SETMAR functions in illegitimate DNA recombination and non-homologous end joining

Michael Tellier*[#] and Ronald Chalmers*

School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK

* Email: michael.tellier@path.ox.ac.uk; ronald.chalmers@nottingham.ac.uk

[#] Current Address: Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RF, UK

Abstract

In anthropoid primates, SETMAR is a fusion between a methyltransferase gene and a domesticated DNA transposase. SETMAR has been found to be involved in several cellular functions including regulation of gene expression, DNA integration and DNA repair. These functions are thought to be mediated through the histone methyltransferase, the DNA binding and the nuclease activities of SETMAR. To better understand the cellular roles of SETMAR, we generated several U2OS cell lines expressing either wild type SETMAR or a truncated or mutated variant. We tested these cell lines with *in vivo* plasmid-based assays to determine the relevance of the different domains and activities of SETMAR in DNA integration and repair. We found that expressing the SET and MAR domains, but not wild type SETMAR, partially affect DNA integration and repair. The methyltransferase activity of SETMAR is also needed for an efficient DNA repair whereas we did not observe any requirement for the putative nuclease activity of SETMAR. Overall, our data support a non-essential function for SETMAR in DNA integration and repair.

Introduction

SETMAR is an anthropoid primate-specific fusion between a histone methyltransferase gene, connected to dimethylation of histone H3 lysine 36 (H3K36me2), and a domesticated Hsmar1 transposase (1-3). The transposase domain is 94% identical to the Hsmar1 transposase consensus sequence but three mutations, including the DDD to DDN mutation of the catalytic triad, completely abolish the transposition activity of SETMAR (4-6). Although some activity, particularly 5'-end nicking, is recovered *in vitro* in the presence of DMSO and Mn^{2+}, it is likely to not be significant in physiological conditions (4). Nevertheless, the transposase DNA-binding domain of SETMAR is under purifying selection and retains robust transposon end binding and the ability to form dimer (2, 4, 7). It has recently been shown that SETMAR could regulate gene expression in human cells through the combination of its binding to the Hsmar1 transposon ends scattered in the human genome and its methyltransferase activity (7).

Earlier experiments with SETMAR revealed that it was involved in illegitimate DNA integration and DNA repair through the non-homologous end joining repair (NHEJ) pathway (1, 8). NHEJ is one of the four pathways used by the cell to repair DNA double-strand breaks (DSBs) and the primary repair pathway throughout the cell cycle (9). NHEJ is a template-independent DNA repair pathway, which relies on Ku proteins to bind the DNA free ends, on nucleases, such as Artemis, or polymerases to trim or fill the DNA overhangs and on the DNA ligase IV complex to ligate together the two blunt ends (9).

Illegitimate DNA recombination and lentivirus cDNA integration are dependent of the NHEJ pathway but the mechanism responsible for plasmid integration, which cannot

3

75   rely on an integrase, remains uncertain (10, 11). The current model states that the

76   circular plasmid needs to be linearized by a DSB for recruiting DNA repair proteins on

77   the plasmid ends. For genomic integration to happen, one plasmid end needs to be in

78   the vicinity of a genomic lesion for the NHEJ proteins to use the linearized plasmid DNA

79   to repair the genomic DSB (11).

80   One of the difficulties in understanding the functions of SETMAR in DNA repair is that it

81   produced a response in a number of different assays, suggesting that it was involved in

82   many different aspects of DNA metabolism. For example, its overexpression promotes

83   classical NHEJ, the random integration of transfected plasmid DNA and the restart of

84   stalled replication forks (1, 12). Based on *in vitro* analysis, it has been hypothesized that

85   purified SETMAR could act as an endonuclease like Artemis (13, 14). However,

86   SETMAR endonuclease activity has only been established *in vitro* and recent papers

87   question its relevance *in vivo* (14, 15). In contrast to Artemis, which promotes both

88   trimming of DNA overhangs and DNA repair in cell extract assays, SETMAR did not

89   stimulate DNA repair and only promotes trimming in one assay.

90   The SET methylase-domain of SETMAR was shown to interact with PRPF19, also

91   known as PSO4, which is a protein involved in the classical NHEJ and the spliceosome,

92   and with DNA ligase IV, which is responsible for ligating the blunt ends in NHEJ (16,

93   17). The interaction with PRPF19 was predicted to target SETMAR to double strand

94   DNA breaks where the SET domain could dimethylate the histone H3 lysine 36 of

95   neighbouring nucleosomes (18). This epigenetic mark recruits and stabilizes the

96   anchoring of Ku70 and NBS1, two early acting NHEJ factors, to the DNA ends (18).

97   Two other papers linked the increase in H3K36me2 following DSBs to the inhibition of

4

98   KDM2A and KDM2B, two histone demethylases involved in the removal of H3K36

99   methylation (19, 20). However, a recent study did not observe an increase in

100   H3K36me2 around DSB sites (21).

101   To better understand the functions of SETMAR in NHEJ, we produced several U2OS

102   cell lines expressing different SETMAR constructs to test the role of the SET and MAR

103   domains and the methyltransferase, DNA binding and nuclease activities of SETMAR in

104   illegitimate DNA integration and repair. We found that expression of the SET and MAR

105   domains, but not of wild type SETMAR, affect DNA integration and repair. SETMAR

106   methyltransferase activity is required for an efficient DNA repair but we did not observe

107   any role for the putative nuclease activity of SETMAR. We hypothesize that the

108   dimerization of SETMAR imposed by the MAR transposase domain could have interfere

109   with the pre-fusion functions of the SET domain.

110

Materials and Methods

Media and growth conditions

113   The T-Rex-U2OS cell lines were maintained in complete Dulbecco's modified Eagle's

114   medium (DMEM, Sigma) supplemented with 10% heat inactivated Foetal Bovine Serum

115   (FBS), 100 u/ml of streptomycin, 100 µg/ml of penicillin, and 5 µg/ml of blasticidin at

116   37°C with 5% CO_2. The medium of T-Rex-U2OS cell lines stably expressing a gene of

117   interest from an integrated pcDNA4TO plasmid was supplemented with 400 µg/ml of

118   zeocin.

119

120 **Plasmids**

121 An artificial codon-optimized version of SETMAR was synthesized by Gene Art (Thermo

122 Fischer) and cloned into pcDNA4TO at the EcoRI/NotI restriction sites. The truncated

123 and mutant (N210A, R432A and D483A) versions of SETMAR were produced by PCR.

124 pRC1712 was constructed by cloning a neomycin resistance gene into pBluescript

125 SKII+ (Agilent) at the BamHI restriction site.

126

127 **Stable transfection of T-Rex-U2OS cells**

128 For each transfection, 2.5×10^5 of cells were seeded in a 6-well plate and grown

129 overnight in DMEM supplemented with 10% FBS. The plasmids were transfected using

130 Lipofectamine 2000 (Invitrogen), following manufacturer's instruction. After 24 hours, a

131 quarter of the cells were transferred to 100 mm dishes and the medium supplemented

132 with 400 µg/ml of zeocin (Invivogen). After 2 weeks of selection, single foci were picked

133 and grown in a 24-well plate. The expression of the gene of interest was verified in each

134 cell line by inducing the PCMV promoter with doxycycline at a final concentration of 1

135 µg/ml for 24 hours. The list of cell lines used in this study is presented in Table 1.

136

137 **Table 1: Mammalian cell lines used in this study**

T-Rex-U2OS	Human osteosarcoma cell line stably expressing the tetracycline repressor protein.
T-Rex-U2OS-TO	T-Rex-U2OS cell line stably transfected with an empty pcDNA4TO.
T-Rex-U2OS-	T-Rex-U2OS cell line stably expressing SETMAR.

SETMAR	
T-Rex-U2OS-TO-FLAG	T-Rex-U2OS cell line stably transfected with an empty pcDNA4TO-FLAG.
T-Rex-U2OS-SET-FLAG	T-Rex-U2OS cell line stably expressing the FLAG-tagged exons 1 and 2 of SETMAR (= SET domain).
T-Rex-U2OS-SET-FLAG N210A	T-Rex-U2OS cell line stably expressing the FLAG-tagged exons 1 and 2 of SETMAR (= SET domain) with the mutation N210A abolishing the methyltransferase activity of SET.
T-Rex-U2OS-MAR-FLAG	T-Rex-U2OS cell line stably expressing the FLAG-tagged exon 3 of SETMAR (= MAR domain).
T-Rex-U2OS-SETMAR-FLAG	T-Rex-U2OS cell line stably expressing a FLAG-tagged SETMAR.
T-Rex-U2OS-SETMAR N210A-FLAG	T-Rex-U2OS cell line stably expressing a FLAG-tagged SETMAR with the mutation N210A abolishing the methyltransferase activity of SETMAR.
T-Rex-U2OS-SETMAR R432A-FLAG	T-Rex-U2OS cell line stably expressing a FLAG-tagged SETMAR with the mutation R432A decreasing the affinity of SETMAR for the transposon end.
T-Rex-U2OS-SETMAR D483A-	T-Rex-U2OS cell line stably expressing a FLAG-tagged SETMAR with the D483A mutation abolishing the catalytic activity of

FLAG	transposase domain of SETMAR.

138

Western blotting

140 Whole cell extracts were harvested from cultures at ~90% confluency in six-well plates.

141 Briefly, cells were washed two times with ice-cold PBS then pelleted for 5 minutes at

142 3000 x g at 4°C. Samples were resuspended in 100 µl of Radio ImmunoPrecipitation

143 Assay (RIPA) buffer (10 mM Tris-HCl pH8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS,

144 1% Triton X-100, 0.1% sodium deoxycholate) with freshly added protease inhibitor

145 cocktail (Roche Applied Science) and incubated on ice for 30 minutes, with a vortexing

146 every 10 minutes. Cell lysates were centrifuged for 15 minutes at 14000 x g at 4°C and

147 the protein in the supernatants was quantified by the Bradford assay.

148 For each western blot, 20 µg of proteins were mixed with 2X SDS loading buffer, boiled

149 for 5 minutes, and electrophoresed on a 10% SDS-PAGE gel. Proteins were transferred

150 to a polyvinylidene difluoride (PVDF) membrane, which was blocked in 5% milk or BSA

151 (Roche) and incubated with specific primary antibodies at 4°C overnight. After washing,

152 membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary

153 antibodies for one hour at room temperature, washed, and signals were detected with

154 the ECL system (Promega) and Fuji medical X-ray film (Fujifilm).

155 The following antibodies were used: anti-beta Tubulin (rabbit polyclonal IgG, 1:500

156 dilution, ab6046, Abcam), anti-Hsmar1 antibody (goat polyclonal, 1:500 dilution,

157 ab3823, Abcam), anti-FLAG (rabbit, 1:500 dilution, F7425, Sigma). The secondary

158 antibodies were horseradish peroxidase-conjugated anti-goat (rabbit polyclonal, 1:5000

159 dilution, ab6741, Abcam) and anti-rabbit (goat polyclonal, 1:5000-1:10000, ab6721,

160 Abcam).

161

Growth rate

163 At day 0, $2x10^4$ cells were seeded in eight 6 cm dishes for each cell line and one dish

164 was count every day for eight days using a hemocytometer.

165

Illegitimate DNA integration assay

167 For integration assays in the T-Rex-U2OS cell lines, $8x10^5$ cells were seeded onto 6-

168 well plates with 2.5 µg of circular or linearized pRC1712 and 5 µl of Lipofectamine 2000

169 (Invitrogen). Twenty-four hours later, cells were trypsinized and $5x10^4$ cells of each

170 transfection were seeded onto 10 cm dishes in medium containing 800 µg/ml of G418

171 (Sigma). After two weeks of selection, surviving foci were fixed for 15 min with 10%

172 formaldehyde in PBS, stained for 30 min with methylene blue buffer (1% methylene

173 blue, 70% ethanol), washed with water, air dried, and photographed. The transfection

174 efficiency was tested by transfecting a pEGFP plasmid. After 24 hours, the live cells

175 were observed using a Carl Zeiss Axiovert S100 TV Inverted Microscope with an HBO

176 100 illuminator. The transfection efficiency was found to be similar between the different

177 cell lines.

178

Non-homologous end-joining assay and FACS analyses

180 Prior to transfection, the pEGFP-Pem1-Ad2 plasmid was digested overnight with *HindIII*

181 or *I-SceI*. The digested plasmids were heat-inactivated and column-purified before

182 being co-transfected with a pRFP plasmid for controlling the transfection efficiency. A

183 day before transfection, $8x10^5$ cells were seeded in 60 mm dishes for obtaining a ~70 %

184 confluency on the transfection day. Transfections were performed with 3 µg of linear

185 pEGFP-Pem1-Ad2, 3 µg of pRFP and 14 µl of Lipofectamine 2000 (Invitrogen),

186 according to manufacturer's instructions. After 24 hours, green (GFP) and red (RFP)

187 fluorescence was measured by fluorescence-activated flow cytometry (FACS). For

188 FACS analysis cells were harvested with Accutase (Sigma), washed once in 1X PBS

189 and fixed in 2% formaldehyde (Sigma). FACS analysis was performed on a Coulter

190 FC500 (Beckman Coulter). The numbers of repaired events are reported as the ratio of

191 green and red positive cells over the total number of red positive cells. This ratio

192 normalizes the numbers of repaired events to the transfection efficiency. The values for

193 all the cell lines are reported as a percent of the control cell lines.

194

195 **Results**

196 **SETMAR overexpression does not promote cell proliferation in the U2OS cell line**

197 It was previously observed that SETMAR overexpression increases the growth rate of

198 the HEK293 and HEK293T cell lines (22). Conversely, SETMAR depletion by RNA

199 interference or CRISPR/Cas9 knock-out was found to decrease the growth rate of THP-

200 1 and DLD-1 cancer cells, respectively (23, 24). We previously shown that a U2OS cell

201 line mildly overexpressing SETMAR that SETMAR is involved in the regulation of the

202 expression of a broad set of genes (7). However, we did not found an enrichment for

203 genes involved in cell cycle (7). To determine whether altering SETMAR expression

204 level also affects the growth rate of the U2OS cell line, we tested three stable T-Rex-

205 U2OS cell lines overexpressing at different level Flag-tagged version of SETMAR and

10

206 one cell line expressing the SET domain only (Fig 1A). The expression level of the SET

207 domain or SETMAR was determined by western blotting using an anti-FLAG antibody to

208 allow the comparison between the cell lines. The growth rate was determined by

209 counting the number of cells across a period of eight days (Fig 1B). A small but

210 significant decrease in cell proliferation was observed for most of the cell lines

211 overexpressing SET or SETMAR after 5 to 6 days.

212 **Fig 1. The overexpression of SET or SETMAR do not promote cell proliferation in**

213 **an U2OS genetic background.**

214 **A**, Western blot for the FLAG-tagged SETMAR in the U2OS, SETF and SMF cell lines.

215 The western blot was performed with anti-FLAG and anti-β-tubulin antibodies. **B**,

216 Growth rate of U2OS, SETF and SMF cell lines. At day 0, $2.0x10^4$ cells were seeded in

217 eight dishes and one dish was counted every day for eight days. Average ± S.E.M. of 3

218 to 5 biological replicates. Statistical test: t-test with Holm-Sidak correction, * p-value <

219 0.05, ** p-value < 0.01, *** p –value < 0.001

220

221 **Characterization of the different SETMAR constructs**

222 To improve our understanding of SETMAR roles in illegitimate DNA integration and the

223 NHEJ pathway, we produced several U2OS cell lines stably overexpressing wild type,

224 truncated or mutant version of SETMAR (Fig 2A and Table 2). SETMAR probably

225 functions as a dimer in the cell with the transposase domain providing the whole dimer

226 interface (25). The endogenous concentration of SETMAR in the U2OS cell line is low,

227 with less than 500 molecules per cell (26). The overexpression of a SETMAR mutant

228   should therefore produce dimers of two mutant monomers and dimers containing wild

229   type and mutant monomers. Dimers with two wild type monomers are the less likely so

230   SETMAR activity in the cell should be hindered by the overexpression of the mutants.

231   The expression level of each cell line was determined by western blotting using anti-

232   SETMAR and anti-FLAG antibodies (Fig 2B). An anti-FLAG antibody was used for the

233   cell lines containing an F (for FLAG-tag) in their names. SM2 and 3 overexpress a

234   version of SETMAR without any FLAG-tag so an antibody against the last nine amino

235   acids of SETMAR was used to determine their expression level.

236

237   **Fig 2. U2OS cell lines used in the *in vivo* DNA repair assay.**

238   **A**, Schematic representation of SETMAR, SET and MAR and the location of the

239   different mutations. **B**, Western blot for the FLAG-tagged SETMAR in the U2OS, SM,

240   SETF, MARF and SMF cell lines. The western blot was performed with anti-Hsmar1,

241   anti-FLAG and anti-β-tubulin antibodies. The cell lines are described in Table 2.

242

243   **Table 2: U2OS cell lines used in the *in vivo* DNA repair assay.**

Full name	Abbreviation	Expression level
T-Rex-U2OS-TO	TO	Null
T-Rex-U2OS-SETMAR	SM2	Low
	SM3	Very high

12

T-Rex-U2OS-TO-FLAG	TOF	Null
T-Rex-U2OS-SET-FLAG	SETF1	Low
	SETF2	Medium
T-Rex-U2OS-SET-N210A- FLAG	SETF-N210A	Low
T-Rex-U2OS-MAR-FLAG	MARF	Medium
T-Rex-U2OS-SETMAR-FLAG	SMF2	Medium
	SMF3	Medium
T-Rex-U2OS-SETMAR-N210A-FLAG	SMF-N210A	Medium
T-Rex-U2OS-SETMAR R432A-FLAG	SMF-R432A	Medium
T-Rex-U2OS-SETMAR D483A-FLAG	SMF-D483A	Medium

244

245   The two control cell lines, TO and TOF, express only the endogenous SETMAR. We

246   used four cell lines overexpressing wild type SETMAR at either low level, SM2, medium

247   level, SMF2 and SMF3, or at very high level, SM3. The SET domain is overexpressed in

248   SETF1 and SETF2 at low and medium level, respectively, whereas the MAR domain,

13

249 which encodes the domesticated Hsmar1 transposase, is overexpressed at medium

250 level. We also inserted three different mutations to abrogate specific functions of

251 SETMAR. The N210A mutation, which is located in the key NHSC motif of the SET

252 domain, abolishes the methyltransferase activity of SETMAR (7). To investigate the

253 relative contribution of SETMAR binding to Hsmar1 transposon ends (inverted terminal

254 repeat, ITR), we inserted the R432A mutation, which decreases the affinity of SETMAR

255 to the Hsmar1 transposon ends (25, 27). To test the requirement of SETMAR's trimming

256 activity, we inserted the D483A mutation. The D483A mutant is catalytically defective

257 because of the mutation of the first D of the DDD triad, which is necessary for the

258 incorporation of one the Mg^{2+} ion (14, 25).

259

260 **The SET and MAR domains but not SETMAR promote DNA integration**

261 SETMAR was previously shown to promote illegitimate integration in the genome. We

262 used the different Flag-tagged SETMAR constructs to gain a better understanding of

263 which SETMAR domains and activities are involved in DNA integration. For integration

264 to happened, two conditions are required (Fig 3A). First, a plasmid need to be linearized

265 by a DSB and it needs to be in the vicinity of a genomic DSB since integration is

266 mediated by the NHEJ pathway (11). However, illegitimate integration is one of the

267 three possible outcomes for a linearized plasmid because it can be either re-circularized

268 or degraded (Fig 3A). The illegitimate plasmid integration rate was determined by

269 transfecting a plasmid encoding a neomycin resistance marker before challenging the

270 cells with G418 for two weeks. Cells in which the plasmid has been integrated into the

271 genome could develop into foci. The foci were counted after staining with methylene

272 blue.

273

Fig 3. The SET and MAR domains increase the frequency of illegitimate DNA

integration.

A, Representation of the integration assay. Cells are transfected with a circular plasmid

encoding a neomycin resistance gene. For integration to occur through the NHEJ

pathway, the plasmid needs to be linearized by a DSB and a plasmid free end has to be

in close vicinity of a genomic DSB. The linearized plasmid can also be repaired, which

re-circularized the plasmid, or be degraded. Following G418 treatment for two weeks,

surviving cells form foci which can be detected by methylene blue staining. **B**, Number

of illegitimate integration events in the genome of a circular plasmid encoding a

neomycin resistance gene. Average ± S.E.M. of 3 biological replicates. Statistical test:

paired t-test, * p-value < 0.05, ** p-value < 0.01. **C**, Representative pictures of

integration plates. The integration rate for each cell line is indicated below each picture.

286

We first determined whether the topology of the plasmid influences its integration

frequency. We therefore transfected a circular or linear version of the same plasmid in

the control cell line TOF and observed a ~3-fold decrease in the integration of the

linearized form compared to the circular plasmid (S1 Fig). We decided to use the

circular plasmid for testing the different SETMAR constructs.

15

292   We performed two sets of plasmid integration experiments with the Flag-tagged cell

293   lines (Fig 3B and C). A significant increase in illegitimate plasmid DNA integration was

294   observed with a medium overexpression of the SET domain, the MAR domain and

295   SETMAR N210A and R483A mutants. A low overexpression of the SET domain or

296   medium overexpression of wild type SETMAR or SETMAR R432A mutant did not affect

297   the plasmid integration rate. A representative selection of the integration plates of each

298   cell line and their respective integration frequency are presented in Fig 3C. These

299   results agree with published work showing that the efficiency of illegitimate

300   recombination in most cell lines is less than 1% (11). We only observed an increase of

301   the efficiency to 2% with a medium overexpression of the SET domain.

302

303   **The SET and MAR domains have an opposite effect on DNA repair**

304   To gain a better understanding of SETMAR functions in the NHEJ pathway, we used a

305   previously described *in vivo* DNA repair assay (28). This assay is based on two

306   plasmids, one encoding the reporter gene (pEGFP) and the other serving as a

307   transfection control (pRFP). The reporter plasmid encodes an *EGFP* gene interrupted

308   by a 2.4 kb intron derived from the rat *Pem1* gene. An exon from the adenovirus (Ad2)

309   has been integrated in the intron abolishing the GFP activity (Fig 4A). The Ad2 exon is

310   flanked by HindIII and I-SceI restriction sites. Cleavage with HindIII or I-SceI yields

311   compatible or incompatible ends, respectively (Fig 4B). These two types of ends require

312   different steps for repair. Compatible ends can be directly ligated while incompatible

313   ends need to be trimmed before the ligation step can occur. The repair of the linearized

314   plasmid by the NHEJ pathway restores the GFP ORF making the cell green (Fig 4C).

315   The repair events were detected by flow cytometry measuring at least 10,000 cells per

316   assay. The repair efficiency was calculated as the ratio of green and red cells over the

317   total number of red cells, thus normalizing the transfection efficiency between cell lines.

318

319   **Fig 4. The SET and the MAR domains of SETMAR have an opposite effect on DNA**

320   **repair by the NHEJ pathway.**

321   **A**, The reporter construct, pEGFP-Pem1-Ad2, is composed of a GFP cassette flanked

322   by a PCMV promoter and a SV40 poly(A) sequence. The GFP coding sequence is

323   interrupted by a 2.4 kb intron containing an adenovirus exon (Ad). The Ad exon is

324   flanked by HindIII and I-SceI restriction sites. The donor (DS) and acceptor (AS) splicing

325   sites are shown. **B**, HindIII and I-SceI restriction sites are respectively composed of a

326   palindromic 6-bp and a non-palindromic 18-bp sequence. Digestion of the reporter

327   construct by HindIII or I-SceI generates respectively compatible and incompatible ends.

328   **C**, The presence of the Ad exon in the GFP ORF inactivates the GFP activity thus

329   making the cell GFP negative. Removal of the Ad exon by HindIII or I-SceI followed by a

330   successful intracellular repair will restore the GFP expression that can be quantified by

331   flow cytometry. Adapted from (28). **D**, DNA repair efficiency of a linearized plasmid with

332   compatible (HindIII) or incompatible (I-SceI) ends in the different cell lines relative to the

333   control cell line (TO or TOF). Average ± S.E.M. of 3 biological replicates. Statistical test:

334   paired t-test, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

335

336   For normalizing between the biological replicates, we calculated for each replicate the

337   ratio of the repair efficiency of each cell line relative to their respective control cell line.

338   The average ratio from three independent experiments is presented in Fig 4D. The SM2

339   and 3 cell lines were compared to the TO control cell line whereas the remaining cell

340   lines were compared to TOF to take into account any possible effect of the Flag tag.

341   Unsurprisingly, there was no significant difference in the repair efficiency of both types

342   of ends between the two control cell lines (S2 Fig). In all cell lines, except SETF1 and 2,

343   a ~10% decrease of the repair efficiency of incompatible ends compared to compatible

344   ends was observed because of the necessary trimming of the DNA overhangs before

345   the plasmid ends could be ligated (S2 Fig), in agreement with previously published

346   results for this assay (28).

347   The overexpression of SETMAR, irrespective of the level of expression, does not

348   promote the repair of either compatible or incompatible ends confirming the recent

349   results obtained in the cell-extract assays (Fig 4D) (14). A slight but non-significant

350   decrease in the repair of compatible ends is observed when the SET domain is

351   expressed at a low level whereas a medium overexpression significantly decreases by

352   ~20% the repair efficiency of both compatible and incompatible ends, suggesting a

353   concentration-dependent effect. A low overexpression of the SETF N210A

354   methyltransferase deficient mutant does not affect DNA repair of both type of ends. In

355   contrast, medium overexpression of the MAR domain increases the repair efficiency of

356   both types of ends by ~20% compared to the control cell line. The overexpression of

357   SETMAR N210A reduces the repair efficiency of both type of ends by ~15% but only the

358   compatible ends decrease is statistically significant, supporting a role for SETMAR

18

359 methyltransferase activity for an efficient DNA repair (18). In contrast, SETMAR D483A

360 mutant increases by ~15% the repair efficiency of both types of ends, which confirms

361 that SETMAR nuclease activity is not required for DNA repair *in vivo* (14). No difference

362 is observed with SETMAR R432A mutant for the repair of both types of ends confirming

363 that the ITR binding activity is not involved in DNA repair.

364

Discussion

366 It has been proposed that SETMAR could be involved in DNA repair through the NHEJ

367 pathway. In this study, we investigated whether SETMAR, the SET or MAR domains

368 and SETMAR methyltransferase, ITR binding and nuclease activities were involved in

369 NHEJ and illegitimate DNA integration *in vivo*. We found that the SET and the MAR

370 domains have an effect on DNA repair and integration but not wild type SETMAR. In

371 addition, SETMAR proposed nuclease activity, which has been observed *in vitro*, does

372 not seem to be functional *in vivo*. In contrast, SETMAR methyltransferase activity is

373 required for an efficient DNA repair.

374 Previous publications have associated SETMAR expression to cell proliferation in

375 different cell lines (22, 23). This association was not observed in the U2OS cell line (Fig

376 1B). Indeed, a modest overexpression of the SET domain or SETMAR only slightly

377 reduces the growth rate after six to seven days. This could possibly be a non-specific

378 effect, as protein overexpression per se is known to reduce growth rate (29). It is also

379 possible that a strong reduction or increase in SETMAR expression is required to

380 observe a change in cellular proliferation. Indeed, our cell lines have been selected for a

381 modest overexpression of SETMAR rather than a strong one. Furthermore, we did not

382 observe an enrichment in cell cycle related genes in the genes differentially expressed

383 upon SETMAR modest overexpression (7).

384 The DNA repair and the DNA integration assays used in this study rely on the NHEJ

385 pathway. The current model of plasmid integration is based on the group of proteins

386 from classical NHEJ (11). This is supported by the idea that for the cell, DNA repair or

387 DNA integration of a linear plasmid produces the same result, i.e. the removal of free

388 ends, which could induce apoptosis or cell cycle arrest. The choice between the

389 outcomes is dependent of the presence or not of a plasmid free end in the vicinity of a

390 genomic DSB. The connection between DNA repair and DNA integration is also

391 supported experimentally with cells depleted for a DNA repair protein which are unable

392 to integrate DNA (30, 31).

393 Linearization of a plasmid is required before its integration in the genome (11).

394 However, a linearized plasmid is more likely to be repaired than to be integrated.

395 Indeed, the probability for the DNA repair complex to find the other plasmid end is

396 higher than being near a genomic lesion due to physical continuity of the plasmid DNA,

397 which ensures that the two ends can never be very far apart. Therefore, promotion of

398 DNA repair could reduce the amount of linear plasmid in the cell by re-circularizing them

399 and through it, decreasing the frequency of integration. The inefficiency of plasmid

400 integration in mammalian cells comes from the combination of several limitations. The

401 major ones are the low frequency of plasmid linearization, the re-circularization of the

402 majority of linearized plasmids by the NHEJ pathway or their degradation, and the low

403 probability of having a genomic and a plasmid end near each other. The proportion of

404 linearized plasmid degraded by the cell is unknown but it is likely to be high. Indeed, the

405 transfection of a linearized plasmid decreases by 3-fold the integration frequency

406 compared to its circular form (S1 Fig). From the data obtained in our NHEJ assay we

407 can estimate the frequency of re-circularization as at least 65% (S1 Fig). Since the

408 integration frequency is below 1%, we can estimate that around 35% of the linear

409 plasmid are degraded. A fraction of the transfected cells may also die from apoptosis if

410 the free ends are not repaired in time. However, in the present experiment this is

411 controlled by using a counted aliquot of the living cells for the G418 selection.

412 Previous works on the function of SETMAR in NHEJ claim specific roles for the SET

413 and the MAR domains (18, 26). The SET domain dimethylates H3K36 of nucleosomes

414 near DSBs. This epigenetic mark recruits and stabilizes the binding of Ku70 and NBS1

415 to the DNA ends (18). The MAR domain trims damaged and undamaged DNA

416 overhangs before other NHEJ proteins ligate the ends (13). It has also been claimed

417 that SETMAR activity is regulated by several interactions with other proteins involved in

418 the NHEJ such as PRPF19 and DNA ligase IV (16, 17). Only a direct interaction

419 between the SET domain and PRPF19 has been confirmed and is supposed to promote

420 the recruitment of SETMAR to DSBs (16).

421 Overexpression of the wild type SETMAR did not affect DNA repair and integration in

422 our *in vivo* assays. We found however that a medium overexpression of the SET

423 domain, but not a low overexpression, decreases DNA repair efficiency and increases

424 illegitimate DNA integration (Fig 3B and 4D). In our assays, both DNA repair and

425 integration are supposed to be dependent on the NHEJ pathway, consistent with a role

426 for the ancestral SET gene in this pathway. However, the mechanism by which the SET

427 domain favours DNA integration over DNA repair is unclear. The decrease in re-

428 circularization with both compatible and incompatible ends found in the DNA repair

429 assay could delay the re-circularization of plasmids, increasing the window of

430 opportunity for a plasmid end to be in the vicinity of a genomic end and therefore

431 promoting its genomic integration.

432 The overexpression of the MAR domain stimulates both DNA repair and integration (Fig

433 3B and 4D). These results seem to indicate that the proportion of linear plasmid

434 degraded is reduced upon overexpression of the MAR domain. It has been previously

435 proposed that the MAR domain of SETMAR could bind DNA free ends to trim the DNA

436 overhangs (13, 14). However, if the MAR domain have a trimming activity, we would

437 expect to observe a larger increase in the repair of incompatible ends versus compatible

438 ends. In fact, the increase in the efficiency of repair is similar for both types of ends,

439 which does not support a trimming activity (Fig 4D). Also, overexpression of SETMAR

440 D483A mutant, which should abolish any remaining catalytic activity of the MAR

441 domain, does not decrease the DNA repair efficiency (Fig 4D). In fact, we observe an

442 increase in DNA integration and in DNA repair with both compatible and incompatible

443 ends (Fig 3B and 4D). This seems to indicate that the MAR domain of SETMAR does

444 not trim DNA overhangs *in vivo* but could however bind to DNA free ends to protect

445 them from degradation, increasing the probability of integrating the linearized plasmid or

446 re-circularizing the plasmid through the NHEJ pathway. The increase in DNA repair and

447 integration could thus be mediated through the interactions of SETMAR with other

448 NHEJ factors. It remains unknown whether these interactions are solely dependent on

449 the SET domain or could also be mediated by the MAR domain. It is however important

22

450   to remember that in our system the endogenous SETMAR is still expressed and the

451   expression of the MAR domain could thus result in MAR dimers and in SETMAR-MAR

452   dimers since SETMAR dimerization is mediated by the MAR domain. The SETMAR-

453   MAR dimers could therefore be the only protein bringing the NHEJ factors to the DNA

454   free ends if the interactions are dependent on the SET domain.

455   Unsurprisingly, the ITR binding activity of SETMAR is not required for DNA repair and

456   integration (Fig 3B and 4D). In contrast, a medium overexpression of the

457   methyltransferase defective mutant, SETMAR N210A, decreases DNA repair and

458   increases DNA integration whereas overexpression of the wild type SETMAR does not

459   affect DNA repair and integration (Fig 3B and 4D). The absence of effect of the N210A

460   mutation in the SET domain construct is likely due to its low expression similar to

461   SETF1, which does not affect DNA repair and integration. SETMAR N210A supports a

462   role for the methyltransferase activity for an efficient DNA repair. It remains however

463   unclear whether this is mediated by the deposition of H3K36me2 or by the methylation

464   of another factors. Two studies, which also observe an increase in H3K36me2 at DSB

465   sites, linked this increase to the removal of histone demethylases from the chromatin

466   rather than active methylation (19, 20). In contrast, a recent study did not found any

467   increase in H3K36me2 at a DSB site but found instead an increase in H3K36me3 (21).

468   We recently found that SETMAR N210A mutant was decreasing the bulk level of

469   H3K36me2 by western blot and also observed a decrease of H3K36me3 at some

470   genomic positions, possibly because of a decreased H3K36me2 level which is required

471   by SETD2 for adding the third methyl group (7). The decreased DNA repair activity with

472   SETMAR N210A could therefore be due to this reduced H3K36me2/me3 level which

23

473 would affect the repair efficiency by the NHEJ pathway. We must however stress that

474 our analysis is based on a plasmid assay whereas previous observations were done on

475 genomic DSBs.

476 An interesting question is why the SET and the MAR domains have an effect on DNA

477 repair and integration but not the wild type SETMAR? A possibility is that the

478 endogenous level of SETMAR in U2OS cells is already sufficient for an efficient DNA

479 repair and therefore increasing wild type SETMAR level will not affect the NHEJ

480 pathway. Another possibility is that SETMAR is a dimer in solution whereas almost all

481 mammalian histone methyltransferase function as monomers (32, 33). The only known

482 exception is vSET, a viral histone methyltransferase, which is active only as a dimer

483 (34). The crystal structure of the SET domain is also a monomer strengthening the

484 hypothesis that the pre-fusion SET gene was operating as a monomer (35). The MAR

485 domain enforces the dimerization of SETMAR so even though the SET domain does not

486 dimerize sensu stricto, the proximity between the two SET domains could however

487 affect their methyltransferase activity or their interactions.

488 An interesting observation supporting this hypothesis is the presence of a SETMAR

489 isoform encoding a defective histone methyltransferase monomer because of a splicing

490 event which removes the majority of the SET and post-SET domains in the second

491 exon. Interestingly, this SETMAR isoform is specific to the species where the SET gene

492 is fused to the Hsmar1 transposase (Fig 5). The 5' donor site is present in primates and

493 other several mammals but the acceptor site is only found in anthropoid primates,

494 except for the old-world monkeys where a single mutation in their common ancestor

495 abolishes the acceptor site. The marmosets also lost their 5' donor site but another less

24

496   conserved site is present 20 nucleotides away. This isoform is expressed in most

497   human tissues but at a lower level than the main isoform encoding the active

498   methyltransferase monomer (36). This means that some SETMAR dimers should

499   contain only one SET domain and could therefore function differently from SETMAR

500   dimers with two SET domains.

501

502   **Fig 5. SETMAR second isoform is specific to the anthropoid primates.**

503   **A**, The human *SETMAR* gene encodes two major isoforms, the full-length protein

504   (isoform 1) and a truncated protein (isoform 2). The second isoform is methyltransferase

505   deficient because of the deletion of the majority of the SET and post-SET domains.

506   Canonical donor site (DS), lariat branch points (LBP), and acceptor splicing site (ASS)

507   are present in the second exon of SETMAR. The top brackets represents the exon

508   codons. **B**, Phylogenetic tree of SETMAR second exon in several mammals. The 5'

509   donor site is found in all primates except for the marmoset (see 4) but should have

510   appeared before the appearance of primates because of his presence in several non-

511   primates mammals. The 3' acceptor site is specific to anthropoid primates except for the

512   old-world monkeys which lost it with a single point mutation.

513

514   **Author Contributions**

515   Conceived and designed the experiments: MT. Performed the experiments: MT.

516   Analyzed the data: MT. Contributed reagents/materials/analysis tools: MT RC. Wrote

517   the paper: MT.

25

References

1.   Lee SH, Oshige M, Durant ST, Rasila KK, Williamson EA, Ramsey H, et al. The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc Natl Acad Sci U S A. 2005;102(50):18075-80.

2.   Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci U S A. 2006;103(21):8101-6.

3.   Tellier M, Bouuaert CC, Chalmers R. Mariner and the ITm Superfamily of Transposons. Microbiol Spectr. 2015;3(2):MDNA3-0033-2014.

4.   Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R. The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol Cell Biol. 2007;27(3):1125-32.

5.   Miskey C, Papp B, Mates L, Sinzelle L, Keller H, Izsvak Z, et al. The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. Mol Cell Biol. 2007;27(12):4589-600.

6.   Tellier M, Chalmers R. Development of a papillation assay using constitutive promoters to find hyperactive transposases. bioRxiv. 2018.

7.   Tellier M, Chalmers R. Human SETMAR is a DNA sequence-specific histone-methylase with a broad effect on the transcriptome. Nucleic Acids Res. 2018.

8.   Williamson EA, Farrington J, Martinez L, Ness S, O'Rourke J, Lee SH, et al. Expression levels of the human DNA repair protein metnase influence lentiviral genomic integration. Biochimie. 2008;90(9):1422-6.

541 9.   Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end

542 joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol.

543 2017;18(8):495-506.

544 10.   Li L, Olvera JM, Yoder KE, Mitchell RS, Butler SL, Lieber M, et al. Role of the

545 non-homologous DNA end joining pathway in the early steps of retroviral infection.

546 EMBO J. 2001;20(12):3272-81.

547 11.   Wurtele H, Little KC, Chartrand P. Illegitimate DNA integration in mammalian

548 cells. Gene Ther. 2003;10(21):1791-9.

549 12.   Hromas R, Williamson EA, Fnu S, Lee YJ, Park SJ, Beck BD, et al. Chk1

550 phosphorylation of Metnase enhances DNA repair but inhibits replication fork restart.

551 Oncogene. 2012;31(38):4245-54.

552 13.   Beck BD, Lee SS, Williamson E, Hromas RA, Lee SH. Biochemical

553 characterization of metnase's endonuclease activity and its role in NHEJ repair.

554 Biochemistry. 2011;50(20):4360-70.

555 14.   Mohapatra S, Yannone SM, Lee SH, Hromas RA, Akopiants K, Menon V, et al.

556 Trimming of damaged 3' overhangs of DNA double-strand breaks by the Metnase and

557 Artemis endonucleases. DNA Repair (Amst). 2013;12(6):422-32.

558 15.   Rath A, Hromas R, De Benedetti A. Fidelity of end joining in mammalian

559 episomes and the impact of Metnase on joint processing. BMC Mol Biol. 2014;15:6.

560 16.   Beck BD, Park SJ, Lee YJ, Roman Y, Hromas RA, Lee SH. Human Pso4 is a

561 metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. J

562 Biol Chem. 2008;283(14):9023-30.

563   17.   Hromas R, Wray J, Lee SH, Martinez L, Farrington J, Corwin LK, et al. The

564   human set and transposase domain protein Metnase interacts with DNA Ligase IV and

565   enhances the efficiency and accuracy of non-homologous end-joining. DNA Repair

566   (Amst). 2008;7(12):1927-37.

567   18.   Fnu S, Williamson EA, De Haro LP, Brenneman M, Wray J, Shaheen M, et al.

568   Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-

569   joining. Proc Natl Acad Sci U S A. 2011;108(2):540-5.

570   19.   Jiang Y, Qian X, Shen J, Wang Y, Li X, Liu R, et al. Local generation of fumarate

571   promotes DNA repair through inhibition of histone H3 demethylation. Nat Cell Biol.

572   2015;17(9):1158-68.

573   20.   Cao LL, Wei F, Du Y, Song B, Wang D, Shen C, et al. ATM-mediated KDM2A

574   phosphorylation is required for the DNA damage repair. Oncogene. 2016;35(3):301-13.

575   21.   Clouaire T, Rocher V, Lashgari A, Arnould C, Aguirrebengoa M, Biernacka A, et

576   al. Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks

577   Deciphers Repair Pathway Chromatin Signatures. Mol Cell. 2018;72(2):250-62 e6.

578   22.   De Haro LP, Wray J, Williamson EA, Durant ST, Corwin L, Gentry AC, et al.

579   Metnase promotes restart and repair of stalled and collapsed replication forks. Nucleic

580   Acids Res. 2010;38(17):5681-91.

581   23.   Wray J, Williamson EA, Sheema S, Lee SH, Libby E, Willman CL, et al. Metnase

582   mediates chromosome decatenation in acute leukemia cells. Blood. 2009;114(9):1852-

583   8.

584   24.   Carlson SM, Moore KE, Sankaran SM, Reynoird N, Elias JE, Gozani O. A

585   Proteomic Strategy Identifies Lysine Methylation of Splicing Factor snRNP70 by the

586   SETMAR Enzyme. J Biol Chem. 2015;290(19):12040-7.

587   25.   Roman Y, Oshige M, Lee YJ, Goodwin K, Georgiadis MM, Hromas RA, et al.

588   Biochemical characterization of a SET and transposase fusion protein, Metnase: its

589   DNA binding and DNA cleavage activity. Biochemistry. 2007;46(40):11369-76.

590   26.   Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The

591   quantitative proteome of a human cell line. Mol Syst Biol. 2011;7:549.

592   27.   Claeys Bouuaert C, Chalmers R. Hsmar1 transposition is sensitive to the

593   topology of the transposon donor and the target. PLoS One. 2013;8(1):e53690.

594   28.   Fattah F, Lee EH, Weisensel N, Wang Y, Lichter N, Hendrickson EA. Ku

595   regulates the non-homologous end joining pathway choice of DNA double-strand break

596   repair in human somatic cells. PLoS Genet. 2010;6(2):e1000855.

597   29.   Wurm FM. Production of recombinant protein therapeutics in cultivated

598   mammalian cells. Nat Biotechnol. 2004;22(11):1393-8.

599   30.   van Attikum H, Bundock P, Hooykaas PJ. Non-homologous end-joining proteins

600   are required for Agrobacterium T-DNA integration. EMBO J. 2001;20(22):6550-8.

601   31.   Zhong Q, Boyer TG, Chen PL, Lee WH. Deficient nonhomologous end-joining

602   activity in cell-free extracts from Brca1-null fibroblasts. Cancer Res. 2002;62(14):3966-

603   70.

604   32.   Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D.

605   Histone methyltransferase activity associated with a human multiprotein complex

606   containing the Enhancer of Zeste protein. Genes Dev. 2002;16(22):2893-905.

607   33.   Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, et al. Regulation

608   of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol.

609   2006;13(8):713-9.

610   34.   Wei H, Zhou MM. Dimerization of a viral SET protein endows its function. Proc

611   Natl Acad Sci U S A. 2010;107(43):18433-8.

612   35.   Schapira M. Structural Chemistry of Human SET Domain Protein

613   Methyltransferases. Curr Chem Genomics. 2011;5(Suppl 1):85-94.

614   36.   Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et

615   al. Proteomics. Tissue-based map of the human proteome. Science.

616   2015;347(6220):1260419.

617

618   **Supporting Information**

619   **S1 Fig. A circularized plasmid is more efficiently integrated than a linearized**

620   **plasmid. A**, Comparison of the integration efficiency in the control cell line, TOF, of a

621   circularized or linearized plasmid encoding a neomycin resistance gene. Average ±

622   S.E.M of three biological replicates. Statistical test: paired t-test, ** p-value < 0.01. **B**,

623   Representative pictures of integration plates. The integration rate of each cell line is

624   indicated below each picture.

625   **S2 Fig. Representative FACS profiles of each cell lines.** FACS profiles of a pRFP

626   plasmid used together with the pEGFP-Pem1-Ad2 reporter substrate are shown for

627   each cell line. The profiles were generated using HindIII- (H3) and I-SceI- (SI) linearized

628   plasmids. The proportion of repaired substrate is indicated in the lower right-hand

629   corner of each profile. The percent is calculated from the number of cells that were

630   doubly EGFP (horizontal) and RFP (vertical) positive versus the number of RFP

631   positive.

632

A

B
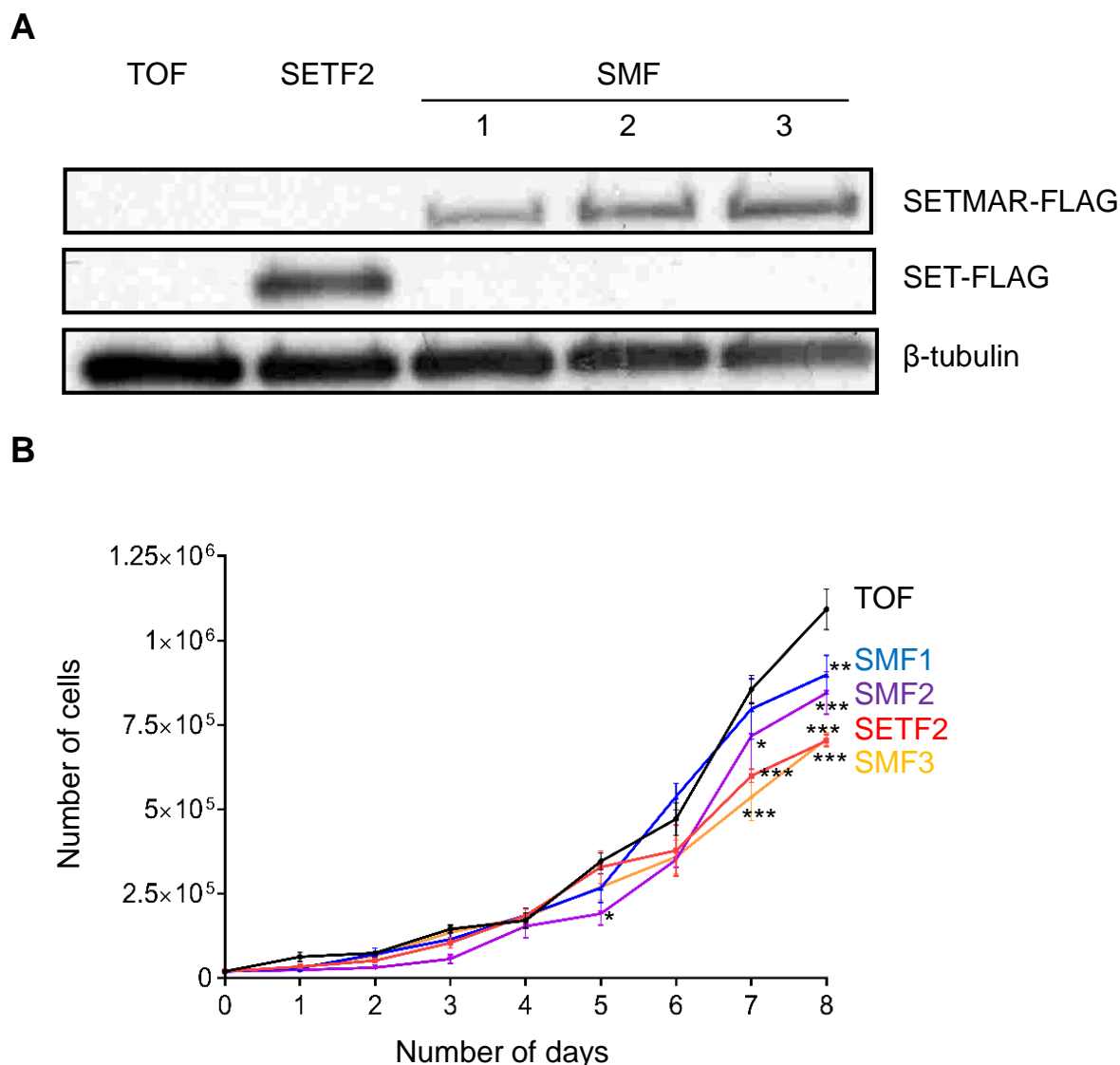
Fig 1. The overexpression of SET or SETMAR do not promote cell proliferation in an U2OS genetic background.

A, Western blot for the FLAG-tagged SETMAR in the U2OS, SETF and SMF cell lines. The western blot was performed with anti-FLAG and anti-β-tubulin antibodies.

B, Growth rate of U2OS, SETF and SMF cell lines. At day 0, 2.0x104 cells were seeded in eight dishes and one dish was counted every day for eight days. Average ± S.E.M. of 3 to 5 biological replicates. Statistical test: t-test with Holm-Sidak correction, * p-value < 0.05, ** p-value < 0.01, *** p –value < 0.001

Figure 1
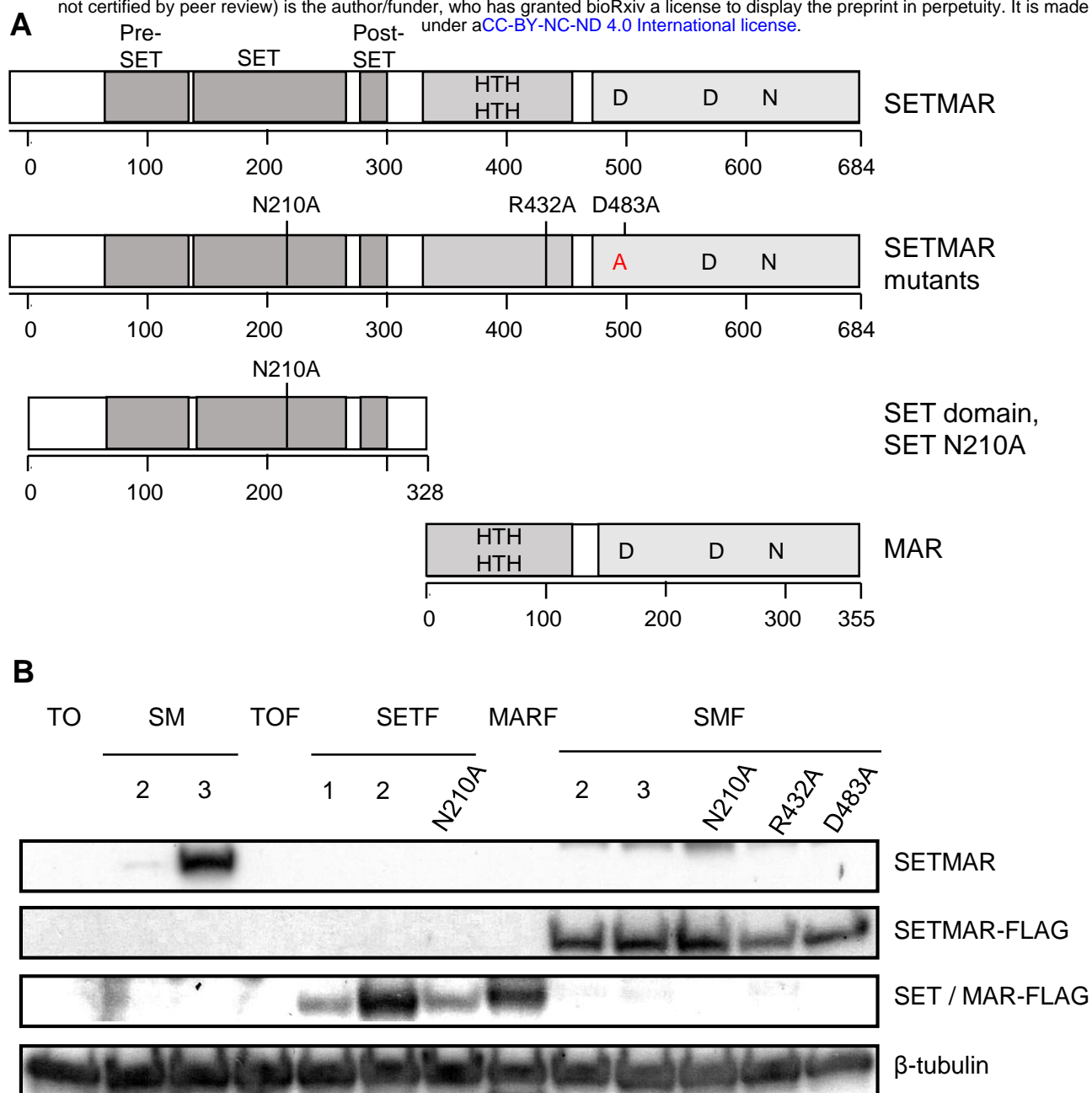
Fig 2. U2OS cell lines used in the in vivo DNA repair assay.
A, Schematic representation of SETMAR, SET and MAR and the location of the different mutations.
B, Western blot for the FLAG-tagged SETMAR in the U2OS, SM, SETF, MARF and SMF cell lines. The western blot was performed with anti-Hsmar1, anti-FLAG and anti-β-tubulin antibodies. The cell lines are described in Table 2.
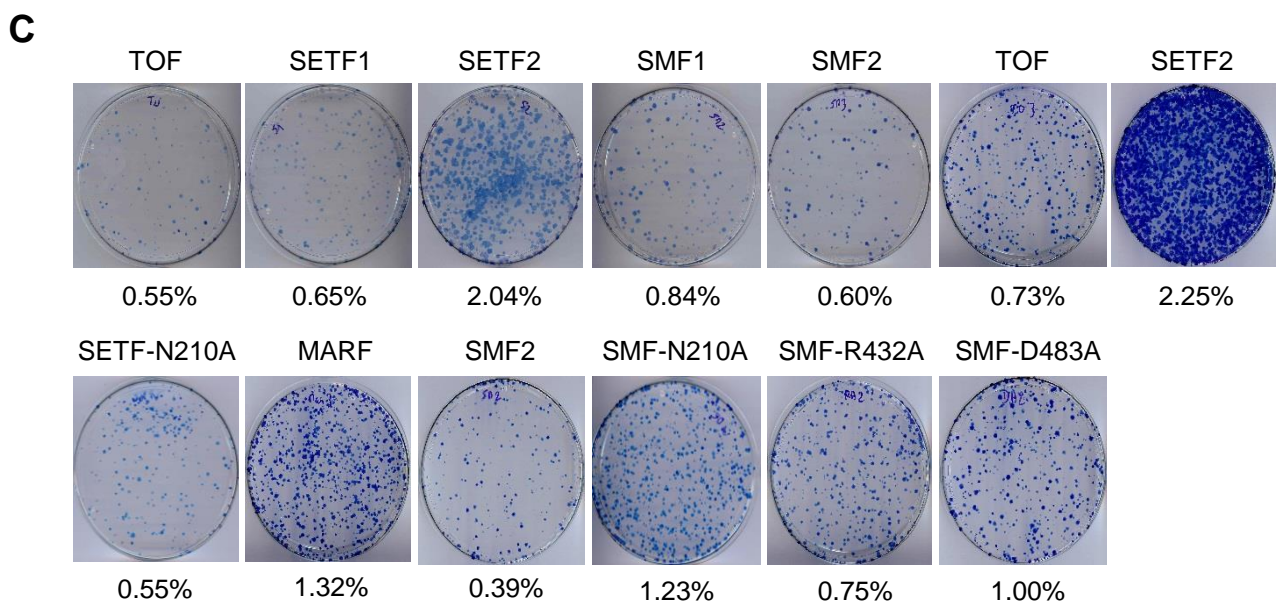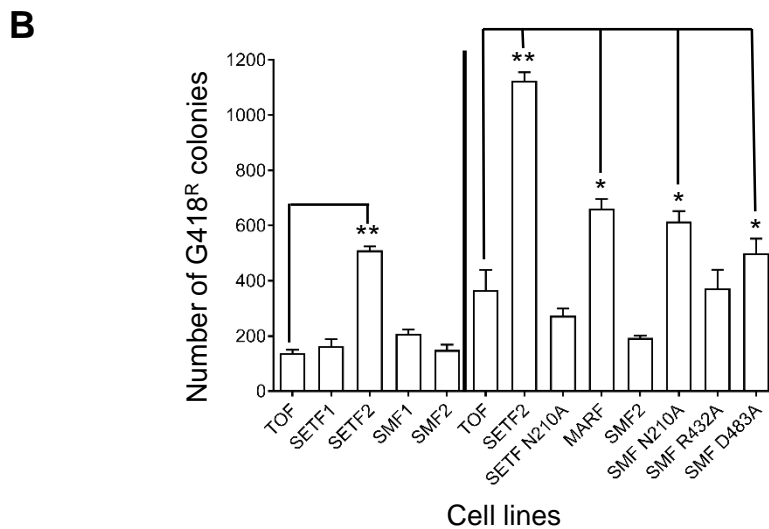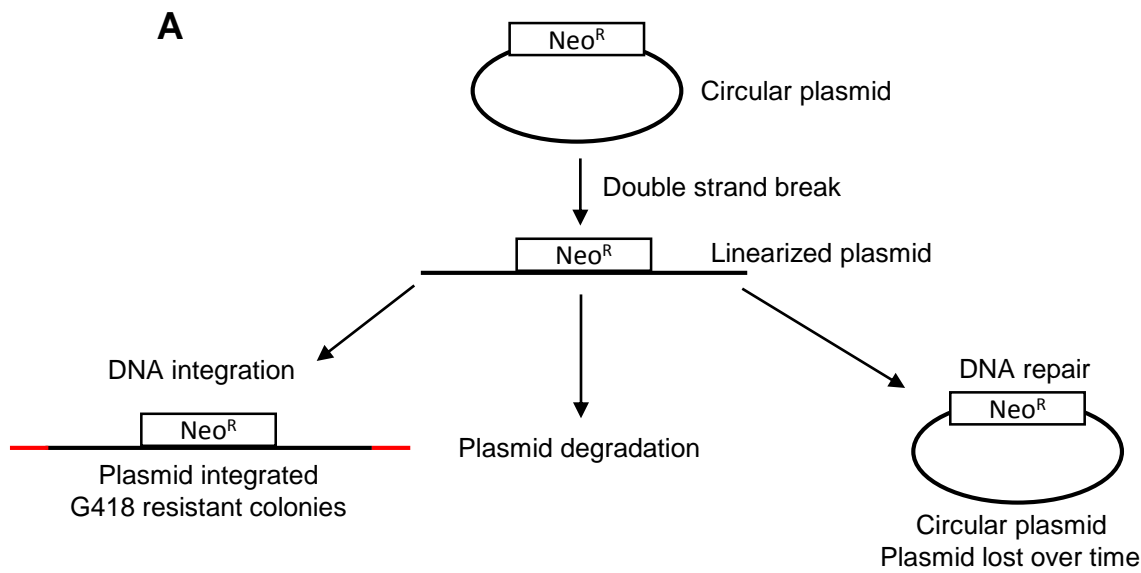
Figure 2

Figure 3

Fig 3. The SET and MAR domains increase the frequency of illegitimate DNA integration.

A, Representation of the integration assay. Cells are transfected with a circular plasmid encoding a neomycin resistance gene. For integration to occur through the NHEJ pathway, the plasmid needs to be linearized by a DSB and a plasmid free end has to be in close vicinity of a genomic DSB. The linearized plasmid can also be repaired, which re-circularized the plasmid, or be degraded. Following G418 treatment for two weeks, surviving cells form foci which can be detected by methylene blue staining.

B, Number of illegitimate integration events in the genome of a circular plasmid encoding a neomycin resistance gene. Average ± S.E.M. of 3 biological replicates. Statistical test: paired t-test, * p-value < 0.05, ** p-value < 0.01.

C, Representative pictures of integration plates. The integration rate for each cell line is indicated below each picture.

A

P_{CMV} HindIII/I-SceI   I-SceI/HindIII

SV40 poly(A)

G   Ad   FP

DS AS   DS AS

B

HindIII (compatible ends and 5' overhangs):

——A   AGCTT ——
——TTCGA   A ——

I-SceI (incompatible ends and 3' overhangs):

——TAGGGATAA   CCCTA ——
——ATCCC   AATAGGGAT ——

C

G — Ad — FP   Splicing   G Ad FP ——→   GFP -

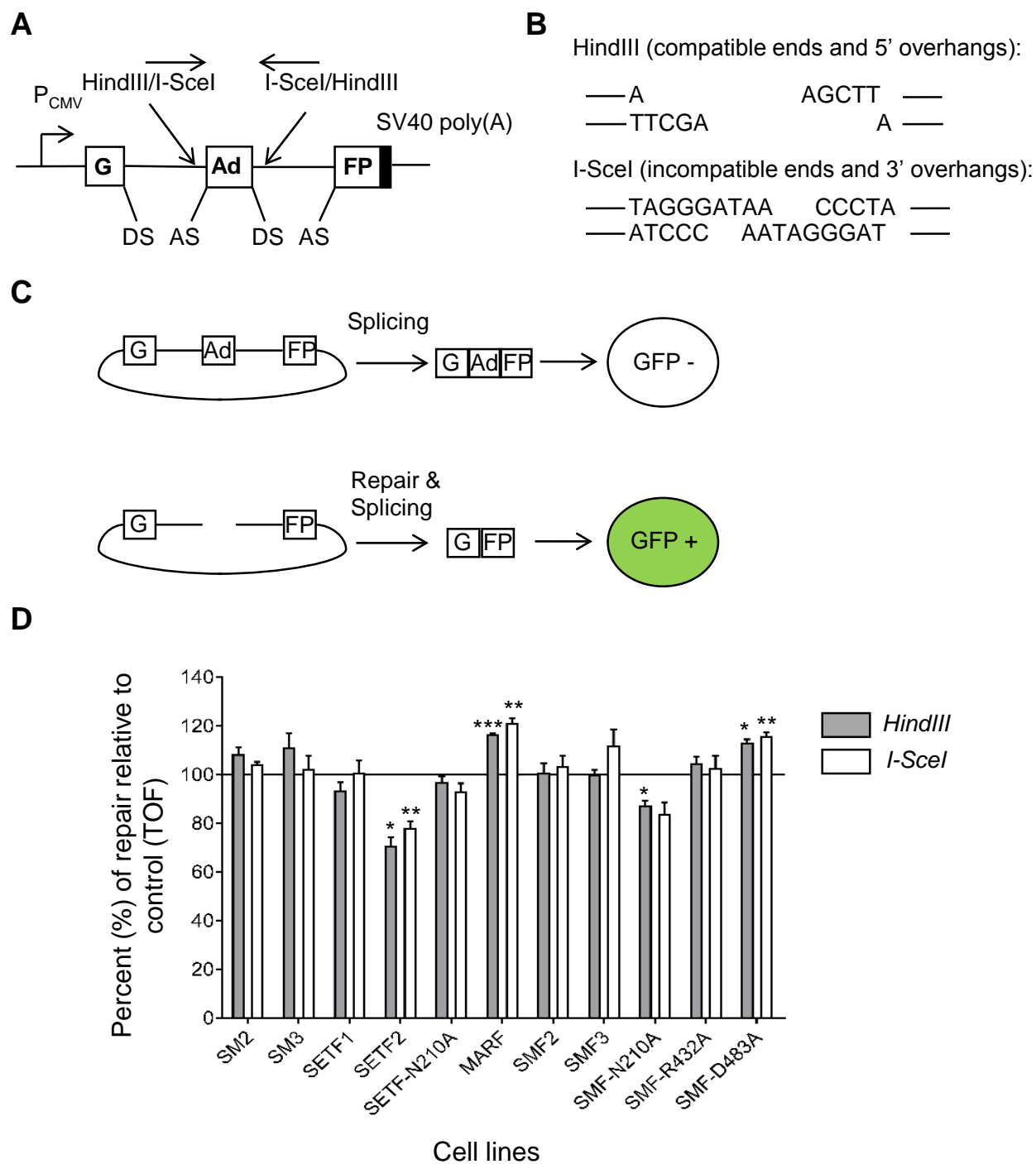G — — FP   Repair &
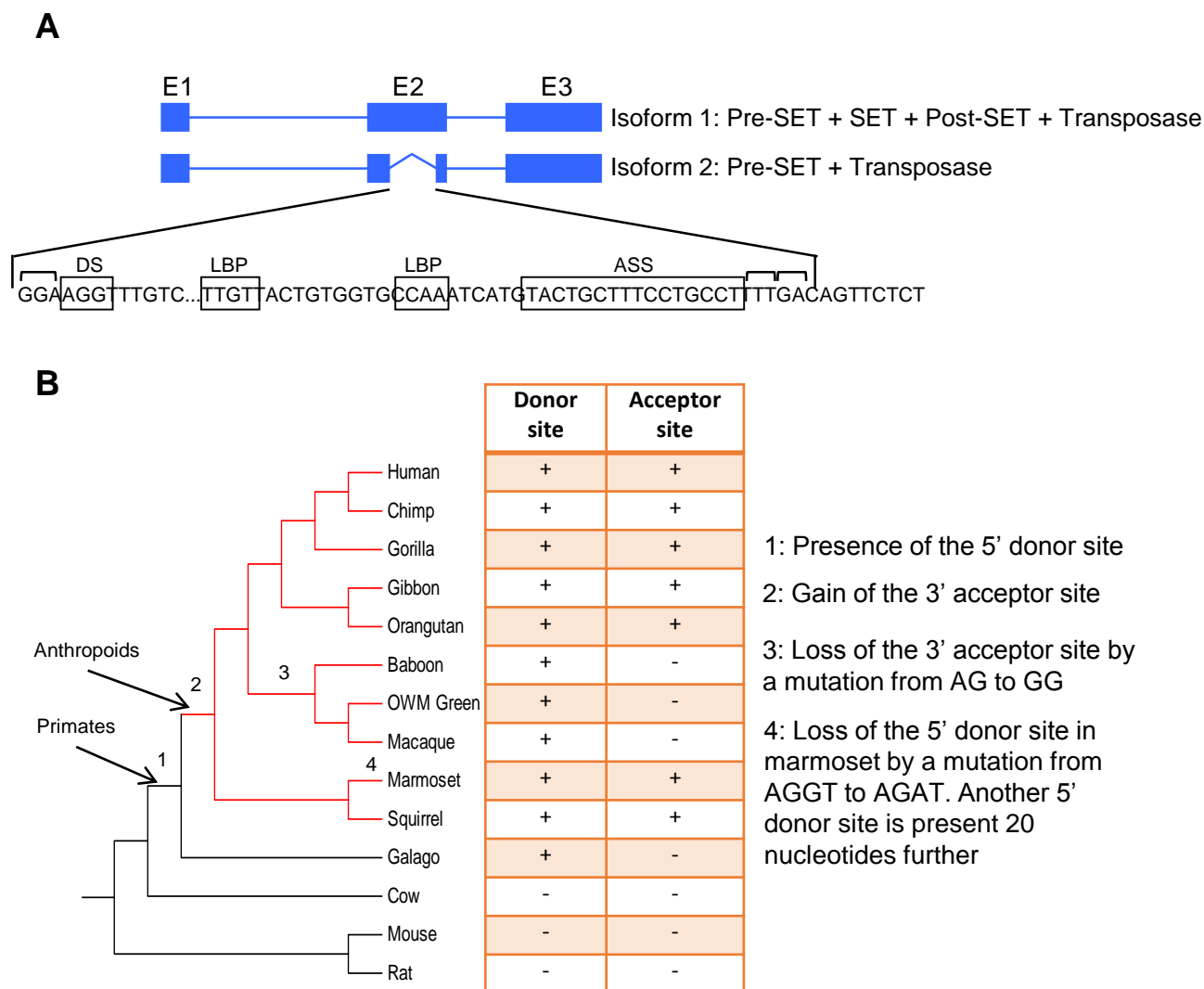Splicing   G FP ——→   GFP +

D

Cell lines

Fig 4. The SET and the MAR domains of SETMAR have an opposite effect on DNA repair by the NHEJ pathway.

A, The reporter construct, pEGFP-Pem1-Ad2, is composed of a GFP cassette flanked by a PCMV promoter and a SV40 poly(A) sequence. The GFP coding sequence is interrupted by a 2.4 kb intron containing an adenovirus exon (Ad). The Ad exon is flanked by HindIII and I-SceI restriction sites. The donor (DS) and acceptor (AS) splicing sites are shown.

B, HindIII and I-SceI restriction sites are respectively composed of a palindromic 6-bp and a non-palindromic 18-bp sequence. Digestion of the reporter construct by HindIII or I-SceI generates respectively compatible and incompatible ends.

C, The presence of the Ad exon in the GFP ORF inactivates the GFP activity thus making the cell GFP negative. Removal of the Ad exon by HindIII or I-SceI followed by a successful intracellular repair will restore the GFP expression that can be quantified by flow cytometry. Adapted from (28).

D, DNA repair efficiency of a linearized plasmid with compatible (HindIII) or incompatible (I-SceI) ends in the different cell lines relative to the control cell line (TO or TOF). Average ± S.E.M. of 3 biological replicates. Statistical test: paired t-test, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

A

E1   E2   E3

Isoform 1: Pre-SET + SET + Post-SET + Transposase

Isoform 2: Pre-SET + Transposase

DS   LBP   LBP   ASS
GGAAGGTTTGTC...TTGTTACTGTGGTGCCAAATCATGTACTGCTTTCCTGCCTTTTGACAGTTCTCT

B

	Donor site	Acceptor site
Human	+	+
Chimp	+	+
Gorilla	+	+
Gibbon	+	+
Orangutan	+	+
Baboon	+	-
OWM Green	+	-
Macaque	+	-
Marmoset	+	+
Squirrel	+	+
Galago	+	-
Cow	-	-
Mouse	-	-
Rat	-	-

Anthropoids

Primates

1: Presence of the 5' donor site

2: Gain of the 3' acceptor site

3: Loss of the 3' acceptor site by a mutation from AG to GG

4: Loss of the 5' donor site in marmoset by a mutation from AGGT to AGAT. Another 5' donor site is present 20 nucleotides further

Fig 5. SETMAR second isoform is specific to the anthropoid primates.

A, The human SETMAR gene encodes two major isoforms, the full-length protein (isoform 1) and a truncated protein (isoform 2). The second isoform is methyltransferase deficient because of the deletion of the majority of the SET and post-SET domains. Canonical donor site (DS), lariat branch points (LBP), and acceptor splicing site (ASS) are present in the second exon of SETMAR. The top brackets represents the exon codons.

B, Phylogenetic tree of SETMAR second exon in several mammals. The 5' donor site is found in all primates except for the marmoset (see 4) but should have appeared before the appearance of primates because of his presence in several non-primates mammals. The 3' acceptor site is specific to anthropoid primates except for the old-world monkeys which lost it with a single point mutation.

Figure 5