposite to G in a CRE gene regulatory element induce transcriptional repression

When it became apparent, that BER of the three investigated, oxidatively induced DNA modifications 8-oxoG, 5-fC and 5-caC in the CRE and GC box gene regulatory element repressed transcription, it was further assessed if such repression is a common BER-consequence of any DNA modification. To investigated this hypothesis, the functional consequences of the common BER substrates Tg, U and 5-hmU (cf. chapter 4.3.3) were analysed in the CRE gene regulatory element. The DNA modifications were placed opposite to guanine at the central CpG dinucleotide of CRE in the minimal promoter CRE-UNO. The T:G miss-match repair (MMR) substrate was used to enable the differentiation between BER and MMR induced functional consequences of modified bases.

CRE-UNO reporters carrying a single 5-hmU, U, Tg, T or the respective C control at the central CpG dinucleotide in the NTS of CRE (Figure 7-38 A) were generated following the strand exchange procedure as described in chapter 6.30. pCRE-UNO-ZA-W was nicked by Nb.BsrDI and gapped using the plasmid specific complementary oligonucleotide (Table 6-32). The gapped plasmid DNA was ligated with synthetic oligonucleotides carrying the desired base modification or the C control, listed in Table 6-29 rows 1 and 6-9. Agarose gel electrophoresis (cf. chapter 6.12) of the modified reporters showed efficient Nb.BsrDI-nicking, gapping and ligation with the synthetic oligonucleotides resulting in 90% of cc-form plasmid DNA for oligonucleotides containing 5-hmU, U, T or C (Figure 7-38 B). Plasmid ligation with Tg containing oligonucleotides only formed 70% of cc-form plasmid DNA. DNA helix distortion by Tg[299] is assumed to cause the reduced ligation efficiency of Tg containing oligonucleotides compared to oligonucleotides containing 5-hmU, U, T or C which do not strongly influence the structure of DNA duplexes. The presence of all DNA modifications was verified by analytical AatII digestion as described in chapter 6.17, since AatII cleavage in the CRE motive is inhibited by the presence of modified nucleobases. Agarose gel electrophoresis (cf. chapter 6.12) of AatII treated CRE-UNO reporters demonstrated the expected inhibition of the AatII activity in constructs containing 5-hmU, U, T and Tg but not the C control (Figure 7-38 C). The presence of

Tg was additionally verified by an analytical digestion with EndoIII as described in chapter 6.17. EndoIII is the *E.coli* homologue of human NTHL1 and can recognise and excise Tg from double-stranded DNA. As expected, agarose gel electrophoresis of EndoIII treated reporters demonstrated significant plasmid cleavage only for Tg containing constructs (Figure 7-38 D). The outcomes of reporter treatment with AatII and EndoIII together with MALDI-TOF analysis of the synthetic oligonucleotides verify the successful generation of 5-hmU, U, T and Tg containing reporters.

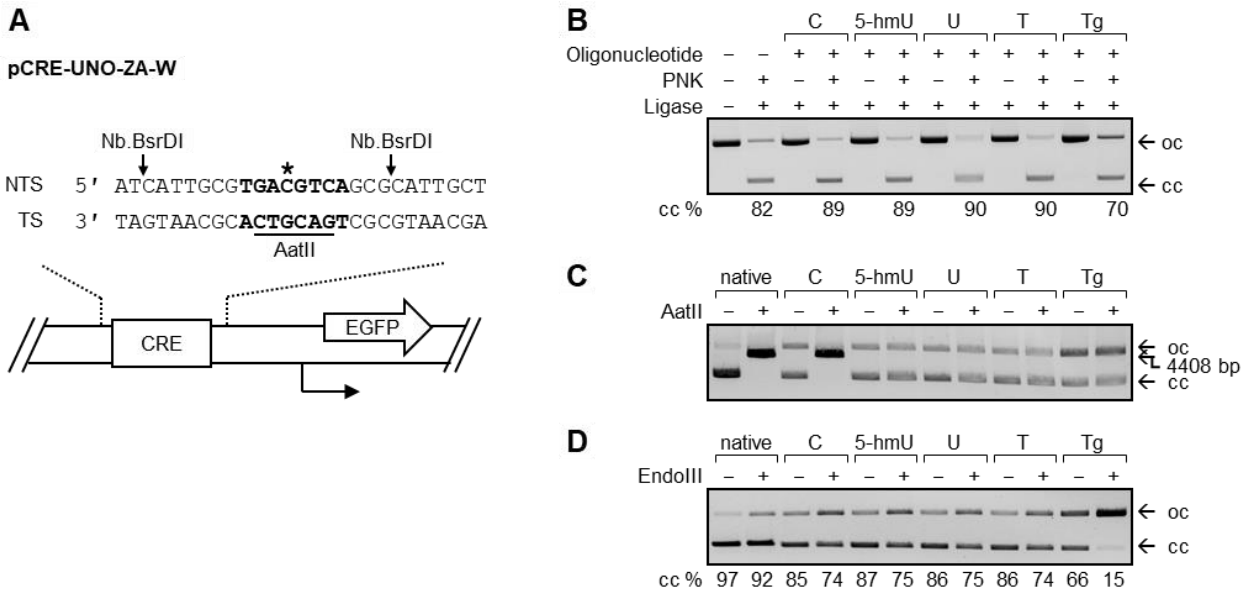


Figure 7-38: Construction of CRE-UNO reporters containing single 5-hmU, U, T or Tg opposite to G at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element

(A) Scheme of the standalone CRE gene regulatory element within the pCRE-UNO-ZA-W reporter: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking sites (black arrows) and position of 5-hmU, U, T or Tg in the incorporated oligonucleotides (asterisk). (B) Ligation of Nb.BsrDI-nicked and gapped CRE-UNO constructs with synthetic oligonucleotides containing the specified DNA modification or the C control in the presence and absence of PNK. (C) Incubation of the generated constructs with the restriction enzyme AatII to verify the presence of all DNA modifications based on the inhibition of the plasmid cleavage. Arrows indicate migration positions of the linearised vector (4408 bp) and the cc- and oc-form plasmid DNA. (D) Incubation of the generated constructs with EndoIII to verify the presence of Tg by EndoIV nicking.

Quantitative expression analysis of HeLa cells (chapter 6.32) showed that 5-hmU, U, Tg and T DNA modifications inhibited the reporter activity 6 hours after transfection, resulting in a residual EGFP expression of 80-88% (Figure 7-39). The negative effects of 5-hmU, U and Tg on transcription intensified with time, reducing the EGFP expression down to 16% (5-hmU), 21% (U) and 18% (Tg) at the 48 hours' time point. In contrast, the EGFP expression from constructs containing the MMR substrate T:G remained steady over 48 hours. The different transcriptional outcomes of BER and MMR substrates indicate that 5-hmU, U, Tg and T affect gene expression by a mechanism which is repair-pathway specific.

The EGFP expression analysis of modified CRE-UNO constructs showed, that from the four investigated DNA lesions, only the three BER substrates 5-hmU, U and Tg induce a drastic reduction of the gene expression over time, which was absent for the T:G miss match.

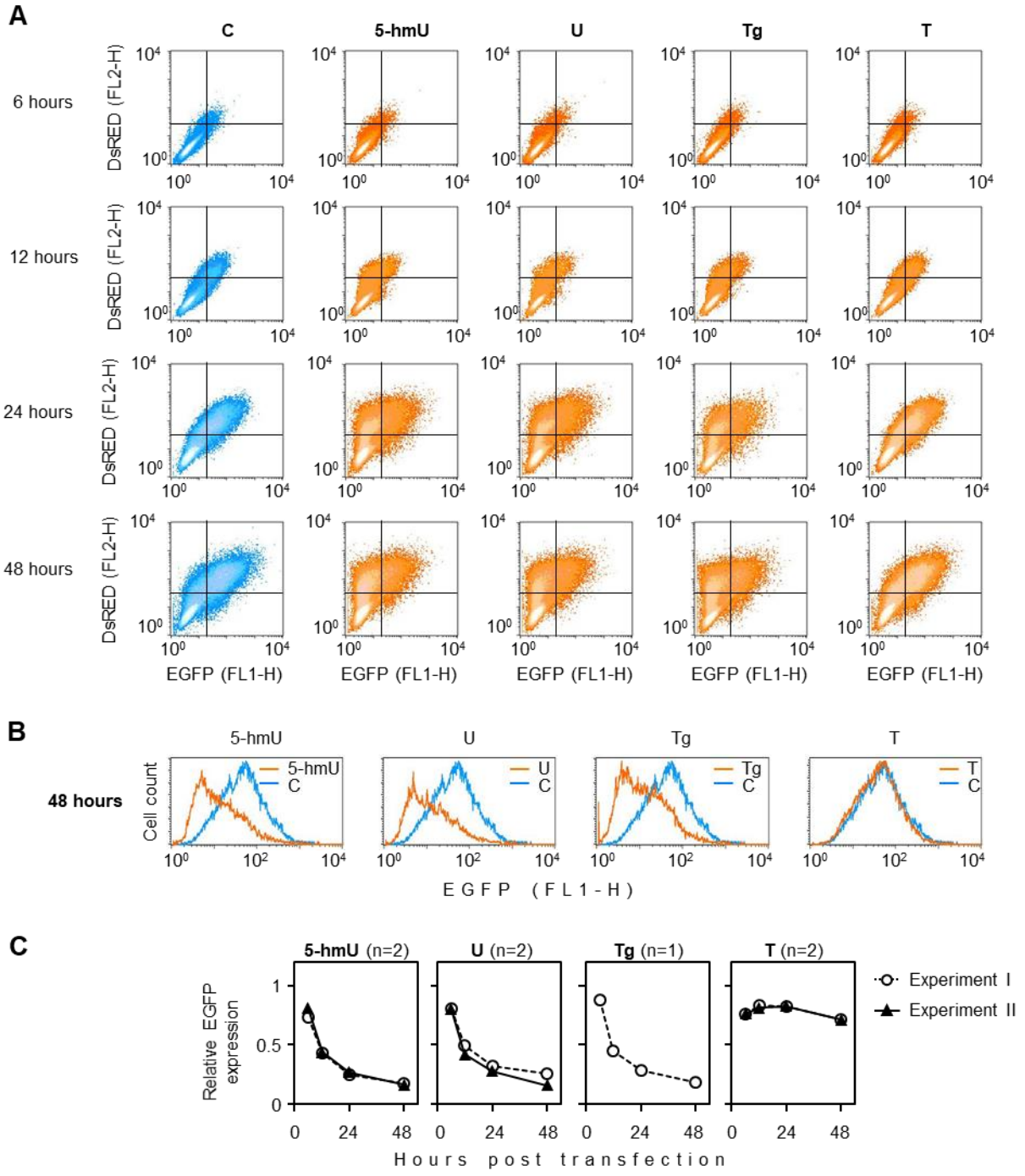


Figure 7-39: Expression of CRE-UNO reporters containing a single 5-hmU, U, Tg or T residue opposite to G at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element in HeLa cells

Quantitative EGFP expression analysis of HeLa cells transfected with pGCbox-ZA-W reporters containing a single 5-hmU, U, Tg or T residue or the respective C control opposite to G at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE upstream regulatory element. Representative (A) scatter plots, (B) fluorescent distribution plots of HeLa cells 6-, 12-, 24- and 48-hours after transfection. (C) Quantification of the relative EGFP expression of transfected HeLa cells in two or one independent experiments.

7.3.2 Establishment of a stable SMUG1 knockout in HeLa cells

The strength and progression of the transcriptional repression induced by 5-hmU, U and Tg in CRE was comparable to the functional consequences caused by BER of 5-fC and 5-caC at this position in CRE-UNO (Figure 7-39 versus Figure 7-2). Since transcriptional repression by 5-fC and 5-caC was induced by TDG-initiated BER (cf. chapter 7.1), it was investigated whether repression of the gene expression by 5-hmU, U and Tg was also induced by a BER-dependent mechanism. Therefore, the transcriptional effects of 5-hmU, U and Tg were analysed in the absence of BER, after generating stable CRISPR-CAS9-mediated DNA glycosylase knockout cell lines. Knockout cell lines of all DNA glycosylases potentially contributing to the repair of 5-hmU, U and Tg should be generated, including TDG (5-hmU), NTHL1 (Tg), SMUG1 (5-hmU, U) and MBD4 (5-hmU). Since U induced transcriptional repression had already been proven to be completely dependent on UNG initiated BER[280,300] and stable TDG and NTHL1 knockout cell lines were already at hand (TDG knockout cf. chapter 7.1.6, NTHL1 knockout generated by Aalaa Farag under my guidance[295]), knockout attempts were focused on SMUG1 and MBD4.

SMUG1 and MBD4 knockout generation followed the same procedure as TDG knockout described earlier (cf. chapter 7.1.6). Both genes were simultaneously targeted at two sgRNA sites enclosing the active site coding nucleotides. Total gene deletion (T) as well as active site deletion approaches (A) were tested for their knockout efficiency. *SMUG1* sgRNAs were designed as described in chapter 6.20 to enclose the nucleotides coding for essential proline 97 amino acid by targeting positions -742 and +1953 (sgRNA set T) or +153 and +1546 (sgRNA set A) as schematically depicted in the upper panel of Figure 7-40 A. The sgRNAs for total gene deletion additionally enclosed the nucleotides coding for histidine 239 active site residue[301], whose significance for base removal was established shortly after the SMUG1 knockout was performed. *MBD4* sgRNAs were selected to target the positions -339 and +8652 (sgRNA set T) or +7651 and +8479 (sgRNA set A) to enclose the nucleotides coding for aspartic acid 560 active site residue as schematically depicted in the upper panel of Figure 7-40 B. Synthetic oligonucleotides used to generate the sgRNA coding DNA insert are listed in Table 6-15 columns 9-24, p. 60. After cloning the corresponding knockout constructs as described in chapter 6.21 (for sequences cf. Appendix III), PCR-based detection of non-rearranged and rearranged *SMUG1* and *MBD4* was established for knockout approaches A and T. The positions of the sgRNA specific

PCR primers (chapter 6.26) within the *SMUG1* and *MBD4* genes are depicted in the middle panel of Figure 7-40 A and B and optimisation and verification of PCR product amplification is shown in Appendix I 6 and Appendix I 7. HeLa cells were transfected with the gene specific knockout construct pairs as described in chapter 6.22 and the *SMUG1* and *MBD4* editing efficiency was analysed on the gene level. Genomic DNA of transfected and untransfected HeLa cells was extracted 24-, and 48-hours after transfection as described in chapter 6.24 and the presence of rearranged and non-rearranged *SMUG1* and *MBD4* genes was verified by PCR (chapter 6.27). Two separate PCR reactions were performed to assess the presence of non-rearranged or rearranged genomic *SMUG1 and MBD4* in HeLa cells. Detection of rearranged and non-rearranged genes was possible in the same PCR run, because the S7 Fusion Polymerase was able to amplify short products from rearranged genes as well as long PCR products from non-rearranged genes in the same reaction using the same primer pair.

The total gene deletion approach efficiently generated rearranged *SMUG1* alleles in HeLa cells, as demonstrated by the presence of the corresponding 425 bp long PCR fragment from primer pair SMUG1_-865F/2243R within the 24 hours sample (Figure 7-40 A+B, lower panel, 24h). gDNA from untransfected HeLa cells was used as PCR template to assess if primers specifically amplified the desired *SMUG1* sequence, thereby verifying that the 425 bp PCR product could not amplify in the absence of gene editing. PCR with gDNA of untransfected cells yielded the expected 3109 bp PCR product from non-rearranged *SMUG1* genes(Figure 7-40 A, lane 2, 3109 bp fragment). No signal was detected at 400 bp, verifying that this PCR product was specific for rearranged *SMUG1*. The absence of non-rearranged PCR products in samples containing gDNA from HeLa cells sampled 48 hours after transfection (Figure 7-40 A, lane 1) indicates that the DNA template was lost during the gDNA preparation procedure and the sample were disregarded for further analysis. PCR assay using the primer pair SMUG1_-158F/2243R failed to demonstrate the *SMUG1* editing success by active site deletion. Thus, a PCR product of a similar length as the 1019 bp PCR product of rearranged SMUG1 was already amplified from gDNA of untransfected cells (Figure 7-40 A, lanes 6). It is therefore impossible to determine if *SMUG1* editing was induced by the active site deletion approach. PCR product of non-rearranged *SMUG1* (2401 bp) was observed in all sample containing gDNA, verifying successful PCR set-up.

Gene editing of *MBD4* using the total gene deletion- or active site deletion approach was detected by PCR using primer pairs MBD4_-409F/8929R and MBD4_7171F/8731R respectively. Agarose gel electrophoresis (cf. chapter 6.12) of PCR samples containing gDNA of transfected HeLa cells failed to show any signal of rearranged *MBD4* at the expected product size of 347 bp and 725 bp respectively (Figure 7-40 B). PCR product of non-rearranged *MBD4* (1559 bp) was only observed for cells targeted for *MBD4* active site deletion but not total gene deletion. It is assumed that the lack of non-rearranged *MBD4* detection by primers MBD4_-409F/8929R was due to the big PCR product length of 9338 bp, which may not be amplifiable by S7 Fusion Polymerase. Successful PCR set-up could thus only be confirmed for active site deletion approach. To assess total gene deletion efficiency of *MBD4,* the PCR has needs to be repeated using a different primer combination for the detection of non-rearranged *MBD4.* To successfully knock out MBD4 in HeLa cells, the MBD4 knockout approach requires optimisation either by testing a different combination of primers to detect rearranged *MBD4* (e.g. MBD4_8479F/+ 8929R), combine a different set of existing knockout constructs or designing completely new sgRNAs to efficiently induced *MBD4* editing.

In summary, the PCR analysis of HeLa cells transfected with a pair of sgRNA-CAS9 expression vectors showed efficient *SMUG1* editing by the total gene deletion approach. Due to the lack of time, *MBD4* editing could not be optimised, wherefore the generation of DNA glycosylase knockout cell lines was focused on SMUG1.

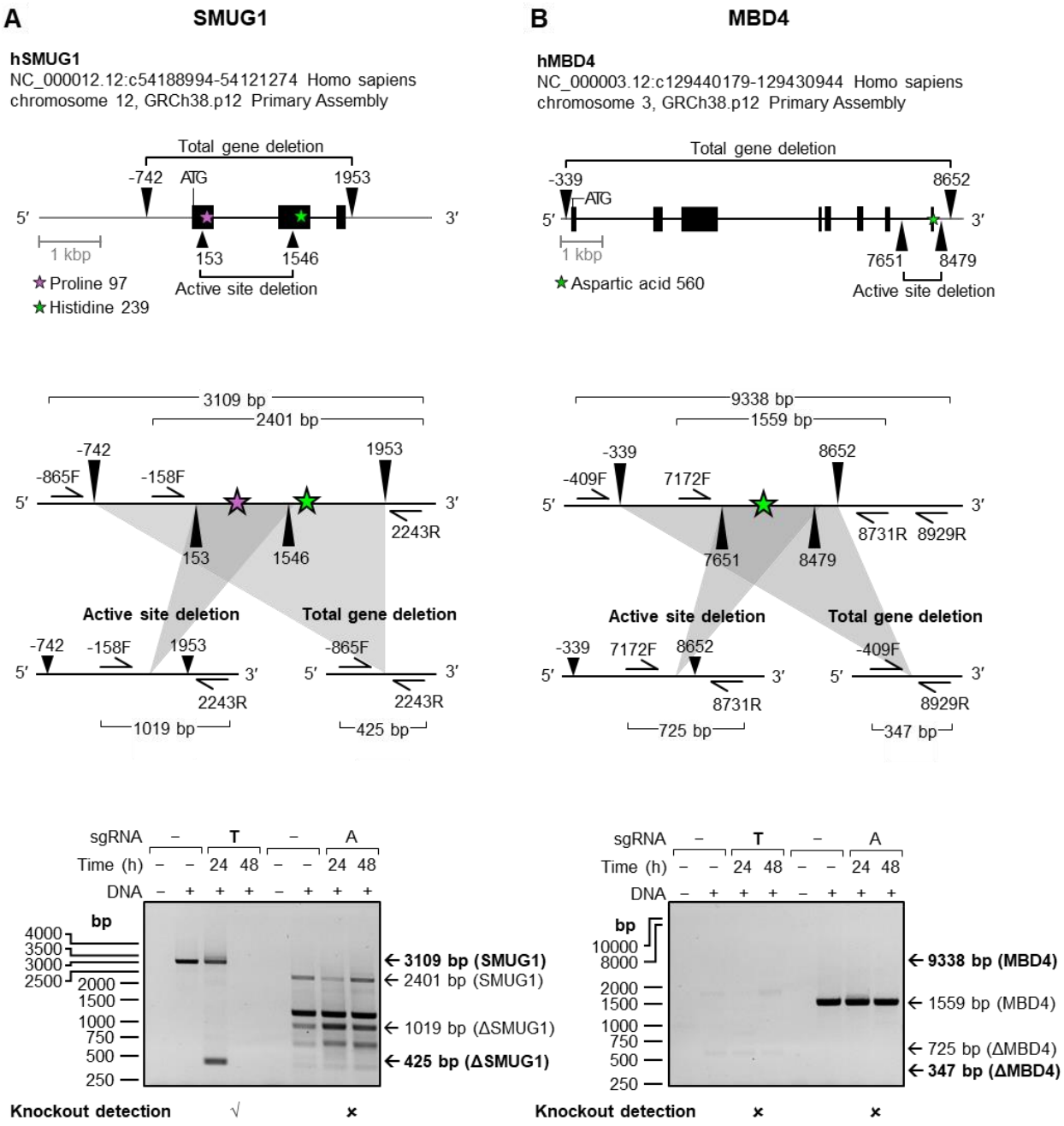


Figure 7-40: CRISPR-CAS9-mediated knockout of SMUG1 and MBD4 by simultaneous targeting of two gene sites

Establishment of a HeLa derived SMUG1 (A) and MBD4 knockout cell line (B) by either deleting the total gene (T) or specifically deleting the active site coding exon (A) by CRISPR-CAS9-mediated gene editing using a combination of two sgRNAs. (Upper panel) True to scale scheme of the human *SMUG1* and *MBD4* sequence: upstream/downstream DNA sequence (grey line), gene coding region (black line), position of ATG start codon (broken line), exons (black boxes), active site coding nucleotides (green star), essential amino acid coding nucleotides (purple star), sgRNA targets (triangle with distance from the TSS indicated above). (Middle panel) Schematic representation of the non-rearranged and rearranged SMUG1 and MBD4 gene after active site deletion or total gene deletion: active site coding sequence (green and purple star), sgRNA targets (triangle) and PCR primer sites (half arrow with distance from the TSS indicated above) with the according PCR product length. (Lower panel) Detection of the SMUG1 and MBD4 knockout in genomic DNA by PCR screening. gDNA from cells transfected with different sets of sgRNA coding constructs was extracted 24- and 48-hours post transfection and used as a PCR template with gDNA of untransfected cells as a control. Total gene deletion of SMUG1 is detected using primers SMUG1_-865F/2243R (425 bp PCR fragment, labelled bold), whilst active site knockout is detected using primers SMUG1_-158F/2243R (1019bp PCR fragment). Total gene knockout of MBD4 is detected using primers MBD4_-409F/8929R (347 bp PCR fragment, labelled bold), whilst active site knockout is detected using primers MBD4_7172F/8731R (725 bp PCR fragment). Non-rearranged genes were detected with the same primer pairs generating products of 3109 bp/2401 bp and 9338 bp/1559 bp from non-rearranged *SMUG1* and *MBD4* respectively. The migration positions of the PCR products within the agarose gel are indicated by arrows.

Based on its higher gene editing efficiency, the total gene deletion approach was used to generated a HeLa derived SMUG1 knockout cell line. HeLa cells were co-transfected with pX330-spCAS9-HF_SMUG1-742, pX330-spCAS9-HF_SMUG1+1953 and the pZAJ transfection marker as described in chapter 6.22 and depicted in Figure 7-41. The cells were sorted approximately 24 hours after transfection (chapter 6.23) as exemplified for TDG in chapter 7.1.6. Top 5% EGFP fluorescent HeLa cells were selected from the pool of living, transfected cells as shown in Figure 7-41 and sorted as single cells with a total amount of 192 clones. The single cell clones were cultivated in 96-well plates for two weeks and 28 of the surviving single cell clones were selected fur further expansion followed by SMUG1 knockout validation on the genome level.

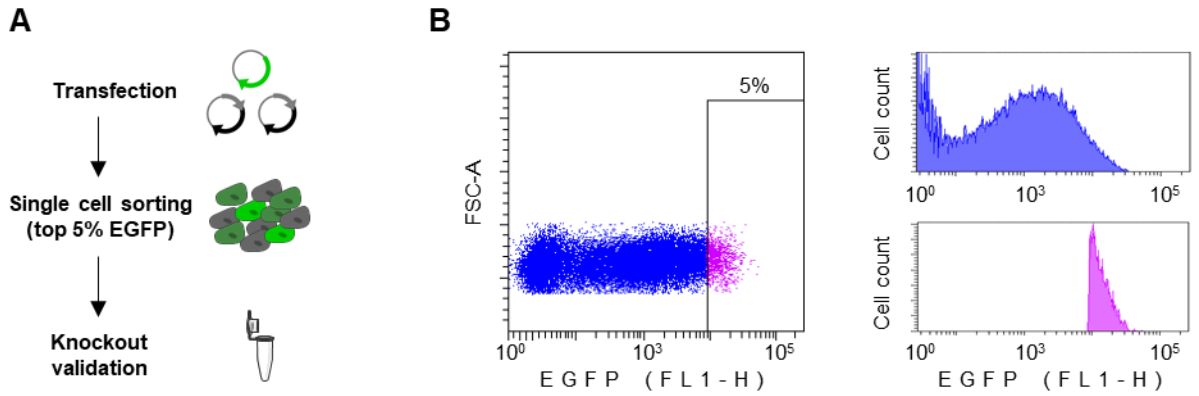


Figure 7-41: Single cell sorting of HeLa cells with potential CRISPR-CAS9-mediated SMUG1 knockout

Single cell sorting of HeLa cells with potential SMUG1 knockout 24 hours after transfection with knockout constructs pX330-spCAS9-HF_SMUG1-742 and pX330-spCAS9-HF_SMUG1+1953 and the pZAJ EGFP transfection marker. (A) Schematic presentation of the approach to generate and select HeLa cells with high SMUG1 knockout potential. HeLa cells are transfected with a pair of knockout constructs and an EGFP transfection marker and sorted as single cells according to top 5% EGFP fluorescence. The single cell clones are expanded and used for knockout validation by PCR and western blot. (B) Quantitative EGFP expression analysis of transfected HeLa during the sorting procedure, selecting cells with top 5% of EGFP fluorescence of the living cell population: representative scatter plot (left panel), fluorescent distribution plot of living cells (top right panel) and fluorescent distribution plot of the selected population with top 5% EGFP fluorescence (lower right panel).

To validate the SMUG1 knockout in selected HeLa derived single cell clones on the gene level, “quick” cell extracts were prepared as described in chapter 6.25 and used as template for PCR. PCR screening for the desired 3109 bp deletion within the *SMUG1* gene was performed using the primer pair SMUG1_-865F/2243R, whereas non-rearranged *SMUG1* alleles were detected using primers SMUG1_1872F and SMUG1_2243R. The expected 425 bp PCR product from rearranged *SMUG1* was detected in 19 of the 28 analysed single cell clones (Figure 7-42, upper gel). Genomic DNA and cell lysate from untransfected and transfected HeLa cells only showed rearranged product amplification in samples of transfected cells, thereby validating the primer specificity. PCR reactions with “quick” cell lysate from the 6 single cell clones

1-B1, 1-C9, 1-E11, 1-F4, 2-C7 and 2-D3 did not amplify the 371 bp product from non-rearranged *SMUG1* (Figure 7-42, lower gel, labelled bold), indicating that gene editing took place in all four *SMUG1* alleles. As expected, control samples containing DNA from untransfected HeLa cells efficiently amplified non-rearranged PCR product, thus validating the primer specificity.

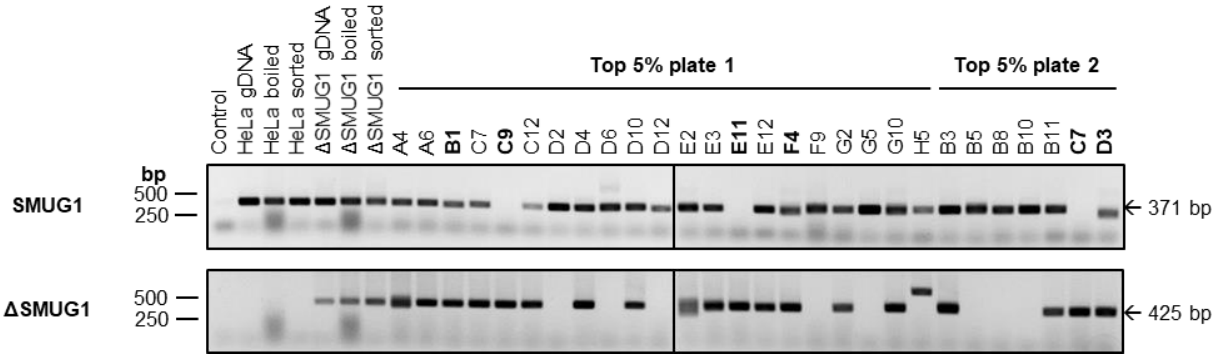


Figure 7-42: Validation of the SMUG1 knockout in selected HeLa single cell clones on the gene level via PCR

Screening of the selected HeLa derived single cell clones for the desired SMUG1 knockout by PCR. 28 of the sorted single cell clones with potential SMUG1 knockout were selected for PCR screening, to detect rearranged (lower panel) and non-rearranged SMUG1 genes (upper panel). Quick-cell extracts were prepared from the clones and a PCR test was performed to detect non-rearranged SMUG1 (371 bp PCR fragment) using primers SMUG1_1872F/2243R. A separate PCR was performed with primers SMUG1_-865F/2243R to verify the desired DNA deletion in SMUG1 genes (425 bp PCR fragment). The migration positions of the PCR products within the agarose gel are indicated by arrows. Clones with promising PCR results are labelled bold.

Potential SMUG1 knockout within clones 1-B1, 1-C9, 1-E11, 1-F4, 2-C7 and 2-D3 was further validated in the next step. Due to the lack of a working SMUG1 specific antibody for western blot analysis[300], the knockout was verified by quantifying the SMUG1 activity of single cell clones by HCR assay. The SMUG1 activity analysis is based on previous findings, showing that 5-hmU opposite to A at position R+2 within the CRE sequence of the CRE-UNO reporter progressively represses the gene expression in a completely SMUG1-dependent manner[97,300]. Thus, SMUG1 activity of HeLa cells can be quantified by transfection with 5-hmU:A containing CRE-UNO reporters and subsequent expression analysis.

The strand exchange method (cf. chapter 6.30) was used to generate CRE-UNO constructs containing a single 5-hmU:A base pair at position R+2 within the CRE sequence (Figure 7-43 A; for construct generation cf. Appendix I 8). Afterwards, the six potential SMUG1 knockout clones were transfected with CRE reporters containing 5-hmU:A or the respective T:A control (Figure 7-43 B) and the EGFP expression was quantified (chapter 6.32). Parental SMUG1 proficient HeLa cells and SMUG1 knockdown cells were used as a control. As established previously[97,300], 5-hmU:A reduced the gene expression in a time-dependent manner, from initially 96%- (6 hours)

down to 52% (24 hours) of residual EGFP expression (Figure 7-43 C and D). Transcriptional repression by 5-hmU:A was completely abolished in clones 1-C9, 1-E11, 2-C7 and 2-D3 verifying the efficient SMUG1 knockout in those cells. Clones 1-B1 and 1-F4 showed a significant but incomplete inhibition of the transcriptional repression by 5-hmU:A, resulting in a ~20% reduction of the EGFP fluorescence 24 hours after transfection. Since SMUG1 knockdown minimised the transcriptional repression by 5-hmU:A to a similar extend, it is assumed that clones 1-B1 and 1-F4 contained residual amounts of SMUG1.

In summary, the quantitative EGFP expression analysis verified the successful generation of four different HeLa derived single cell clones with the desired CRISPR-CAS9-mediated SMUG1 knockout and clone 1-E11 was selected to be used in further experiments.

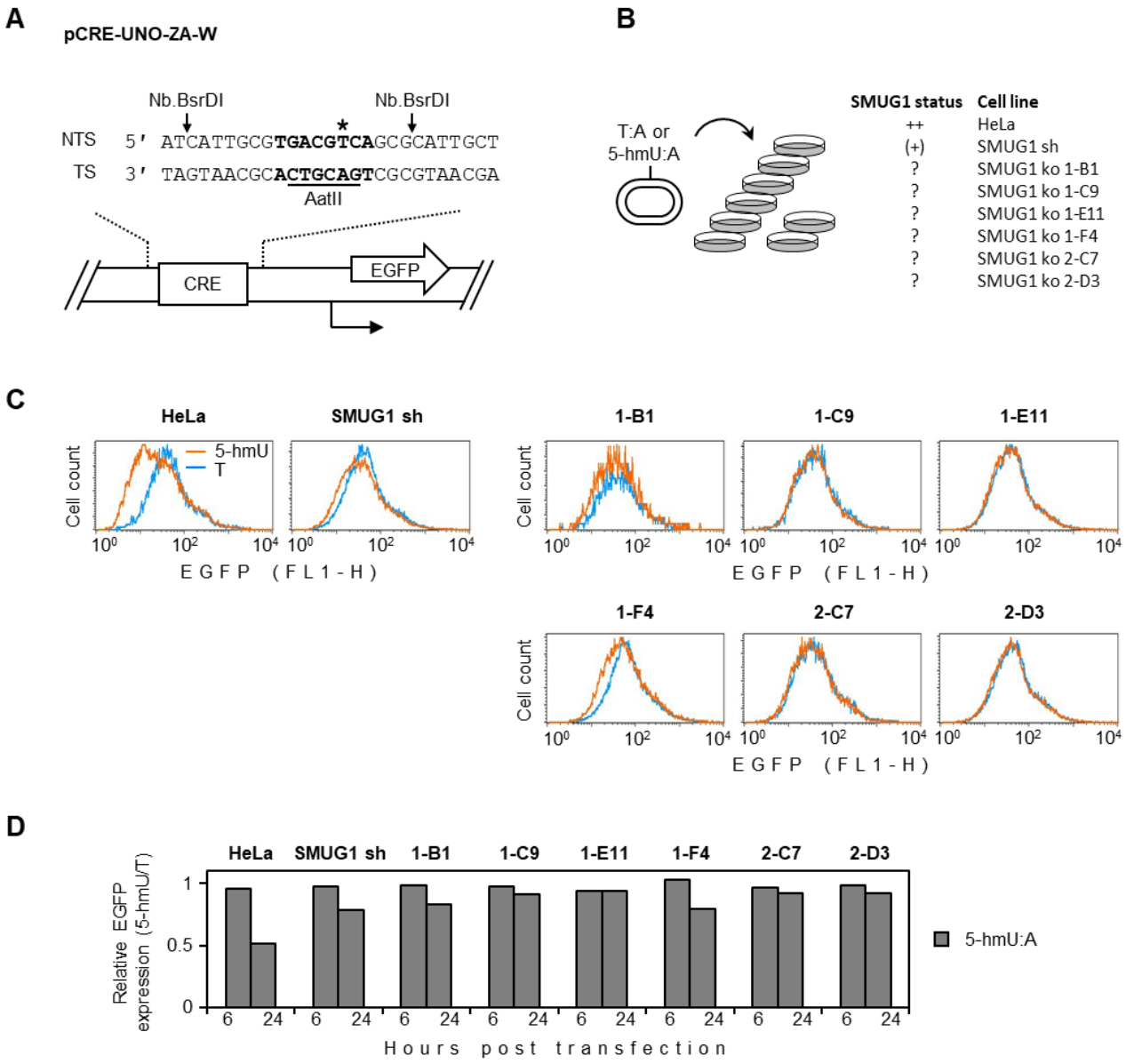


Figure 7-43: Expression of CRE-UNO reporters containing 5-hmU opposite to A within a standalone CRE gene regulatory element in selected SMUG1 knockout clones

Quantitative EGFP expression analysis of HeLa derived SMUG1 knockout clones, the control HeLa cell line (no ko) and a stable SMUG1 knockdown cell line transfected with pCRE-UNO-ZA-W reporters containing single 5-hmU opposite to A or the respective T control at position R+2 (counted from central CpG dinucleotide) within the non-transcribed strand of the standalone CRE upstream regulatory element. (A) Scheme of the standalone CRE gene regulatory element within the pCRE-UNO-ZA-W reporter: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking sites (black arrows) and position of 5-hmU or the respective T control in the incorporated oligonucleotides (asterisk). (B) Transfection scheme of HeLa cells with different SMUG1 status with CRE-UNO reporters containing 5-hmU or the respective T control. (C) Fluorescent distribution plots (24 hours only) and (C) quantitative EGFP expression of cells 6- and 24-hours after transfection (n=1).

7.3.3 The transcriptional repression by 5-hmU, U and Tg but not T opposite to guanine is potentially BER-dependent

To assess the impact of BER on the functional consequences of 5-hmU, U, T, and Tg, the HeLa derived cells with stable knockout of the DNA glycosylases SMUG1 (clone 1-E11), TDG (clone 2-F3) and NTHL1 (clone 1-F5) as well as the parental, DNA glycosylase proficient HeLa cell line were transfected with pCRE-UNO-ZA-W constructs containing 5-hmU:A, 5-hmU:G, U:G, T:G, Tg:G or the respective unmodified control (chapter 6.32). Quantitative EGFP expression analysis over 24 hours showed the expected transcriptional repression by 5-hmU:G, 5-hmU:A, U:G and Tg:G, reducing the EGFP expression to 34%, 52%, 40%, and 29% after 24 hours respectively (Figure 7-44 B and C). Knockout of NTHL1 completely reverted the transcriptional repression induced by Tg, verifying that NTHL1 was essential for the regulation mechanism.

Interestingly, SMUG1 knockout did not detectably affect the transcriptional repression induced by 5-hmU opposite to G, whereas it completely reverted the repressive effects of 5-hmU opposite to A as seen previously[97]. TDG and NTHL1 knockout reverted the 5-hmU:G induced transcriptional by 7%, however additional experiments need to be performed to verify the significance of this effect. Based on the expression analysis, it is assumed that transcriptional repression by 5-hmU:A was mediated by SMUG1 dependent mechanism, whereas transcriptional repression by 5-hmU:G did not solely depend on the presence of SMUG1, TDG or NTHL1 in human cells. The transcriptional repression by 5-hmU:G, 5-hmU:A and Tg in CRE progressed in a similar fashion as shown for 5-fC and 5-caC and equally dependent on the presence of lesion specific DNA-glycosylases, supporting the hypothesis that also the transcription regulation mechanism is similar. It is therefore assumed that repression of the gene expression by 5-hmU:A and Tg was caused by SMUG1 and NTHL1-initiated BER respectively. Following this notion, complete absence of transcriptional silencing by Tg in NTHL1 ko cells indicates that NTHL1 is the only DNA glycosylase to excise Tg:G in CRE. A similar DNA glycosylase dependency was established for 5-hmU:A, indicate that SMUG1 is the only DNA glycosylase to remove 5-hmU opposite to A in vivo, in accordance with pervious findings[248]. Since SMUG1 knockout did not affect transcriptional repression by 5-hmU:G, it can be assumed that several back-up pathways exist to repair 5-hmU opposite to G in the absence of SMUG1 as described previously[302]. Candidates for in vivo excision of 5-hmU opposite to G are TDG and MBD4, which were shown to act on

5-hmU:G containing DNA [251,252,254]. Assessment of the functional consequences of 5-hmU:G in cell lines with double and triple knockout of SMUG1, TDG and MBD4 could be used to determine the contribution of the respective DNA glycosylases to BER of 5-hmU:G. As expected, SMUG1, NTHL1, and TDG knockout had no influence on the transcriptional repression induced by U (residual EGFP expression which was 3-7% higher than in the corresponding glycosylase proficient cells) and on the stable EGFP expression by T:G containing constructs.

The quantitative EGFP expression analysis of 5-hmU, U, Tg and T containing reporters in DNA glycosylase deficient cells showed that the transcriptional repression induced by 5-hmU:A and Tg:G was not caused by the DNA lesion itself but by a SMUG1 and NTHL1 dependent mechanism. 5-hmU:G and U:G induced an equal degree of transcriptional repression, which due to its resemblance with the outcomes of 5-hmU:A, Tg (above), 5-fC, 5-caC (cf. chapter 7.1) and 8-oxoG[97] for gene expression is believed to be induced by BER. Detected for five different base modifications in two different UREs (CRE and GC box), the presented results indicate that indirect, BER-mediate transcriptional repression is a common mechanism of transcription regulation by modified nucleobases.

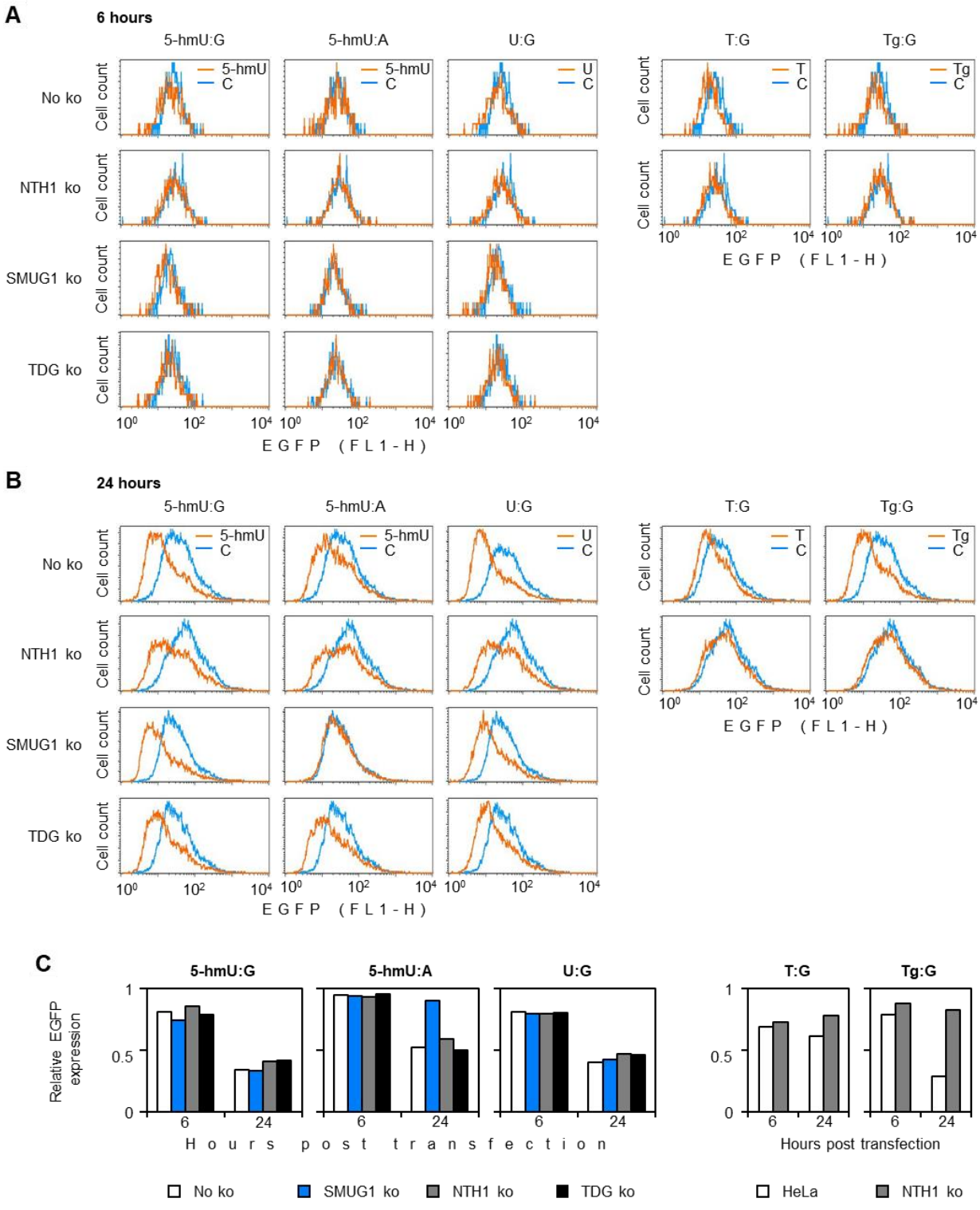


Figure 7-44: Expression of CRE-UNO reporters containing 5-hmU, U, T or Tg in the standalone CRE gene regulatory element in isogenic cell lines with stable SMUG1, TDG or NTHL1 knockout

Quantitative EGFP expression analysis of clonal HeLa cell lines with stable SMUG1 (1-E11), TDG (2-F3) or NTH1 (1-F5) knockout and the control HeLa cell line (no ko) transfected with pCRE-UNO-ZA-W reporters containing single 5-hmU, U, T or Tg opposite to G, the respective C control at the central CpG dinucleotide within the non-transcribed strand of the standalone CRE gene regulatory element or a control reporter containing 5-hmU or T opposite to A at position R+2. (A) Fluorescent distribution plots of HeLa and HeLa derived SMUG1, TDG or NTHL1 knockout cells 6- and (B) 24 hours after transfection. (C) Quantification of the relative EGFP expression of HeLa and HeLa derived SMUG1, TDG or NTHL1 knockout cells 6- and 24 hours after transfection (n=1).

8 Discussion

Alterations of the canonical DNA bases are frequently found in the human genome and are linked to many common diseases including cancer. Astonishingly, modifications of the nucleobases can influence the affected gene's expression already at an occurrence of only one modified residue per promoter sequence[5,276,278,280]. Transcription regulation by a single modified nucleobase has been reported previously in this lab for the oxidatively induced DNA modifications 5-fC, 5-caC and 8-oxoG, which significantly reduce the reporter gene expression in the CRE gene regulatory element[5]. Since all three modifications are present in the human genome under physiological conditions[1,59,303–305] where they may alter transcription of the affected genes, it is very important to determine the impact of 5-fC, 5-caC and 8-oxoG on gene expression and understand its cause. However, the knowledge about transcriptional consequences of 5-fC and 5-caC is yet very limited due to the base modifications' rare occurrence in human DNA. Although studied in more detail, basic functional consequences of 8-oxoG are still discussed controversially, since the transcriptional outcomes of the major guanine oxidation product vary strongly depending on the DNA sequence context and 8-oxoG repair[126,138,140,306,307]. In this project the functional consequences of the oxidatively induced DNA modifications 5-fC, 5-caC and 8-oxoG were investigated in the already studied CRE and the newly selected GC box gene regulatory element, with the aim to determine the effects of the primary base modifications and separate them from repair induced influences on the gene expression.

8.1 5-fC and 5-caC in CRE and GC box reporters negatively influence the gene expression by a dual mechanism of direct promoter inhibition and BER-mediated gene silencing

The influence on the gene expression of a single 5-fC and 5-caC residue within a CRE and GC box gene regulatory element was assessed in the first section of this work. The conducted experiments were based on previous findings of Julia Allgayer (AK Khobta), which showed that 5-fC and 5-caC in a minimal CRE promoter induced transcriptional repression that intensified with time and correlated with cellular levels of the repair protein TDG[97]. Experimental setup was however insufficient to establish a causal connection between TDG levels and transcriptional repression by 5-fC and

5-caC to determine the effects of the primary base modifications on the gene expression in human cells.

Hence one aim of this project was to validate the functional consequences of 5-fC and 5-caC as primary base modifications in the absence of repair. This was achieved using chemically stabilised derivatives of 5-fC and 5-caC and assessing their effects on the gene expression in CRE-UNO reporters. Expression analysis of reporters containing BER-resistant 2′-fluorinated derivatives of 5-fC and 5-caC at the central CpG dinucleotide of a strand alone CRE gene regulatory element showed that both primary modifications negatively affected transcription (cf. chapter 7.1.1). The prompt expression reduction in both elements already detectable after 6 hours, indicates that 5-fC and 5-caC directly diminish the promoter activation. It is unlikely that the promoter inhibition is caused by a blockage of RNAP2 because 5-fC and 5-caC are positioned 86 bp upstream from TSS (cf. Appendix II and[97]). More probable is that 5-fC and 5-caC directly diminish the binding of distant TFs and thereby reduce promoter activation. Indeed, cell free assays showed that single 5-fC and 5-caC residues within the CRE gene regulatory element partially inhibit the binding of the CREB transcription factor[5]. Transferring these results to a cellular context, direct CRE inhibition by 5-fC and 5-caC likely results from abolished CREB-binding, which usually would promote CRE-UNO activation.

Interestingly, 5-fC and 5-caC in the strong CRE promoter CMV-1111 did not directly reduce the promoter activity(cf. chapter 7.1.3). Considering that diminished TF binding by 5-fC and 5-caC may be the cause for CRE-UNO inhibition, it is reasonable to conclude that in CMV-1111, diminished promoter activation by TF binding inhibition was too weak to be detected since the promoter contains four instead of one CRE sites. Thus, 5-fC or 5-caC at one of the four CRE sites in CMV-1111 only diminishes TF binding at the affected CRE, leaving the three other CRE sites available for TF binding, which enhances the CMV-1111 activity.

To assess the functional consequences of 5-fC and 5-caC in a gene regulatory element with different GC-content, the GC box was selected as a representative GC-rich URE. Quantitative EGFP expression analysis of minimal GC box reporters containing a single 5-fC or 5-caC residue at the central CpG dinucleotides of a standalone GC box showed, that BER-resistant 5-caC and to a lesser extent also 5-fC had a negative impact on promoter activity (cf. chapter 7.1.11 and illustrated in Figure 8-1).

Comparable to CRE, the negative impact of 5-fC and 5-caC on GC box activity was imminent and independent from repair. As a similarly BER-independent, imminent promoter inhibition was shown for 5-fC and 5-caC in the minimal CRE promoter (cf. chapter 7.1.1), it is assumed that both effects are caused by the same mechanism, namely by TF binding inhibition. Transcription factors which bind to GC box consensus sequences are for example proteins of the SP family, including SP1[286,308], SP2[309,310] and SP3[311]. The effect of cytosine methylation on SP1 and SP2 binding to their target sequence has been controversially discussed, since 5-mC was shown to have both negative[88,91,92] and positive effects on DNA-protein interaction[87]. Although the influences of 5-fC and 5-caC on TF binding have not been the target of a detailed investigation till date, it is probable that 5-fC and 5-caC have a more drastic negative effect on SP-GC box binding than 5-mC. This assumption is based on findings on other TFs, showing that 5-fC and 5-caC reduce ERG1-DNA binding more drastically than 5-mC, owing to the greater steric hinderance and electrostatic repulsion of the formyl- and more severely of the carboxy group[95,96]. Following this notion, SP1 was shown to established several hydrogen bonds and hydrophilic interactions with the DNA bases of the purine-rich strand of the GC box, which may be abolished by the presence of modified nucleobases. Cγ and Cδ of glutamic acid 583 of SP1 forms hydrophobic interactions with C5 and C6 of cytosine at position R-1 of the GC box (CpG dinucleotide in purine-rich strand of the GC box)[312]. When a formyl- or a carboxy group is present at C6 of cytosine instead of a hydrogen atom (5-fC/5-caC instead of cytosine), the hydrophilic groups probably disrupt the hydrophobic interactions with SP1. Furthermore, the hydrophilic formyl- and carboxy group are likely to repel the negatively charged glutamic acid 583 of SP1, thereby destabilising the DNA-protein binding. SP1-GC box interaction was also observed between aspartic acid 610 of SP1 and cytosine at position Y+1, indicating that also in the pyrimidine-rich DNA strand, the presence of 5-fC and 5-caC may destabilise the DNA-protein interaction. Since the carboxy group of 5-caC is bigger and more polar than the formyl group of 5-fC, its negative effects on SP1 binding are assumed to be bigger in magnitude than those of 5-fC. This also explains the relatively small impact of BER-resistant 5-fC on the GC box activity compared to 5-caC, which inhibited the gene expression much more drastically (cf. chapter 7.1.11, 7.1.13 and Figure 8-1, compare F-fC and F-caC). In contrast to 5-fC and 5-caC, the base modifications 5-mC and 5-hmC in either DNA strand did not significantly alter the GC box activity (cf. chapter 7.1.11, 7.1.13 and Figure 8-1). It is

thus assumed, that the effects of 5-mC and 5-hmC on TF binding were neutral for GC box activity in cells. Future studies using electrophoretic mobility shift assays with purified proteins or cell extracts can provide additional insights into the effects of 5-mC, 5-hmC, 5-fC and 5-caC on TF binding to the GC box.

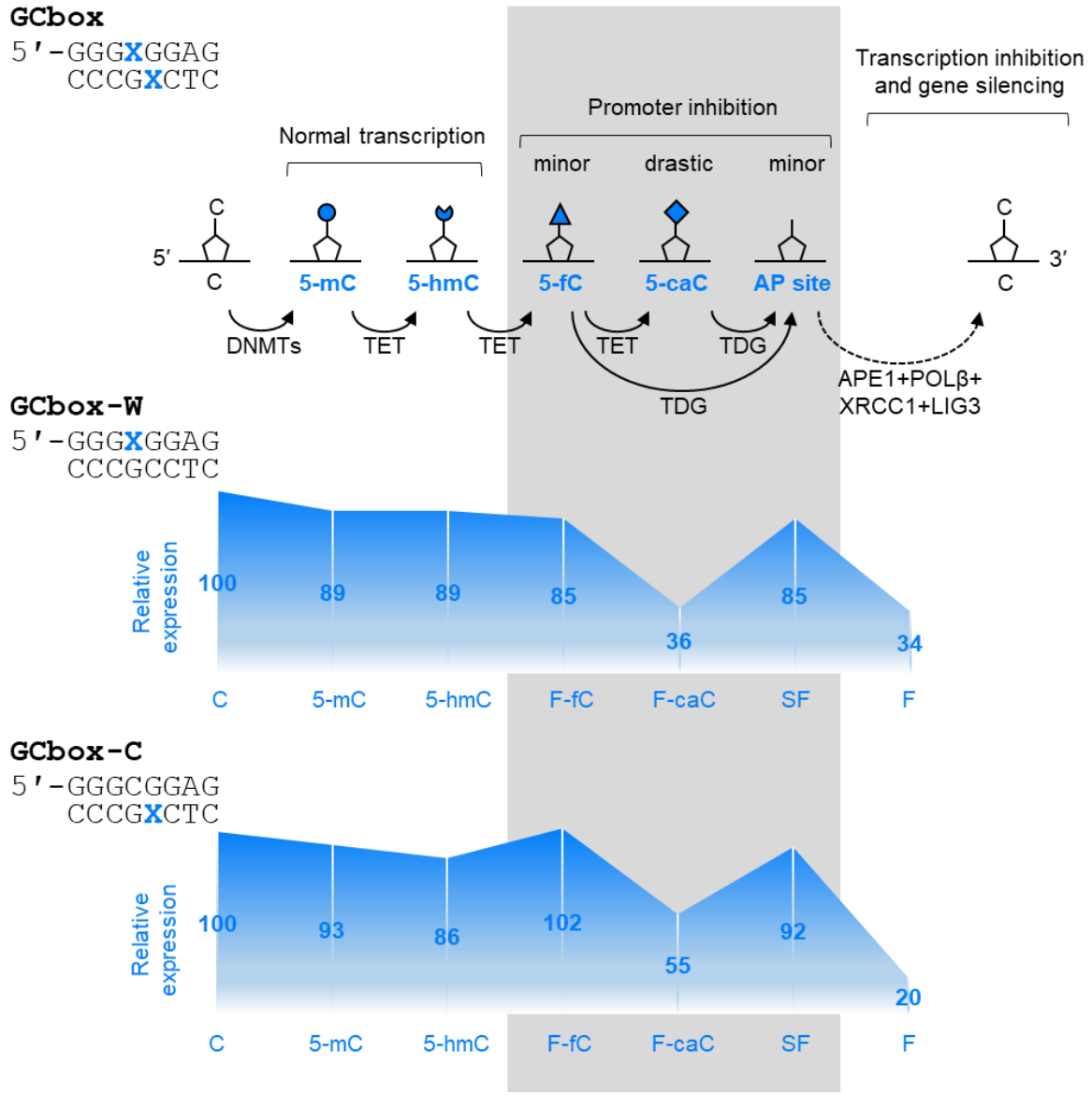


Figure 8-1: Overview of the transcriptional consequences of 5-mC, 5-hmC, 5-fC and 5-caC in the GC box gene regulatory element

Transcriptional effects of the cytosine modifications 5-mC, 5-hmC, 5-fC, 5-caC within the GC box gene regulatory element. (Upper panel) Model of the TET-TDG-dependent active DNA demethylation via 5-hmC, 5-fC, 5-caC and AP site intermediates, with the effects of the respective DNA modification on gene expression indicated above. Dotted arrow indicates reactions which have not been investigated in detail. (Middle and lower panel) Relative reporter expression of HeLa cells transfected with constructs containing either a single 5-mC and 5-hmC (cf. chapters 7.1.9 and 7.1.10), BER-resistant 5-fC or 5-caC (chapter 7.1.11), BER-resistant or BER-sensitive AP site (SF or F, chapter 7.1.15) or the dC control at the central CpG dinucleotide within the purine-rich strand (middle panel) and pyrimidine-rich strand (lower panel) of the GC box.

After determining the functional consequences of the primary base modifications 5-fC and 5-caC in CRE and GC box gene regulatory elements, the impact of BER on transcription regulation by 5-fC and 5-caC was investigated. Quantitative expression analysis of GC box reporters containing BER-resistant and BER-sensitive 5-caC in either DNA strand revealed that promoter inactivation by the primary base modification was diminished in the presence of repair (compare expression of F-caC and 5-caC containing GC box reporters 6 hours after transfection, chapter 7.1.11). Since chemical stabilisation of 5-caC by 2′-fluorination selectively impeded base removal rather than general properties of the nucleobase[294], it is assumed that the different functional outcomes of 5-caC and F-caC are caused by selective processing of 5-caC by BER. Reactivation of the gene expression by 5-caC-repair was also abolished by TDG knockout (cf. chapter 7.1.13), demonstrating that TDG mediated base removal of 5-caC during BER plays an essential part in reactivating the GC box promoter.

Monitoring the gene expression of 5-caC and F-caC containing GC box reporters over the course of 24 hours, showed that BER of 5-caC induced transcriptional repression in the long term (cf. chapter 7.1.11 and Figure 8-2). A similar transcriptional repression was observed for 5-fC in both strands of the GC box as well as 5-fC and 5-caC in CRE (chapters 7.1.11 and 7.1.1) resulting in expression levels well below the basal promoter activity (chapters 7.1.1 and 7.1.11 compare 5-fC/5-caC containing GC box reporter expression with pCRE-ZERO expression). HCR with TDG proficient cells showed that TDG was a prerequisite for the transcriptional silencing of CRE and GC box promoters (cf. chapters 7.1.7 and 7.1.13). Transcriptional consequences of BER-resistant 5-caC and 5-fC in GC box reporters were similar in TDG proficient- and TDG knockout cells (cf. chapter 7.1.14), verifying that mere presence of TDG protein was not sufficient to induce transcriptional repression but needed base removal. These findings also verify that TDG is the only DNA glycosylase to initiate BER of 5-fC and 5-caC in the GC box and induce transcriptional repression in human cells. Since TDG knockout completely abolished the time-dependent reduction of the gene expression by 5-fC and 5-caC repair (cf. chapters 7.1.7 and 7.1.13) whilst SMUG1- and UNG knockdown did not significantly influence this transcriptional repression (cf. chapters 7.1.4 and 7.1.12), it is unlikely that a backup DNA glycosylase exists for TDG to initiate BER of 5-fC and 5-caC. These results thus confirm previous findings in cell free assays, which showed that TDG is the only known human DNA glycosylase to excise 5-fC and 5-caC from the DNA[231,313]. The similar transcriptional consequences of chemically stabilised F-fC

and F-caC and their BER-sensitive 2′deoxy counterparts 5-fC and 5-caC in BER deficient cells (cf. chapters 7.1.11 and 7.1.14) proved that TDG knockout and 2′-fluorination of the nucleotides can be equally used to abolish BER.

Interestingly, the repression magnitude by 5-fC and 5-caC in CMV-1111 was only slightly bigger than in the weak CRE-UNO promoter[5] (chapter 7.1.1 CRE-UNO versus 7.1.3 CMV-1111), indicating that promoter strength does not influence the repression. Transcriptional repression by BER of 5-fC and 5-caC in CMV-1111 further demonstrates that repair-dependent repression signalling spread beyond the local modified CRE, covering the whole promoter. It is thus assumed that an epigenetic silencing mechanism is involved in the establishment of transcriptional repression by BER of 5-fC and 5-caC.
Comparing the effect of 5-fC/5-caC in the NTS versus the TS of the GC box gene regulatory element showed reduction of the gene expression with similar intensities (cf. chapter 7.1.9 and 7.1.10) excluding any strand biases. This is in line with previous findings, showing that transcriptional repression by 5-fC and 5-caC in CRE-UNO reporters is independent from the DNA strand[97].
It is worth noting that the BER-dependent transcriptional silencing by 5-fC/5-caC in either strand of the GC box was significantly more pronounced for 5-fC than for 5-caC (steeper slope of EGFP expression over time cf. chapters 7.1.9 and 7.1.10). A possible explanation for the increased impact of 5-fC is the higher excision efficiency of TDG towards 5-fC compared to 5-caC[170], resulting in a more intense induction of BER and the associated transcriptional repression.

To determine the BER protein or DNA repair intermediate which eventually induced transcriptional repression by 5-fC and 5-caC repair, the consequences of the apyrimidinic site intermediate were investigated in both gene regulatory elements. Comparing the transcriptional consequences of BER-resistant and BER-sensitive AP sites at the position of 5-fC/5-caC in the CRE UNO and GC box promoters, it was demonstrated that APE1-mediated strand incision is essential to induce transcriptional repression (cf. chapter 7.1.2, 7.1.15 and Figure 8-1). It is thus assumed, that 5-fC and 5-caC in the CRE-UNO and GC box promoter repressed the gene expression by an indirect, repair dependent mechanism. This mechanism requires removal of the 5-fC and 5-caC base by TDG (cf. discussion above) and DNA-incision at the generated AP

site by APE1. An analogous silencing mechanism is assumed to cause the transcriptional repression observed for 5-fC and 5-caC in the strong CMV-1111 promoter, since the repression progressed in a similar fashion and was equally BER-dependent (cf. chapter 7.1.3).

Interestingly, the negative impact of a BER-resistant AP site on GC box promoter activity was much less pronounced than the impact of BER-resistant 5-caC (compare F-caC with SF in Figure 8-1, for details cf. chapters 7.1.11 and 7.1.15). At the same time, BER-sensitive 5-caC initially induced the same amount of gene expression as a BER-resistant AP site (compare 5-caC chapter 7.1.11 with SF chapter 7.1.15). These findings demonstrate that BER of 5-caC initially reactivates the GC box promoter and that this reactivation was probably caused by TDG-initiated base removal resulting in the formation of an apyrimidinic site at the position of 5-caC (illustrated in Figure 8-1), which per se has a less pronounced negative effect on the gene expression than 5-caC

8.2 8-oxoG negatively influences the GC box activity in a position dependent manner, by direct promoter inhibition and BER-mediated gene silencing

The second section of this work focused on dissecting the transcriptional consequences of the major guanine oxidation product 8-oxoG in the GC box. The common GC box gene regulatory element was selected as a representative URE to dissect the effects of 8-oxoG on gene expression in a GC-rich promoter. A minimal GC box promoter was used in order to reduce the number of variables of transcription regulation to a minimum. This approach enabled the assessment of modification specific effects on gene expression in the absence of any structure related consequences, which for example had been observed in complex PQS containing promoters (cf. chapter 4.2.2). The functional consequences of 8-oxoG at three selected positions in the purine-rich strand (R) (positions R-3, R-2 and R+1) and one position in the pyrimidine-rich strand of the GC box (position Y-1) were analysed to determine potential position biases within the GC-rich, PQS-free GC box gene regulatory element, and the impact of BER on transcription regulation by 8-oxoG was investigated.

Quantitative EGFP expression analysis in BER deficient HeLa cells showed that 8-oxoG per se only directly inhibited the GC box activity in the purine-rich (positions R-3, R-2 and R+1) but not in the pyrimidine-rich DNA strand(position Y-1), indicating a

strand biased transcription regulation mechanism (cf. chapter 7.2.6 and Figure 8-2). Further comparison of the expression levels revealed that GC box promoter inhibition by 8-oxoG not only differed between the different DNA strands but also between the different positions within the same DNA strand (GC box inhibition by 8-oxoG: R+1 = R-3 > R-2), verifying that promoter inhibition by the primary base modification was position-dependent. This is in line with previous findings showing that 8-oxoG at positions R+1, R-3 and R-4 of a selected GC box within the PQS containing VEGF promoter deprived of secondary structure formation reduced the gene expression in a position dependent manner in BER deficient cells[7,140]. The results of this study thus verify that even within the core sequence of the GC box, variation of the 8-oxoG-position by only one nucleotide can significantly change the base modification's effects on the gene expression.

Repression of the gene expression by 8-oxoG in the purine-rich strand of the GC box in BER deficient cells was already detectable 6 hours after transfection (cf. chapter 7.2.6), indicating that 8-oxoG directly diminished the promoter activity. A comparable direct promoter inhibition by 8-oxoG had been demonstrated for the minimal CRE-UNO promoter, were the negative effect of 8-oxoG was attributed to abolished TF binding[97]. Due to analogous imminent establishment and long-term maintenance of promoter inhibition by 8-oxoG in minimal CRE promoters and the purine-rich strand of the minimal GC box promoter, it is likely that abolished TF binding caused promoter inhibition in both cases. This hypothesis is supported by previous findings showing that SP1 binding to the GC box is abolished by guanine oxidation at all positions investigated in this project[127,128]. Although 8-oxoG at position Y-1 equally abolished SP1 transcription factor binding under cell free conditions[127,128], the presented in vivo data indicate that the primary base modification at this position is neutral for the reporter activity (cf. chapters 7.2.1, 7.2.2, 7.2.6, 7.2.7 and Figure 8-2). It is therefore assumed that in the cellular context, guanine oxidation in the pyrimidine-rich strand only minorly affects GC box-TF binding and accordingly the promoter activity, whereas guanine oxidation in the purine-rich strand reduces the GC box-TF binding to different degrees.

Strand-dependent effects of 8-oxoG on the GC box activity may be explained by position-dependent interactions of DNA bases in the GC box consensus sequence with TF binding. It was for example demonstrated that during GC box-SP1 binding all guanine bases in the purine-rich strand of the GC box consensus sequence interact

with SP1 whilst the guanine base at position Y-1 in the pyrimidine-rich strand was excluded from direct SP1 interactions[312]. It is therefore likely, that in contrast to G, the electrostatically higher charged and sterically bigger 8-oxoG may diminish DNA-SP1 interactions at GC box positions where guanine bases interact with the protein (R-3, R-2 and R+1), whilst DNA-protein binding is unaffected by 8-oxoG at positions without SP1-DNA base interactions (Y-1). Furthermore, it was shown that SP1 is not be the only TF whose binding to the GC box is affected by 8-oxoG in vivo. In addition to SP1, several other TFs have been described to bind to GC box sequences, including but not limited to the SP family members SP3/SP4[314,315] and the basic transcription element binding protein 1[316]. In contrast to SP1, the influence of guanine oxidation on the binding of such TFs is largely unknown and may differently affect promoter activation. Occupying the same target sequence, it is likely that these proteins compete for GC box binding in vivo and that the effects of 8-oxoG at different positions in the GC box are cumulative consequences of the enhanced or abolished binding of multiple TFs.

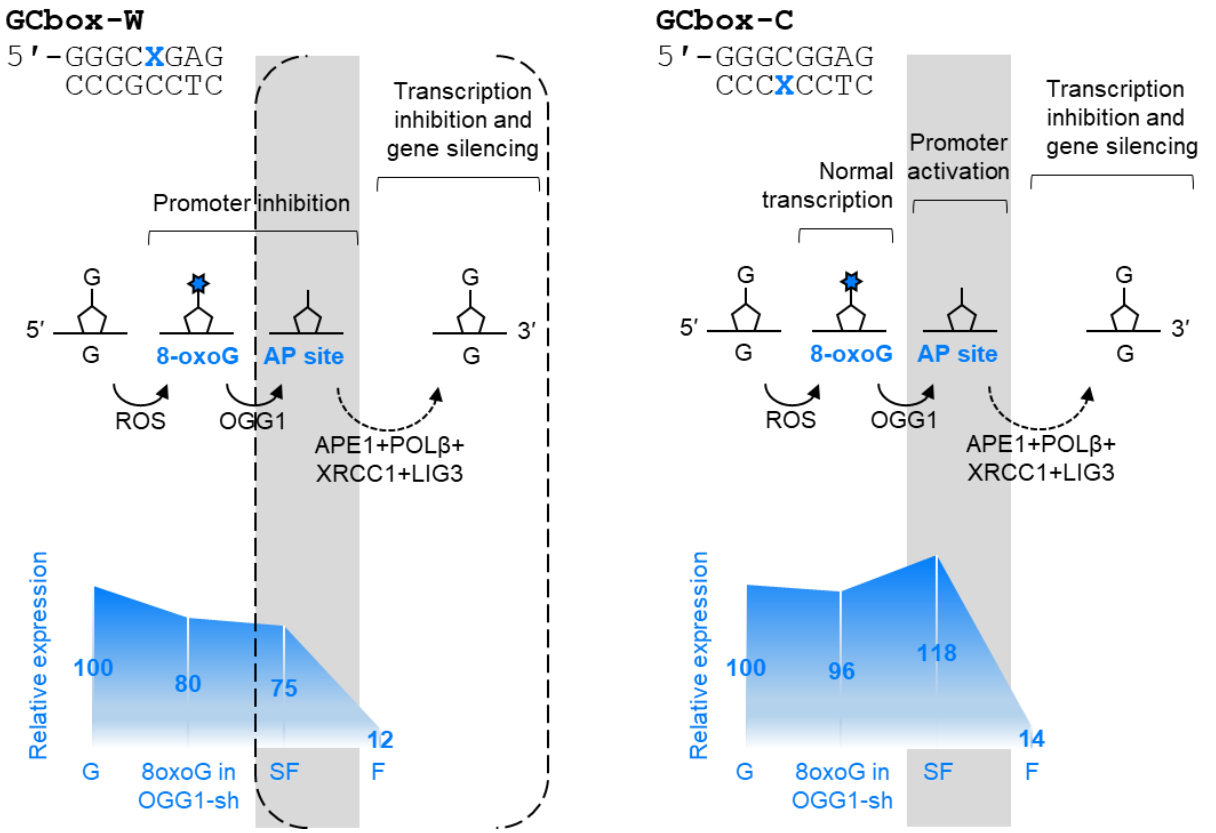


Figure 8-2: Overview of the transcriptional consequences of 8-oxoG in the GC box gene regulatory element

Transcriptional effects of the major guanine oxidation product 8-oxoG at the central CpG dinucleotide of the GC box gene regulatory element. (Upper panel) Model of OGG1-initiated BER of 8-oxoG via AP site intermediates, with effects on gene expression indicated above. Dotted arrow indicates reactions which have not been investigated in detail. (Middle and lower panel) Relative mean EGFP expression values of HeLa- or derived OGG1 knockdown cells (indicated as OGG1-sh) transfected with reporter constructs containing either a single 8-oxoG (chapter 7.2.2), BER-resistant or BER-sensitive AP sites (SF or F, chapter 7.2.5) or the dG control at the central CpG dinucleotide within the purine-rich strand (left panel) and pyrimidine-rich strand (right panel) of the GC box. Dotted brackets indicate reactions which might not take place in cells.

In addition to direct promoter inhibition by 8-oxoG, 8-oxoG at positions R-2 and Y-1 of the GC box caused a time-dependent repression of the gene expression in the presence of repair (cf. chapters 7.2.2 and 7.2.7). Transcriptional repression was reverted by OGG1-knockdown, indicating that 8-oxoG reduced the gene expression by a mechanism which was mediated by OGG1-initiated BER (cf. chapters 7.2.2 and 7.2.7). Analysing the functional consequences of the apurinic site repair intermediate at position Y-1 of the GC box verified that APE1-mediated DNA strand cleavage was essential for the induction of transcriptional repression (cf. chapter 6.2.5 and Figure 8-2). A similar BER-dependent transcriptional repression had been previously reported for 8-oxoG removal from the 5′UTR and EGFP gene body[136], indicating that transcriptional silencing by BER is a common consequence of 8-oxoG.

Interestingly, OGG1-dependent transcriptional repression was only observed for positions R-2 and Y-1 of the GC box (cf. chapters 7.2.2 and 7.2.7). To determine whether absence of OGG1-dependent transcriptional repression by 8-oxoG at

positions R-3 and R+1 of the GC box was caused by absent BER or by a different BER-outcome, the functional consequences of downstream repair processes were compared. It was shown that BER-sensitive apurinic sites at positions R+1 and Y-1 of the GC box were efficiently cleaved by APE1 in vitro (cf. chapter 6.2.4) and equally induced BER-dependent transcriptional repression in human cells (cf. chapter 6.2.5 and Figure 8-2). Considering that reporters containing 8-oxoG at position R+1 lacked transcriptional silencing in cells, whilst BER of apurinic sites at the same position drastically reduced the gene expression, it is anticipated that OGG1 did not excise 8-oxoG from position R+1 of the GC box (and by analogy also R-3) in the cellular context. Following this notion, cleavage assays with pure human OGG1 and human cell extracts showed a strong cleavage bias towards 8-oxoG at positions R-2 and Y-1 compared to 8-oxoG at positions R-3 and R+1 (cf. chapters 7.2.3 and 7.2.8). Correlation between preferred 8-oxoG removal from positions R-2 and Y-1 in the GC box and the OGG1-dependent transcriptional repression at these positions, indicates that OGG1 biases govern the magnitude of gene silencing by 8-oxoG in vivo, with presumably absent 8-oxoG removal at positions R-3 and R+1.

Based on the local nucleotide composition of the GC box, it is inferred that OGG1 prefers 8-oxoG removal from a 5′-XGG sequence context (position of the modified base underlined, with 8-oxoG at R-3: 5′-GGG and R+1: 5′-CGG) over base removal from 5′-XGC (8-oxoG at R-2: 5′-GGC and Y-1: 5′-CGC) (chapters 6.2.3 and 6.2.8). Sequence biases of OGG1 may be explained by the different influences of 8-oxoG flanking DNA bases on base removal. Thus, spatially small cytosine bases have less potential to interfere with the removal of the 8-oxoG base by OGG1 than the comparably bulky guanine bases. Since only DNA bases directly adjacent to 8-oxoG get into contact with the active centre of OGG1[317], these nucleobases are likely to have the greatest impact on OGG1 activity. A correlation between the excision efficiency of 8-oxoG by OGG1 and the local purine/pyrimidine composition has also been reported in the 5′UTR and gene coding region of EGFP[97,276], further supporting the hypothesis that the OGG1 activity is regulated by the local DNA sequence. Additionally, published evidence demonstrates that 8-oxoG residues is most frequently present in the middle of 5′-XGG trinucleotides[318] in yeast genomes, which based on the findings of this study can not only be attributed to higher oxidation sensitivity of the specific G as speculated previously[319–322], but also to differential 8-oxoG repair.

8.3 Biological consequence of BER substrates and putative mechanism of gene silencing

Comparing the functional consequences of 5-fC, 5-caC and 8-oxoG in the CRE and GC box gene regulatory element, it was shown that the base modifications influenced the gene expression by both direct and indirect mechanisms, which are assumed to be similar for 5-fC, 5-caC and 8-oxoG. Potential mechanisms of transcriptional regulation by the oxidatively induced base modifications are discussed below and are graphically summarised in Figure 8-3.

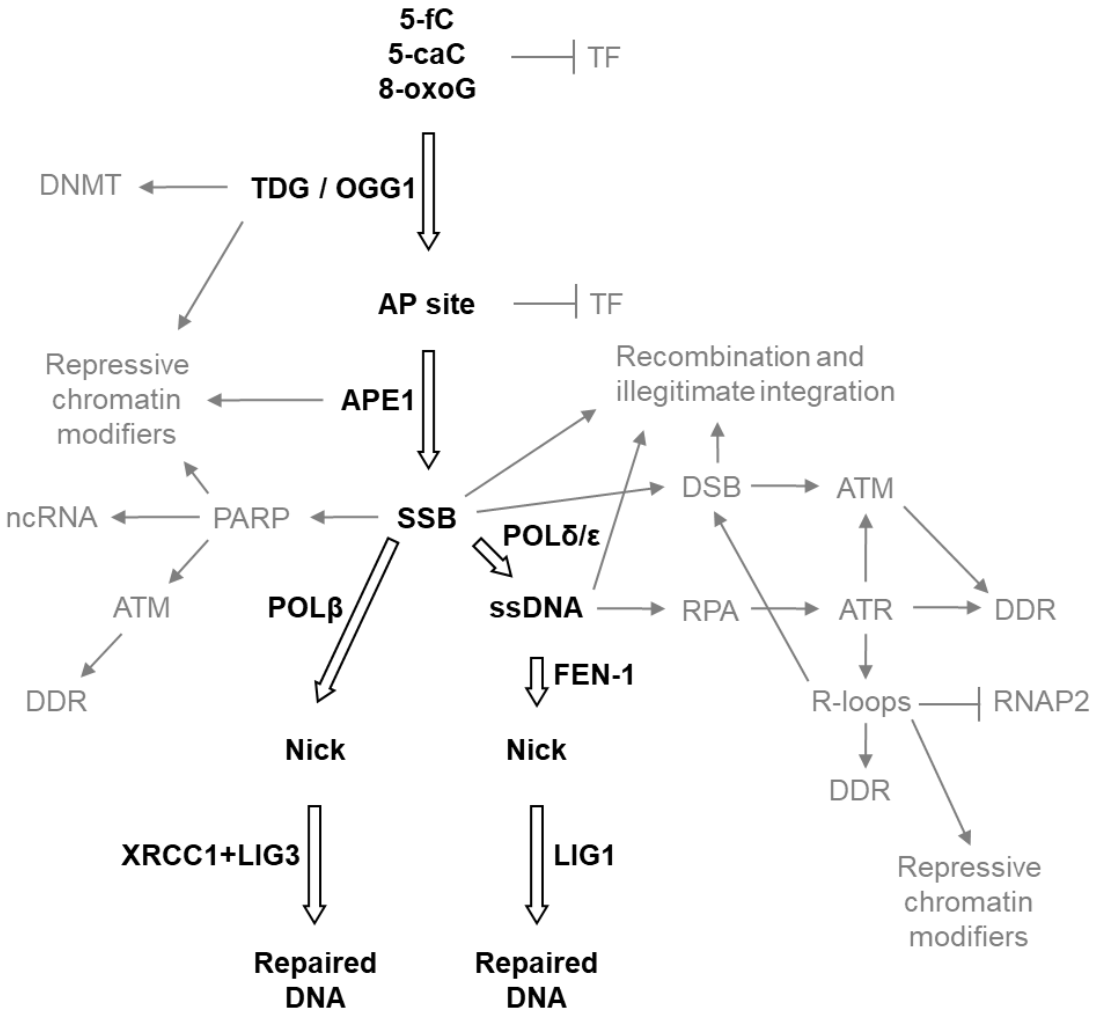


Figure 8-3: Overview of potential mechanisms of transcriptional regulation by the oxidatively induced DNA modifications 5-fC, 5-caC and 8-oxoG

Schematic presentation of the potential mechanisms of transcriptional regulation by the primary base modifications 5-fC, 5-caC and 8-oxoG and their processing by BER, showing the major participating enzymes and structural repair intermediates: proteins, repair intermediates and enzymatic reactions directly involved in BER (black), recruitment/activation (arrows) or repression (crossed line) of effectors mediating transcriptional regulation (grey).

The three primary base modifications 5-fC, 5-caC and 8-oxoG at the central CpG dinucleotide of the standalone CRE gene regulatory element of CRE-UNO reporters inhibited the promoter activity to ~15 % (chapter 7.1.1 and[97]). Stable CRE-UNO inhibition by 5-fC, 5-caC and 8-oxoG was attributed to the inhibition of CREB transcription factor binding, based on DNA-protein binding assays performed earlier[97]. Interestingly, the primary DNA modifications did not have the same effect on promoter activity when placed in a GC box gene regulatory element. In minimal GC box promoters the direct impact of 5-fC, 5-caC and 8-oxoG on promoter activity differed drastically, with 50% reduction by 5-caC whilst effects of 5-fC and 8-oxoG were much smaller (Figure 8-1 and Figure 8-2). The reason for the different expression outcomes of the three primary base modifications is likely to be found in their different effects on the binding of GC box specific transcription factors, as previously shown for CRE. One of the most prominent TFs targeting GC box sequences is SP1, which was shown to directly interact with different bases in the GC box sequence (cf. discussion chapter 8.2). It is hypothesised that modified nucleobases can inhibit the DNA-SP1 interactions by steric- and electrostatic repulsion, which explains the more drastic effects of 5-caC, having a bigger and more electrostatically charged chemical group than 5-fC and 8-oxoG. The position dependent consequences of 8-oxoG for GC box activity may equally be caused by different interactions between SP1 amino acids and nucleobases of the GC box sequence

SP1 binding modulation by the presence of modified nucleobases may also explain the GC box reactivation by TDG-mediated removal of the 5-caC base (cf. Figure 8-1 and chapters 7.1.11, 7.1.13 and 7.1.15). In contrast to 5-caC, the generated AP-site does not contain any sterically big or electrostatically charged groups, which might make the nucleoside almost neutral for SP1 binding. The slightly reduced GC box activity by BER-resistant AP sites observed in HeLa cells may result from absent hydrophilic SP1 interactions with the DNA base at positions R-1 and Y+1 (position of 5-fC and 5-caC, cf. discussion chapter 8.1).

Comparing the functional consequences of 8-oxoG and AP sites at position Y-1 of the GC box in the absence of repair, it was demonstrated that gene expression levels are 20% increased for AP site containing reporters (cf. Figure 8-2 right panel, and chapters 7.2.1 and 7.2.5). It is thus assumed that OGG1-mediated base removal of 8-oxoG positively affects the GC box activity, comparable to 5-caC in this promoter. Based on DNA-protein binding analysis it is hypothesised, that the positive effect of 8-oxoG base

removal on gene expression is caused by a tighter binding of SP1 to the AP site containing GC box sequence compared to 8-oxoG and G (cf. discussion chapter 8.2), thereby enhancing SP1 recruitment and promoter activity. Comparably, binding of other GC box specific TF may involve the guanine base at position Y-1, where 8-oxoG and apurinic sites can directly influence DNA-protein binding.

Effects of the primary base modifications 5-fC, 5-caC and 8-oxoG on the gene expression were far surpassed in their intensity by the consequences of BER. It was shown that BER of 5-fC and 5-caC in CRE and GC box promoters drastically reduced the gene expression (cf. chapters 7.1.7, 7.1.11, 7.1.13 and[97]). Transcriptional repression was also observed for BER of 8-oxoG in both UREs (cf. chapters 7.2.2, 7.2.7 and[97]). Similarities between BER-mediated transcriptional repression by 8-oxoG and 5-fC and 5-caC, suggest a common underlying mechanism, which involved APE1-mediated strand incision. Furthermore, investigations on the functional consequences of the structurally unrelated BER substrates U, 5-hmU and Tg in the minimal CRE promoter showed that the base modifications reduced the gene expression in a time- and BER-dependent manner (cf. chapters 7.3). Observed for six independent base modifications and in two different gene regulatory elements, it is probable that BER-dependent transcriptional repression is a common functional outcome of BER substrates in gene promoters. Interestingly, U and 8-oxoG at varying positions within the EGFP coding sequence[6,136,280], 5′-UTR and 3′-UTR of the EGFP gene[97] also induce transcriptional repression in an APE1-dependent manner. Combined with these findings, the results indicate that processing of nucleobase modifications by BER generally represses the gene expression, irrespectively of the nucleotide position, DNA strand, URE or gene region the altered nucleobase is positioned in (cf. chapters 7.1.2, 7.1.15 and 7.2.5)[97,136,280].

Repression of the gene expression by BER of 5-fC, 5-caC, 8-oxoG, 5-hmU, U and Tg can be induced by different mechanisms summarised in Figure 4-1. One possibility is that recruitment of BER proteins to the modified nucleobase sterically competes with recruitment of the transcription machinery. This hypothesis is supported by previously findings, showing that already OGG1 without additional BER machinery competes with CREB transcription factor for binding to an 8-oxoG containing CRE gene regulatory element[323]. Since experiments with CMV-1111 reporters showed that transcriptional

repression can be induced by BER of 5-fC and 5-caC 344 bp upstream from the TSS, inhibition of the transcription machinery would however need to act in great dissonance from TSS. BER proteins are assumed to leave the damage site after restoring the canonical DNA sequence, yet transcriptional repression by BER of 5-fC, 5-caC, 8-oxoG, 5-hmU, U and Tg persisted over 24-48 hours (cf. chapters 7.1.1, 7.1.3, 7.1.9, 7.1.10, 7.2.1 and 7.2.6). It is therefore deduced, that competition between BER and transcription initiation cannot be the only cause of the observed time-dependent reduction of the gene expression.

Another means of abolishing gene expression is the recruitment of repressor proteins which inhibit the binding of transcription activators to their local target sites. However, it is unlikely that the BER-mediated transcriptional silencing was induced by a sequence-specific repressor protein, since it was observed in various different sequence contexts and by different DNA modifications (cf. chapters 7.1, 7.2, 7.3)[97,136,280]. Furthermore, the negative effect of BER on gene expression was not locally constricted to the affected URE but spread over the whole promoter, as seen for BER of 5-fC and 5-caC in the CMV-1111 promoter (cf. chapter 7.1.3), indicating that an epigenetic silencing mechanism was involved in transcriptional repression. Involvement of an epigenetic silencing mechanism in transcription regulation is supported by previous experiments in our lab on the functional consequences of 8-oxoG in the pZA EGFP reporter (precursor of all EGFP reporters used in this work). It was shown that inhibition of histone deacetylases completely reverted the transcriptional repression by BER of 8-oxoG in the 5′-UTR and 3′-UTR of the pZA reporters[97,136]. These findings indicate that HDACs are involved in the establishment of transcriptional silencing by BER of 8-oxoG, furthering the notion of an underlying epigenetic silencing mechanism.

To induce transcriptional repression by BER, the silencing signal needs to be transmitted from the processed lesion to repressive chromatin-modifiers such as HDACs by e.g. protein-protein interactions. Such transmission cascade has been established for the transcriptional silencing by BER of 8-oxoG in eight tumour suppressors including CDK inhibitor 2A[324]. For these eight genes, ROS induced generation of 8-oxoG in the GC-rich promoter attracts OGG1, which indirectly recruits repressive chromatin modifiers including euchromatic histone lysine methyl-transferase 2 and DNMTs. DNMTs establish local *de novo* methylation at the gene promoters, whereas euchromatic histone lysine methyl-transferase 2 induces

repressive histone modifications H3K27me3 and H3K9me2, all mediating gene silencing[324]. TDG occupancy on the DNA was also reported to be linked to local modulation of histone acetylation patterns potentially affecting gene expression[325]. In the present work, induction of gene silencing by DNA glycosylases can neither be confirmed nor excluded, however such direct effect would need to be linked to DNA glycosylase activity, since BER-resistant 5-fC and 5-caC did not induce transcriptional repression (cf. chapters 7.1.1 and 7.1.11). Furthermore, strand incision at synthetic AP sites by APE1 also induced transcriptional repression, which was independent from DNA glycosylase activity (compare F in Figure 7-23 and Appendix I 5), indicating that the transcriptional repression was induced by a BER-intermediate process or downstream from base removal.

The step in BER after base removal is strand incision by APE1, a protein which has also been linked to chromatin modifiers. For example, APE1 has been shown to be directly involved in the downregulation of the parathyroid hormone (PTH) expression by binding to negative calcium responsive elements present in the PTH promoter[196]. More precisely, APE1 stably interacts with HDAC1, HDAC2 and HDAC3 forming a repressor complex. APE1 acetylation via the histone acetyltransferase p300 drastically increases the recruitment of those APE1-HDAC complexes to the PTH promoter, which abolishes the target gene expression by deacetylating histones in the promoter region[196]. Linking gene silencing to repressive chromatin modifiers via APE1 would explain the essentiality of DNA strand incision for transcriptional silencing by BER of 5-fC, 5-caC, 8-oxoG and U repair in various gene promoters (cf. chapters 7.1.2, 7.1.15 and 7.2.5)[5,135,136,276,280]. It also explains the irrelevance of the BER substrate type for the establishment of the transcriptional silencing. However, recruitment of repressive chromatin proteins by APE1 would need to be linked to APE1 activity since transcriptional silencing was absent for BER-resistant AP sites (cf. chapters 7.1.2, 7.1.15 and 7.2.5).

Alternatively, transcriptional repression by BER of 5-fC, 5-caC, 8-oxoG, 5-hmU, U and Tg may be induced by a silencing mechanism downstream from APE1 mediated strand incision. Such mechanism could for example be initiated by the presence of SSBs, the product of strand incision by APE1. SSBs activate poly ADP-ribose polymerases (PARPs), which add poly ADP-ribose (PAR) moieties to protein substrates (PARylation), thereby regulating transcription. Protein located PAR-chains are strong signalling molecules initiating the recruitment of repressive chromatin complexes and

chromatin modifiers[326]. PARylation also causes makro H2A1.1 formation, which triggers the DNA Damage Response (DDR) via the protein kinase ataxia-telangiectasia mutated (ATM)[326]. ATM functions as a signal amplifier, inducing H2Ak119 ubiquitination and chromatin condensation, as well as phosphorylation of H2AX and other substrates including Chk2 and p53. Substrate phosphorylation initiates ATM signalling cascades, which upregulate the expression of DDR-responsive genes to enhance repair and abolish transcription[327]. Furthermore, both ATM signalling and PARylation can activate the expression of small non-coding RNAs[326], which can guiding RNA-induced silencing complexes to target mRNA transcripts, degrade the and thereby abolish gene expression.

To repair SSBs in the DNA, short- or long patch BER can be applied, as described in chapter 4.3. Processing of an SSB by long patch BER generates a stretch of single stranded DNA (ssDNA) as repair intermediate. This ssDNA is recognised and bound by replication protein A (RPA), providing another platform for transcription regulation by various different processes. RPA coated ssDNA can activated DDR via recruitment of ATM- and Rad3-Related (ATR) protein, which phosphorylates a variety of protein substrates inducing ATR-signalling cascades[327]. ATR can also activate ATM and facilitate the formation of R-loops[327,328]. R-loops are DNA-RNA hybrids formed by invasion of a stretch of ssRNA into their dsDNA template under the displacement of a ssDNA stretch. Formation of R-loops is facilitated at the position of SSBs as well as ssDNA due to facilitated strand invasion. At gene promoters, R-loops can be formed by invasion of antisense transcripts[329], which are frequently generated opposite to transcribed genes in human cells. Already by their mere size of 100-2000 bp can R-loops abolish the recruitment and binding of TFs, RNAP2 and other proteins that are essential for transcription initiation[328]. Furthermore, R-loops can recruit repressive chromatin modifiers and initiate DDR[328]. Nuclease activity at R-loops can result in strand cleavage at the ssDNA excluded from the DNA-RNA loop[328]. If strand cleavage occurs opposite or in close proximity to the SSB generated by BER it can cause the formation of a DNA double strand break (DSB), activating DDR via ATM. DSBs may also be generated by topoisomerase 1 induced strand incision opposite to the BER-generated SSB.

In addition to the already mentioned mechanisms of transcriptional repression, SSBs as well as ssDNA and DSBs can enhance the probability of recombination- or illegitimate integration events of plasmid DNA into the HeLa genome[330]. Integration of

reporter DNA stretches into the HeLa genome can disrupt the promoter or coding region of the EGFP gene encoded on the plasmid, thereby abolishing reporter expression. Furthermore, significant DNA integration into the genomic context is expected to largely disrupt normal chromatin structure and function[330], thereby silencing the expression of any integrated DNA.

Further investigations are needed to determine which transcription regulation mechanism caused the transcription inhibition by BER of 5-fC, 5-caC and 8-oxoG in the CRE and GC box promoter and established gene silencing. Transcriptional silencing by APE1-mediated strand incision or downstream repair processes can also be advantageous under physiological circumstances, considering that BER complexes and the transcription machinery would otherwise compete for DNA binding at the damaged site. This competition could arrest both processes or worse reduce their fidelity, thereby increasing the probability of mutations in the genome. Inhibiting transcription initiation of damaged genes ensures that BER reconstitutes the correct DNA sequence before transcription resumes, thereby ensuring the highest possible fidelity of both processes.

In summary, my work provides important insights into interactions of the oxidatively induced base modifications 5-fC, 5-caC and 8-oxoG with promoter activation and the functional consequences of their repair. It was demonstrated, that 5-fC and 5-caC as primary base modifications negatively affected the gene expression in a URE-dependent manner, showing greatly exacerbated inhibition of the GC box activity by 5-caC compared to CRE. Fascinatingly, the presented results supply first evidence, that removal of the 5-caC base by TDG reactivates the expression of GC box reporters. These findings evidence the important role of TDG in transcription regulation, on the one hand reactivating transcription of GC box reporters by removing the 5-caC base, whilst on the other hand causing transcriptional silencing by progression of the TDG-initiated BER process. The presented experiments reveal a notable transcription regulation complexity by 8-oxoG even in the simplest GC box promoter. It was shown that shifting of the 8-oxoG position in the GC box sequence by only one nucleotide drastically altered the negative impact of the primary base modification on promoter activity. The study further revealed that OGG1 essentially influenced the effects of 8-oxoG on the gene expression. Thus, transcriptional repression by BER of 8-oxoG was only induced by 8-oxoG at positions with preferential OGG1 activity, indicating that

OGG1 governs the transcriptional consequences of 8-oxoG in human cells. BER-dependent transcriptional silencing was not only observed for 5-fC, 5-caC and 8-oxoG in gene promoters but also for the three independent DNA base modifications 5-hmU, U and Tg, indicating a common transcription regulation mechanism.

9 References

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715; 10.1038/362709a0 (1993).
2. Bauer, N. C., Corbett, A. H. & Doetsch, P. W. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Res* **43,** 10083–10101; 10.1093/nar/gkv1136 (2015).
3. Johnson, T. B. & Coghill, R. D. RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS 1. *J. Am. Chem. Soc.* **47,** 2838–2844; 10.1021/ja01688a030 (1925).
4. W. G. Ruppel. Zur Chemie der Tuberrelbacillen. *Zeitschrift für Physiol. Chem.,* 218–232 (1898).
5. Kitsera, N. *et al.* Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. *Nucleic Acids Res* **45,** 11033–11042; 10.1093/nar/gkx718 (2017).
6. Kitsera, N. *et al.* 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res* **39,** 5926–5934; 10.1093/nar/gkr163 (2011).
7. 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* **114,** 2604–2609; 10.1073/pnas.1619809114 (2017).
8. Bjelland, S. & Seeberg, E. Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutation research* **531,** 37–80; 10.1016/j.mrfmmm.2003.07.002 (2003).
9. Klungland, A. *et al.* Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 13300–13305; 10.1073/pnas.96.23.13300 (1999).
10. Cheadle, J. P. & Sampson, J. R. MUTYH-associated polyposis--from defect in base excision repair to clinical genetic testing. *DNA Repair* **6,** 274–279; 10.1016/j.dnarep.2006.11.001 (2007).
11. Kraemer, K. H., Lee, M. M. & Scotto, J. Xeroderma pigmentosum. Cutaneous, ocular, and neurologic abnormalities in 830 published cases. *Archives of dermatology* **123,** 241–250; 10.1001/archderm.123.2.241 (1987).
12. Andrews, A. D., Barrett, S. F. & Robbins, J. H. Xeroderma pigmentosum neurological abnormalities correlate with colony-forming ability after ultraviolet radiation. *Proc Natl Acad Sci USA* **75,** 1984–1988; 10.1073/pnas.75.4.1984 (1978).
13. Mallery, D. L. *et al.* Molecular analysis of mutations in the CSB (ERCC6) gene in patients with Cockayne syndrome. *American journal of human genetics* **62,** 77–85; 10.1086/301686 (1998).
14. Xie, Y. *et al.* Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer research* **64,** 3096–3102; 10.1158/0008-5472.can-03-3834 (2004).

15. Doğru-Abbasoğlu, S. *et al.* The Arg194Trp polymorphism in DNA repair gene XRCC1 and the risk for sporadic late-onset Alzheimer's disease. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* **28,** 31–34; 10.1007/s10072-007-0744-x (2007).

16. Kavli, B. *et al.* B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *The Journal of Experimental Medicine* **201,** 2011–2021; 10.1084/jem.20050042 (2005).

17. Boorstein, R. J., Hilbert, T. P., Cunningham, R. P. & Teebor, G. W. Formation and stability of repairable pyrimidine photohydrates in DNA. *Biochemistry* **29,** 10455–10460; 10.1021/bi00498a004 (1990).

18. Tremblay, S. & Wagner, J. R. Dehydration, deamination and enzymatic repair of cytosine glycols from oxidized poly(dG-dC) and poly(dI-dC). *Nucleic Acids Res* **36,** 284–293; 10.1093/nar/gkm1013 (2008).

19. Jena, N. R. & Mishra, P. C. Formation of ring-opened and rearranged products of guanine: mechanisms and biological significance. *Free radical biology & medicine* **53,** 81–94; 10.1016/j.freeradbiomed.2012.04.008 (2012).

20. Dizdaroglu, M. & Jaruga, P. Mechanisms of free radical-induced damage to DNA. *Free radical research* **46,** 382–419; 10.3109/10715762.2011.653969 (2012).

21. Steenken, S. & Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* **119,** 617–618; 10.1021/ja962255b (1997).

22. Burrows, C. J. & Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission. *Chemical reviews* **98,** 1109–1152; 10.1021/cr960421s (1998).

23. Barnes, D. E. & Lindahl, T. Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annu. Rev. Genet.* **38,** 445–476; 10.1146/annurev.genet.38.072902.092448 (2004).

24. Halliwell, B. Free Radicals and Other Reactive Species in Disease. In *Encyclopedia of life sciences,* edited by J. W. &. S. Ltd (Wiley, Chichester, England, 2005), pp. 1–9.

25. Vallyathan, V. & Shi, X. The role of oxygen free radicals in occupational and environmental lung diseases. *Environmental health perspectives* **105 Suppl 1,** 165–177; 10.1289/ehp.97105s1165 (1997).

26. Klaunig, J. E., Kamendulis, L. M. & Hocevar, B. A. Oxidative stress and oxidative damage in carcinogenesis. *Toxicologic pathology* **38,** 96–109; 10.1177/0192623309356453 (2010).

27. Cadet, J., Mouret, S., Ravanat, J.-L. & Douki, T. Photoinduced damage to cellular DNA: direct and photosensitized reactions. *Photochemistry and photobiology* **88,** 1048–1065; 10.1111/j.1751-1097.2012.01200.x (2012).

28. Ames, B. N., Shigenaga, M. K. & Hagen, T. M. Oxidants, antioxidants, and the degenerative diseases of aging. *Proc Natl Acad Sci USA* **90,** 7915–7922; 10.1073/pnas.90.17.7915 (1993).

29. Albright, C. D., Salganik, R. I., Craciunescu, C. N., Mar, M.-H. & Zeisel, S. H. Mitochondrial and microsomal derived reactive oxygen species mediate apoptosis induced by transforming growth factor-beta1 in immortalized rat hepatocytes. *Journal of cellular biochemistry* **89,** 254–261; 10.1002/jcb.10498 (2003).

30. Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell research* **18,** 27–47; 10.1038/cr.2008.8 (2008).

31. Schieber, M. & Chandel, N. S. ROS function in redox signaling and oxidative stress. *Current biology : CB* **24,** R453-62; 10.1016/j.cub.2014.03.034 (2014).

32. Gedik, C. M. & Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **19,** 82–84; 10.1096/fj.04-1767fje (2005).

33. Neeley, W. L. & Essigmann, J. M. Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chemical research in toxicology* **19,** 491–505; 10.1021/tx0600043 (2006).

34. Zuo, S., Boorstein, R. J. & Teebor, G. W. Oxidative damage to 5-methylcytosine in DNA. *Nucleic Acids Res* **23,** 3239–3243; 10.1093/nar/23.16.3239 (1995).

35. Valinluck, V. *et al.* 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* **32,** 4100–4108; 10.1093/nar/gkh739 (2004).

36. Bienvenu, C., Wagner, J. R. & Cadet, J. Photosensitized Oxidation of 5-Methyl-2‘-deoxycytidine by 2-Methyl-1,4-naphthoquinone: Characterization of 5-(Hydroperoxymethyl)-2‘-deoxycytidine and Stable Methyl Group Oxidation Products §. *J. Am. Chem. Soc.* **118,** 11406–11411; 10.1021/ja962073h (1996).

37. Privat, E. & Sowers, L. C. Photochemical deamination and demethylation of 5-methylcytosine. *Chemical research in toxicology* **9,** 745–750; 10.1021/tx950182o (1996).

38. Wagner, J. R. & Cadet, J. Oxidation reactions of cytosine DNA components by hydroxyl radical and one-electron oxidants in aerated aqueous solutions. *Accounts of chemical research* **43,** 564–571; 10.1021/ar9002637 (2010).

39. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5,** e15367; 10.1371/journal.pone.0015367 (2010).

40. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10,** 2709–2721; 10.1093/nar/10.8.2709 (1982).

41. Bird, A. DNA methylation patterns and epigenetic memory. *Genes & Development* **16,** 6–21; 10.1101/gad.947102 (2002).

42. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes & Development* **23,** 781–783; 10.1101/gad.1787609 (2009).

43. Jeltsch, A. Molecular enzymology of mammalian DNA methyltransferases. *Current topics in microbiology and immunology* **301,** 203–225; 10.1007/3-540-31390-7_7 (2006).

44. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99,** 247–257; 10.1016/s0092-8674(00)81656-6 (1999).

45. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* **279,** 48350–48359; 10.1074/jbc.M403427200 (2004).

46. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187,** 226–232 (1975).

47. Hsieh, C. L. In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Molecular and cellular biology* **19,** 8211–8218; 10.1128/mcb.19.12.8211 (1999).

48. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics* **19,** 219–220; 10.1038/890 (1998).

49. Qin, W. *et al.* DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell research* **25,** 911–929; 10.1038/cr.2015.72 (2015).

50. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908–912; 10.1038/nature06397 (2007).

51. Liu, X. *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* **4,** 1563; 10.1038/ncomms2562 (2013).

52. Nishiyama, A. *et al.* Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502,** 249–253; 10.1038/nature12488 (2013).

53. Ben-Hattar, J. & Jiricny, J. Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene* **65,** 219–227; 10.1016/0378-1119(88)90458-1 (1988).

54. Iguchi-Ariga, S. M. & Schaffner, W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & Development* **3,** 612–619; 10.1101/gad.3.5.612 (1989).

55. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40,** 91–99; 10.1016/0092-8674(85)90312-5 (1985).

56. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes & Development* **25,** 1010–1022; 10.1101/gad.2037511 (2011).

57. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481; 10.1038/nature12433 (2013).

58. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523,** 212–216; 10.1038/nature14465 (2015).

59. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303; 10.1126/science.1210597 (2011).

60. Cadet, J. & Wagner, J. R. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **764-765,** 18–35; 10.1016/j.mrgentox.2013.09.001 (2014).

61. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153,** 692–706; 10.1016/j.cell.2013.04.002 (2013).

62. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935; 10.1126/science.1170116 (2009).

63. Pfaffeneder, T. *et al.* The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem. Int. Ed.* **50,** 7008–7012; 10.1002/anie.201103899 (2011).

64. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nature chemical biology* **11,** 555–557; 10.1038/nchembio.1848 (2015).

65. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angewandte Chemie (International ed. in English)* **49,** 5375–5377; 10.1002/anie.201002033 (2010).

66. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348; 10.1038/nature10066 (2011).

67. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular Cell* **42,** 451–464; 10.1016/j.molcel.2011.04.005 (2011).

68. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome biology* **12,** R54; 10.1186/gb-2011-12-6-r54 (2011).

69. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149,** 1368–1380; 10.1016/j.cell.2012.04.027 (2012).

70. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8,** 200–213; 10.1016/j.stem.2011.01.008 (2011).

71. Perillo, B. *et al.* DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* **319,** 202–206; 10.1126/science.1147674 (2008).

72. Cortázar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470,** 419–423; 10.1038/nature09672 (2011).

73. Imai, K. *et al.* Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nature immunology* **4,** 1023–1028; 10.1038/ni974 (2003).

74. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14,** 204–220; 10.1038/nrg3354 (2013).

75. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13,** 484–492; 10.1038/nrg3230 (2012).

76. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics* **14,** 9–25; 10.1159/000130315 (1975).

77. Baylin, S. B. & Jones, P. A. Epigenetic Determinants of Cancer. *Cold Spring Harbor Perspectives in Biology* **8**; 10.1101/cshperspect.a019505 (2016).

78. Fetahu, I. S. *et al.* Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease. *Science advances* **5,** eaaw2880; 10.1126/sciadv.aaw2880 (2019).

79. Blanch, M., Mosquera, J. L., Ansoleaga, B., Ferrer, I. & Barrachina, M. Altered Mitochondrial DNA Methylation Pattern in Alzheimer Disease-Related Pathology and in Parkinson Disease. *The American journal of pathology* **186,** 385–397; 10.1016/j.ajpath.2015.10.004 (2016).

80. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326; 10.1038/nature14192 (2015).

81. Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature genetics* **24,** 88–91; 10.1038/71750 (2000).

82. Geiman, T. M. *et al.* DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochemical and biophysical research communications* **318,** 544–555; 10.1016/j.bbrc.2004.04.058 (2004).

83. Ludwig, A. K., Zhang, P. & Cardoso, M. C. Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease. *Frontiers in Genetics* **7,** 115; 10.3389/fgene.2016.00115 (2016).

84. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389; 10.1038/30764 (1998).

85. Lunyak, V. V. *et al.* Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298,** 1747–1752; 10.1126/science.1076469 (2002).

86. Nikitina, T. *et al.* Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Molecular and cellular biology* **27,** 864–877; 10.1128/MCB.01593-06 (2007).

87. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**; 10.1126/science.aaj2239 (2017).

88. Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Current opinion in genetics & development* **3,** 226–231; 10.1016/0959-437x(93)90027-m (1993).

89. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159; 10.1016/j.cell.2013.02.004 (2013).

90. Tian, H.-P. *et al.* DNA Methylation Affects the SP1-regulated Transcription of FOXF2 in Breast Cancer Cells. *J. Biol. Chem.* **290,** 19173–19183; 10.1074/jbc.M114.636126 (2015).

91. Michelotti, G. A. *et al.* Epigenetic regulation of human alpha1d-adrenergic receptor gene expression: a role for DNA methylation in Sp1-dependent regulation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **21,** 1979–1993; 10.1096/fj.06-7118com (2007).

92. Jiang, D. *et al.* Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma. *Cell death & disease* **10,** 1; 10.1038/s41419-018-1236-z (2018).

93. Pang, R. T.-K., Lee, L. T.-O., Ng, S. S.-M., Yung, W.-H. & Chow, B. K.-C. CpG methylation and transcription factors Sp1 and Sp3 regulate the expression of the human secretin receptor gene. *Molecular endocrinology (Baltimore, Md.)* **18,** 471–483; 10.1210/me.2003-0245 (2004).

94. 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* **85,** 2066–2070; 10.1073/pnas.85.7.2066 (1988).

95. Hashimoto, H. *et al.* Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes & Development* **28,** 2304–2313; 10.1101/gad.250746.114 (2014).

96. Zandarashvili, L., White, M. A., Esadze, A. & Iwahara, J. Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. *FEBS letters* **589,** 1748–1753; 10.1016/j.febslet.2015.05.022 (2015).

97. Julia Allgayer. Verschiedene Modi der Transkriptionshemmung durch Verschiedene Modi der Transkriptionshemmung durch 8-Oxo-7,8-dihydroguanin (8-oxoG) in der DNA. Dissertation. Johannes Gutenberg-University, 2016.

98. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature chemistry* **6,** 1049–1055; 10.1038/nchem.2064 (2014).

99. Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angewandte Chemie (International ed. in English)* **54,** 12511–12514; 10.1002/anie.201502722 (2015).

100. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–1430; 10.1016/j.cell.2012.11.022 (2012).

101. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem. Int. Ed.* **55,** 11797–11800; 10.1002/anie.201605994 (2016).

102. Sato, K., Kawamoto, K., Shimamura, S., Ichikawa, S. & Matsuda, A. An oligodeoxyribonucleotide containing 5-formyl-2'-deoxycytidine (fC) at the CpG site forms a covalent complex with DNA cytosine-5 methyltransferases (DNMTs). *Bioorganic & medicinal chemistry letters* **26,** 5395–5398; 10.1016/j.bmcl.2016.10.040 (2016).

103. Raiber, E.-A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nature chemistry* **10,** 1258–1266; 10.1038/s41557-018-0149-x (2018).

104. Wang, D. *et al.* MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res* **45,** 2396–2407; 10.1093/nar/gkw1184 (2017).

105. Impey, S. *et al.* Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119,** 1041–1054; 10.1016/j.cell.2004.10.032 (2004).

106. Zhang, X. *et al.* Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc Natl Acad Sci USA* **102,** 4459–4464; 10.1073/pnas.0501076102 (2005).

107. Lonze, B. E. & Ginty, D. D. Function and Regulation of CREB Family Transcription Factors in the Nervous System. *Neuron* **35,** 605–623; 10.1016/s0896-6273(02)00828-0 (2002).

108. Sakamoto, K., Karelina, K. & Obrietan, K. CREB: a multifaceted regulator of neuronal plasticity and protection. *Journal of neurochemistry* **116,** 1–9; 10.1111/j.1471-4159.2010.07080.x (2011).

109. Giampà, C. *et al.* Striatal modulation of cAMP-response-element-binding protein (CREB) after excitotoxic lesions: implications with neuronal vulnerability in Huntington's disease. *The European journal of neuroscience* **23,** 11–20; 10.1111/j.1460-9568.2005.04545.x (2006).

110. Shibutani, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349,** 431–434; 10.1038/349431a0 (1991).

111. Hsu, G. W., Ober, M., Carell, T. & Beese, L. S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **431,** 217–221; 10.1038/nature02908 (2004).

112. Maga, G. *et al.* 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature* **447,** 606–608; 10.1038/nature05843 (2007).

113. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C--T:A mutations in colorectal tumors. *Nature genetics* **30,** 227–232; 10.1038/ng828 (2002).

114. Sakumi, K. *et al.* Ogg1 knockout-associated lung tumorigenesis and its suppression by Mth1 gene disruption. *Cancer research* **63,** 902–905 (2003).

115. Arai, T. *et al.* Cell proliferation in liver of Mmh/Ogg1-deficient mice enhances mutation frequency because of the presence of 8-hydroxyguanine in DNA. *Cancer research* **63,** 4287–4292 (2003).

116. Lonkar, P. & Dedon, P. C. Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates. *International journal of cancer* **128,** 1999–2009; 10.1002/ijc.25815 (2011).

117. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421; 10.1038/nature12477 (2013).

118. Isogawa, A. Functional cooperation of Ogg1 and Mutyh in preventing G: C--T: a transversions in mice. *Fukuoka igaku zasshi = Hukuoka acta medica* **95,** 17–30 (2004).

119. Pascucci, B. *et al.* Reconstitution of the base excision repair pathway for 7,8-dihydro-8-oxoguanine with purified human proteins. *Nucleic Acids Res* **30,** 2124–2130; 10.1093/nar/30.10.2124 (2002).

120. Krokan, H. E. & Bjoras, M. Base Excision Repair. *Cold Spring Harbor Perspectives in Biology* **5,** a012583-a012583; 10.1101/cshperspect.a012583 (2013).

121. Krokan, H. E., Standal, R. & Slupphaug, G. DNA glycosylases in the base excision repair of DNA. *The Biochemical journal* **325 (Pt 1),** 1–16; 10.1042/bj3250001 (1997).

122. Ohno, M. *et al.* A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. *Genome research* **16,** 567–575; 10.1101/gr.4769606 (2006).

123. Ding, Y., Fleming, A. M. & Burrows, C. J. Sequencing the Mouse Genome for the Oxidatively Modified Base 8-Oxo-7,8-dihydroguanine by OG-Seq. *Journal of the American Chemical Society* **139,** 2569–2572; 10.1021/jacs.6b12604 (2017).

124. Poetsch, A. R., Boulton, S. J. & Luscombe, N. M. Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome biology* **19,** 215; 10.1186/s13059-018-1582-2 (2018).

125. Mangerich, A. *et al.* Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc Natl Acad Sci USA* **109,** E1820-9; 10.1073/pnas.1207829109 (2012).

126. Pastukh, V. *et al.* An oxidative DNA "damage" and repair mechanism localized in the VEGF promoter is important for hypoxia-induced VEGF mRNA expression. *American journal of physiology. Lung cellular and molecular physiology* **309,** L1367-75; 10.1152/ajplung.00236.2015 (2015).

127. Ramon, O. *et al.* Effects of 8-oxo-7,8-dihydro-2'-deoxyguanosine on the binding of the transcription factor Sp1 to its cognate target DNA sequence (GC box). *Free radical research* **31,** 217–229; 10.1080/10715769900300781 (1999).

128. Ghosh, R. & Mitchell, D. L. Effect of oxidative DNA damage in promoter elements on transcription factor binding. *Nucleic Acids Res* **27,** 3213–3218; 10.1093/nar/27.15.3213 (1999).

129. Hailer-Morrison, M. K., Kotler, J. M., Martin, B. D. & Sugden, K. D. Oxidized guanine lesions as modulators of gene transcription. Altered p50 binding affinity and repair shielding by 7,8-dihydro-8-oxo-2'-deoxyguanosine lesions in the NF-kappaB promoter element. *Biochemistry* **42,** 9761–9770; 10.1021/bi034546k (2003).

130. Blake, M. C., Jambou, R. C., Swick, A. G., Kahn, J. W. & Azizkhan, J. C. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Molecular and cellular biology* **10,** 6632–6641; 10.1128/mcb.10.12.6632 (1990).

131. Kim, H. J. *et al.* Transcription factor Sp1 regulates basal transcription of the human DRG2 gene. *Biochimica et biophysica acta* **1809,** 184–190; 10.1016/j.bbagrm.2011.01.004 (2011).

132. Ross, S., Tienhaara, A., Lee, M.-S., Tsai, L.-H. & Gill, G. GC box-binding transcription factors control the neuronal specific transcription of the cyclin-dependent kinase 5 regulator p35. *J. Biol. Chem.* **277,** 4455–4464; 10.1074/jbc.M110771200 (2002).

133. Ishii, S. *et al.* Binding of the Sp1 transcription factor by the human Harvey ras1 proto-oncogene promoter. *Science* **232,** 1410–1413; 10.1126/science.3012774 (1986).

134. Ishii, S., Merlino, G. T. & Pastan, I. Promoter region of the human Harvey ras proto-oncogene: similarity to the EGF receptor proto-oncogene promoter. *Science* **230,** 1378–1381; 10.1126/science.2999983 (1985).

135. Khobta, A., Kitsera, N., Speckmann, B. & Epe, B. 8-Oxoguanine DNA glycosylase (Ogg1) causes a transcriptional inactivation of damaged DNA in the absence of functional Cockayne syndrome B (Csb) protein. *DNA Repair* **8,** 309–317; 10.1016/j.dnarep.2008.11.006 (2009).

136. 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* **44,** 7267-7280; 10.1093/nar/gkw473 (2016).

137. Antoniali, G. *et al.* SIRT1 gene expression upon genotoxic damage is regulated by APE1 through nCaRE-promoter elements. *Molecular biology of the cell* **25,** 532–547; 10.1091/mbc.E13-05-0286 (2014).

138. Cogoi, S., Ferino, A., Miglietta, G., Pedersen, E. B. & Xodo, L. E. The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription. *Nucleic Acids Res* **46,** 661–676; 10.1093/nar/gkx1142 (2018).

139. Ba, X. & Boldogh, I. 8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions. *Redox biology* **14,** 669–678; 10.1016/j.redox.2017.11.008 (2018).

140. Fleming, A. M., Zhu, J., Ding, Y. & Burrows, C. J. 8-Oxo-7,8-dihydroguanine in the Context of a Gene Promoter G-Quadruplex Is an On-Off Switch for Transcription. *ACS Chemical Biology* **12,** 2417–2426; 10.1021/acschembio.7b00636 (2017).

141. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43,** 8627–8637; 10.1093/nar/gkv862 (2015).

142. Lindahl, T. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad Sci USA* **71,** 3649–3653; 10.1073/pnas.71.9.3649 (1974).

143. Carter, R. J. & Parsons, J. L. Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Molecular and cellular biology* **36,** 1426–1437; 10.1128/MCB.00030-16 (2016).

144. Lindahl, T. & Wood, R. D. Quality control by DNA repair. *Science* **286,** 1897–1905; 10.1126/science.286.5446.1897 (1999).

145. Jacobs, A. L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121,** 1–20; 10.1007/s00412-011-0347-4 (2012).

146. Hans, F., Senarisoy, M., Bhaskar Naidu, C. & Timmins, J. Focus on DNA Glycosylases-A Set of Tightly Regulated Enzymes with a High Potential as Anticancer Drug Targets. *International journal of molecular sciences* **21**; 10.3390/ijms21239226 (2020).

147. Hilbert, T. P., Chaung, W., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. Cloning and Expression of the cDNA Encoding the Human Homologue of the DNA Repair Enzyme, Escherichia coli Endonuclease III. *J. Biol. Chem.* **272,** 6733–6740; 10.1074/jbc.272.10.6733 (1997).

148. Wilson, D. M. & Barsky, D. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research/DNA Repair* **485,** 283–307; 10.1016/S0921-8777(01)00063-5 (2001).

149. Srivastava, D. K. *et al.* Mammalian abasic site base excision repair. Identification of the reaction sequence and rate-determining steps. *J. Biol. Chem.* **273,** 21203–21209; 10.1074/jbc.273.33.21203 (1998).

150. Beard, W. A. & Wilson, S. H. Structural design of a eukaryotic DNA repair polymerase: DNA polymerase β. *Mutation Research/DNA Repair* **460,** 231–244; 10.1016/s0921-8777(00)00029-x (2000).

151. Beard, W. A., Prasad, R. & Wilson, S. H. Activities and Mechanism of DNA Polymerase β. *Methods in enzymology* **408,** 91–107; 10.1016/S0076-6879(06)08007-4 (2006).

152. Kim, Y.-J. & Wilson, D. M. Overview of base excision repair biochemistry. *Current molecular pharmacology* **5,** 3–13; 10.2174/1874467211205010003 (2012).

153. Fortini, P. *et al.* Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells. *Biochemistry* **37,** 3575–3580; 10.1021/bi972999h (1998).

154. Stucki, M. *et al.* Mammalian base excision repair by DNA polymerases delta and epsilon. *Oncogene* **17,** 835–843; 10.1038/sj.onc.1202001 (1998).

155. Klungland, A. & Lindahl, T. Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *The EMBO journal* **16,** 3341–3348; 10.1093/emboj/16.11.3341 (1997).

156. Tsutakawa, S. E. *et al.* Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **145,** 198–211; 10.1016/j.cell.2011.03.004 (2011).

157. Manvilla, B. A., Maiti, A., Begley, M. C., Toth, E. A. & Drohat, A. C. Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *Journal of Molecular Biology* **420,** 164–175; 10.1016/j.jmb.2012.04.028 (2012).

158. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401,** 301–304; 10.1038/45843 (1999).

159. Petronzelli, F. *et al.* Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J. Cell. Physiol.* **185,** 473–480; 10.1002/1097-4652(200012)185:3<473::AID-JCP19>3.0.CO;2-# (2000).

160. O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279,** 9750–9757; 10.1074/jbc.M312232200 (2004).

161. McGoldrick, J. P., Yeh, Y. C., Solomon, M., Essigmann, J. M. & Lu, A. L. Characterization of a mammalian homolog of the Escherichia coli MutY mismatch repair protein. *Molecular and cellular biology* **15,** 989–996; 10.1128/mcb.15.2.989 (1995).

162. Hazra, T. K. *et al.* Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci USA* **99,** 3523–3528; 10.1073/pnas.062053799 (2002).

163. Hazra, T. K. *et al.* Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.* **277,** 30417–30420; 10.1074/jbc.C200355200 (2002).

164. Liu, M. *et al.* The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proc Natl Acad Sci USA* **107,** 4925–4930; 10.1073/pnas.0908307107 (2010).

165. Marenstein, D. R. *et al.* Substrate Specificity of Human Endonuclease III (hNTH1). *J. Biol. Chem.* **278,** 9005–9012; 10.1074/jbc.M212168200 (2003).

166. Hilbert, T. P. *et al.* Purification of a mammalian homologue of Escherichia coli endonuclease III: identification of a bovine pyrimidine hydrate-thymine glycol DNAse/AP lyase by irreversible cross linking to a thymine glycol-containing oligoxynucleotide. *Biochemistry* **35,** 2505–2511; 10.1021/bi952516e (1996).

167. Radicella, J. P., Dherin, C., Desmaze, C., Fox, M. S. & Boiteux, S. Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **94,** 8010–8015; 10.1073/pnas.94.15.8010 (1997).

168. Kavli, B. *et al.* hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **277,** 39926–39936; 10.1074/jbc.M207107200 (2002).

169. Neddermann, P. & Jiricny, J. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.* **268,** 21218–21224 (1993).

170. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem.* **286,** 35334–35338; 10.1074/jbc.C111.284620 (2011).

171. Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: Protein mimicry of DNA. *Cell* **82,** 701–708; 10.1016/0092-8674(95)90467-0 (1995).

172. Rosenquist, T. A., Zharkov, D. O. & Grollman, A. P. Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase. *Proc Natl Acad Sci USA* **94,** 7429–7434; 10.1073/pnas.94.14.7429 (1997).

173. Dalhus, B. *et al.* Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase. *Structure (London, England : 1993)* **19,** 117–127; 10.1016/j.str.2010.09.023 (2011).

174. Hill, Jeff K. Hazra, Tapas, K, Izumi, T. & Mitra, S. Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *Nucleic Acids Res* **29,** 430–438; 10.1093/nar/29.2.430 (2001).

175. Williams, S. D. & David, S. S. Evidence that MutY is a monofunctional glycosylase capable of forming a covalent Schiff base intermediate with substrate DNA. *Nucleic Acids Res* **26,** 5123–5133; 10.1093/nar/26.22.5123 (1998).

176. Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends in genetics : Trends Genet.* **9,** 246–249; 10.1016/0168-9525(93)90089-z (1993).

177. Markkanen, E. Not breathing is not an option: How to deal with oxidative DNA damage. *DNA Repair* **59,** 82–105; 10.1016/j.dnarep.2017.09.007 (2017).

178. Bhakat, K. K., Mokkapati, S. K., Boldogh, I., Hazra, T. K. & Mitra, S. Acetylation of human 8-oxoguanine-DNA glycosylase by p300 and its role in 8-oxoguanine repair in vivo. *Molecular and cellular biology* **26,** 1654–1665; 10.1128/MCB.26.5.1654-1665.2006 (2006).

179. Kang, L., Zhao, W., Zhang, G., Wu, J. & Guan, H. Acetylated 8-oxoguanine DNA glycosylase 1 and its relationship with p300 and SIRT1 in lens epithelium cells from age-related cataract. *Experimental eye research* **135,** 102–108; 10.1016/j.exer.2015.02.005 (2015).

180. Hu, J., Imam, S. Z., Hashiguchi, K., Souza-Pinto, N. C. de & Bohr, V. A. Phosphorylation of human oxoguanine DNA glycosylase (alpha-OGG1) modulates its function. *Nucleic Acids Res* **33,** 3271–3282; 10.1093/nar/gki636 (2005).

181. Dantzer, F., Luna, L., Bjørås, M. & Seeberg, E. Human OGG1 undergoes serine phosphorylation and associates with the nuclear matrix and mitotic chromatin in vivo. *Nucleic Acids Res* **30,** 2349–2357; 10.1093/nar/30.11.2349 (2002).

182. Bravard, A. *et al.* Redox regulation of human OGG1 activity in response to cellular oxidative stress. *Molecular and cellular biology* **26,** 7430–7436; 10.1128/MCB.00624-06 (2006).

183. Cividini, F. *et al.* O-GlcNAcylation of 8-Oxoguanine DNA Glycosylase (Ogg1) Impairs Oxidative Mitochondrial DNA Lesion Repair in Diabetic Hearts. *J. Biol. Chem.* **291,** 26515–26528; 10.1074/jbc.M116.754481 (2016).

184. Noren Hooten, N., Kompaniez, K., Barnes, J., Lohani, A. & Evans, M. K. Poly(ADP-ribose) polymerase 1 (PARP-1) binds to 8-oxoguanine-DNA glycosylase (OGG1). *J. Biol. Chem.* **286,** 44679–44690; 10.1074/jbc.M111.255869 (2011).

185. Vidal, A. E., Hickson, I. D., Boiteux, S. & Radicella, J. P. Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step. *Nucleic Acids Res* **29,** 1285–1292; 10.1093/nar/29.6.1285 (2001).

186. Marsin, S. *et al.* Role of XRCC1 in the coordination and stimulation of oxidative DNA damage repair initiated by the DNA glycosylase hOGG1. *J. Biol. Chem.* **278,** 44068–44074; 10.1074/jbc.M306160200 (2003).

187. Campalans, A. *et al.* Interaction with OGG1 is required for efficient recruitment of XRCC1 to base excision repair and maintenance of genetic stability after exposure to oxidative stress. *Molecular and cellular biology* **35,** 1648–1658; 10.1128/MCB.00134-15 (2015).

188. Ramdzan, Z. M. *et al.* RAS transformation requires CUX1-dependent repair of oxidative DNA damage. *PLoS biology* **12,** e1001807; 10.1371/journal.pbio.1001807 (2014).

189. Pal, R. *et al.* CUX2 protein functions as an accessory factor in the repair of oxidative DNA damage. *J. Biol. Chem.* **290,** 22520–22531; 10.1074/jbc.M115.651042 (2015).

190. Busso, C. S., Iwakuma, T. & Izumi, T. Ubiquitination of mammalian AP endonuclease (APE1) regulated by the p53-MDM2 signaling pathway. *Oncogene* **28,** 1616–1625; 10.1038/onc.2009.5 (2009).

191. Busso, C. S., Lake, M. W. & Izumi, T. Posttranslational modification of mammalian AP endonuclease (APE1). *Cellular and molecular life sciences : CMLS* **67,** 3609–3620; 10.1007/s00018-010-0487-3 (2010).

192. Yacoub, A., Kelley, M. R. & Deutsch, W. A. The DNA repair activity of human redox/repair protein APE/Ref-1 is inactivated by phosphorylation. *Cancer research* **57,** 5457–5459 (1997).

193. Huang, E. *et al.* The role of Cdk5-mediated apurinic/apyrimidinic endonuclease 1 phosphorylation in neuronal death. *Nature cell biology* **12,** 563–571; 10.1038/ncb2058 (2010).

194. Fritz, G. & Kaina, B. Phosphorylation of the DNA repair protein APE/REF-1 by CKII affects redox regulation of AP-1. *Oncogene* **18,** 1033–1040; 10.1038/sj.onc.1202394 (1999).

195. Hsieh, M. M., Hegde, V., Kelley, M. R. & Deutsch, W. A. Activation of APE/Ref-1 redox activity is mediated by reactive oxygen species and PKC phosphorylation. *Nucleic Acids Res* **29,** 3116–3122; 10.1093/nar/29.14.3116 (2001).

196. Bhakat, K. K., Izumi, T., Yang, S.-H., Hazra, T. K. & Mitra, S. Role of acetylated human AP-endonuclease (APE1/Ref-1) in regulation of the parathyroid hormone gene. *The EMBO journal* **22,** 6299–6309; 10.1093/emboj/cdg595 (2003).

197. Roychoudhury, S. *et al.* Human Apurinic/Apyrimidinic Endonuclease (APE1) Is Acetylated at DNA Damage Sites in Chromatin, and Acetylation Modulates Its DNA Repair Activity. *Molecular and cellular biology* **37**; 10.1128/MCB.00401-16 (2017).

198. Chattopadhyay, R. *et al.* Regulatory role of human AP-endonuclease (APE1/Ref-1) in YB-1-mediated activation of the multidrug resistance gene MDR1. *Molecular and cellular biology* **28,** 7066–7080; 10.1128/MCB.00244-08 (2008).

199. Qu, J., Liu, G.-H., Huang, B. & Chen, C. Nitric oxide controls nuclear export of APE1/Ref-1 through S-nitrosation of cysteines 93 and 310. *Nucleic Acids Res* **35,** 2522–2532; 10.1093/nar/gkl1163 (2007).

200. Thakur, S., Dhiman, M., Tell, G. & Mantha, A. K. A review on protein-protein interaction network of APE1/Ref-1 and its associated biological functions. *Cell biochemistry and function* **33,** 101–112; 10.1002/cbf.3100 (2015).

201. Vidal, A. E., Boiteux, S., Hickson, I. D. & Radicella, J. P. XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *The EMBO journal* **20,** 6530–6539; 10.1093/emboj/20.22.6530 (2001).

202. Ranalli, T. A., Tom, S. & Bambara, R. A. AP endonuclease 1 coordinates flap endonuclease 1 and DNA ligase I activity in long patch base excision repair. *J. Biol. Chem.* **277,** 41715–41724; 10.1074/jbc.M207207200 (2002).

203. Yamamori, T. *et al.* SIRT1 deacetylates APE1 and regulates cellular base excision repair. *Nucleic Acids Res* **38,** 832–845; 10.1093/nar/gkp1039 (2010).

204. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403,** 501–502; 10.1038/35000656 (2000).

205. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell stem cell* **15,** 447–459; 10.1016/j.stem.2014.08.003 (2014).

206. Ma, D. K. *et al.* Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323,** 1074–1077; 10.1126/science.1166859 (2009).

207. Cedar, H. & Bergman, Y. Choreography of Ig allelic exclusion. *Current opinion in immunology* **20,** 308–317; 10.1016/j.coi.2008.02.002 (2008).

208. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492; 10.1126/science.1083558 (2003).

209. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature genetics* **27,** 31–39; 10.1038/83730 (2001).

210. Hervouet, E. *et al.* Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells. *PLoS ONE* **5,** e11333; 10.1371/journal.pone.0011333 (2010).

211. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *The EMBO journal* **32,** 340–353; 10.1038/emboj.2012.331 (2013).

212. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279,** 52353–52360; 10.1074/jbc.M407695200 (2004).

213. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105; 10.1038/nature08829 (2010).

214. Dominguez, P. M. *et al.* DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell reports* **12,** 2086–2098; 10.1016/j.celrep.2015.08.036 (2015).

215. Zhu, B. *et al.* 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* **28,** 4157–4165; 10.1093/nar/28.21.4157 (2000).

216. Millar, C. B. *et al.* Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297,** 403–405; 10.1126/science.1073354 (2002).

217. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445,** 671–675; 10.1038/nature05515 (2007).

218. Schmitz, K.-M. *et al.* TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Molecular Cell* **33,** 344–353; 10.1016/j.molcel.2009.01.015 (2009).

219. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135,** 1201–1212; 10.1016/j.cell.2008.11.042 (2008).

220. Métivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452,** 45–50; 10.1038/nature06544 (2008).

221. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *J. Biol. Chem.* **288,** 9084–9091; 10.1074/jbc.M112.445585 (2013).

222. Bochtler, M., Kolano, A. & Xu, G.-L. DNA demethylation pathways: Additional players and regulators. *BioEssays : news and reviews in molecular, cellular and developmental biology* **39,** 1–13; 10.1002/bies.201600178 (2017).

223. Schomacher, L. & Niehrs, C. DNA repair and erasure of 5-methylcytosine in vertebrates. *BioEssays : news and reviews in molecular, cellular and developmental biology* **39**; 10.1002/bies.201600218 (2017).

224. Bogdanović, O. *et al.* Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nature genetics* **48,** 417–426; 10.1038/ng.3522 (2016).

225. Dawlaty, M. M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental cell* **29,** 102–111; 10.1016/j.devcel.2014.03.003 (2014).

226. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133; 10.1038/nature09303 (2010).

227. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307; 10.1126/science.1210944 (2011).

228. Ikeda, Y. & Kinoshita, T. DNA demethylation: a lesson from the garden. *Chromosoma* **118,** 37–41; 10.1007/s00412-008-0183-3 (2009).

229. Pidugu, L. S. *et al.* Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry* **55,** 6205–6208; 10.1021/acs.biochem.6b00982 (2016).

230. Weber, A. R. *et al.* Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat Commun* **7**; 10.1038/ncomms10806 (2016).

231. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology* **8,** 328–330; 10.1038/nchembio.914 (2012).

232. Hardeland, U., Steinacher, R., Jiricny, J. & Schär, P. Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *The EMBO journal* **21,** 1456–1464; 10.1093/emboj/21.6.1456 (2002).

233. Coey, C. T. *et al.* E2-mediated small ubiquitin-like modifier (SUMO) modification of thymine DNA glycosylase is efficient but not selective for the enzyme-product complex. *J. Biol. Chem.* **289,** 15810–15819; 10.1074/jbc.M114.572081 (2014).

234. Steinacher, R. & Schär, P. Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Current biology : CB* **15,** 616–623; 10.1016/j.cub.2005.02.054 (2005).

235. Tini, M. *et al.* Association of CBP/p300 Acetylase and Thymine DNA Glycosylase Links DNA Repair and Transcription. *Molecular Cell* **9,** 265–277; 10.1016/S1097-2765(02)00453-7 (2002).

236. Waters, T. R., Gallinari, P., Jiricny, J. & Swann, P. F. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **274,** 67–74; 10.1074/jbc.274.1.67 (1999).

237. Ho, J. J., Cattoglio, C., McSwiggen, D. T., Tjian, R. & Fong, Y. W. Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. *Genes & Development* **31,** 830–844; 10.1101/gad.295741.116 (2017).

238. Shimizu, Y., Iwai, S., Hanaoka, F. & Sugasawa, K. Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. *The EMBO journal* **22,** 164–173; 10.1093/emboj/cdg016 (2003).

239. Schomacher, L. *et al.* Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat Struct Mol Biol* **23,** 116–124; 10.1038/nsmb.3151 (2016).

240. Slyvka, A., Mierzejewska, K. & Bochtler, M. Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Scientific reports* **7,** 9001; 10.1038/s41598-017-07458-4 (2017).

241. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434; 10.1016/j.cell.2011.03.022 (2011).

242. Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260,** 1816–1819; 10.1126/science.8511591 (1993).

243. Muramatsu, M. *et al.* Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* **102,** 553–563; 10.1016/s0092-8674(00)00078-7 (2000).

244. Revy, P. *et al.* Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). *Cell* **102,** 565–575; 10.1016/s0092-8674(00)00079-9 (2000).

245. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146,** 67–79; 10.1016/j.cell.2011.06.020 (2011).

246. Yoon, J.-H., Iwai, S., O'Connor, T. R. & Pfeifer, G. P. Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. *Nucleic Acids Res* **31,** 5399–5404; 10.1093/nar/gkg730 (2003).

247. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature chemical biology* **10,** 574–581; 10.1038/nchembio.1532 (2014).

248. Boorstein, R. J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276,** 41991–41997; 10.1074/jbc.M106953200 (2001).

249. Matsubara, M. *et al.* Identification and characterization of mammalian 5-formyluracil-DNA glycosylase. *Nucleic acids research. Supplement (2001),* 233–234; 10.1093/nass/3.1.233 (2003).

250. Matsubara, M. *et al.* Mammalian 5-formyluracil-DNA glycosylase. 1. Identification and characterization of a novel activity that releases 5-formyluracil from DNA. *Biochemistry* **42,** 4993–5002; 10.1021/bi027322v (2003).

251. Rusmintratip, V. & Sowers, L. C. An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. *Proc Natl Acad Sci USA* **97,** 14183–14187; 10.1073/pnas.97.26.14183 (2000).

252. Liu, P., Burdzy, A. & Sowers, L. C. Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair* **2,** 199–210; 10.1016/s1568-7864(02)00198-2 (2003).

253. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* **40,** 4841–4849; 10.1093/nar/gks155 (2012).

254. Hashimoto, H., Zhang, X. & Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res* **40,** 8276–8284; 10.1093/nar/gks628 (2012).

255. Olinski, R., Starczak, M. & Gackowski, D. Enigmatic 5-hydroxymethyluracil: Oxidatively modified base, epigenetic mark or both? *Mutation research. Reviews in mutation research* **767,** 59–66; 10.1016/j.mrrev.2016.02.001 (2016).

256. Aller, P., Rould, M. A., Hogg, M., Wallace, S. S. & Doublié, S. A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol. *Proc Natl Acad Sci USA* **104,** 814–818; 10.1073/pnas.0606648104 (2007).

257. Cadet, J., Douki, T. & Ravanat, J.-L. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspects and formation in cells. *Accounts of chemical research* **41,** 1075–1083; 10.1021/ar700245e (2008).

258. Dizdaroglu, M. Oxidative damage to DNA in mammalian chromatin. *Mutation Research/DNAging* **275,** 331–342; 10.1016/0921-8734(92)90036-O (1992).

259. Frenkel, K., Goldstein, M. S. & Teebor, G. W. Identification of the cis-thymine glycol moiety in chemically oxidized and gamma-irradiated deoxyribonucleic acid by high-pressure liquid chromatography analysis. *Biochemistry* **20,** 7566–7571; 10.1021/bi00529a035 (1981).

260. McNulty, J. M., Jerkovic, B., Bolton, P. H. & Basu, A. K. Replication inhibition and miscoding properties of DNA templates containing a site-specific cis-thymine glycol or urea residue. *Chemical research in toxicology* **11,** 666–673; 10.1021/tx970225w (1998).

261. Breimer, L. H. & Lindahl, T. DNA glycosylase activities for thymine residues damaged by ring saturation, fragmentation, or ring contraction are functions of endonuclease III in Escherichia coli. *J. Biol. Chem.* **259,** 5543–5548 (1984).

262. Takao, M. *et al.* A Back-up Glycosylase in Nth1 Knock-out Mice Is a Functional Nei (Endonuclease VIII) Homologue. *J. Biol. Chem.* **277,** 42205–42213; 10.1074/jbc.M206884200 (2002).

263. Andersen, S. *et al.* Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis* **26,** 547–555; 10.1093/carcin/bgh347 (2005).

264. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H. E. Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA Repair* **6,** 505–516; 10.1016/j.dnarep.2006.10.014 (2007).

265. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418,** 99–103; 10.1038/nature00862 (2002).

266. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158; 10.1038/nature05610 (2007).

267. Doseth, B. *et al.* Uracil-DNA glycosylase in base excision repair and adaptive immunity: species differences between man and mouse. *J. Biol. Chem.* **286,** 16669–16680; 10.1074/jbc.M111.230052 (2011).

268. Akbari, M. *et al.* Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells. *Nucleic Acids Res* **32,** 5486–5498; 10.1093/nar/gkh872 (2004).

269. Haushalter, K. A., Todd Stukenberg, P., Kirschner, M. W. & Verdine, G. L. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Current Biology* **9,** 174–185; 10.1016/S0960-9822(99)80087-6 (1999).

270. Nilsen, H. *et al.* Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *The EMBO journal* **20,** 4278–4286; 10.1093/emboj/20.15.4278 (2001).

271. Kemmerich, K., Dingler, F. A., Rada, C. & Neuberger, M. S. Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res* **40,** 6016–6025; 10.1093/nar/gks259 (2012).

272. Visnes, T., Akbari, M., Hagen, L., Slupphaug, G. & Krokan, H. E. The rate of base excision repair of uracil is controlled by the initiating glycosylase. *DNA Repair* **7,** 1869–1881; 10.1016/j.dnarep.2008.07.012 (2008).

273. Doseth, B., Ekre, C., Slupphaug, G., Krokan, H. E. & Kavli, B. Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. *DNA Repair* **11,** 587–593; 10.1016/j.dnarep.2012.03.003 (2012).

274. Waters, T. R. & Swann, P. F. Kinetics of the action of thymine DNA glycosylase. *J. Biol. Chem.* **273,** 20007–20014; 10.1074/jbc.273.32.20007 (1998).

275. Mouret, J.-F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chemico-biological interactions* **77,** 187–201; 10.1016/0009-2797(91)90073-g (1991).

276. Allgayer, J., Kitsera, N., Lippen, C. von der, Epe, B. & Khobta, A. Modulation of base excision repair of 8-oxoguanine by the nucleotide sequence. *Nucleic Acids Res* **41,** 8559–8571; 10.1093/nar/gkt620 (2013).

277. Lühnsdorf, B. *et al.* Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Analytical Biochemistry* **425,** 47–53; 10.1016/j.ab.2012.03.001 (2012).

278. 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* **38,** 4285–4295; 10.1093/nar/gkq170 (2010).

279. Campalans, A., Amouroux, R., Bravard, A., Epe, B. & Radicella, J. P. UVA irradiation induces relocalisation of the DNA repair protein hOGG1 to nuclear speckles. *Journal of cell science* **120,** 23–32; 10.1242/jcs.03312 (2007).

280. Lühnsdorf, B., Epe, B. & Khobta, A. Excision of Uracil from Transcribed DNA Negatively Affects Gene Expression. *J. Biol. Chem.* **289,** 22008–22018; 10.1074/jbc.M113.521807 (2014).

281. Bauer, D. E., Canver, M. C. & Orkin, S. H. Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9. *JoVE*; 10.3791/52118 (2014).

282. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42,** W401-7; 10.1093/nar/gku410 (2014).

283. 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* **44,** W272-6; 10.1093/nar/gkw398 (2016).

284. Schröder, A. S. *et al.* Synthesis of (R)-Configured 2′-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides. *Org. Lett.* **18,** 4368–4371; 10.1021/acs.orglett.6b02110 (2016).

285. Zhou, H. *et al.* Frequency and distribution of AP-1 sites in the human genome. *DNA research : an international journal for rapid publication of reports on genes and genomes* **12,** 139–150; 10.1093/dnares/12.2.139 (2005).

286. Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* **234,** 47–52; 10.1126/science.3529394 (1986).

287. Bucher, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of Molecular Biology* **212,** 563–578; 10.1016/0022-2836(90)90223-9 (1990).

288. Dynan, W. S., Sazer, S., Tjian, R. & Schimke, R. T. Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature* **319,** 246–248; 10.1038/319246a0 (1986).

289. Angel, P. *et al.* Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell* **49,** 729–739; 10.1016/0092-8674(87)90611-8 (1987).

290. Chinenov, Y. & Kerppola, T. K. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* **20,** 2438–2452; 10.1038/sj.onc.1204385 (2001).

291. Toone, W. M., Morgan, B. A. & Jones, N. Redox control of AP-1-like factors in yeast and beyond. *Oncogene* **20,** 2336–2346; 10.1038/sj.onc.1204384 (2001).

292. Mechta-Grigoriou, F., Gerald, D. & Yaniv, M. The mammalian Jun proteins: redundancy and specificity. *Oncogene* **20,** 2378–2389; 10.1038/sj.onc.1204381 (2001).

293. McBride, K. & Nemer, M. The C-terminal domain of c-fos is required for activation of an AP-1 site specific for jun-fos heterodimers. *Molecular and cellular biology* **18,** 5073–5081; 10.1128/mcb.18.9.5073 (1998).

294. Schröder, A. S. *et al.* 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chemical communications (Cambridge, England)* **52,** 14361–14364; 10.1039/c6cc07517g (2016).

295. Aalaa Farag. Tools for Analysis of Overlapping DNA Repair Pathways for Thymine Glycol. Master thesis. Johannes Gutenberg-University, 2019.

296. Cappelli, E., Degan, P. & Frosina, G. Comparative repair of the endogenous lesions 8-oxo-7, 8-dihydroguanine (8-oxoG), uracil and abasic site by mammalian cell extracts: 8-oxoG is poorly repaired by human cell extracts. *Carcinogenesis* **21,** 1135–1141 (2000).

297. Barzilay, G. *et al.* Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. *Nat Struct Mol Biol* **2,** 561–568; 10.1038/nsb0795-561 (1995).

298. Erzberger, J. P. & Wilson, D. M. The role of Mg2+ and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis. *Journal of Molecular Biology* **290,** 447–457; 10.1006/jmbi.1999.2888 (1999).

299. Kao, J. Y., Goljer, I., Phan, T. A. & Bolton, P. H. Characterization of the effects of a thymine glycol residue on the structure, dynamics, and stability of duplex DNA by NMR. *J. Biol. Chem.* **268,** 17787–17793 (1993).

300. Bork Lühnsdorf. Einfluss von Uracil in der DNA auf die Genexpression: die Rolle der Basen-Exzisions-Reparatur. Dissertation. Johannes Gutenberg-University, 2015.

301. Iakovlev, D. A., Alekseeva, I. V., Kuznetsov, N. A. & Fedorova, O. S. Role of Arg243 and His239 Residues in the Recognition of Damaged Nucleotides by Human Uracil-DNA Glycosylase SMUG1. *Biochemistry. Biokhimiia* **85,** 594–603; 10.1134/S0006297920050089 (2020).

302. Sjolund, A. B., Senejani, A. G. & Sweasy, J. B. MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles. *Mutation research* **743-744,** 12–25; 10.1016/j.mrfmmm.2012.11.001 (2013).

303. Hegde, M. L., Izumi, T. & Mitra, S. Oxidized base damage and single-strand break repair in mammalian genomes: role of disordered regions and posttranslational modifications in early enzymes. *Progress in molecular biology and translational science* **110,** 123–153; 10.1016/B978-0-12-387665-2.00006-7 (2012).

304. Radak, Z. & Boldogh, I. 8-Oxo-7,8-dihydroguanine: links to gene expression, aging, and defense against oxidative stress. *Free radical biology & medicine* **49,** 587–596; 10.1016/j.freeradbiomed.2010.05.008 (2010).

305. Fraga, C. G., Shigenaga, M. K., Park, J. W., Degan, P. & Ames, B. N. Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine. *Proc Natl Acad Sci USA* **87,** 4533–4537; 10.1073/pnas.87.12.4533 (1990).

306. Spivak, G. & Hanawalt, P. C. Host cell reactivation of plasmids containing oxidative DNA lesions is defective in Cockayne syndrome but normal in UV-sensitive syndrome fibroblasts. *DNA Repair* **5,** 13–22; 10.1016/j.dnarep.2005.06.017 (2006).

307. Pastoriza-Gallego, M., Armier, J. & Sarasin, A. Transcription through 8-oxoguanine in DNA repair-proficient and Csb(-)/Ogg1(-) DNA repair-deficient mouse embryonic fibroblasts is dependent upon promoter strength and sequence context. *Mutagenesis* **22,** 343–351; 10.1093/mutage/gem024 (2007).

308. Dynan, W. S. & Tjian, R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35,** 79–87; 10.1016/0092-8674(83)90210-6 (1983).

309. Moorefield, K. S., Fry, S. J. & Horowitz, J. M. Sp2 DNA binding activity and trans-activation are negatively regulated in mammalian cells. *J. Biol. Chem.* **279,** 13911–13924; 10.1074/jbc.M313589200 (2004).

310. Terrados, G. *et al.* Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. *Nucleic Acids Res* **40,** 7844–7857; 10.1093/nar/gks544 (2012).

311. Ihn, H. & Trojanowska, M. Sp3 is a transcriptional activator of the human alpha2(I) collagen gene. *Nucleic Acids Res* **25,** 3712–3717; 10.1093/nar/25.18.3712 (1997).

312. Oka, S. *et al.* NMR structure of transcription factor Sp1 DNA binding domain. *Biochemistry* **43,** 16027–16035; 10.1021/bi048438p (2004).

313. Hashimoto, H., Zhang, X. & Cheng, X. Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *Journal of Molecular Biology* **425,** 971–976; 10.1016/j.jmb.2013.01.013 (2013).

314. Hagen, G., Müller, S., Beato, M. & Suske, G. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic Acids Res* **20,** 5519–5525; 10.1093/nar/20.21.5519 (1992).

315. Hagen, G., Müller, S., Beato, M. & Suske, G. Sp1-mediated transcriptional activation is repressed by Sp3. *The EMBO journal* **13,** 3843–3851 (1994).

316. Imataka, H. *et al.* Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. *The EMBO journal* **11,** 3663–3671 (1992).

317. Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403,** 859–866; 10.1038/35002510 (2000).

318. Wu, J., McKeague, M. & Sturla, S. J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *Journal of the American Chemical Society* **140,** 9783–9787; 10.1021/jacs.8b03715 (2018).

319. Senthilkumar, K., Grozema, F. C., Guerra, C. F., Bickelhaupt, F. M. & Siebbeles, L. D. A. Mapping the sites for selective oxidation of guanines in DNA. *J. Am. Chem. Soc.* **125,** 13658–13659; 10.1021/ja037027d (2003).

320. Kawanishi, S., Oikawa, S., Murata, M., Tsukitome, H. & Saito, I. Site-specific oxidation at GG and GGG sequences in double-stranded DNA by benzoyl peroxide as a tumor promoter. *Biochemistry* **38,** 16733–16739; 10.1021/bi990890z (1999).

321. Yoshioka, Y. *et al.* Experimental and Theoretical Studies on the Selectivity of GGG Triplets toward One-Electron Oxidation in B-Form DNA. *J. Am. Chem. Soc.* **121,** 8712–8719; 10.1021/ja991032t (1999).

322. Saito, I. *et al.* Mapping of the Hot Spots for DNA Damage by One-Electron Oxidation: Efficacy of GG Doublets and GGG Triplets as a Trap in Long-Range Hole Migration. *J. Am. Chem. Soc.* **120,** 12686–12687; 10.1021/ja981888i (1998).

323. Moore, S. P. G., Kruchten, J., Toomire, K. J. & Strauss, P. R. Transcription Factors and DNA Repair Enzymes Compete for Damaged Promoter Sites. *J. Biol. Chem.* **291,** 5452–5460; 10.1074/jbc.M115.672733 (2016).

324. Xia, L. *et al.* CHD4 Has Oncogenic Functions in Initiating and Maintaining Epigenetic Suppression of Multiple Tumor Suppressor Genes. *Cancer cell* **31,** 653-668.e7; 10.1016/j.ccell.2017.04.005 (2017).

325. Henry, R. A. *et al.* Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by p300. *Biochemistry* **55,** 6766–6775; 10.1021/acs.biochem.6b00841 (2016).

326. Capozzo, I., Iannelli, F., Francia, S. & Di d'Adda Fagagna, F. Express or repress? The transcriptional dilemma of damaged chromatin. *The FEBS journal* **284,** 2133–2147; 10.1111/febs.14048 (2017).

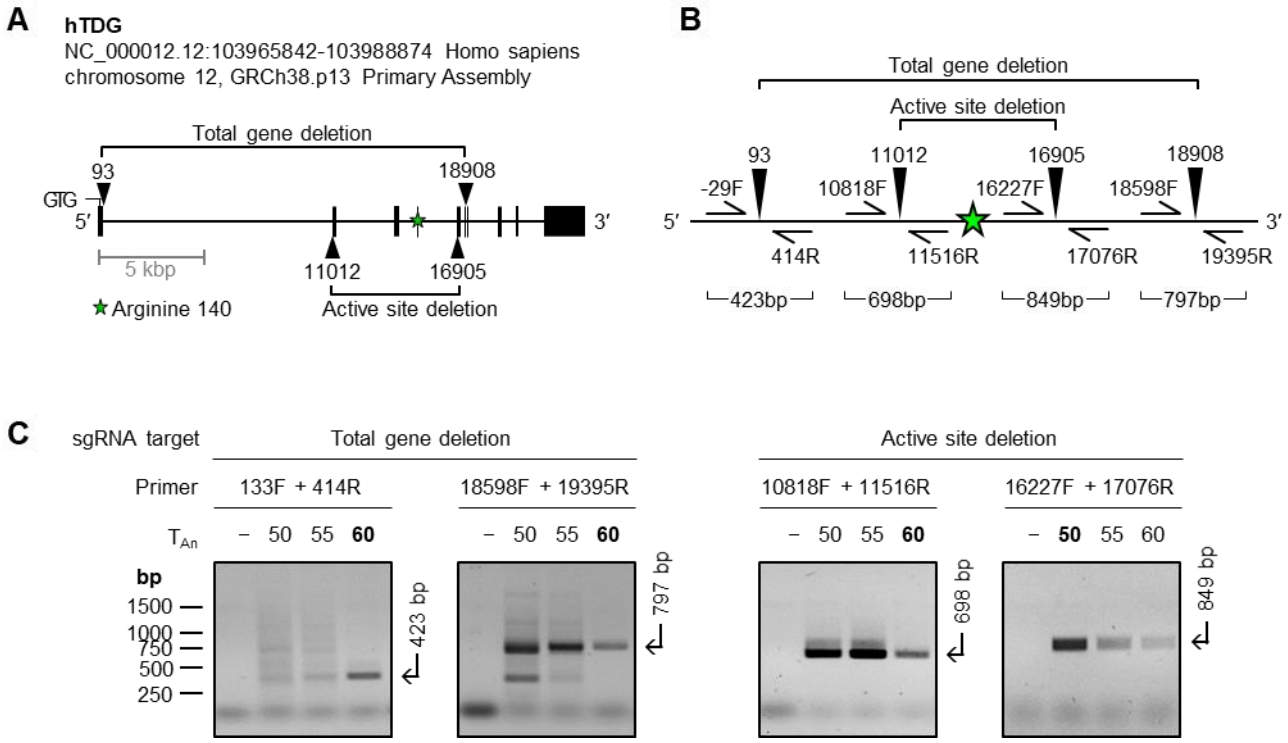
327. Maréchal, A. & Zou, L. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harbor Perspectives in Biology* **5**; 10.1101/cshperspect.a012716 (2013).

328. Rinaldi, C., Pizzul, P., Longhese, M. P. & Bonetti, D. Sensing R-Loop-Associated DNA Damage to Safeguard Genome Stability. *Frontiers in Cell and Developmental Biology* **8,** 618157; 10.3389/fcell.2020.618157 (2020).

329. Niehrs, C. & Luke, B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nature reviews. Molecular cell biology* **21,** 167–178; 10.1038/s41580-019-0206-3 (2020).

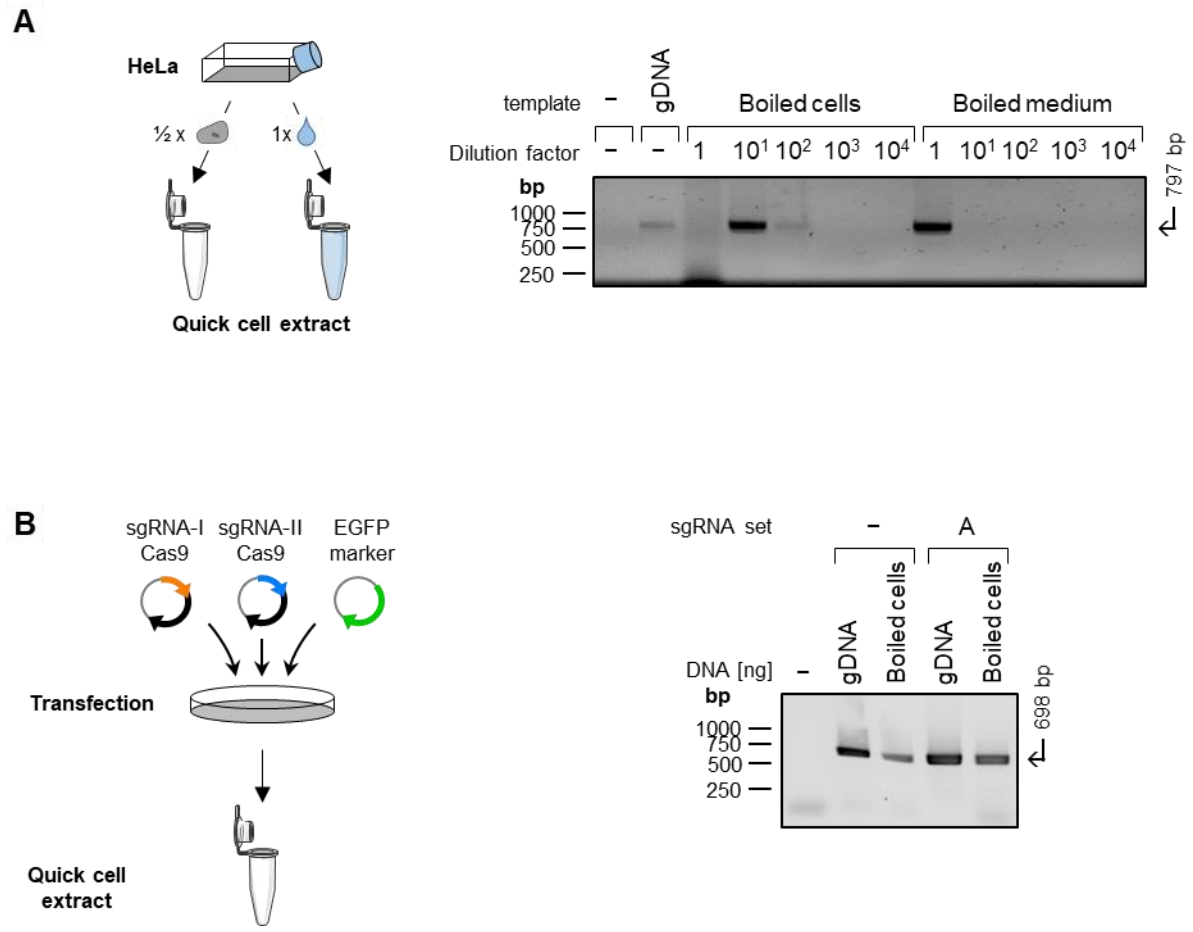
330. Würtele, H., Little, K. C. E. & Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene therapy* **10,** 1791–1799; 10.1038/sj.gt.3302074 (2003).

10 Appendix I



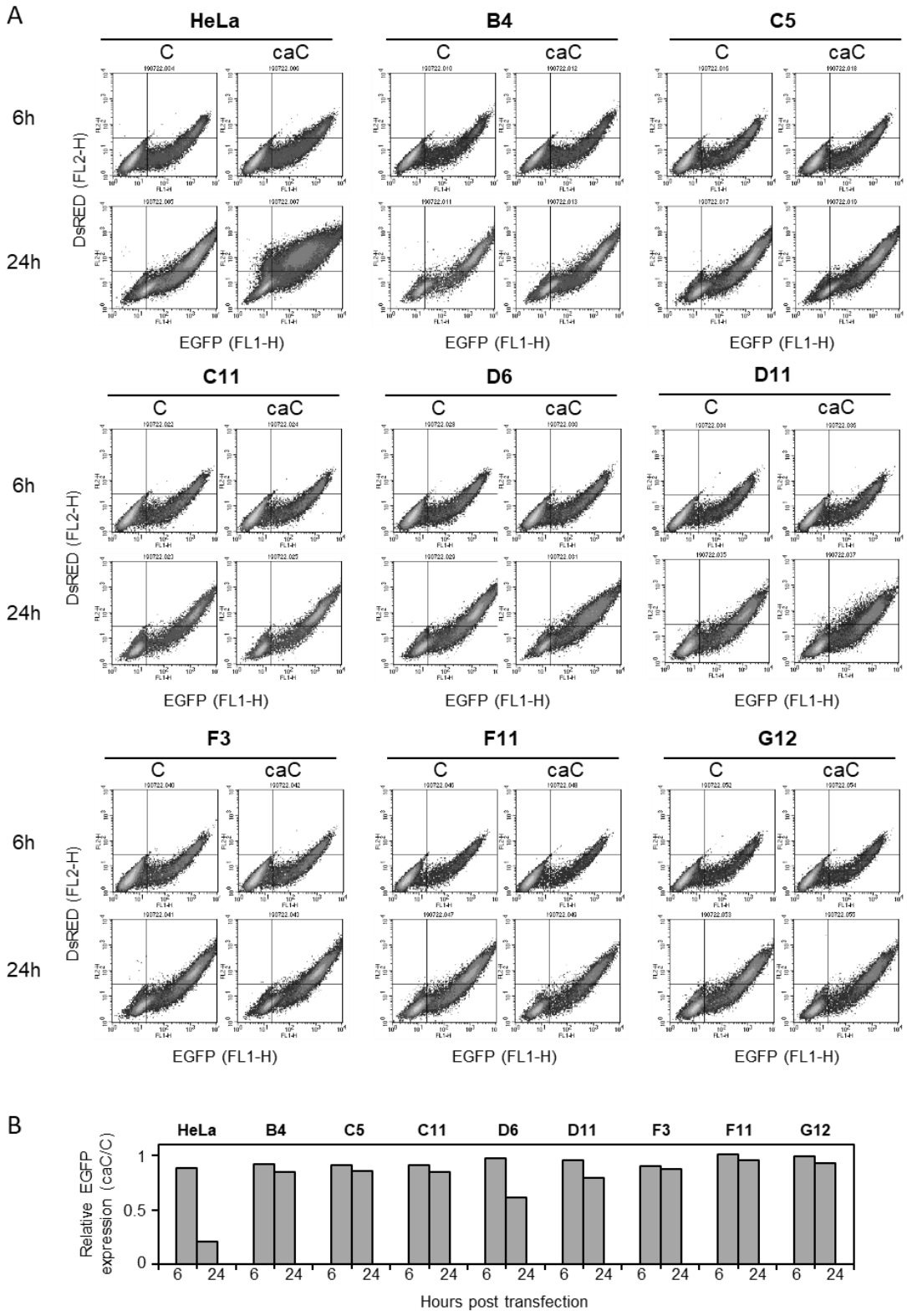
Appendix I 1: PCR amplification of human non-rearranged TDG genes by four different primers pairs

Amplification of non-rearranged TDG genes by *Taq* DNA Polymerase at three different annealing temperatures to optimise the PCR conditions. (A) True to scale scheme of the human TDG sequence: gene coding region (black line), upstream/downstream DNA sequence (greyline), exons (black boxes), active site coding exon (green and purple star), sgRNA targets (triangle with distance from the TSS indicated above). (B) Schematic representation of the non-rearranged and rearranged TDG gene after active site deletion or total gene deletion: active site coding sequence (green and purple star), sgRNA targets (triangle) and PCR primer sites (half arrow with distance from the TSS indicated above) with the according PCR product length. (C) Amplification of selected sequences from the TDG gene by PCR using 10 ng of gDNA from HeLa cells as template and three different annealing temperatures (T_{Anneal}). The optimal T_{Anneal} is indicated bold and the migration positions of the respective PCR products within the agarose gel are indicted by arrows to the right.



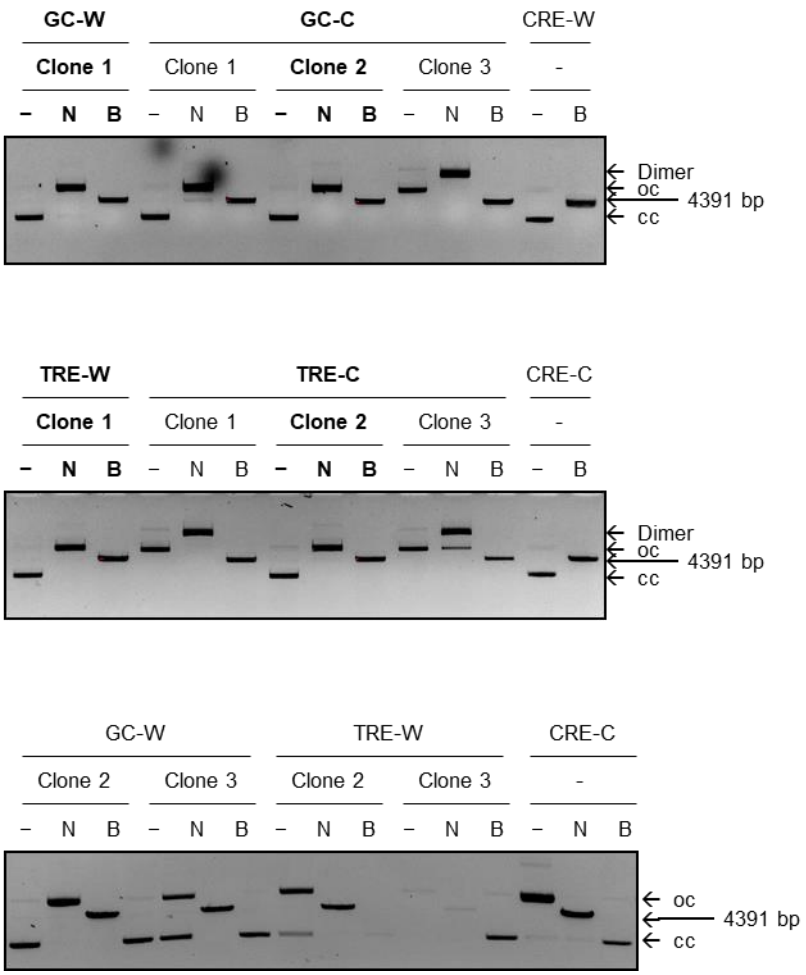
Appendix I 2: PCR amplification of human non-rearranged TDG genes from quick cell extracts

Amplification of non-rearranged TDG genes by S7 Fusion DNA polymerase using boiled cells or their medium as template to optimise the PCR conditions from quick cell extracts (chapter 6.25). (A) Amplification of a non-rearranged TDG sequence from quick cell extracts of parental HeLa cells. (Left panel) Schematic representation of the experimental setup to generate the quick cell extracts from exponentially growing HeLa cells at 70% confluency in 25 cm^2 flasks. (Right panel) Amplification of a 797 bp long TDG sequence by PCR using different dilutions of boiled HeLa cells or their medium generated by quick cell extracts with primers TDG_18598F/19395R. 10 ng of gDNA was used as a control and the product amplification was analysed by agarose gel electrophoresis. As quick cell extracts from boiled HeLa cells at a 1 to 10 dilution induced highest PCR product amplification this condition was used for all further PCR reactions. (B) Amplification of a rearranged TDG sequences from quick cell extracts of HeLa cells transfected with the two TDG active site deletion knockout constructs and the pZAJ transfection marker. (Left panel) Schematic representation of the experimental setup to knockout TDG in HeLa cells by co-transfection of knockout constructs coding for the targeting sgRNAs I or II and the CAS9 protein and an EGFP transfection marker: knockout constructs (circles with orange/blue arrows indicating the sgRNA sequence and black arrows indicating the CAS9 sequence), EGFP transfection marker (circle with green arrow indicating EGFP sequence). (Right panel) Amplification of a 689 bp long sequence from non-rearranged TDG by PCR using different dilutions a 1 to 10 dilution of boiled transfected HeLa cells generated by quick cell extracts with primers TDG_16227F/17076R. Quick cell extracts of untransfected cells were used as a control and as well as 10 ng of gDNA of untransfected and transfected cells and the product amplification was analysed by agarose gel electrophoresis. Similar DNA signal intensities verify the efficient PCR product amplification from both gDNA and quick cell extracts of transfected cells.



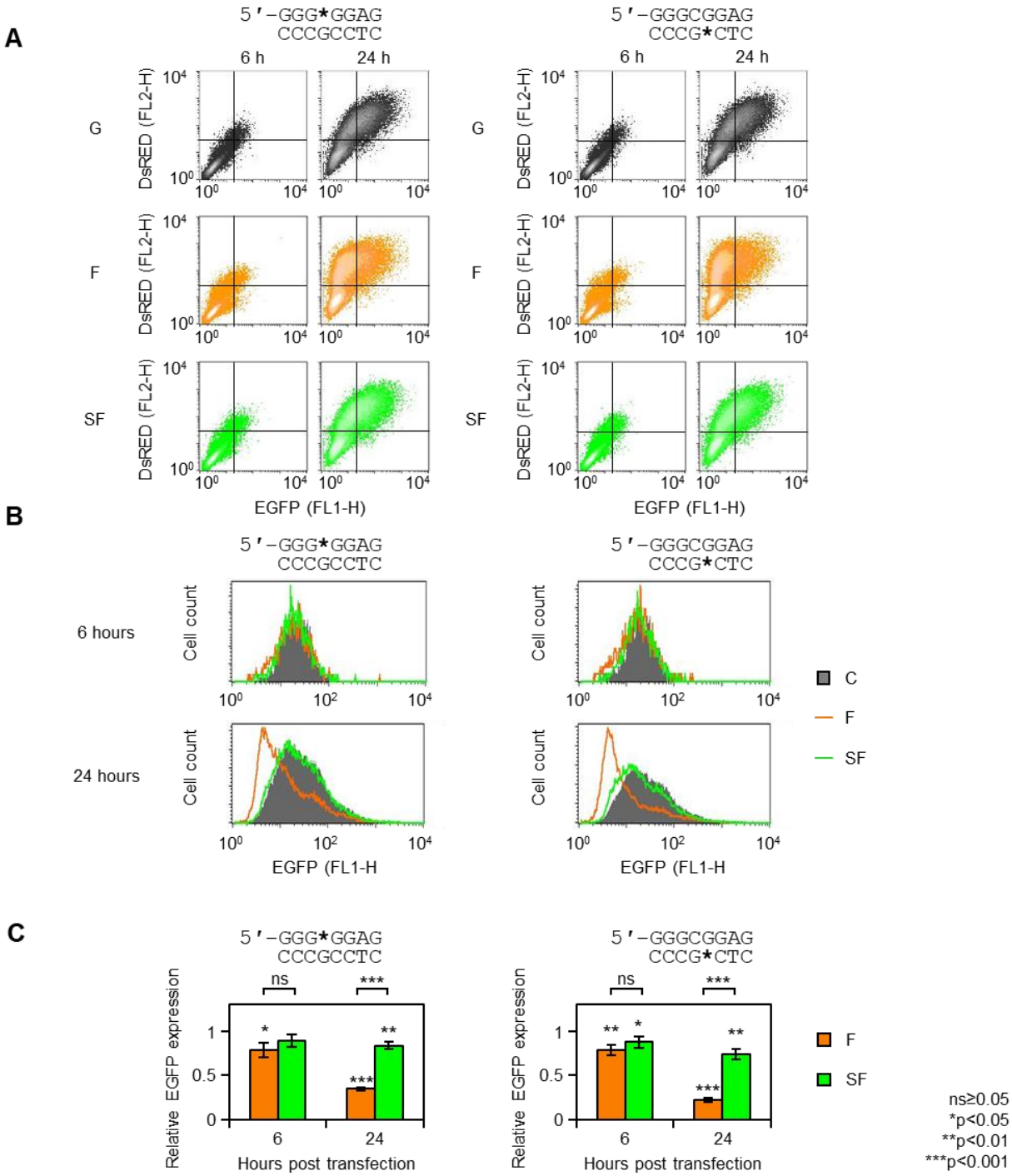
Appendix I 3: Expression of CMV-1111 reporters containing a single 5-caC at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter in eight selected TDG knockout single cell clones

Quantitative EGFP expression analysis of HeLa and eight derived TDG knockout single cell clones transfected with pCMV-1111-ZA-W reporters containing single 5-caC residue at the central CpG dinucleotide within the non-transcribed strand of a selected CRE element of the CMV-1111 promoter. (A) Fluorescent distribution plots and (B) relative EGFP expression of transfected HeLa and HeLa derived TDG knockout clones 6- and 24 hours after transfection (n=1).



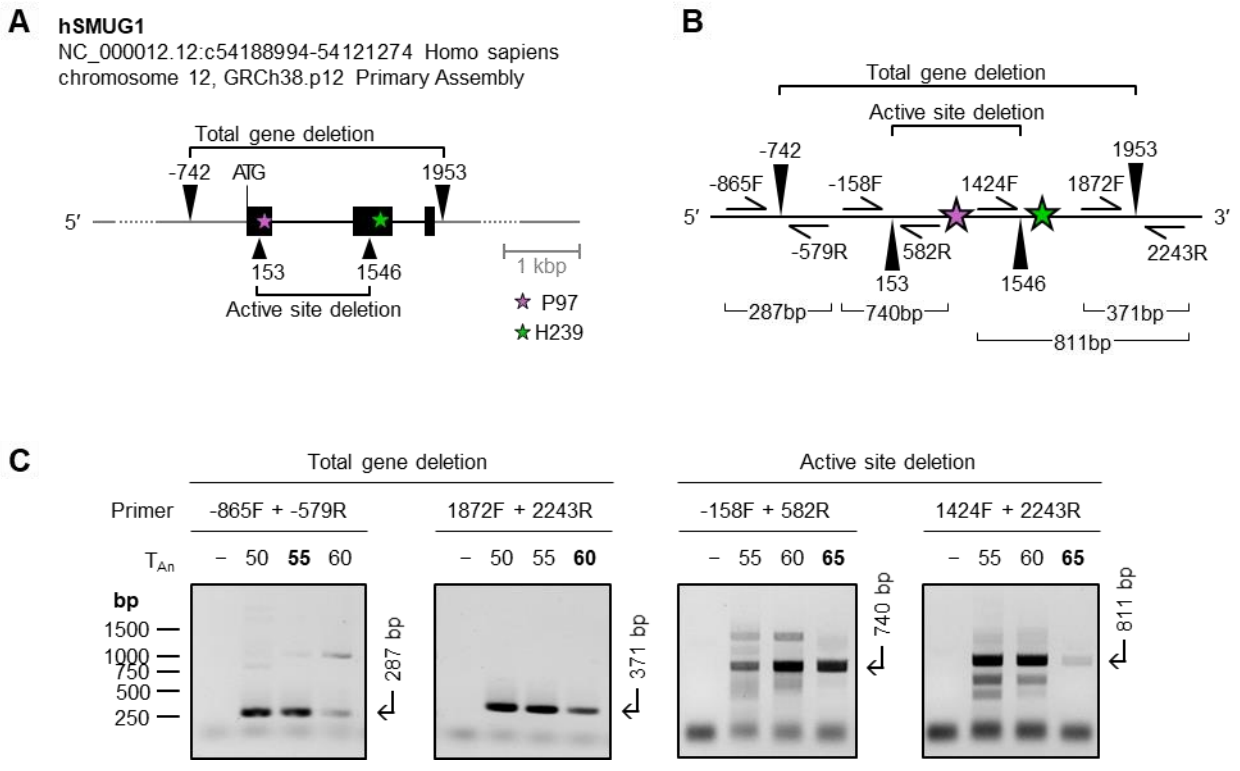
Appendix I 4: Analytical digestion of cloned minimal GC box and TRE reporters

Minimal GC box and TRE reporters were generated from CRE-UNO constructs by cloning procedure. Cloned vectors were amplified in E.coli and three colonies per reporter were selected and used for plasmid mini-preparation. Analytical digestion assays were performed with the extracted plasmid DNA using BsrDI and Nb.BsrDI enzymes, followed by agarose gel electrophoresis. BsrDI and Nb.BsrDI share the same recognition sites, which are located at either side of the GC box/TRE sequences based on parental CRE-UNO reporter design. Treatment of the GC box, TRE or parental CRE-UNO vectors by Nb.BsrDI converts the cc-form plasmids into oc-form plasmids. BsrDI cleaves the reporter DNA, resulting in a big 4391bp fragment and a short 18 bp fragment (undetectable with this approach). Arrows indicate migration positions of the linearised vector (4391 bp), dimers and of the cc- and oc-form vector in agarose gels. Plasmid dimers run slower than monomers in their cc- and oc-form, thereby enabling their differentiation. BsrDI treatment only generates fragments of 4391bp if the GC box and TRE coding inserts are correctly inserted into the plasmid during the cloning procedure, verifying successful GC box and TRE reporter generation in combination with DNA sequencing (data not shown). Clones which were selected for further experiments (bold) were expanded and used for plasmid mega-preparation.



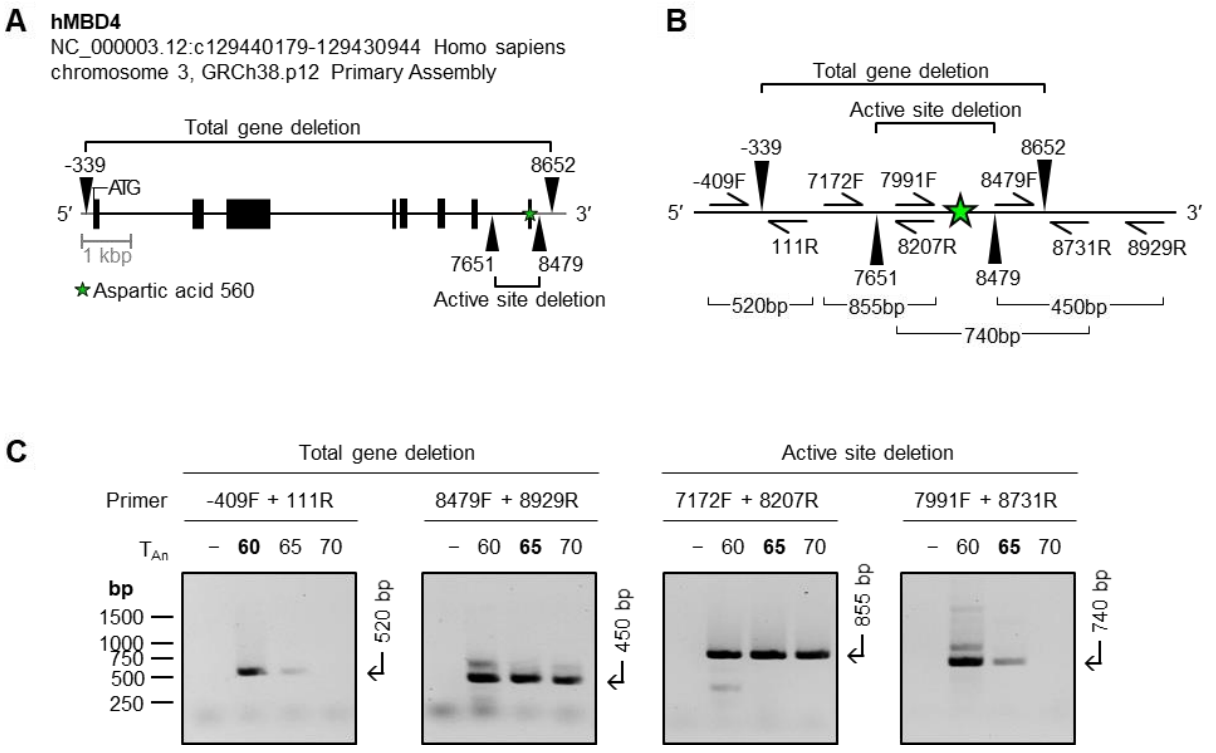
Appendix I 5: Expression of GC box reporters containing a single apyrimidinic site in either strand of the central CpG dinucleotide of the GC box gene regulatory element in HeLa derived TDG knockout cells

Quantitative EGFP expression analysis of HeLa derived TDG knockout cells transfected with pGCbox-ZA-W (left panel) and pGCbox-ZA-C reporters (right panel) containing either a single apyrimidinic site analogue F, its endonuclease protected derivative SF or the G control in either strand of the central CpG dinucleotide of the GC box upstream regulatory element (position of the modified base: asterisk). Representative (A) scatter plots and (B) fluorescent distribution plots of HeLa derived TDG knockout cells 6- and 24 hours after transfection. C) Quantification of the relative EGFP expression of transfected TDG knockout HeLa cells in four independent experiments (mean ± SD).



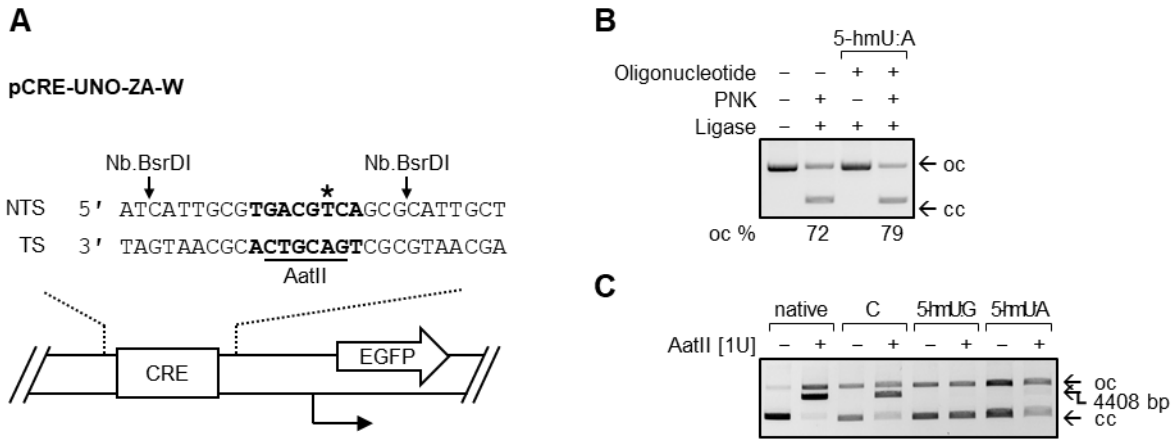
Appendix I 6: PCR amplification of non-rearranged SMUG1 genes by four different primers pairs

Amplification of non-rearranged SMUG1 genes by *Taq* DNA Polymerase at three different annealing temperatures to optimise the PCR conditions. (A) True to scale scheme of the human SMUG1 sequence: gene coding region (black line), upstream/downstream DNA sequence (greyline), exons (black boxes), histidine 239 active site coding nucleotides (green star), proline 97 essential amino acid coding nucleotides (purple star), sgRNA targets (triangle with distance from the TSS indicated above). (B) Schematic representation of the non-rearranged and rearranged SMUG1 gene after active site deletion or total gene deletion: active site coding sequence (green and purple star), sgRNA targets (triangle) and PCR primer sites (half arrow with distance from the TSS indicated above) with the according PCR product length. (C) Amplification of selected sequences from the SMUG1 gene by PCR using 10 ng of gDNA from HeLa cells as template and three different annealing temperatures (T_{Anneal}). The migration positions of the respective PCR products within the agarose gel are indicted by arrows to the right.



Appendix I 7: PCR amplification of non-rearranged MDB4 genes by four different primers pairs

Amplification of non-rearranged MBD4 genes by *Taq* DNA Polymerase at three different annealing temperatures to optimise the PCR conditions. (A) True to scale scheme of the human MBD4 sequence: gene coding region (black line), upstream/downstream DNA sequence (greyline), exons (black boxes), aspartic acid 560 active site coding nucleotides (green star), sgRNA targets (triangle with distance from the TSS indicated above). (B) Schematic representation of the non-rearranged and rearranged MBD4 gene after active site deletion or total gene deletion: active site coding sequence (green star), sgRNA targets (triangle) and PCR primer sites (half arrow with distance from the TSS indicated above) with the according PCR product length. (C) Amplification of selected sequences from the MBD4 gene by PCR using 10 ng of gDNA from HeLa cells as template and three different annealing temperatures (T_{Anneal}). The migration positions of the respective PCR products within the agarose gel are indicted by arrows to the right.



Appendix I 8: Construction of CRE-UNO reporters containing 5-hmU opposite to A at position +2 within the non-transcribed strand of the standalone CRE gene regulatory element

(A) Scheme of the standalone CRE gene regulatory element within the pCRE-UNO-ZA-W reporter: EGFP coding sequence (white arrow), TSS (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking sites (black arrows) and position of 5-hmU in the incorporated oligonucleotides (asterisk). (B) Ligation of Nb.BsrDI-nicked and gapped CRE-UNO constructs with synthetic oligonucleotides containing 5-hmU or the "dT" control at position +2 (counted from central CpG dinucleotide) within the non-transcribed strand of the standalone CRE upstream regulatory element in the presence and absence of PNK. (C) Incubation of the generated construct and a pCRE-UNO-ZA-W reporter containing 5-hmU opposite to G at the central CpG dinucleotide of the non-transcribed strand of the standalone CRE gene regulatory element (chapter 7.3) as control with the restriction enzyme AatII (chapter 6.17) to verify the presence of the modifications based on the inhibition of the plasmid cleavage. Deviating from the standard protocol 1U instead of 3U of AatII were used with 100 ng of plasmid DNA resulting in incomplete unmodified plasmid cleavage. Arrows indicate migration positions of the linearised vector (4408 bp) and of the covalently closed (cc) and open circular (oc) form plasmids.

11 Appendix II

Sequences of EGFP reporters in FASTA format

pGCbox-ZA-W (4408 bp)

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATCATTGCATGGGCGGAGCGCATTGCTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA
CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCA
GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCCTAGCA
GAAGCACAGGCTGCAGGGTGACGGTCCATCCCGCTCTCCTGGGCACAAGGCATGGGCAGCGTGCCATC
ATCCTGCTCCTCCACCTCCGGCGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAGTACTCAGATC
TCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAAC
TGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCC
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTT
TGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTT
AAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAA
AAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTG
GACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTA
ATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCT
AGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT
ACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGC
TCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCC
AGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATT
TTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTT
TTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAA
CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACA
ACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTG
TCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCC
ACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATT
GGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CTGATGCTATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACAT
CGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCA

TCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCG
TCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATC
GACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGA
AGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGC
GCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACC
AAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTC
GGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGC
CCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGG
CAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAG
GGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTC
CCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTG
CCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATC
TAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGC
GTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT
TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAG
CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTAT
AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAG
CCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

pGCbox-ZA-C (4408 bp)

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATGCAATGCGTGGGCGGAGCGCAATGTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA
CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCA
GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCCTAGCA
GAAGCACAGGCTGCAGGGTGACGGTCCATCCCGCTCTCCTGGGCACAAGGCATGGGCAGCGTGCCATC
ATCCTGCTCCTCCACCTCCGGCGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAGTACTCAGATC
TCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAAC
TGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCC
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTT
TGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTT
AAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAA

AAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTG
GACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTA
ATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCT
AGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT
ACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGC
TCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCC
AGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATT
TTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTT
TTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAA
CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACA
ACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTG
TCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCC
ACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATT
GGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CTGATGCTATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACAT
CGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCA
TCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCG
TCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATC
GACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGA
AGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGC
GCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACC
AAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTC
GGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGC
CCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGG
CAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAG
GGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTC
CCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTG
CCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATC
TAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGC
GTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT
TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAG
CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTAT
AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAG
CCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

pTREC-ZA-W (4407 bp)

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATCATTGCATGAGTCAGCGCATTGCTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCCA
CCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCCTAGCAG
AAGCACAGGCTGCAGGGTGACGGTCCATCCCGCTCTCCTGGGCACAAGGCATGGGCAGCGTGCCATCA
TCCTGCTCCTCCACCTCCGGCGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAGTACTCAGATCT
CGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACT
GATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTAC
AAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT
GTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTA
AATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAA
AGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGG
ACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAA
TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAG
AGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTA
GGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTA
CAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT
ACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA
AGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCT
CCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCA
GGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCT
AACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT
TTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAAC
AAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAA
CAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGT
CAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCA
CGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTG
GGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC
TGATGCTATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATC
GCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCAT
CAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGT
CGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCG
ACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAA
GAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCG
CATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCA
AGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCG
GAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCC

CACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGC
AATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGG
GCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCC
CCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGC
CATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCT
AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCG
TCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGC
CACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGC
GGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA
TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT
AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATA
GTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

pTREC-ZA-C (4407 bp)

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATGCAATGCGCTGACTCATGCAATGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCCA
CCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCCTAGCAG
AAGCACAGGCTGCAGGGTGACGGTCCATCCCGCTCTCCTGGGCACAAGGCATGGGCAGCGTGCCATCA
TCCTGCTCCTCCACCTCCGGCGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAGTACTCAGATCT
CGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACT
GATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTAC
AAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT
GTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTA
AATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAA
AGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGG
ACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAA
TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAG
AGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTA
GGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTA
CAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT
ACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA
AGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCT
CCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCA
GGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCT
AACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT

TTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAAC
AAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAA
CAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGT
CAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCA
CGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTG
GGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC
TGATGCTATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATC
GCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCAT
CAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGT
CGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCG
ACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAA
GAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCG
CATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCA
AGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCG
GAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCC
CACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGC
AATAAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGG
GCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCC
CCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGC
CATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCT
AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCG
TCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGC
CACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGC
GGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA
TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT
AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATA
GTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

12 Appendix III

FASTA formatted sequences of pX330-SpCAS9-HF derived vectors for the induction of CRISPR-CAS9-mediated gene silencing.

Nucleotide sequences of the parental pX330-SpCAS9-HF vector were obtained from ADDGENE website (https://www.addgene.org/108301/sequences/).

pX330-SpCAS9-HF1 (8506 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTT
AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG
TGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCC
AATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA
CCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG
GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACTTT
CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTG
CTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTT
GTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTAT
GGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCT
GCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTA
CTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGG
GTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCACTT
TTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATA
TTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTC
CCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGAT
CACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCA
AGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGA
ACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGA
GATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGA
AGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACC
ATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGC
CCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCG
ACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAAC
GCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCT
GATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCC
TGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGC
CAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCC
CCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTC
GTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGG
CTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGG
ACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTT

TTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACG
TGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACC
CCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCGC
CTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGC
GAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAA
AGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCA
ACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAG
GAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGAT
ACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACA
ATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGACAG
CCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACA
TTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAG
CTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGAC
CACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTAC
TACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGA
TGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAA
GCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTAC
TGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAG
AGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGATCA
CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATC
CGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCG
CCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTG
CGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAA
CATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCG
AGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTG
CTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACG
GCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCC
AAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA
AGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAG
GGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACG
AGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTG
CTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTT
TACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT
ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACA
CGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAA
AAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATC
TGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAAT
AAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTAGT
GATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCC
GACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGTAC

GCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGC
CAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCC
GTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAA
AAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC
GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTCGG
GCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTTAA
CAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACAAT
CTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGG
CTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG
TTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAA
TGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA
TTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTT
CAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGC
GGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGT
TGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGA
CGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG
TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCA
CAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACG
ACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTA
CTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCT
GCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCGCG
GTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGA
GCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG
CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTA
GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGC
TGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC
AGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTT
ATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG
AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCA
CATGT

pX330-SpCAS9-HF1_ΔTDG+93 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGAACGCGGGCAGGTAATACCGGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC

TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG

TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA
GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA
GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC
ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC

CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT
GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔTDG+11012 (8507 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGATGGCTGAAGCTCCTAATAGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC
TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT
TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC

GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG
AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC

TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT

CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGT

pX330-SpCAS9-HF1_ΔTDG+16905 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCTACCAGGGAAGTATGGTATGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC

ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA
GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA
GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC
ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC

CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT
GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔTDG+18908 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGACGAAATATGGACGTTCAAGGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC

CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA

GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA
GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC
ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT

GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔSMUG1-742 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCATCTAAGGCAAGATGGCGTGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC

ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA
GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA

GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC
ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT
GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG

CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔSMUG1+153 (8507 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGCATCATCTACAATCCCGGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC
TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT
TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC
GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC

CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG
AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC
TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG

ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGT

pX330-SpCAS9-HF1_ΔSMUG1+1546 (8507 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAGTAAGGTTGCGCCCGCTGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC

TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTTAATTATT
TTGTGCAGCGATGGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT
TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC
GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG

AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC
TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC

TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGT

pX330-SpCAS9-HF1_ΔSMUG1+1953 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGAATACGTTTCCCAGCGACCGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC
TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT
TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC
GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG

CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG
AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC
TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG

TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGT

pX330-SpCAS9-HF1_ΔMBD4-339 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCCGAGCGCGCATGTCCGAAAGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT

CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA
GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA
GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC
ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG

AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT
GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔMBD4+7651 (8507 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTACACCACTACAGGAAAGCGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC
TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT

TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC
GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG
AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG

TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC
TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG

GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGT

pX330-SpCAS9-HF1_ΔMBD4+8479 (8508 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCAGTAGCAAAGATCCACCATGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGC
GCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTT
TATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGC
GCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCA
CTTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATG
ATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCA
TCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAG
AGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCT
GGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACA
GCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC

CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGG
CCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGAC
CGCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAG
ACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGA
CAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGC
TGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA
GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGA
TCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTG
ATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTT
TTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAG
CAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA
TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGT
ACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAG
TCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
GAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG
AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAA
GCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAA
GTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCT
GTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC
AAAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC

ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCT
AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCG
CCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGC
CTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAG
TACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCT
CGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
AGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGT
TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTT
GCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAA
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACT
TCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCC
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTG
GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC
TCACATGT

pX330-SpCAS9-HF1_ΔMBD4+8652 (8507 bp)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTGGGCCCCTAGCTTTAGCAGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCG
CCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTC
TGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGGGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT
ATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCG
CTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGT
TACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAA
GGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCAC
TTTTTTTCAGGTTGGACCGGTGCCACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGA
TATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCAT
CAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGA
GAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAAC
GAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGAT
GGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT
TTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCA
CCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACC
GCCTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTG
AAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTT
CAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC

GAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAG
ATACACCGGCTGGGGCGCGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGGCGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCA
CATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACG
AGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTAC
GATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAG
AGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGGCGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGA
TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAAC
CGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT
CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCC
TAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA
AGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA
CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG
TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
CACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCA
AAAAAGAAAAAGTAAGAATTCCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTA
GTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGC
CCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCC
TGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGT
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC
CCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTC
GGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA

GGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTT
AATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGC
TTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACC
AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTT
CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGAAGCCG
CGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACA