Article

EGFP Reporters for Direct and Sensitive Detection of Mutagenic Bypass of DNA Lesions

Marta Rodriguez-Alvarez [1], Daria Kim [2,3] and Andriy Khobta [1,*

[1] Unit "Responses to DNA Lesions", Institute of Toxicology, University Medical Center of the Johannes Gutenberg University Mainz, Obere Zahlbacher Str. 67, 55131 Mainz, Germany; mrodrigu@uni-mainz.de
[2] Novosibirsk State University, 1 Pirogova St., 630090 Novosibirsk, Russia; kim.daria.nsk@gmail.com
[3] Laboratory of Genome and Protein Engineering, SB RAS Institute of Chemical Biology and Fundamental Medicine, 8 Lavrentieva Ave., 630090 Novosibirsk, Russia
* Correspondence: khobta@uni-mainz.de

check for
updates

Abstract: The sustainment of replication and transcription of damaged DNA is essential for cell survival under genotoxic stress; however, the damage tolerance of these key cellular functions comes at the expense of fidelity. Thus, translesion DNA synthesis (TLS) over damaged nucleotides is a major source of point mutations found in cancers; whereas erroneous bypass of damage by RNA polymerases may contribute to cancer and other diseases by driving accumulation of proteins with aberrant structure and function in a process termed "transcriptional mutagenesis" (TM). Here, we aimed at the generation of reporters suited for direct detection of miscoding capacities of defined types of DNA modifications during translesion DNA or RNA synthesis in human cells. We performed a systematic phenotypic screen of 25 non-synonymous base substitutions in a DNA sequence encoding a functionally important region of the enhanced green fluorescent protein (EGFP). This led to the identification of four loss-of-fluorescence mutants, in which any ulterior base substitution at the nucleotide affected by the primary mutation leads to the reversal to a functional EGFP. Finally, we incorporated highly mutagenic abasic DNA lesions at the positions of primary mutations and demonstrated a high sensitivity of detection of the mutagenic DNA TLS and TM in this system.

Keywords: DNA damage; DNA damage tolerance; damage bypass; translesion synthesis (TLS); transcriptional mutagenesis; mutation assay; reporter assay; host cell reactivation (HCR); enhanced green fluorescent protein (EGFP)

1. Introduction

Damage to DNA is usually repaired in cells within minutes; however, repair efficiency can be limited by a number of factors (for instance, peculiar structure of DNA modifications, their accessibility to DNA repair proteins, oversaturation of the cellular repair capacity by excessive damage loads, or impairment of the relevant repair pathway by a genetic, epigenetic or pharmacological mechanism). Therefore, damage tolerance is essential for cell survival, particularly under genotoxic stress conditions [1,2]. One of the most important DNA damage tolerance mechanisms is the ability of DNA polymerases to bypass DNA damage during replication [3]. As replicative DNA polymerases have very low capacities of DNA synthesis past the damaged bases, dedicated translesion synthesis (TLS) enzymes are required for copying damaged DNA [4,5]. However, since TLS capacity is incompatible with proofreading exonuclease activity, all DNA polymerases specialized in damage bypass have elevated error rates. This makes TLS an important source of DNA damage-induced mutagenesis and the major cause of point mutations found in cancers [6,7].

As replication machinery needs to deal with a variety of structurally different DNA lesions, higher organisms possess multiple TLS enzymes with characteristic DNA damage bypass efficiency and fidelity profiles, as exemplified by biochemical data available for human Y-family DNA polymerases [8–12]. However, physiological roles in the bypass of defined types of DNA modifications and the mechanism of selection of a particular TLS polymerase have been explored in far less detail. The analyses of mutations induced by specific types of DNA damage in the cellular context are extremely challenging or impossible in organisms with large genomes. Most of the relevant knowledge has come from artificially constructed TLS templates, which were delivered to suitable host cells for accomplishment of DNA synthesis over the damaged nucleotide. Thus, vectors carrying defined synthetic DNA modifications were designed either for integration into chromosomal DNA [3] or, more commonly, for allowing extrachromosomal TLS and repair followed by shuttling to bacteria and sequencing of mutations in the survived fraction of transformation-competent vector DNA [13–16]. The efficiency of detection of both faithful and mutagenic TLS can be immensely enhanced by accommodating the DNA lesion within a reporter gene [17,18]; however, for the fidelity assessment, TLS reporters described in the literature also require shuttling into bacteria [19]. Thus, a reporter system for the direct detection of mutagenic TLS in mammalian cells would bring extensive benefits to TLS research.

Another damage tolerance mechanism of emerging importance is the bypass of DNA lesions by transcribing RNA polymerase complexes [20,21]. Mounting evidence strongly suggests that persistent transcription blockage by DNA damage is implicated in degenerative disease [20]. Additionally, transcriptional bypass may also play a role as a causal factor of disease if accompanied with a high error rate. Studies demonstrated that unrepaired miscoding lesions lead to accumulation of mutant transcripts in the process called "transcriptional mutagenesis" (TM) [22,23]. When critical cancer genes were affected, TM caused functional alterations in the cellular growth and DNA damage response pathways, even in the absence of genetic mutations [22,24]. Consequently, it was proposed that transient phenotypic fluctuations in the crucial genome stability pathways may increase the chance of mutation by stimulating premature replication and, thus, promote carcinogenesis [25,26].

Plasmid constructs carrying defined DNA lesions at specific nucleotide positions were instrumental for revealing the phenotypic outcomes of TM in yeast and mammalian cells by activation of either the encoded indicator proteins [27–29] or the functional components of intrinsic cellular signaling pathways [22,24]. We therefore hoped that an actionable reporter approach would be equally useful for characterization of mutations introduced by DNA TLS. We recently described a new enhanced green fluorescent protein (EGFP) c.613C>T point mutation, which leads to the synthesis of a non-fluorescent EGFP Q205* truncated protein, and showed that any subsequent nucleotide substitution at the affected position results in reversion to a fluorescent phenotype, thus offering a sensitive reporter for detection of TM [30]. Importantly, the mutated base pair in the EGFP Q205* gene is flanked by tandem Bpu10I sites retained from the original EGFP coding sequence. This allows for a highly efficient targeted incorporation of synthetic DNA modifications into either transcribed (TS) or non-transcribed (NTS) DNA strands using commercially available pair of the strand-specific Bpu10I-derived nicking endonucleases [31,32]. In this work, we report application of the EGFP Q205* reporter to sensitive detection of mutagenic DNA synthesis over a single apurinic/apyrimidinic (AP) DNA lesion, incorporated opposite to a gap in the TS. Furthermore, based on screening of further 24-point mutations, we selected and validated a set of EGFP mutants suitable for direct analyses of DNA and RNA synthesis errors caused by damage affecting different DNA bases.

2. Materials and Methods

2.1. Reporter Constructs for Detection of TM

For the sake of coherence between the EGFP amino acid and nucleotide sequences, we used here the open reading frame-based rather than the protein-based numbering, with the initiator methionine numbered as 1, the A of the ATG codon numbered as c.1 and the complementary T as ts.1. This corresponds

to the GenBank entries AAB08064 (protein) and U57609 (DNA). The pZAJ_5c vector encoding the functional EGFP protein [33] and the derived pEGFP_Q205* vector carrying the c.613C>T point mutation coding for the non-fluorescent truncated EGFP 1-204 protein, were described previously [30]. The procedure for site-specific incorporation of synthetic oligonucleotides into a single-stranded gap generated in the TS with the Nb.Bpu10I nicking endonuclease (Thermo Fisher Scientific Inc., St. Leon-Rot, Germany) was described previously [30]. Oligonucleotides 5′-TCAGGGCGGACT[THF]GGTGC-3′ and 5′-TCAGGGCGGACT[S-THF]GGTGC-3′ containing the synthetic AP lesion tetrahydrofuran with either phosphodiester (THF) or the nuclease-resistant phosphorothioate (S-THF) 5′-linkage were from BioSpring GmbH (Frankfurt am Main, Germany). The respective unmodified oligonucleotide 5′-TCAGGGCGGACTAGGTGC-3′ was from Eurofins Genomics (Ebersberg, Germany). The original pZAJ_5c vector accommodating the matched 5′-TCAGGGCGGACTGGGTGC-3′ synthetic oligonucleotide (Eurofins Genomics) was used as a reference for the original fluorescent EGFP. The quality of generated constructs carrying the specified modifications was controlled as described previously [30]. Representative results are shown in Supplementary Figure S1. Incorporation of AP lesions into the newly generated pEGFP_A207P vector was performed by the same procedure, but synthetic oligonucleotides were 5′-TCAGGG[THF/S-THF]GGACTGGGTGC-3′ (with AP lesions) and 5′-TCAGGGGGGACTGGGTGC-3′ (unmodified control).

2.2. Reporter Constructs for Detection of Mutagenic TLS Templated by the AP Lesion

To construct a reporter for detection of mutagenic TLS, the Nt.Bpu10I nicking endonuclease (Thermo Fisher Scientific Inc., St. Leon-Rot, Germany) was used to generate the 18-nt gap in the pEGFP_Q205* vector (note the opposite strand specificity of the gapping reaction with respect to the procedure described for TM above). The procedure was essentially as described previously for pZAJ vector [32], except that the sequence of competitor oligonucleotide used to deplete the excised DNA fragment was 5′-GGGCGGACTAGGTGCTCA-3′. The resulting gap in the NTS was used to accommodate synthetic oligonucleotide 5′-TGAGCACC[THF]AGTCCGCCC-3′ containing the THF AP lesion (BioSpring GmbH) or the respective 5′-TGAGCACCTAGTCCGCCC-3′ unmodified oligonucleotide (Eurofins Genomics). Next, the opposite DNA strand was nicked at two sites with Nb.Bpu10I (2 U/microgram vector DNA) and the excised stretch of the TS depleted by incubation with excess of the complementary oligonucleotide 5′-GCACCTAGTCCGCCCTGA-3′ to generate 18-nt gap opposite to the strand carrying the lesion (Supplementary Figure S1). Finally, gapped vector constructs were cleaned up using Amicon Ultra Ultracel 30 centrifugation devices (Millipore, Schwalbach am Taunus, Germany), as described previously [32]. The reference construct used to determine the gap repair efficiency in transfected cells was generated from the pZAJ_5c vector by the same procedure, but the synthetic oligonucleotide accomodated in the NTS was 5′-TGAGCACCCAGTCCGCCC-3′, which corresponds to the original nucleotide sequence of fully functional EGFP. Sequences of accessory oligonucleotides used for the strand depletion steps were the same as described previously [32].

For generation of the TLS reporter construct from the pEGFP_A207P vector, the procedure was the same as for pEGFP_Q205*, but in this case synthetic oligonucleotides were: 5′-GGGGGGACTGGGTGCTCA-3′ (depletion of the excised NTS fragment); 5′-TGAGCACCCAGTCC[THF]CCC-3′ (TLS template with AP lesion); 5′-TGAGCACCCAGTCCCCCC-3′ (unmodified control) and 5′-GCACCCAGTCCCCCCTGA-3′ (final depletion of the TS fragment).

2.3. Phenotypic Screening of the EGFP Mutants

The procedure for efficient integration of synthetic oligonucleotides with mismatched bases into circular DNA was described previously [34]. We started from the pZAJ_5c vector [33], which harbors the EGFP coding sequence from the pEGFP-C3 expression vector (GenBank accession number U57609) originally purchased from Clontech (Saint-Germain-en-Laye, France). Tandem nicks were generated in the TS with Nb.Bpu10I nicking endonuclease and the excised fragment was eliminated by incubation with a complementary competitor strand, as described previously [32]. The resulting 18-nt gap in

the TS was filled by seamless ligation of 26 different synthetic oligonucleotides (all purchased from Eurofins Genomics): the matching 5′-TCAGGGCGGACTGGGTGC-3′ 18-mer (which reconstituted the original EGFP sequence) and 18-mers containing 25 non-synonymous single nucleotide substitutions at the underlined positions (which created a set of hybrid constructs with original EGFP sequence in the NTS and mutated TS). Integrity of the generated constructs and correct incorporation of the synthetic strands were verified by inhibition of ligation in the aliquots incubated in the absence of polynucleotide kinase, as described previously [32]. The respective controls are shown in Supplementary Figure S3. The resulting constructs were transfected to DLD1 cells to analyze the EGFP fluorescence intensity.

2.4. Phenotypic Validation of Newly Identified EGFP Mutants

Vectors pEGFP_Q205P (containing the c.614A>C mutation), pEGFP_S206A (c.616T>G), pEGFP_S206Y (c.617C>A), and pEGFP_A207P (c.619G>C) were generated from the pZAJ_5c vector encoding the functional EGFP protein [33] by introducing the respective point mutations. Fluorescence intensities of mutant EGFP variants were measured in transfected HeLa cells, relative to the intensity of the original EGFP (measured in cells transfected with pZAJ_5c vector in parallel). As a non-fluorescent reference, cells were transfected with the pEGFP_Q205* vector encoding the truncated EGFP 1-204 variant described previously [30]. To assess phenotypes of secondary mutations affecting the selected nucleotides in the non-fluorescent mutants, the vectors pEGFP_Q205*, pEGFP_Q205P, pEGFP_S206Y, and pEGFP_A207P were nicked at the conserved tandem Nb.Bpu10I sites and the excised fragments of the TS were substituted in each mutant with synthetic oligonucleotides containing four different bases (N = G/A/C/T) at the primary mutation site. The procedure for strand substitution was the same as described under 2.3. The sequences of the oligonucleotides used for depletion of the excised TS fragment were 5′-GCACCTAGTCCGCCCTGA-3′ (in the case of pEGFP_Q205*); 5′-GCACCCCGTCCGCCCTGA-3′ (pEGFP_Q205P); 5′-GCACCCAGTACGCCCTGA-3′ (pEGFP_S206Y); and 5′-GCACCCAGTCCCCCCTGA-3′ (pEGFP_A207P). The sequences of the inserted oligonucleotides were 5′-TCAGGGCGGACTNGGTGC-3′ (in the case of pEGFP_Q205*); 5′-TCAGGGCGGACNGGGTGC-3′ (pEGFP_Q205P); 5′-TCAGGGCGNACTGGGTGC-3′ (pEGFP_S206Y); and 5′-TCAGGGNGGACTGGGTGC-3′ (pEGFP_A207P). Phenotypes of the secondary mutations were determined by transfection of the constructs into DLD1 cells.

2.5. Cell Culture, Transfection, and EGFP Expression Analysis

HeLa (human cervical carcinoma) and DLD1 (human colorectal carcinoma) cell lines were used as transfection hosts for EGFP expression analyses. Cells were propagated in DMEM high glucose medium supplemented with 10% fetal bovine serum. Twenty-four hours before transfection cells were seeded on 6-well plates (Nunc, Wiesbaden, Germany) at 300,000 HeLa cells/well and 400,000 DLD1 cells/well. Exponentially growing cells were transfected with the help of Effectene (Qiagen, Hilden, Germany). Equal amounts (400 ng) of the specified EGFP expression constructs and the pDsRed-Monomer-N1 vector (Clontech, Saint-Germain-en-Laye, France) were combined for transfections. Cells were fixed using 1% formaldehyde 24 h post-transfection and analyzed by flow cytometry, as described previously [35], using FACSCalibur[TM] and the CellQuest[TM] Pro software (Beckton Dickinson GmbH, Heidelberg, Germany). Untransfected cells were cut off by selective gating based on the DsRed expression prior to generation of the EGFP fluorescence (FL1-H) distribution plots and average EGFP signal per cell determined as the median of the distribution. For quantification of relative brightness of the EGFP mutants and for measurement of TM, the signal was normalized relative to the median fluorescence of the reference construct, coding for a fully functional EGFP, recorded in the same experiment. For evaluation of the TLS mutation rates, a template encoding the conventional EGFP was used to calculate the repair efficiency of the gap in the TS expressed as a ratio between the number of EGFP-positive cells and the total number of transfected cells. In parallel, the mutation rates were determined by the same approach in the mutant EGFP constructs containing the gap and either unmodified nucleotides or THF at the sites of primary mutations in the opposite DNA strand.

3. Results

3.1. Detection of Mutagenic Bypass of Abasic Site During RNA and DNA Synthesis Using EGFP Q205* Reporter

To generate a reporter for the direct detection of transcriptional mutagenesis (TM) and mutagenic DNA translesion synthesis (TLS) in human cells, we adopted the c.613C>T loss-of-function EGFP mutant encoding the truncated EGFP Q205* protein. Its DNA sequence allows for the strand-specific incorporation of oligonucleotides containing synthetic DNA lesions at defined positions (as explained in Materials and Methods and Supplementary Figure S1) and, especially important, any single nucleotide substitution at the position 613 of the coding sequence leads to an amino acid which reconstitutes the EGFP fluorescence [30]. We used these features to design a transfection-based assay for direct detection of the TM and TLS occurring in human cells (Figure 1a).

For the detection of TM, a potentially miscoding DNA lesion needs to be inserted on the place of dA in the transcribed DNA strand (TS). Thereby, any misincorporated nucleotide in the messenger RNA opposite to the lesion would result in reversion to a fluorescent EGFP phenotype (Figure 1a, left arm). To ensure that the phenotypic change is driven by a nucleotide change in RNA, and not by mutation of DNA sequence acquired during propagation of vector DNA in the host cells, we used a vector that does not replicate in HeLa cells [31]. We reasoned that the same vector would be adaptable also for detection of TLS errors, but in this case the lesion has to be accommodated on the place of dT in the non-transcribed DNA strand (NTS) and a fragment of the opposite DNA strand needs to be removed to generate a single-stranded gap. Thereby, production of full-length EGFP transcripts in the host cells is only possible if preceded by re-synthesis of the missing stretch of the TS, which requires the lesion bypass (Figure 1a, right arm). After accomplishment of the gap repair, only expression constructs that had gained (by mutagenic TLS) a nucleotide substitution at the lesion site would yield a fluorescent EGFP protein.

Since abasic sites are extremely common and a highly mutagenic type of DNA damage, we used the AP lesion as a model DNA lesion for investigation of the bypass fidelity in HeLa cells. To detect TM, we analyzed EGFP fluorescence in cells transfected with constructs accommodating two types of synthetic AP lesions at the nucleotide 613 of the EGFP Q205* gene (Figure 1b). We used a synthetic tetrahydrofuran (THF) AP lesion and also its counterpart flanked by a 5′ phosphorothioate bond (S-THF), which confers resistance to the key base excision repair (BER) enzyme AP endonuclease 1 (APE1), as reported previously [30,36]. As expected, EGFP signal in cells transfected with the THF construct was negligible, because this type of AP lesion undergoes a very efficient BER resulting in reconstitution of the A:T base pair. In contrast, BER-resistant AP lesion (S-THF) showed a large proportion of reversion to a fluorescent EGFP phenotype in comparison to dA. This result indicates that transcriptional bypass of S-THF takes place with high rate of ribonucleotide misincorporation, which erases the premature termination codon and thereby restores the protein fluorescence. Thus, transcriptional mutagenesis over an abasic site can be readily detected in fully repair-proficient cells using the EGFP Q205* reporter.
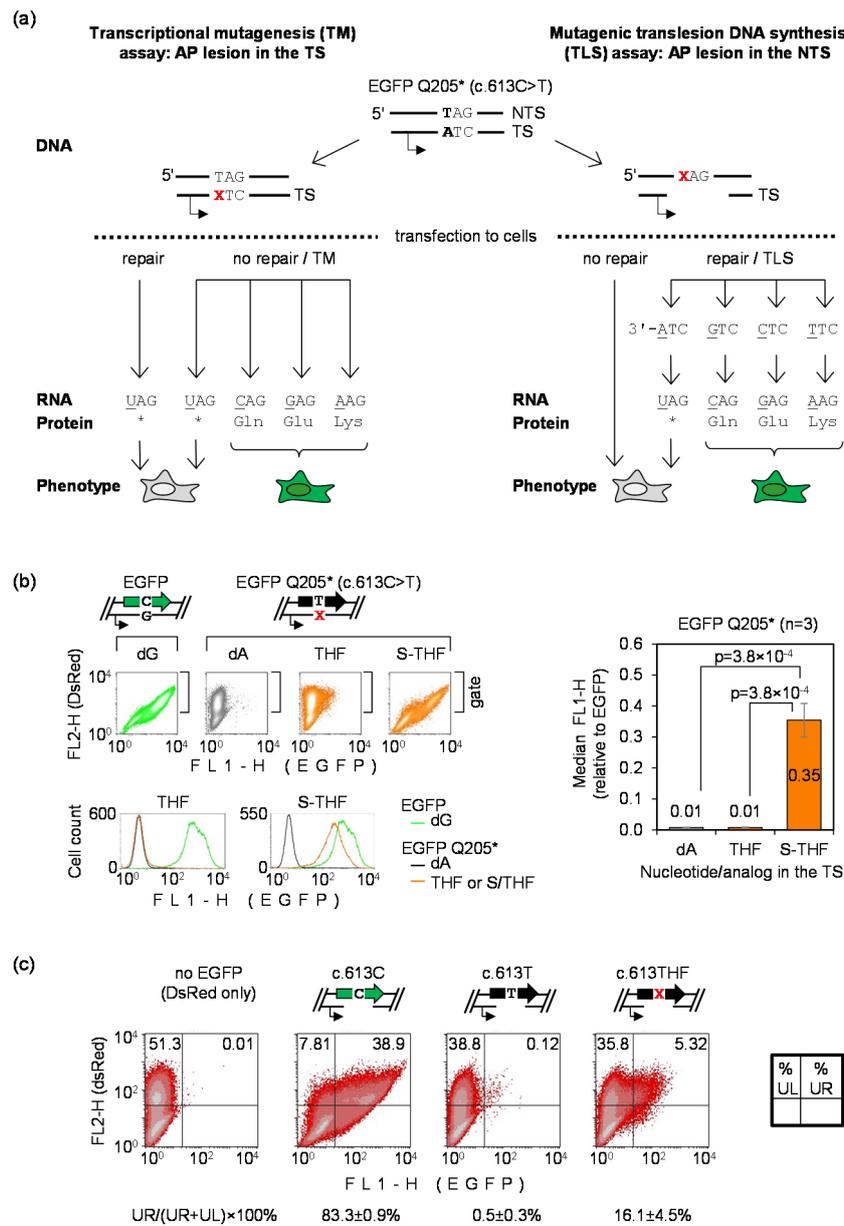
Figure 1. Transfection-based reporter assay for direct detection of transcriptional mutagenesis (TM) and mutagenic translesion DNA synthesis (TLS) templated by synthetic DNA lesions incorporated into vector DNA: (**a**) Scheme of the enhanced green fluorescent protein (EGFP) Q205* vector and of the derived constructs containing synthetic apurinic/apyrimidinic (AP) lesions (X) at the specified positions; (**b**) Detection of TM by flow cytometry of HeLa cells analyzed 24 h after transfection with constructs containing dA, THF and S-THF in double stranded DNA at the nucleotide 613 in TS. Scatter plots were gated by the DsRed fluorescence to generate the EGFP fluorescence distribution plots and calculate the median FL1-H. Grouped plots on the left show data of a representative experiment. Bar chart of the right shows quantification of the EGFP expression gain for *n* = 3 independent experiments (mean±SD); (**c**) Detection of mutagenic TLS. Cells transfected with constructs containing dT or THF at the nucleotide 613 in NTS opposite to a 18-nt gap were analyzed by flow cytometry 24 h post transfection. Mutant frequency was calculated as a ratio of EGFP-positive cells (in the upper right quadrant, UR) to total transfected cell count (UR+UL) for *n* = 3 independent experiments (mean ± SD).

To detect mutagenic DNA TLS, we substituted the nucleotide c.613T in the EGFP Q205* coding sequence for a synthetic abasic site (THF) and generated an 18-nucleotide gap in the transcribed DNA strand opposite to the lesion, as described under Materials and Methods (Section 2.2). For the

TLS analyses, it was not necessary to use the phosphorothioate-protected (S-THF) lesion, because human APE1 does not efficiently cleave AP sites in single-stranded DNA (Supplementary Figure S2). To monitor repair efficiency of the gap in the absence of damage in the single-stranded stretch of DNA, we generated and analyzed in parallel a wild-type EGFP construct containing an identical 18-nt gap in the TS. As judged by its expression, the gap repair efficiencies were heterogeneous across the cell population, with EGFP signal detected in approximately 83% of transfected cells (Figure 1c). We assumed that the repair efficiency of the gap must be the same in the EGFP Q205* construct encoding the non-fluorescent truncated protein. Nonetheless, the fraction of EGFP-positive cells in this case remained almost negligibly small (0.5 ± 0.3%), in accordance with the expectation of a very low rate of base substitutions spontaneously generated at the nucleotide 613 during re-synthesis of the missing DNA strand. Of note, the fraction of cells showing reversal to a fluorescent phenotype was strongly (by >30-fold) increased in the case of the c.613THF construct, as compared to c.613T, thus indicating that TLS over the THF lesion occurs with a high rate of nucleotide misincorporation. With these reporter constructs, we further were able to detect mutagenic TLS in several tumor as well as non-cancerous cell lines (data not shown and Supplementary Figure S2). Intriguingly, albeit the reversion to fluorescent EGFP was universally observed, the mutant frequencies showed substantial variability between non-isogenic cell lines, as demonstrated by a representative data (Supplementary Figure S2). This is likely to reflect not only differences in the TLS mechanism, but rather a variability of a broad range of factors, including DNA repair, recombination and the damage tolerance mechanisms.

In summary, the results shown in Figure 1 demonstrate usefulness of the EGFP Q205* reporter vector to efficiently detect both transcriptional mutagenesis and mutagenic translesion DNA synthesis in human cells.

3.2. Screening the EGFP Sequence for Candidate Single Nucleotide Substitutions Leading to the Loss of Fluorescence

Even though the EGFP Q205* reporter allows for a very sensitive detection of lesion-specific mutagenicity, its usefulness is limited to modifications affecting thymine (for detection of TLS) and adenine (for detection of TM). Moving towards a reporter system of a broader applicability, we sought other inactivating EGFP point mutations that would yield different types of base pairs in different NTS:TS orientations. Of note, the nucleotide sequence available for incorporation of synthetic DNA lesions, as confined by the available Bpu10I nicking sites, encodes a stretch of amino acids within the beta-strand 10. The target region lies in a close proximity to fluorophore in the folded EGFP, with at least two residues (T204 and S206) directly involved in functional molecular interaction with the fluorophore [37]. We therefore hoped that the likelihood of finding deleterious mutations by random mutagenesis of this region would be high. We compiled all single nucleotide substitutions possible within the sequence fragment flanked by the Bpu10I sites (Table 1). Of the 33 potential single nucleotide substitutions within the available 11-nucleotide window (four codons), seven did not confer an amino acid change and one was the c.613C>T nonsense transition mutation identified previously [30]. To rapidly examine whether any of the remaining 25 non-silent mutations would eliminate the EGFP fluorescence, we employed a simplified approach for the phenotypic screening of the candidate mutants. We took advantage of the Nb.Bpu10I nicking endonuclease to excise the TS fragment comprising the nucleotides 608–625 of the EGFP coding sequence in a vector encoding a functional EGFP protein and substituted it for synthetic DNA strands covering the whole spectrum of potentially useful mutations identified in Table 1, as described under Material and Methods (Section 2.3). In this way, we generated the set of hybrid EGFP constructs containing the innate NTS and various substituted nucleotides in the TS (Figure 2a). To prevent correction of the TS sequence by the DNA repair system in cells, we chose a mismatch repair-deficient DLD1 cell line as a suitable transfection host for the EGFP expression analyses.

Besides the 613C:A construct (which was expected to lose the fluorescence), seven other mismatches out of the 25 tested displayed more than a five-fold reduction of the EGFP fluorescence levels and thereby were considered as potential candidates for the loss-of-function mutations (Figure 2b). The repertoire of

identified base substitutions with a loss-of-function potential covered the whole spectrum of nucleotides at the positions affected by the mutations—four with C:G (NTS:TS), one more with T:A (in addition to c.613C>T mutant, which was already available), and one each with G:C and A:T base pairs. One of the candidates (ts.610A>T) unintentionally dropped out of the screen. However, further phenotypic characterization of this nucleotide substitution turned out to be no longer necessary for the following reasons: firstly, the equivalent T204S amino acid substitution (conferred by the 611C:C mismatch) showed a residual EGFP signal, and secondly, the target T:A (NTS:TS) nucleotide pair was already covered by the available c.613C>T mutant.

For subsequent validation of the identified loss-of-fluorescence mutants, we prioritized those in which subsequent random nucleotide substitutions at the position affected by the primary mutation were more likely to restore the fluorescent phenotype. This could be predicted from reasonably high signals displayed by the respective neighbors in the preliminary screen (Figure 2b). We grouped mutations based on the resulting base pairs and identified the potential top candidates within each of the four groups (highlighted in bold in Table 1).

Table 1. Analysis of the EGFP coding sequence fragment flanked by the available Bpu10I sites and the results of preliminary screening based on the expression of mismatch-containing constructs in DLD1 cells. Candidates selected for validation and the Q205* mutation characterized previously are highlighted in bold typeface.

Codon No. and Amino Acid.	Codon (5′→3′)	Nucleotide Position and the Nature of the Mismatch (NTS:TS)	Expected Amino Acid Change	% EGFP Expression
204 Thr	ACC	610A:C	T204A	20–75%
		610A:G	T204P	<20%
		610A:A	T204S	no data [1]
		611C:T	T204N	20–75%
		611C:C	T204S	20–75%
		611C:A	T204I	20–75%
		612C:T	=	not screened
		612C:C	=	not screened
		612C:A	=	not screened
205 Gln	CAG	613C:T	Q205K	>75%
		613C:C	Q205E	20–75%
		613C:A	**Q205 ***	**<20%**
		614A:C	Q205R	>75%
		614A:G	**Q205P**	**<20%**
		614A:A	Q205L	20–75%
		615G:T	=	not screened
		615G:G	Q205H	>75%
		615G:A	Q205H	>75%
206 Ser	UCC	616T:T̲	S206T	>75% [2]
		616T:C̲	**S206A**	**<20%**
		616T:G̲	S206P	<20%
		617C:T̲	**S206Y**	**<20%**
		617C:C̲	S206C	20–75%
		617C:A̲	S206F	<20%
		618C:T̲	=	not screened
		618C:C̲	=	not screened
		618C:A̲	=	not screened
207 Ala	GCC	619G:T̲	A207T	>75%
		619G:G̲	**A207P**	**<20%**
		619G:A̲	A207S	>75%
		620C:T̲	A207D	20–75%
		620C:C̲	A207G	20–75%
		620C:A̲	A207V	20–75%

[1] The same T204S amino acid change is tolerated in the 611C:C construct; [2] Recorded fluorescence higher than in the original EFGP construct; (*) Termination codon; (=) No amino acid change expected.
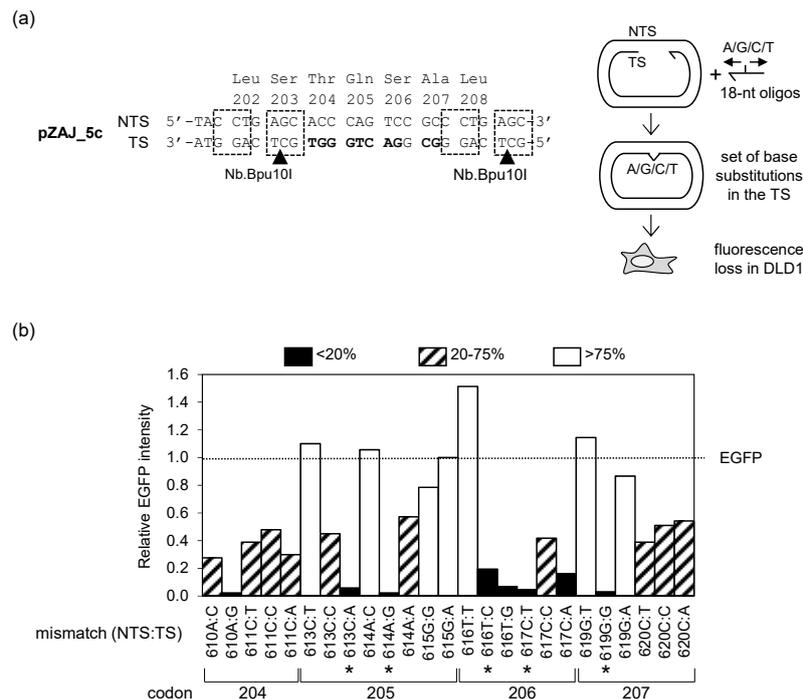
Figure 2. Phenotypic screening of mismatches with the indicated nucleotide substitutions in the TS to identify potential loss-of-fluorescence EGFP mutations: (**a**) Schematic representation of the screening strategy. The pZAJ_5c vector sequence fragment (on the left) shows tandem Nb.Bpu10I recognition sequences (boxes) and the nicking sites in the TS (arrowheads). Nucleotides targeted by single substitutions are marked bold; (**b**) Fluorescence intensities (relative to the original EGFP) measured in the DLD1 cells transfected with constructs containing the specified single-nucleotide mismatches (NTS:TS). Asterisks show candidate mutations selected for further validation.

3.3. Validation of EGFP Mutants Carrying Inactivating Single Nucleotide Substitutions

We cloned four selected EGFP mutants (Q205P, S206Y, S206A and A207P) to validate their phenotypes by comparing brightness of the resulting proteins with conventional EGFP (encoded within the original pZAJ_5c vector) and with the Q205* variant identified previously as non-fluorescent. Three of the four new mutants (Q205P, S206Y and A207P) displayed no EGFP fluorescence in transfected HeLa cells. The S206A variant showed residual fluorescence, albeit of diminished brightness (relative expression = 0.331) (Figure 3a). We thus reasoned that the S206A mutant would be of limited usefulness for the detection of mutagenic TM or TLS; however, all other mutants were considered potentially suitable for this purpose.

For the efficient detection of mutagenesis at a target nucleotide in the TM or TLS assays, it would be essential to verify that subsequent base substitutions at the affected position efficiently restore the EGFP fluorescence. This was expected based on the results of the preliminary screening (Figure 2); however, the former approach was potentially prone to false positives, because construction of mismatched templates for the screen then departed from the fluorescent EGFP version. To test all 16 sequence variants more rigorously, we used a single-strand substitution approach applied previously, but inverted the detection principle by starting from inherently non-fluorescent EGFP. In each of the identified EGFP loss-of-fluorescence mutants, we replaced the TS fragment comprising nucleotides 608–625 for different synthetic strands, containing each of the four nucleotides at the positions of the primary base substitutions in the respective mutants (as described in the Methods Section 2.4.). An analysis of the EGFP fluorescence in DLD1 cells showed that all nucleotide substitutions in the TS at the primary mutation sites led to the at least partial restoration of the EGFP fluorescence (Figure 3b), whereas strand replacements not leading to nucleotide substitutions (NTS:TS 613T:A in Q205*; 614C:G in Q205P; 617A:T in S206Y; and 619C:G in A207P) all retained the loss-of-fluorescence phenotype. Based on the almost complete regain of EGFP fluorescence in the Q205*, Q205P and A207P mutants (Figure 3b and Table 2), we unequivocally predict that any base substitution

mutation affecting the specified base pairs would be detectable with very high sensitivities in these mutants. In the case of the S206Y mutant, two of the secondary base substitutions (ts.617T>C and especially ts.617T>A) resulted in only partial recovery of the fluorescent signal, which would decrease detection sensitivity of these mutations proportionally. Still, even in the cases of base substitutions yielding the lowest levels of EGFP fluorescence, the signals clearly exceeded the EGFP S206Y background (by the factors of >40 and >6, for ts.617T>C and ts.617T>A, respectively). With these minor limitations, we conclude that all four newly generated EGFP mutants would be suitable as reporters for the direct detection of erroneous bypass of DNA lesions during transcription and DNA synthesis. The range of base pairs covered by the available mutants is applicable to analyses of mutagenic TLS at DNA damage affecting A, C or T in the coding DNA strand (Table 2). Conversely, TM can be analyzed for DNA lesions affecting A, G or T in the transcribed strand.
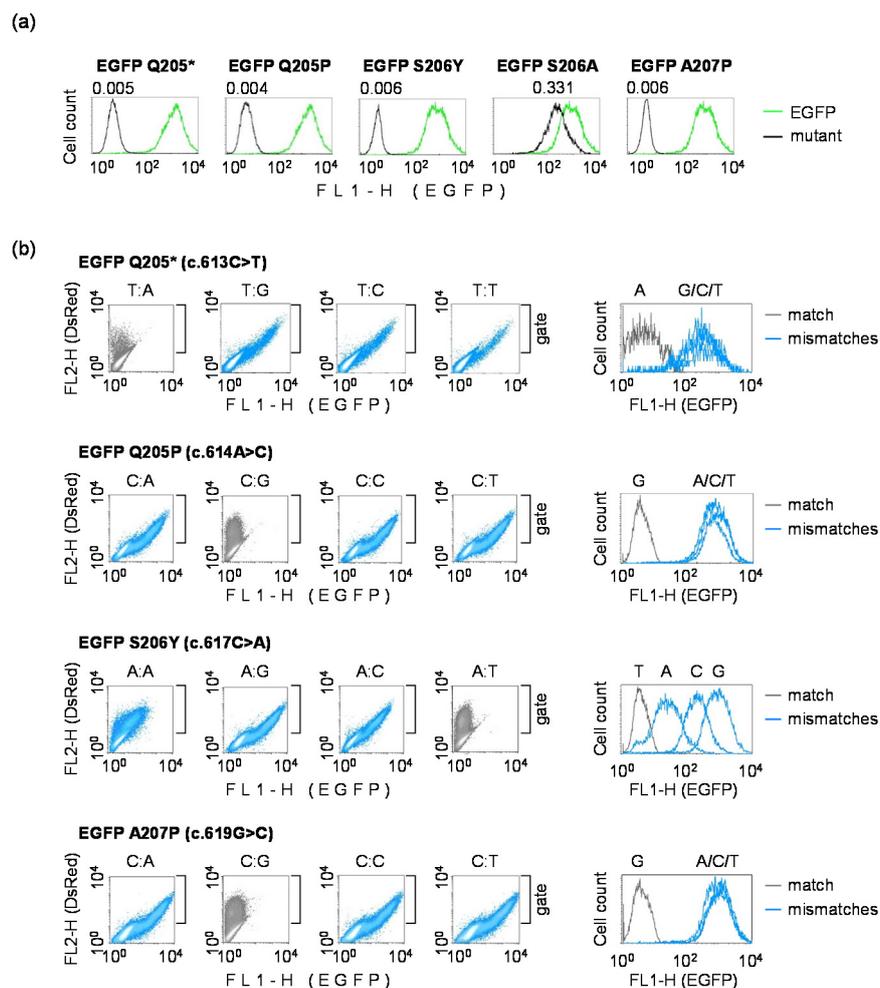
Figure 3. Validation of the candidate EGFP loss-of-function variants containing different nucleotide pairs at the mutation site: (**a**) Quantitative flow cytometry analyses of cloned EGFP mutants selected based on the low fluorescence signal in the preliminary screen. HeLa cells were transfected with vectors encoding the specified EGFP mutants or the original EGFP, in the presence of DsRed-Monomer as a transfection marker to generate the overlaid fluorescence distribution plots, as explained in Figure 1. Values above the peaks report fluorescence intensities relative to the original EGFP (calculated from the median FL1-H); (**b**) Reversal to the fluorescent phenotype by targeted single nucleotide substitutions in the TS. Expression vectors for the indicated EGFP mutants were used to generate all possible mismatches (blue color) or a correct base pair (grey color) by ligation of the respective synthetic oligonucleotides. EGFP fluorescence was analyzed in transfected DLD1 cells.

Table 2. Summary of the loss-of-fluorescence EGFP mutants convertible to fluorescent protein species by the indicated single nucleotide substitutions (underlined) and EGFP phenotypes of the analyzed secondary mutations in the TS.

Non-Fluorescent Mutants			Detectable Secondary Mutations			
Protein	DNA	Base Pair (NTS:TS)	DNA TS 3′→5′ (TLS)	RNA 5′→3′ (TM)	Amino Acid Change [1]	Brightness (% EGFP) [2]
Q205 *	c.613C>T	T:A	GTC	CAG	*205Q	100
			CTC	GAG	*205E	~70
			TTC	AAG	*205K	~100
Q205P	c.614A>C	C:G	GTC	CAG	P205Q	100
			GAC	CUG	P205L	~80
			GCC	CGG	P205R	~100
S206Y	c.617C>A	A:T	AGG	UCC	Y206S	100
			AAG	UUC	Y206F	~3
			ACG	UGC	Y206C	~20
A207P	c.619G>C	C:G	CGG	GCC	P207A	100
			AGG	UCC	P207S	~100
			TGG	ACC	P207T	~100

[1] Mutations restoring the original EGFP amino acid sequence are listed for each mutant in the first row; [2] Estimates based on the expression of constructs in which the TS fragment was replaced by synthetic DNA oligonucleotides containing the indicated mismatched nucleotides (underlined in the DNA sequence). For each mutant, signals of constructs with restored TS sequence of the wild-type EGFP were set as 100% and relative brightness of the remaining two mismatched constructs calculated accordingly. The estimates are based on results of parallel transfections performed in duplicates. The relative brightness values of constructs containing matched nucleotides (encoding mutant EGFP variants) were negligibly low in all cases (≤0.5%); (*) Termination codon.

3.4. Detection of TM Using the A207P (c.619G>C) EGFP Mutant

To prove that TM can be sensitively detected at a nucleotide conferring a missense EGFP mutation, we further assessed the outcome of the highly miscoding S-THF AP lesion in the newly obtained EGFP A207P (c.619G>C) mutant. We incorporated synthetic THF and S-THF AP lesions at the nucleotide 619 in the TS (Figure 4a,b) and measured the resulting EGFP signal in transfected HeLa cells. As in the case of the Q205* (c.613C>T) mutant analyzed previously (Figure 1), unrepairable S-THF lesion caused very strong gain of the EGFP fluorescence as a result of TM (Figure 4c). The results thereby confirm extremely high miscoding potential of AP lesions, as previously observed in the Q205* (c.613C>T) reporter (Figure 1b). In both reporter systems, we observed recovery of the fluorescence intensity from zero to the level of 33–35% of the signal produced by fully functional EGFP protein, thus indicating that at least one third of all transcripts contained nucleotide substitutions at the lesion site. Since AP lesions undergo nucleotide excision repair (NER) in human cells, TM rates are even higher in NER deficient cells ([30] and unpublished results). As a demonstration of the extraordinary efficiency of repair of AP lesions in cells, it is important to note that the TM signal was fully eliminated when cells were transfected with the construct containing the BER-sensitive THF lesion. This is documented by the direct comparison with the control construct containing dG at the analyzed position (Figure 4c). In summary, the new reporter does not only allow for the extremely sensitive detection of TM, but it also offers a tool for the assessment of the repair capacity of cells with a very broad (at least 30-fold) dynamic range of detection.
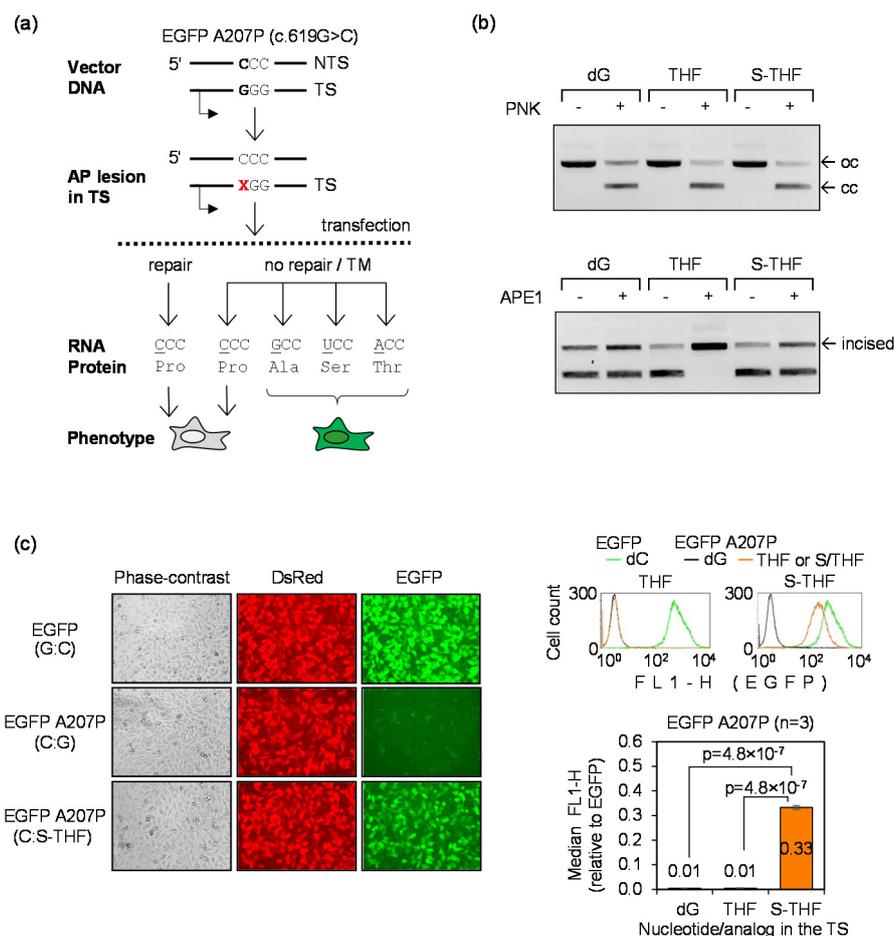
Figure 4. Detection of TM by AP lesion in the EGFP A207P (c.619G>C) reporter: (**a**) Scheme of the EGFP A207P reporter indicating position of the lesion (X) at the nucleotide 619 of the TS and possible outcomes of the transcriptional bypass; (**b**) Verification of incorporation of the lesions into the vector DNA. Upper gel: proof of incorporation of synthetic DNA strands containing dG or the indicated AP lesions (THF, S-THF) into the TS. The vector containing the single-stranded 18-nt gap in the TS migrates as open circular form (oc). Efficient ligation of synthetic oligonucleotides into the gap is confirmed by reconstitution of the covalently closed circular form (cc). Lower gel: verification of incision of the THF lesion by human APE1 and inhibited incision of the S-THF; (**c**) TM visualized by fluorescence microscopy and quantitative flow cytometry as a reversal of the non-fluorescent EGFP phenotype. HeLa cells were transfected with the indicated EGFP constructs in the presence of DsRed-Monomer as a transfection marker.

3.5. Detection of Mutagenic TLS Using the A207P (c.619G>C) EGFP Mutant

The AP site is generally considered a "non-instructive" lesion and specificity of the deoxyribonucleotide incorporation during DNA synthesis over AP lesions in mammalian cells is a matter of controversy [9,14,38,39]. Some results indicated that different nucleotides are incorporated opposite to AP lesions in the template strand in a rather random fashion [38], whereas most reports suggested preferential (up to 85%) incorporation of dA, similar to the "A-rule" previously established for the SOS-induced replication in bacteria [40]. In the present study, we observed a pronounced regain of the EGFP signal by TLS of AP lesion in the EGFP Q205* reporter, as compared to the template containing thymine in the probed position (Figure 1c). This clearly indicated that dA is not the only nucleotide incorporated opposite to the AP lesion during repair synthesis over the 18-nucleotide gap, because adenine at the position 613 in the TS would restore the UAG translation termination codon. Therefore, to test whether the nucleotide incorporation opposite to AP lesions is random in human cells, it was instructive to analyze the mutation rate in a different reporter. We used the EGFP

A207P mutant to generate gapped constructs containing either dC or THF at the nucleotide position 619 and analyzed the rates of reversion to the fluorescent EGFP in transfected HeLa cells, using a procedure analogous to the one used formerly for the EGFP Q205* reporter. As expected, the repair of the gap only occasionally led to reactivation of the EGFP fluorescence in the absence of the lesion (0.11 ± 0.02% cells). However, the fraction of EGFP-positive cells rose to 57.3 ± 2.8% (>500-fold) when dC in the NTS was substituted by the THF lesion (Figure 5). Thereby, the EGFP reactivation rate by AP site in the reporter that detects incorporation of any nucleotide except dG is substantially higher than in the EGFP Q205*, which is reactivated by incorporation of any nucleotide but dA (16.1 ± 4.4%, with an approximately 30-fold increase over the dT template, as shown in Figure 1). The result thus indicates that insertion of nucleotides opposite to AP lesion during DNA repair synthesis takes place in a non-random fashion and that incorporation of dA is strongly favored over dG human cells. Importantly, in the absence of the lesion, the repair rates of the gap were remarkably similar between these experiments (83.3 ± 0.9% and 77.8 ± 6.6%, as calculated based on the construct coding for the fully functional EGFP transfected in parallel). Thus, different EGFP reactivation rates between the A207P and Q205* mutants should be entirely attributed to differential incorporation of dA and dG at the damage site.
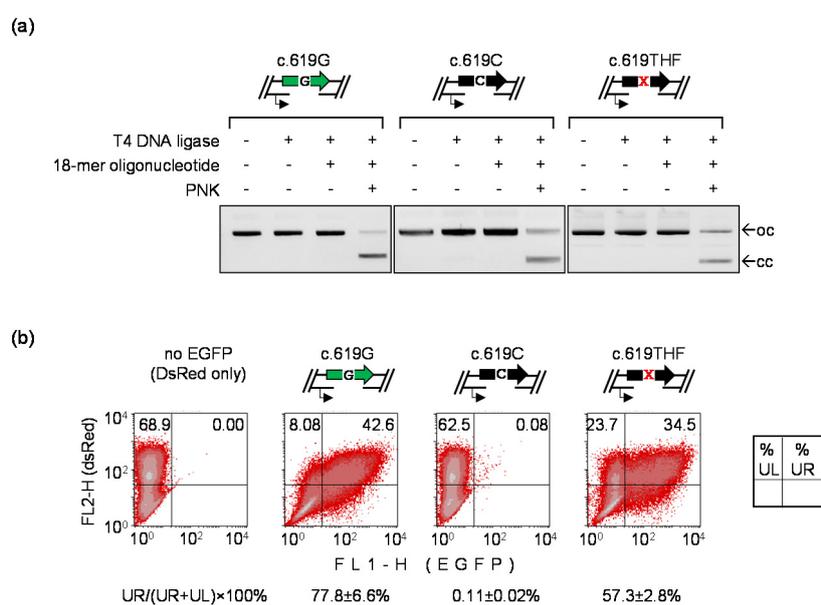

Figure 5. Detection of mutagenic TLS over the synthetic AP lesion in the EGFP A207P (c.619G>C) reporter by reversion to the fluorescent phenotype: (**a**) verification of the presence of gaps in the TS opposite to the specified nucleotides in the TLS reporter constructs by inhibited ligation and its reconstitution with matching synthetic 18-mer oligonucleotides. THF lesion at the nucleotide 619 is marked with "X"; (**b**) Repair of the 18-nt gap by re-synthesis of the missing TS fragment in the construct encoding the functional EGFP (c.619G) and detection of mutagenic TLS on the EGFP A207P template containing no modification (c.619C) or AP lesion (c.619THF) at the nucleotide 619. HeLa cells were transfected with the indicated constructs together with the DsRed-Monomer expression vector as a transfection marker. The fractions of EGFP-positive cells (UR/(UR+UL)×100%) were used as an estimate of the repair and mutation rates, as in Figure 1 (*n* = 2 independent experiments, mean ± range).

4. Discussion

The erroneous bypass of damaged nucleotides during DNA replication is a major source of mutations [6,7]. Some of the mutation signatures found in cancers have been attributed to environmental mutagens. However, as multi-layered resistance and repair mechanisms within the cell (comprising metabolism, DNA repair and DNA damage response) efficiently protect against experimental mutagenesis, many of the annotated mutation types are difficult to trace back to exposure to a particular carcinogen [41–43]. Along with the whole-genome sequencing approaches, the reporter-based strategies

offer promising potential in addressing this fundamental problem [44]. An advantage of exogenously delivered vectors carrying specific types of DNA lesions is their suitability to investigate the outcomes of damage in DNA while avoiding cell exposure to deleterious doses of the damaging agent. This is especially valuable for the lesions that cannot be induced in the chromosomal DNA with high specificity and activate cytotoxic responses if they are present in the genome in significant amounts [45]. Methods have been developed for efficient incorporation of defined DNA lesions into vectors, which can be propagated in cells extrachromosomally [14,38,46–49]; and strategies for integration of DNA modified at precise sites into genomes are emerging [3,50,51]. In conjunction with ever growing availability of synthetic DNA modifications, replication- and transcription-competent extrachromosomal vectors, suitable for incorporation of structurally defined DNA modifications at defined positions, thereby open possibilities for investigation of miscoding properties of virtually any type of DNA lesion.

A straightforward and precise way for the targeted incorporation of synthetic DNA lesions into plasmid-based vectors is ligation of short synthetic DNA strands into gapped circular DNA, which can be conveniently generated using sequence-specific nicking endonucleases [31,52–56]. The use of nicking endonucleases for introducing modifications can be problematic in the protein coding regions of reporter genes, because conserved amino acid sequence restrains the possibilities for accommodation of the required recognition sites. Excitingly, however, we previously found out that the conventional EGFP coding sequence contained pre-existing tandem sites for the pair of nicking endonucleases Nb.Bpu10I and Nt.Bpu10I [32]. We realized that the opposite strand specificities of the available nickases provide an immediate advantage by allowing deliberate modification of either DNA strand. Consequently, the same reversible mutant containing an inactivating base substitution between the Bpu10I sites can be used as a reporter for either erroneous DNA TLS or TM simply by switching the nicking enzymes targeting either the NTS or TS, respectively. Moreover, the availability of nicking enzymes with opposite strand specificities allows precise and efficient depletion of TS following the insertion of a mutagenic lesion into the NTS, which is a superior way to generate a TLS substrate containing a gap in the DNA strand opposite to the lesion (Figure 1a and Supplementary Figure S1).

The protein region encoded by nucleotide sequence confined by the existing Bpu10I nicking sites within the EGFP gene is important for efficient excited-state proton transfer in the immediate fluorophore environment and for correct structural organization of the folded protein [37,57–59]. Not surprisingly, we found that most (17 of 23) of the amino acid substitutions tested in the fluorescence screen had negative effects on the EFGP brightness; however, only six led to a complete loss of fluorescence (Figure 2 and not-shown data). Thus, a complete EGFP inactivation was only observed for mutations leading to a proline residue invariantly breaking the beta-sheet structure (T204P, Q205P, S206P, and A207P), a termination codon (Q205*) or a bulky amino acid residue on the fluorophore side of the beta-strand (S206Y). Consequently, we did not find any fully inactive EGFP mutant containing guanine at a nondegenerate codon site, which would have a potential to lead to an amino acid with changed properties upon a random secondary mutation. Thus, the S206A (c.616T>G) mutant retained approximately one third of the original EGFP brightness (Figure 3a) and was reversible to the wild-type phenotype by only one of the three possible base substitutions (data not shown). Nonetheless, with the exception of guanine, all other bases were covered by the identified set of mutants to meet the requirements for both the complete loss of EGFP fluorescence and the capacity of its restoration by any kind of subsequent base substitution (Figure 3). The newly identified mutants thereby provide a set of reporters for the detection of DNA TLS errors at DNA lesions derived from A, C or T. The same reporter vectors (but with a different procedure) are useful for analyses of TM induced by lesions affecting T, G or A, respectively.

To test the sensitivity of the new reporters, we generated constructs containing synthetic AP lesions at the positions deemed suitable for detection of erroneous DNA TLS and TM. Although it has long been known that the bypass of AP lesions frequently results in mutations [60], it is controversial whether AP sites are non-instructive or rather miscoding during DNA synthesis in mammalian cells. Thus, shuttle vectors recovered from the monkey COS-7 cells indicated a rather random pattern of nucleotide misincorporation opposite to the AP lesion, a slight bias against dG [38], or a preferential—but not

strictly specific—incorporation of either dA or dT [39]. In contrast, studies conducted by different groups in human cells revealed clear preferences to dA followed by dT and dC [14,61]. Remarkably, a rapid quantitative estimate based on very different EGFP reversal rates documented in the Q205* and A207P mutants in our present study (compare Figures 1c and 5) unmistakably indicated that incorporation of dA is strongly favored over dG in human cells. Moreover, since the fluorescence recovery rate in the A207P mutant approaches the maximum repair rate set by the conventional EGFP construct containing a single-stranded gap of the same size, we infer that dA is incorporated in >70% cases, which is coherent with results obtained by others in human cells using the shuttle vector sequencing techniques [14,61]. It is important to take into account that none of human Y-family DNA polymerases have a preference for dAMP when inserting a nucleotide opposite an AP site [9]. Considering that of the B-family polymerases, only pol α and pol δ can bypass AP site with relatively high efficiencies and selectivity to dAMP [9,62,63], the results imply that a great portion of the overall bypass capacity in human cells is attributable to pol δ and a much smaller part of it to the dedicated TLS polymerase(s).

As in the case of erroneous DNA TLS, TM was also detected with excellent sensitivity using the gain-of-function reporter principle. Because AP sites in cells are very efficiently processed by the BER pathway, which would prevent TM, we had to use a structurally analogous synthetic lesion with a non-cleavable linkage to prevent the excision. We previously showed that the erroneous bypass of such a BER-resistant AP lesion during transcription can be detected using the EGFP Q205* (c.613C>T) reporter with a signal strength at the level of 10–80% of the wild-type EGFP reporter, depending on the status of NER as the secondary repair pathway in human cells [30]. Here, we tested the Q205* and A207P mutants, which yield apurinic lesions (in contrast to physiologically less relevant apyrimidinic sites), in HeLa cells. Considering that the transcriptional bypass of AP lesions in human cells occurs predominantly with incorporation of A [64], the resulting base substitutions should be detectable in both Q205* and A207P reporters (Figure 3). Despite the NER proficiency of the host cell line, we detected a signal at the level of >30% wild-type EGFP in both reporter constructs (Figures 1b and 4c), which indicates extremely high detection sensitivity of TM.

5. Conclusions

In summary, by using AP lesions as a model, we demonstrated that nucleotide substitutions, which arise from the erroneous bypass of DNA lesions during either DNA or RNA synthesis, can be efficiently detected in cells using the EGFP Q205* (c.613C>T) mutant as a reporter with a positive readout. The EGFP Q205* reporter is sensitive to TLS errors induced by modifications of thymine in the NTS and to TM caused by modifications of adenine in the TS. In order to create reporters tailored to detection of mutagenic TLS and TM induced by modifications affecting other bases in DNA, we further designed a set of appropriate loss-of-function mutations in the nucleotide sequence in the EGFP region suited for seamless incorporation of custom synthetic DNA modifications into either DNA strand. As a proof of performance of the generated reporters, we analyzed the fidelities of the bypass of mutagenic AP lesions incorporated at the mutation sites. The obtained results provided reliable quantitative estimates of the error rates induced by AP lesions during TLS and TM as well as quick clues about specificities of the arising nucleotide substitutions. Considering the broad use of EGFP as a readily detectable fluorescent probe, and taking into account the ease, robustness and versatility of the experimental procedures described here, we hope that our tools will be useful to the community for the characterization of miscoding properties of defined types of DNA modifications and the elucidation of the specific damage tolerance mechanisms of the cell.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/6/902/s1, Figure S1: Site-specific incorporation of AP lesions (THF or S-THF) into the pEGFP_Q205* to generate reporter constructs for detection of TM and TLS, Figure S2: Resistance of THF in the TLS template to APE1 and mutagenic TLS of synthetic AP lesions (THF and S-THF) in HeLa and MRC5 cell lines, Figure S3: Generation of EGFP expression constructs containing a set of 26 single nucleotide substitutions in the TS.

References

1. Friedberg, E.C. Suffering in silence: The tolerance of DNA damage. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 943–953. [CrossRef] [PubMed]

2. Chatterjee, N.; Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58*, 235–263. [CrossRef] [PubMed]

3. Izhar, L.; Ziv, O.; Cohen, I.S.; Geacintov, N.E.; Livneh, Z. Genomic assay reveals tolerance of DNA damage by both translesion DNA synthesis and homology-dependent repair in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1462–1469. [CrossRef] [PubMed]

4. Lange, S.S.; Takata, K.I.; Wood, R.D. DNA polymerases and cancer. *Nat. Rev. Cancer* **2011**, *11*, 96–110. [CrossRef]

5. Sale, J.E.; Lehmann, A.R.; Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 141–152. [CrossRef]

6. Bielas, J.H.; Loeb, K.R.; Rubin, B.P.; True, L.D.; Loeb, L.A. Human cancers express a mutator phenotype. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18238–18242. [CrossRef]

7. Hoeijmakers, J.H.J. Genome maintenance mechanisms for preventing cancer. *Nature* **2001**, *411*, 366–374. [CrossRef]

8. Maga, G.; Villani, G.; Crespan, E.; Wimmer, U.; Ferrari, E.; Bertocci, B.; Hübscher, U. 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature* **2007**, *447*, 606–608. [CrossRef]

9. Choi, J.Y.; Lim, S.; Kim, E.J.; Jo, A.; Guengerich, F.P. Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α, δ, η, ι, κ, and REV1. *J. Mol. Biol.* **2010**, *404*, 34–44. [CrossRef]

10. Shibutani, S.; Suzuki, N.; Grollman, A.P. Mechanism of frameshift (deletion) generated by acetylaminofluorene-derived DNA adducts in vitro. *Biochemistry* **2004**, *43*, 15929–15935. [CrossRef]

11. Schorr, S.; Schneider, S.; Lammens, K.; Hopfner, K.P.; Carell, T. Mechanism of replication blocking and bypass of Y-family polymerase η by bulky acetylaminofluorene DNA adducts. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20720–20725. [CrossRef] [PubMed]

12. Sherrer, S.M.; Taggart, D.J.; Pack, L.R.; Malik, C.K.; Basu, A.K.; Suo, Z. Quantitative analysis of the mutagenic potential of 1-aminopyrene-DNA adduct bypass catalyzed by Y-family DNA polymerases. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* **2012**, *737*, 25–33. [CrossRef] [PubMed]

13. Yasui, M.; Dong, H.; Bonala, R.R.; Suzuki, N.; Ohmori, H.; Hanaoka, F.; Johnson, F.; Grollman, A.P.; Shibutani, S. Mutagenic properties of 3-(deoxyguanosin-N^2-yl)-2-acetylaminofluorene, a persistent acetylaminofluorene-derived DNA adduct in mammalian cells. *Biochemistry* **2004**, *43*, 15005–15013. [CrossRef]

14. Avkin, S.; Adar, S.; Blander, G.; Livneh, Z. Quantitative measurement of translesion replication in human cells: Evidence for bypass of abasic sites by a replicative DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3764–3769. [CrossRef] [PubMed]

15. Hashimoto, K.; Bonala, R.; Johnson, F.; Grollman, A.P.; Moriya, M. Y-family DNA polymerase-independent gap-filling translesion synthesis across aristolochic acid-derived adenine adducts in mouse cells. *DNA Rep.* **2016**, *46*, 55–60. [CrossRef]

16. Basu, A.K.; Pande, P.; Bose, A. Translesion synthesis of 2-deoxyguanosine lesions by eukaryotic DNA polymerases. *Chem. Res. Toxicol.* **2017**, *30*, 61–72. [CrossRef]

17. Ziv, O.; Zeisel, A.; Mirlas-Neisberg, N.; Swain, U.; Nevo, R.; Ben-Chetrit, N.; Martelli, M.P.; Rossi, R.; Schiesser, S.; Canman, C.E.; et al. Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin. *Nat. Commun.* **2014**, *5*, 5437. [CrossRef]

18. Veaute, X.; Fuchs, R.P.P. Greater susceptibility to mutations in lagging strand of DNA replication in Escherichia coli than in leading strand. *Science* **1993**, *261*, 598–600. [CrossRef]

19. Yagi, T.; Fujikawa, Y.; Sawai, T.; Takamura-Enya, T.; Ito-Harashima, S.; Kawanishi, M. Error-prone and error-free translesion DNA synthesis over site-specifically created DNA adducts of aryl hydrocarbons (3-nitrobenzanthrone and 4-aminobiphenyl). *Toxicol. Res.* **2017**, *33*, 265–272. [CrossRef]

20. Lans, H.; Hoeijmakers, J.H.J.; Vermeulen, W.; Marteijn, J.A. The DNA damage response to transcription stress. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 766–784. [CrossRef]

21. Brégeon, D.; Doetsch, P.W. Transcriptional mutagenesis: Causes and involvement in tumour development. *Nat. Rev. Cancer* **2011**, *11*, 218–227. [CrossRef] [PubMed]

22. Saxowsky, T.T.; Meadows, K.L.; Klungland, A.; Doetsch, P.W. 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18877–18882. [CrossRef] [PubMed]

23. Brégeon, D.; Peignon, P.A.; Sarasin, A. Transcriptional mutagenesis induced by 8-Oxoguanine in mammalian cells. *PLoS Genet.* **2009**, *5*, e1000577. [CrossRef] [PubMed]

24. Ezerskyte, M.; Paredes, J.A.; Malvezzi, S.; Burns, J.A.; Margison, G.P.; Olsson, M.; Scicchitano, D.A.; Dreij, K. O6-methylguanine–induced transcriptional mutagenesis reduces p53 tumor-suppressor function. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4731–4736. [CrossRef] [PubMed]

25. Doetsch, P.W. Translesion synthesis by RNA polymerases: Occurrence and biological implications for transcriptional mutagenesis. *Mutat. Res.* **2002**, *510*, 131–140. [CrossRef]

26. Holmquist, G.P. Cell-selfish modes of evolution and mutations directed after transcriptional bypass. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2002**, *510*, 141–152. [CrossRef]

27. Shaw, R.J.; Bonawitz, N.D.; Reines, D. Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. *J. Biol. Chem.* **2002**, *277*, 24420–24426. [CrossRef]

28. Burns, J.A.; Dreij, K.; Cartularo, L.; Scicchitano, D.A. O^6-Methylguanine induces altered proteins at the level of transcription in human cells. *Nucleic Acids Res.* **2010**, *38*, 8178–8187. [CrossRef]

29. Nagel, Z.D.; Margulies, C.M.; Chaim, I.A.; McRee, S.K.; Mazzucato, P.; Ahmad, A.; Abo, R.P.; Butty, V.L.; Forget, A.L.; Samson, L.D. Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1823–E1832. [CrossRef]

30. Kitsera, N.; Rodriguez-Alvarez, M.; Emmert, S.; Carell, T.; Khobta, A. Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res.* **2019**, *47*, 8537–8547. [CrossRef]

31. Kitsera, N.; Stathis, D.; Lühnsdorf, B.; Müller, H.; Carell, T.; Epe, B.; Khobta, A. 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.* **2011**, *39*, 5926–5934. [CrossRef] [PubMed]

32. Lühnsdorf, B.; Kitsera, N.; Warken, D.; Lingg, T.; Epe, B.; Khobta, A. Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal. Biochem.* **2012**, *425*, 47–53. [CrossRef] [PubMed]

33. Allgayer, J.; Kitsera, N.; Von Der Lippen, C.; Epe, B.; Khobta, A. Modulation of base excision repair of 8-oxoguanine by the nucleotide sequence. *Nucleic Acids Res.* **2013**, *41*, 8559–8571. [CrossRef] [PubMed]

34. Lühnsdorf, B.; Epe, B.; Khobta, A. Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.* **2014**, *289*, 22008–22018. [CrossRef] [PubMed]

35. Khobta, A.; Anderhub, S.; Kitsera, N.; Epe, B. Gene silencing induced by oxidative DNA base damage: Association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res.* **2010**, *38*, 4285–4295. [CrossRef]

36. Allgayer, J.; Kitsera, N.; Bartelt, S.; Epe, B.; Khobta, A. Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Res.* **2016**, *44*, 7267–7280. [CrossRef]

37. Brejc, K.; Sixma, T.K.; Kitts, P.A.; Kain, S.R.; Tsien, R.Y.; Ormö, M.; Remington, S.J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2306–2311. [CrossRef]

38. Gentil, A.; Cabral-Neto, J.B.; Mariage-Samson, R.; Margot, A.; Imbach, J.L.; Rayner, B.; Sarasin, A. Mutagenicity of a unique apurinic/apyrimidinic site in mammalian cells. *J. Mol. Biol.* **1992**, *227*, 981–984. [CrossRef]

39. Takeshita, M.; Eisenberg, W. Mechanism of mutation on DNA templates containing synthetic abasic sites: Study with a double strand vector. *Nucleic Acids Res.* **1994**, *22*, 1897–1902. [CrossRef]

40. Lawrence, C.W.; Borden, A.; Banerjee, S.K.; Leclerc, J.E. Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res.* **1990**, *18*, 2153–2157. [CrossRef]

41. Kucab, J.E.; Zou, X.; Morganella, S.; Joel, M.; Nanda, A.S.; Nagy, E.; Gomez, C.; Degasperi, A.; Harris, R.; Jackson, S.P.; et al. A Compendium of Mutational Signatures of Environmental Agents. *Cell* **2019**, *177*, 821–836. [CrossRef] [PubMed]

42. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef] [PubMed]

43. Hollstein, M.; Alexandrov, L.B.; Wild, C.P.; Ardin, M.; Zavadil, J. Base changes in tumour DNA have the power to reveal the causes and evolution of cancer. *Oncogene* **2017**, *36*, 158–167. [CrossRef] [PubMed]

44. Birnbaum, M.D.; Nemzow, L.; Kumar, A.; Gong, F.; Zhang, F. A Rapid and Precise Mutation-Activated Fluorescence Reporter for Analyzing Acute Mutagenesis Frequency. *Cell Chem. Biol.* **2018**, *25*, 1038–1049. [CrossRef]

45. Fedeles, B.I.; Essigmann, J.M. Impact of DNA lesion repair, replication and formation on the mutational spectra of environmental carcinogens: Aflatoxin B 1 as a case study. *DNA Repair* **2018**, *71*, 12–22. [CrossRef]

46. Koehl, P.; Burnouf, D.; Fuchs, R.P.P. Construction of plasmids containing a unique acetylaminofluorene adduct located within a mutation hot spot. A new probe for frameshift mutagenesis. *J. Mol. Biol.* **1989**, *207*, 355–364. [CrossRef]

47. Levine, R.L.; Yang, I.Y.; Hossain, M.; Pandya, G.A.; Grollman, A.P.; Moriya, M. Mutagenesis induced by a single 1, N6-ethenodeoxyadenosine adduct in human cells. *Cancer Res.* **2000**, *60*, 4098–4104.

48. Wee, H.A.; Brown, W.W.; Lippard, S.J. Preparation of mammalian expression vectors incorporating site-specifically platinated-DNA lesions. *Bioconjug. Chem.* **2009**, *20*, 1058–1063.

49. Brégeon, D.; Doetsch, P.W. Reliable method for generating double-stranded DNA vectors containing site-specific base modifications. *Biotechniques* **2004**, *37*, 760–766. [CrossRef]

50. Pagès, V.; Mazón, G.; Naiman, K.; Philippin, G.; Fuchs, R.P. Monitoring bypass of single replication-blocking lesions by damage avoidance in the Escherichia coli chromosome. *Nucleic Acids Res.* **2012**, *40*, 9036–9043. [CrossRef]

51. Masłowska, K.H.; Laureti, L.; Pagès, V. iDamage: A method to integrate modified DNA into the yeast genome. *Nucleic Acids Res.* **2019**, *47*, e124. [CrossRef] [PubMed]

52. Wang, H.; Hays, J.B. Simple and rapid preparation of gapped plasmid DNA for incorporation of oligomers containing specific DNA lesions. *Mol. Biotechnol.* **2001**, *19*, 133–140. [CrossRef]

53. Baker, D.J.; Wuenschell, G.; Xia, L.; Termini, J.; Bates, S.E.; Riggs, A.D.; O'Connor, T.R. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J. Biol. Chem.* **2007**, *282*, 22592–22604. [CrossRef] [PubMed]

54. Kitsera, N.; Gasteiger, K.; Lühnsdorf, B.; Allgayer, J.; Epe, B.; Carell, T.; Khobta, A. Cockayne syndrome: Varied requirement of transcription-coupled nucleotide excision repair for the removal of three structurally different adducts from transcribed DNA. *PLoS ONE* **2014**, *9*, e94405. [CrossRef]

55. Ji, S.; Park, D.; Kropachev, K.; Kolbanovskiy, M.; Fu, I.; Broyde, S.; Essawy, M.; Geacintov, N.E.; Tretyakova, N.Y. 5-Formylcytosine-induced DNA–peptide cross-links reduce transcription efficiency, but do not cause transcription errors in human cells. *J. Biol. Chem.* **2019**, *294*, 18387–18397. [CrossRef]

56. Piekna-Przybylska, D. Reporter assays for BER pathway. In *Methods in Molecular Biology*; Balakrishnan, L., Stewart, J.A., Eds.; Springer: New York, NY, USA, 2019; pp. 145–160. ISBN 978-1-4939-9500-4.

57. Jung, G.; Wiehler, J.; Zumbusch, A. The photophysics of green fluorescent protein: Influence of the key amino acids at positions 65, 203, and 222. *Biophys. J.* **2005**, *88*, 1932–1947. [CrossRef]

58. Ehrig, T.; O'Kane, D.J.; Prendergast, F.G. Green-fluorescent protein mutants with altered fluorescence excitation spectra. *FEBS Lett.* **1995**, *367*, 163–166. [CrossRef]

59. Stepanenko, O.V.; Kuznetsova, I.M.; Verkhusha, V.V.; Turoverov, K.K. Beta-Barrel Scaffold of Fluorescent Proteins. Folding, Stability and Role in Chromophore Formation. In *International Review of Cell and Molecular Biology*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 302, pp. 221–278.

60. Loeb, L.A.; Preston, B.D. Mutagenesis by Apurinic/Apyrimidinic Sites. *Annu. Rev. Genet.* **1986**, *20*, 201–230. [CrossRef]

61. Weerasooriya, S.; Jasti, V.P.; Basu, A.K. Replicative bypass of abasic site in *Escherichia coli* and human cells: Similarities and differences. *PLoS ONE* **2014**, *9*, e107915. [CrossRef]

62. Villani, G.; Shevelev, I.; Orlando, E.; Pospiech, H.; Syvaoja, J.E.; Markkanen, E.; Hubscher, U.; Le Gac, N.T. Gap-directed translesion DNA Synthesis of an abasic site on circular DNA templates by a human replication complex. *PLoS ONE* **2014**, *9*, e93908. [CrossRef]

63. Mozzherin, D.J.; Shibutani, S.; Tan, C.K.; Downey, K.M.; Fisher, P.A. Proliferating cell nuclear antigen promotes DNA synthesis past template lesions by mammalian DNA polymerase δ. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6126–6131. [CrossRef] [PubMed]

64. Chaim, I.A.; Nagel, Z.D.; Jordan, J.J.; Mazzucato, P.; Ngo, L.P.; Samson, L.D. In vivo measurements of interindividual differences in DNA glycosylases and APE1 activities. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E10379–E10388. [CrossRef] [PubMed]