Model system to analyze RNA-mediated DNA repair in mammalian cells

Lisa Tschage, Eric Kowarz, Rolf Marschalek

Institute of Pharmaceutical Biology, Goethe-University, Frankfurt am Main, Germany.

*** corresponding author:**

Prof. Dr. Rolf Marschalek
Inst. Pharmaceutical Biology
University of Frankfurt
Max-von-Laue-Str. 9
60438 Frankfurt/Main, Germany
Tel: +49-69-798-29647; Fax: +49-69-798-29662
E-mail: rolf.marschalek@em.uni-frankfurt.de

Abstract (196 words)

"RNA-templated/directed DNA repair" is a new biological mechanism that has been experimentally demonstrated in bacteria, yeast and mammalian cells. Different RNAs or artificial DNA/RNA hybrid molecules have been used to study their role in DNA double-strand break (DSB) repair. Recent work also demonstrated that small non-coding RNAs (DDRNAs produced by DICER, Drosha) and/or newly RNAPII transcribed RNA (dilncRNA) are orchestrating the initial steps in DSB repair processes, giving another hint for the involvement of RNA in DNA repair processes. Here we demonstrate that pre-mRNA molecules could be used for DSB repair. Our mammalian cell culture system is based on 3 components: (1) a mutated reporter gene producing an intron-containing pre-mRNA, (2) an sgRNA-guided dCas13b::ADAR RNA editor specific for this pre-mRNA, and (3) I-SceI to create a transient DSB situation. We were able demonstrate that the ADAR-edited pre-mRNA is being used *in cis* for DSB repair, thereby converting the mutated reporter gene into an active reporter gene due to reconstituted splicing. We also used overexpression and knock-down of several proteins to investigate their putative role in this novel RNA-mediated DNA repair (RmDR) pathway. Based on our data, RNA can be used as template for DNA repair processes.

Key words: RNA-mediated DNA repair, CRISPR/Cas13b, RNA-editing, nucleotide transition, DNA damage

1. Introduction

The idea that RNA may influence DNA repair processes originated in the late 90's when several publications demonstrated the existence of "non-genomically encoded fusion transcripts" (NGEFTs) in peripheral mononuclear cells from healthy donors (1,2). These authors demonstrated e.g. the presence of pro-neoplastic fusion transcripts, like *BCR::ABL* or *MLL::AF4*, in the peripheral white blood cells (WBCs) of healthy individuals. None of these cells exhibited a corresponding chromosomal translocation, indicating that these NGEFTs must have been created by a different and yet unknown mechanism.

In the following years, more of these NGEFTs were identified: *MLL::AF4* or *MLL*-partial tandem duplications (PTDs) (3,4), *BCR::ABL* (5), *TEL::AML1* and *AML1::ETO* (6), *PML::RARα* (7), *NPM::ALK* and *ATIC::ALK* (8,9) in WBCs of healthy individuals, or *JAZF1::JJAZ1* in endometrial tissue (10,11) and *SLC45A3::ELK4* in prostate cells (12). So, the question arose, how these pro-neoplastic fusion transcripts are being created in a tissue-specific manner.

A decade later, experimental investigations revealed that all these NGEFTs are arising due to a specific property of these genes involved in chromosomal translocations (CTs): they exhibit the ability to produce premature terminated transcripts (PTTs) that occur in specific introns (e.g. *MLL*-intron 9, *AF4*-intron 3, *AF9*-intron 5, *ELL*-intron 2, *ENL*-intron 3, *ETV6*-intron 5, *EWSR1*-intron 7, *NUP98*-intron 12, *NUP98*-intron 13 and RUNX1-intron 6). Interestingly, all these identified introns were part of already classified "break point cluster regions" (13,14). Noteworthy, other genes that are usually not involved in the formation of CTs did not display this feature (e.g. *GAPDH*, *ACTB*, *HSPCB*, *CCND3, RPL3* or *MLL2*). Moreover, genome-wide studies revealed that about 16% of all human genes (15) have the capability to produce intronic PTTs (reviewed in 16).

Depending on the investigated tissue, these intronic PTTs give rise to specific trans-splicing events that result in the above mentioned pre-neoplastic NGEFTs (13,14). This is due to the fact that transcripts - prematurely terminated within a certain intron - carry an unsaturated splice donor site. This unsaturated splice donor is able to splice either *in cis* or *in trans* into other transcripts that derive from the same (causing transcripts with exon repetitions) or other genes transcribed in vicinity (trans-spliced NGEFTs). Genes found to be involved in the creation of CTs are even co-transcribed in the same transcription factory (17,18). This is due to the fact that chromatin loops, where these genes are residing, are in relatively close proximity within the 3-dimensional architecture of the interphase nucleus (19-21). Chromosome territories (interphase chromosomes) which exhibit these chromatin loops on their surface are shaped in different tissues in different manners, which explains why NGEFTs are always produced in a tissue-specific fashion.

This led to the hypothesis that these trans-spliced NGEFTs could be the molecular source for the onset of CTs in case of a DNA damage situation, and thus, argues for an "RNA-directed DNA repair" as the underlying genetic mechanism for the onset of CTs in case of DSB repair situations (13,14). This has been substantiated in a recent publication that demonstrated that CT only occur during DSB repair - via the NHEJ DNA repair pathway - if these NGEFTs are present (22). Without NGEFTs, DNA lesions were repaired accordingly via NHEJ, and no CT can be identified even when PCR amplifying technologies were used for their potential detection. This important result raised the general question, whether RNA molecules could be *per se* involved in DNA repair processes.

Earlier work seems to support this notion that RNA is a necessary component of repair mechanism when analyzing different DNA repair pathways. DICER and DROSHA are recruited to DSB sites in order to produce "DNA damage-induced small RNAs" (DDRNAs) that are initially required at the DDR foci to initiate a repair process by recruiting the MRE11-RAD50-NBS1 (MRN) complex. This has been shown first in mammalian cells and in the zebra fish system (23), and was later on confirmed in plants (24). In addition, RNA polyemerase II (RNAPII) is recruited to the MRN complex and starts to synthesize "damage-induced long non-coding RNAs" (dilncRNAs) that act a as source for further DDRNA formation, but also for the recruitment of the "p53-binding protein 1" (53BP1) to initiate the DNA damage response (DDR) pathway (25). In the yeast system, RNA can be reverse transcribed into complementary deoxyribonucleic acid (cDNA) and then acting as a template for homologous recombination (HR) (26). In *Schizosaccharomyces pombe* system, RNA-DNA hybrids are necessarily involved in the HR pathway which are then resolved by RNAse H to maintain genome stability (27). Storici and co-workers were able to show in the yeast system that oligo-nucleotides containing hybrid RNA/DNA sequences - complementary to both ends of an induced double-strand break (DSB) - can be directly used as template for homologous DNA repair (RNA-templated DNA repair) (28). Similar results were achieved in bacteria and in a human cell line (29,30). So, the question was posed if RNA is not only recruited as short DDRNAs or long dilncRNA necessary for protein complex recruitment at DSB foci, but can also be used as a direct template for DNA repair processes, and thus, to reverse the biological dogma of information flux. If it could be demonstrated in an experimental fashion that RNA is used as direct template for DNA repair processes, then nuclear RNA is a powerful source for maintaining genome integrity.

In order to validate this new concept of "RNA-mediated DNA repair" (RmDR), we established a mammalian cell culture model system. For this purpose, we designed a

defective EGFP reporter gene that contains an unspliceable intron because of a single point mutation in the 3'-splice acceptor site. This reporter gene was stably integrated into the genome via the Sleeping Beauty transposon technology (31) and expresses a pre-mRNA in a constitutive fashion. We then used an already established dCas13b::ADAR fusion protein to RNA-edit the reporter gene-derived pre-mRNA (32). Subsequently, we could demonstrate that these edited pre-mRNA transcripts were able to revert the mutation in the genomic DNA of the defective reporter gene after inducing DSBs in vicinity by the endonuclease I-SceI (33). Since the RNA-editing process is already leading to a spliceable form of the existing pre-mRNA that is being translated into functional GFP, only a small portion of the unspliced pre-mRNA was still left to participate in the nucleotide transitions process at the genomic DNA. Therefore, we could not expect a high frequency of stable genetic revertants. In our experiments, the observed frequency was in the range of 5 x 10^{-5} for the correction of the mutated splice acceptor site. However, we were able to increase the frequency of reverting this specific mutation within the genomic DNA by about 3 to 8-fold when we overexpressed either RAD50 or RAD51 in combination with PolQ (1.5 - 4 x 10^{-4}). Based on our data, we propose that nuclear hnRNA - or processed forms thereof - is exhibiting an interesting and novel feature, namely to support the fidelity of DNA repair processes by being the *in cis* template. This could be quite important for genome integrity and the generation of SNPs to diversify the gene pool. Our findings will be discussed in the context of existing data from other laboratories.

2. Material and Methods

2.1 Plasmid cloning

The positive control was designed by using two halves (489 bp + 231 bp) of the EGFP gene (Addgene #60511) in which we inserted an artificial mammalian intronic sequence (80 bp) that consisted of the 5'-portion of mouse *Lonp1* intron 17 (encoding the 5'-splice donor site; 30 bp) which was fused to an inverted 18 bp long I-SceI recognition site (5'-ATTACCCTGTTATCCCTA-3') and a slightly modified 3'-portion of mouse *Safb* intron 11 where we created a perfect branch A nucleotide consensus sequence (5'-YTNAY-3'; 34) together with the original splice acceptor site (32 bp). The same construct was used to clone a variant 3'-portion of mouse *Safb* intron 11 in order to insert a single point mutation that destroyed the splice acceptor site of mouse *Safb* intron 11 (last 3 nucleotides of intron: CAG -> CAA). This variant reporter gene was used throughout our studies and produces stably unspliced pre-mRNA.

For the cloning of both intron-containing reporter plasmids (positive control and defective reporter) the following primers were used: GFP.sfi.F 5'-GGCCTCTGAGGC-CACCATGTTGAGCAAGGGCGAGGAGCTGT-3', GFP.sfi.R 5'-GGCCTGACAGGCC-TTACTTGTACAGCTCGTCCATGCCGAGA-3', In.GFP.F 5'-TGACCAATTACCCT-GTTATCCCTATTTCTGACTTCATTTCTTCCCTCTGCTTACAGGTGAACTTCAAGAT CCGCCACA-3', In.GFP.Fm 5'-TGACCAATTACCCTGTTATCCCTATTTCTGACTTC–ATTTCTTCCCTCTGCTTACAAGTGAACTTCAAGATCCGCCACA-3'.
During our optimization experiments with the reporter gene, the Blasticidin resistance gene was cloned behind the EGFP reporter gene, separated by a P2A auto-cleavage domain. This allowed the selection of positive clones. The final construct was cloned into the optimized Sleeping Beauty transposon vector backbone pSBbi-RP (Addgene #60513) via the flanking SfiI sites.

The following primers were used to fuse the BSD resistance gene in-frame to the 3'-end of the EGFP reporter gene: GFP.P2A.R 5'-CGTCACCAGCCTGCTTCAGCAGGC-TGAAGTTAGTAGCTCCACTGCCCTTGTACAGCTCGTCCACGCCGAGAGTGAT-3', BSD.P2A.F 5'-GCCTGCTGAAGCAGGCTGGTGACGTCGAGGAGAATCCTGGCCC-TTTGTCTCAAGAAGAATCCACCCTCATTGA-3', BSD.sfi.R 5'-GGCCTGACAGGCC-TTAGCCCTCCCACACATAACCAGAGG-3'.
The single guide RNA (sgRNA) cassettes for criGFP and crNT were cloned into the PspCas13b crRNA backbone vector (Addgene #103854), digested with *Bbs*I after annealing the respective oligonucleotides with each other: crGG_iGFP.F 5'-CACCG-GATCTTGAAGTTCACCTGTAAGCAGAGGGAAGAAATGAAGTCAGAAATAG-3', crGG_iGFP.R 5'-CAACCTATTTCTGACTTCATTTCTTCCCTCTGCTTACAGGTGAA-CTTCAAGATCC-3', crGG_NT.F 5'-CACCGGTAATGCCTGGCTTGTCGACGCATAG-TCTG-3', crGG_NT.R 5'-CAACCAGACTATGCGTCGACAAGCCAGGCATTACC-3', respectively.
The plasmid dPspCas13b::ADAR2$_{DD}$(E488Q/T375G) (Addgene #103870) was used to create a vector construct expressing the dCas13b::ADAR fusion protein. For our purposes, we replaced the original NES (5'-CTGCCTCCACTTGAAAGACTGACAC-TG-3') by an NLS (5'- CCAGCAGCCAAGCGAGTAAAACTCGAC-3') to guarantee the nuclear localization. of the fusion protein. For the validation that the dCas13b-guided RNA editing system is indeed localizing within the nucleus, we replaced the ADAR portion by a functional GFP gene, resulting in the dPspCas13bNLS::GFP vector construct (see Supplemental **Fig S1**). As backbone for these constructs, we used the Sleeping Beauty transposon vector pSBtet-BH (Addgene #60499).

The I-SceI gene was amplified from the plasmid pCBASceI (Addgene #26477) by using the following oligonucleotides: ISceI.sfi.F 5'-GGCCTCTGAGGCCACCATGAAAAACA-TCAAAAAAAACCAGGTAATGAACCTGGGT-3' and ISceI.sfi.R 5'-GGCCTGACAGG-CCTTATTTCAGGAAAGTTTCGGAGGAGATAGTGTTC-3'. Subsequently, the I-SceI nuclease was cloned upstream of the dPspCas13b-NLS::ADAR gene cassette, again separated by a P2A auto-cleavage domain. For this purpose, the following oligo-nucleotides were used: ISceI.P2A.F 5'-ACCACCTGGGTAACCTGGTAATCAC-3', ISceI.P2A.R: 5'-GTCACCAGCCTGCTTCAGCAGGCTGAAGTTAGTAGCTCCACTG-CCTTTCAGGAAAGTTTCGGAGGAGATAGTG-3', ADAR.P2A.F 5'-AGCCTGCTGAA-GCAGGCTGGTGACGTCGAGGAGAATCCTGGCCCTAACATCCCCGCTCTGGTG-GAAAACCAGAA-3' and ADAR.P2A.R 5'-GTCAGGTCCCTGTACATCTTCAGCA-3'.
This final vector expresses a polycystronic mRNA that encodes all necessary enzymatic components when induced by 1 µg Doxycycline over 24-48 hours. This construct is named "I-SceI/dCas13b::ADAR" throughout the manuscript

To carry out the over expression experiments with RAD51, RAD52 and RAD54, we amplified them from cDNA of NALM-6 cells (DSMZ-No. ACC 128). The correct open reading frame of all transgenes were analyzed by Sanger sequencing before ligating them into the pSBbi-RH backbone via their flanking *Sfi*I sites. RAD50 was purchased from Addgene (Addgene #116784) and also cloned via flanking *Sfi*I sites in the same target vector pSBbi-RH. The oligonucleotides for cloning all 4 RAD genes were the following: Rad50.sfi.F 5'-GGCCTCTGAGGCCACCATGTCCCGGATCGAAAAGATG-AGC-3', Rad50.sfi.R 5'-GGCCTGACAGGCCTTAATGAACATTGAATCCCAGGGAG-CTA-3', Rad51.sfi.F 5'-GGCCTCTGAGGCCaccATGGCAATGCAGATGCAGCTTGA-AG-3', Rad51.sfi.R 5'-GGCCTGACAGGCCTCAGTCTTTGGCATCTCCCACTCCAT-CT-3', Rad52.sfi.F 5'-GGCCTCTGAGGCCaccATGTCTGGGACTGAGGAAGCAAT-TC-3', Rad52.sfi.R 5'-GGCCTGACAGGCCTTAAGATGGATCATATTTCCTTTTCTT-CATGTCCTGG-3', Rad54.sfi.F 5'-GGCCTCTGAGGCCaccATGAGGAGGAGCTTG-

GCTCCCAGCCAGCT-3', Rad54.sfi.R 5'-GGCCTGACAGGCCTCAGCGGAGGCCC-CGCTGCTCCTCATGAGAA-3'.

2.2 Cell cultivation and stable integration of reporter genes

The cultivation of all HEK293T cell lines (ACC 635, DMSZ) was carried out at 37°C, 5% CO_2 and a relative humidity of 95%. Cells were maintained in DMEM Low Glucose medium (DMEM-LPA, Capricorn Scientific) supplemented with 10% (v/v) FCS (FBS-11A, Capricorn Scientific), 2 mM L-Glutamine (STA-B, Capricorn Scientific), and 1% (v/v) PenStrep (PS-B, Capricorn Scientific). Cells were splitted every 2-3 days after treating them with 1 ml Accutase (ACC-1B, Capricorn Scientific) for 5 min at 37°C. The splitting ratio was usually 1:5 to 1:10 and the cell suspension was transferred onto a new culture dish together with fresh medium containing all mentioned supplements.

Stable transfections were carried out at roughly 80% confluency of HEK293T cells (ACC 635, DMSZ) cells by using Metafectene® Pro (T040-1.0, Biontex) according to the manufacture's protocol. Briefly, at least 1h prior to transfection the media had to be changed to an antibiotic free DMEM Low Glucose with L-Glutamine and FCS. For the stable integration of the defective reporter gene or the intact control gene, two solutions were prepared and filled up to 100 µl with PBS (PBS-1A, Capricorn). The first one contained 1 µg of a DNA mixture composed of 10 ng of the respective plasmid DNA, 890 ng of the filler plasmid DNA (not integrated) psfiMCS01ΔBamHI+MCS, kindly provided by Eric Kowarz, and 100 ng Sleeping Beauty transposase vector SB100X (31). The second solution contained 6 µl Metafectene® Pro. Solution 1 was added to solution 2, incubated for 20 min at RT and added carefully on top of the cells. After 6h the transfection medium was removed and fresh DMEM Low Glucose medium with all supplements was added. Starting 24h later, the cells were regularly selected with 2 µg/ml Puromycin (P11-019, PAA Laboratories) for the following two weeks until all cells were red fluorescent. After having established the positive control cell line (PCL), the cell line with the stably transfected defective reporter gene was second-transfected with 900 ng of the enzymatic transgene ISceI/dCas13b::ADAR, mixed with 100 ng sleeping beauty transposase vector SB100X. The selection was carried out with 300 µg/ml Hygromycin B (Invitrogen) until the cells were all red and blue fluorescent. The resulting reporter cell line (RCL^{mut}) was used throughout all the subsequent experiments.

2.3 RmDR (RmDR) and validation of DNA repair at the genomic level

Twenty-four hours after seeding 4 x 10^5 the RCL^{mut*} cells into 6-well plates, transient transfection experiments were carried out by using 2 µg targeting or non-targeting sgRNA expressing vector DNA. Six hours after the transient transfection, the transfection medium was exchanged with fresh growth medium containing 1 µg/ml Doxycyline. The cells were splitted into two wells after 18h into fresh medium without Doxycycline and subsequently microscopically screened for GFP positive cells after another 24h. Briefly, 588 pictures were taken of each well by using the Fluorescence Microscope Observer Z1 (Objective: Epiplan-Neofluar 5x/0.15, Carl Zeiss) and checked for GFP positive cells. The total number of GFP-positive cells within each sample was determined by counting manually all GFP positive cells on each of the 588 pictures. After this screening, the total living cell count was determined for each well after an Accutase treatment in the TC10 automated cell counter (BioRad). The cells were then put back onto a new culture dish and selected with 300 µg/ml Blasticidin for the next 2-3 weeks until only GFP positive cells were present. The reason for the

medium change after 18h is due to the fact that the estimated half-life of I-SceI is 12 - 24h (35)

For the statistical data analysis, all these experiments were carried out at least four times and the number of GFP-positive cells within 10^6 cells was determined for each sample. The numbers were normalized to the sample treated with the nt sgRNA and a T-test was carried out to prove the statistical significance.

2.4 Validation of RmDR at genomic DNA level by sequencing

After Blasticidin selection of the repaired RCL^{mut*} cells for 2-3 weeks, the GFP-positive cells were used to isolate genomic DNA. For this purpose, cells were harvested with an Accutase treatment (see above) and pelleted by a centrifuge step at 800 rpm for 5 minutes. The cell pellet was resuspended in 200 µl RSB buffer (10 mM Tris/HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.5% Triton X100 and incubated for 3 min on ice for cell lysis. The cell nuclei were pelleted by a centrifugation step at 14,000 rpm for 30 sec, and the supernatant was discarded. The pellet of nuclei was resuspended in 200 µl RSB buffer and 1.5 µl RNAse A (10 mg/ml) was added. After incubation for 1h at 37°C, 200µl TENS buffer (100 mM Tris/HCl pH 8.0, 400 mM ETDA pH 8.0, 200 mM NaCl, 1% SDS, 50 µg/ml Proteinase K) was added and incubated for 1h at 55°C. Another 110 µl 5 M NaCl was added and the sample was swiveled for 30 sec. The sample was centrifuged for 10 min at 14,000 rpm and the supernatant was transferred into a new cap and supplemented with 10 µl 3M sodium acetate pH 7.0. The DNA was cleaned with the DNA Clean & Concentrator Kit 25 (#D4006, Zymogen). Briefly, the complete sample was loaded onto a Zymoclean spin-column. Then, the sample was centrifuged for 1 min at 14,000 rpm. The column was washed with 750 µl PE buffer and the column was dried by centrifuging for 1 min at 14,000 rpm. Finally, the DNA was eluted with 30 µl EB buffer from the column.

For proving RmDR at the DNA level, an aliquot of the isolated genomic DNA (500 ng which equals ~150.000 genomes) was directly sequenced by Sanger sequencing with the primer SA.SF 5'-TTCAAGGAGGACGGCAACAT-3'.

2.5 RAD protein overexpression and siRNA knock-down experiments

For the transient overexpression of PolQ and all 4 RAD genes (RAD50, RAD51, RAD52 and RAD54) within the established RCL^{mut*} (ca. 50-60 % confluent), 1µg of all plasmids were transfected with the transfection reagent Metafectene® Pro. Briefly, $4x10^5$ stably transfected reporter cells were seeded into a 6-well plate. After 24h they were transiently transfected with a mixture of 1 µg PolQ plasmid (Addgene #73132) and 1 µg of one respective RAD plasmid. Six hours after that, the cells were transiently transfected with 2 µg of the corresponding sgRNA vectors and following another 6h post transfection the cells were treated with 1 µg/ml Doxycycline to induce I-SceI and the dCas13b::ADAR RNA editor. The cells were expanded onto two wells after 18h and after another 24h the wells were microscopically screened for GFP positive cells and treated as mentioned above.

For the transient knock-down experiments, ON-TARGET plus SMARTpool siRNA's were purchased from Dharmacon and transfected according to the supplier's protocol: ON-TARGETplus Non-targeting Pool (#D-001810-10-05, Dharmacon), ON-TARGETplus Human RAD50 (10111) siRNA - SMARTpool (#L-005232-00-0005, Dharmacon), ON-TARGETplus Human RAD51 (5888) siRNA – SMARTpool (#L-

003530-00-0005, Dharmacon), ON-TARGETplus Human RAD52 (5893) siRNA – SMARTpool (#L-011760-00-0005, Dharmacon), ON-TARGETplus Human RAD54 (25788) siRNA – SMARTpool (#L-010572-00-0005, Dharmacon) The siRNA's were resuspended in 1x siRNA buffer (60 mM KCl, 6 mM HEPES pH 7.5, 0.2 mM $MgCl_2$ in ddH_2O) to gain a 50 µM stock solution. For transfection the medium of the cells was changed to an antibiotic-free one and two solutions were prepared which were filled up to 200 µL with serum-free DMEM Low Glucose with L-Glutamine. The first solution contained 10 µl of the respective 5 µM siRNA, the second included 5 µl DharmaFECT (#T-2001-03, Dharmacon) and both solutions were incubated for 5 min at RT. Following that, they were mixed with each other and incubated for 20 min at RT after which it was filled up to 2 ml with antibiotic-free medium and added on top of the cells. Experiments were carried out with the stable transfected reporter cell line. First, 1 x 10^6 stably transfected cells were being seeded into a 10 cm cell culture dish and transiently transfected after 24h with 6 µg PolQ plasmid. After 6h the cells were detached by an Accutase treatment and 4 x 10^5 cells were then seeded in a 6-well plate where they were being transiently transfected with a siRNA targeting one respective RAD after another 18h. Six hours later, the cells were transiently transfected with 2 µg of the corresponding sgRNA vectors. After another 6h post transfection, cells were treated with 1 µg/ml Doxycycline for inducing the expression of transgenes. The cells were expanded onto two wells after 18h, and after another 24h the wells were screened for GFP positive cells and treated as mentioned above.

For the statistical data analysis, all these experiments were carried out at least four times and the amount of GFP-positive cells within 10^6 cells was determined for each sample. The numbers were normalized to the sample treated with the targeting sgRNA and a T-test was carried out to prove the statistical significance.

2.6 RNA isolation, cDNA synthesis and RT-PCR

Cells were harvested with an Accutase treatment and RNA was isolated with the RNeasy Mini Kit (#74106, Qiagen) according the manufacturer's instructions. Briefly, cells were resuspended in RLN-buffer and incubated on ice for 5 min after which the samples were centrifuged at 300 x g for 2 min. The supernatant was transferred into a new reaction tube and 600 µl RLT-buffer with 10% β-mercaptoethanol was added. After vortexing 430 µl ethanol (96%) was added and gently mixed. The sample was placed on a RNeasy-spin-column and centrifuged for 15 sec at 10,000 rpm. Then, the column is washed with 700 µl RW1-buffer and centrifuged for 1 min at 10,000 rpm. To each column a mixture of 10 µl DNase-stock-solution together with 70 µl RDD-buffer was added and left incubating for 1h at RT. The column was washed twice with 500 µl RPE-buffer and centrifuged for 2 min at 10,000 rpm. Finally, the RNA was eluted with 30 µl RNase-free water from the column by centrifuging for 2 min at 1,000 rpm. If samples were not used directly, they were stored at -80°C.

After the preparation of RNA, cDNA was synthesized by mixing 1 µg RNA filled up to 7 µl with RNase-free water and 1 µl N6 primer (100 pmol) which was then incubated for 10 min at 70°C. Following that, the mixture was placed on ice for 2 min. To this mixture of RNA and N6 primer 1 µl RNasin (N2515, Promega), 4 µl buffer 5x First Strand, 0.25 µl of each dNTP (2.5 mM), 2 µl 100 mM DTT and 1 µl Superscript II (18064-014, Fisher Scientific) were added. The sample was than incubated at RT for 10 min followed by an incubation time of 1h at 42°C. Finally, 30 µl of RNase-free water was added and the sample was placed for 10 min at 70°C.

To prove for differences in gene expression after overexpression or knock-down of PolQ and RAD genes, the following primers were used to amplify the corresponding genes: Rad50.for 5'-GGAAATGGGTCAAATGCAGG-3', Rad50.rev 5'-GGCGAATGA-TGAGTGAGGCT-3', Rad51.for 5'-CCACTGCAACTGAATTCCACC-3', Rad51.rev 5'-GCTACACCAAACTCATCAGCG-3', Rad52.for 5'-TCCAGCTGAAGGATGGTTCA-3', Rad52.rev 5'-CTGTTGTGCGTTGGTCAGC-3', Rad54.for 5'-TGGTTACAGCTCTA-AGGCCC-3', Rad54.rev 5'-AGAAGTGGCGCTCTACATCC-3', PolQ.for 5'-GGAATG-GTGGTTGTGGATGA-3', PolQ.rev 5'-GGATTGGTGAAACCCTCTGA-3'. The PCR-amplified DNA fragments were visualized on an agarose gel.

2.7 Protein expression analysis using Western blot

Fourty-eight hours after transient overexpression or knock down of PolQ ± RAD genes (RAD50, RAD51, RAD52 and RAD54) within the established RCL^{mut*} cell line and Doxyxcycline induction, all cells were detached with Accutase and the protein was extracted from ~$5x10^7$ cells. The cells were lysed with the modified RIPA-buffer (1 x cOmplete Mini protease inhibitor (11836153001, Merck), 1 mM NaF, 0.3 mM Na_3VO_4, 1 mM PMSF, 3 mM β-glycerophosphate, 10 mM Sodium pyrophosphate) for 1h at -80°C. After centrifugation for 15 min at 15,000 x g and 4°C, the protein concentration within the supernatant was determined with a BCA-Assay using PierceTM BCA Protein Assay Kit (23225, Thermo Scientific) according to the manufacturer's protocol. The protein lysate was separated by using Mini-PROTEAN TGX Stain-Free Gels 4-15% (4568085, Bio-Rad). The broad range marker HiMarkTM (LC5699, Thermo Scientific) was used for size quantification. For blotting the proteins onto the Low Fluorescence Western Membranes PVDF (ab133411, Abcam) the Matsudaira buffer (10 mM CAPS and 15% Methanol in ddH2O, pH 11) was used and the tank blot was carried out at 30 V over night at 4°C. The membrane was blocked with 5% BSA in TBS-T for 1h at RT and the primary antibody (diluted 1/1,000) was applied over night at 4°C. After washing the membran with TBS-T and incubating it with HRP-conjugated secondary antibody (diluted 1/10,000) for 1h at RT, the protein was detected with the Clarity™ Western ECL Substrate (170-5061, Bio-Rad) and ChemiDoc™ XRS+ system (1708265, Bio-Rad). The following antibodies were used: anti-alpha tubulin antibody (ab15246, Abcam), anti-RAD50 antibody (ab89, Abcam), anti-RAD51 antibody (ab213, Abcam), recombinant anti-RAD52 antibody [EPR3464(2)] (ab124971, Abcam), anti-RAD54 antibody (ab11055, Abcam), anti-PolQ polyclonal antibody (PA569577, Thermo Scientific), goat anti-mouse IgG H&L (HRP) (ab97023, Abcam), goat anti-rabbit IgG H&L (HRP) (ab6721, Abcam), and monoclonal anti-β-Actin-peroxidase (A3854, Sigma-Adrich), respectively.

3. Results

3.1 Development of a cellular reporter gene assays to investigate RmDR

In order to establish an experimental system that would allow us to investigate the proposed RmDR, we constructed a positive control cell line (PCL) together with a reporter gene cell line (RCL^{mut}). The PCL was stably transfected with an EGFP trans-gene that was interrupted by an artificial intron of 80 bp (see **Fig. 1A**). The estab-lishment of the PCL was necessary in order to validate the correct and efficient splicing of the synthetic intron in the stably transfected cell line. As shown in the Supplementary **Fig. S2**, we could validate that the artificial intron was correctly spliced out, and the visual inspection of the resulting green fluorescence (EGFP) was similar to the

constitutively expressed red fluorescence deriving from the dsTomato protein encoded by pSBbi-RP vector backbone.

The RCL^{mut} was constructed by using the same intron, however, point-mutated at the last nucleotide of the intron (G->A). This point mutation was sufficient to eliminate all GFP activity in corresponding RCL^{mut} cells. For a better selection of positive clones in the experimental setting, the defective GFP reporter gene was additionally fused via a P2A autocleavage domain to the Blasticidine resistance gene. This allowed to differentiate in subsequent experiments between "only RNA-edited" transcripts (which would transiently result in green cells), but also to select for a stable conversion of inactive reporter gene (BSD^{neg}) into an active reporter gene (BSD^{pos}) (see **Fig 1B**) due to a premature translational stop codon inside the intron in case that no splicing occurs.

After stable integration into the genome, cells are constitutively red (dTom and the Puromycin resistance are constitutively expressed by the pSBbi-RP vector backbone), but exhibit no green fluorescence. We have to mention that we used in first experiments all components cloned on single plasmids for the establishment of our test system. In these transient co-transfection experiments several plasmids encoding e.g. dCas13b::ADAR, I-Sce1, non-targeting or targeting guide RNA expressing vectors were all co-transfected, but the measured variability in the amount of GFP-positive cells was too high. Therefore, we designed a second vector (pSBTet-BH) that encodes a Doxycyline-inducible I-SceI separated by a P2A autocleavage domain from the dPspCas13b-longlinker-ADAR2$_{DD}$(E488Q/T375G) fusion gene (REPAIRv2; 32). However, we exchanged the existing NES at the C-terminus of dCas13b portion with an NLS to guarantee the intracellular nuclear localization (see Supplemental **Fig. S1**). The enzymatic RNA-editing activity of the dCas13b::ADAR was estimated to be in the range of ~28% in the original publication (32). This novel vector, coding for an Dox-inducible, polycistronic transcript shortly named "I-SceI/dCas13b::ADAR", encodes constitutively expressed BFP and the selection marker Hygromycin (see **Fig. 1B**). This second vector was used to stably transfect the already existing RCL^{mut} cell line in order to establish the final reporter cell line RCL^{mut*}. The RCL^{mut*} cells exhibit two fluorescent markers (red and blue), is resistant to Puromycin and Hygromycin, but displays neither green fluorescence nor is it Blasticidin resistant. The RCL^{mut*} cells were used for all subsequently performed experiments.

3.2 Conversion of an inactive reporter gene into an active reporter gene by RNA-editing

Upon transient transfection with non-targeting or targeting sgRNAs-expressing vectors and Doxyxcyline induction, single cells in the population became green, indicating that non-spliced EGFP transcripts were now spliced due to a transitional mutation of the A-nucleotide at the 3'-end of the artificial intron into an Inosine-nucleotide (within the first 24h). However, this GFP-positivity faded away over time when the medium was exchanged against fresh medium without Doxycycline. Thus, a stable GFP positivity (after 48h) was only possible if (1) the transcribed RNA of the defective reporter gene was edited by the ADAR enzyme, and (2) the Inosine-containing and unspliced pre-mRNA is then used as a template for a subsequent DNA repair process (3) at the DSB that was transiently induced by the I-SceI restriction enzyme. The I-SceI hydrolysis occurs shortly upstream (38 bp) of the critical nucleotide of the defective splice acceptor site. The mean results of these experiments are depicted in **Fig. 1C**, where we usually observed ~40 ± 9 GFP$^+$ cells/Mio cells in sgRNA-targeted experiments (compared to the ~4 ± 1 GFP$^+$ cells/Mio cells in the non-targeting experiment; n=5).

The complete workflow is depicted in **Fig. 2** (left side), where also example images of control and targeting experiments are shown **Fig. 2** (right side, upper and lower panel). Pictures of the red and blue channel demonstrate the presence of both vector backbones in the RCL^{mut*} cells, while the green channel displays single cells where the successful RmDR is visualized by the green fluorescent cell (see **Fig. 2** lower part). Based on these experiments, the reversion rate of the defective splice acceptor into an active splice acceptor was in the range of 5 x 10^{-5}. Stable green fluorescent cells were subsequently selected in medium containing Blasticidin. The background reversion rate due to only I-SceI induction and error prone repair via the NHEJ repair pathway was in the range of 4 x 10^{-6} (see **Fig. 1C**).

To reassure that green cells are indeed genetically modified we also analyzed and sequenced the genomic DNA isolated from PCL cells, sgRNA-untreated RCL^{mut*} cells, and sgRNA-targeted and selected RCL^{mut*} cells. The results are shown in **Fig. 3**, where the primary DNA sequences is depicted for all three experimental situations. This figure shows a part of the artificial intron containing the translational in-frame stop codon (Stop), the inverted I-SceI recognition site, the branch A nucleotide consensus sequence, as well as the intact or defective splice acceptor site, fused to exon 2 of the disrupted EGPF reporter gene. Since an aliquot (500 ng) of the isolated genomic DNA was directly sequenced without prior PCR amplification, the sequence represents the profile of ~150,000 genome copies in the depicted chromatograms. Since the chromatogram of the DNA sequences does not display any double peaks within the corrected splice acceptor site, we have to conclude that the ADAR-edited RNA template was successfully used to introduce a distinct point mutation in the DNA of our reporter gene. These experiments have been repeated several times (n = 5) with identical results, indicating that we had successfully established a reproducible experimental setting. Since the non-targeting sgRNA experiments displayed reversion rates which were always 10-times lower, we assume that the A->G transition could also take place due to error-prone repair of the DNA lesion, but is at least 10-times higher when the RNA-edited mRNA was available *in cis*, and thus, used for correcting the mutated splice acceptor sequence of the defective reporter gene. Therefore, we propose that the observed gene correction occurred due to an "RmDR" mechanism in the RCL^{mut*} cells.

3.3 Enhancement of the RmDR effect

After having set-up reproducible experimental conditions to study this novel "RmDR" mechanism, we aimed to understand which cellular components are probably involved in this novel pathway. Based on data in the literature, we decided to investigate a few candidate genes. We first chose some RAD proteins, namely RAD50, RAD51, RAD52 and RAD54. In addition, we were also interested in investigating the effects of PolQ (Polymerase theta) encoded by the *PolQ* gene (hereafter named PolQ), a DNA repair polymerase for which it was recently shown to revert cellular RNA into single-stranded DNA (29) for a "theta-mediated end joining" (TMEJ) DNA repair pathway in the Drosophila system (36). Based on data in the literature, PolQ-derived repair activity is independent from the NHEJ DNA repair pathway, and probably also independent from KU70 protein, DNA Lig IV and RAD51 protein in Drosophila (37).

Therefore, we used our RCL^{mut*} cells to transfect them transiently with non-targeting/targeting sgRNA vectors in combinations with plasmids that encode either one of the 4 RAD proteins, PolQ alone or combinations of RAD protein and PolQ. *Vice versa*, we also used a knock-down strategy by using transiently transfected siRNA's against the mRNA produced from these 4 RAD genes.

As summarized in **Fig. 4A**, RT-PCR experiments were performed with an identical amount of cells to assess the transcriptional properties of all 5 genes. *PolQ* transcripts were hardly visible apart from the cells that are overexpressing the vector-encoded *PolQ* gene. In all investigated cells, knock-down (Kd) or overexpression (OE) was nicely visible. We also performed Western blot experiments to validate the over-expression at the protein level (**Fig. 4B**). Based on these expression data, the RCL^{mut*} cells expressed RAD50 and RAD52, but much less RAD51 and RAD54 protein. Overexpression always led to a much higher steady-state amount of all 4 RAD proteins, however, with RAD54 expressed at lower levels. PolQ was slightly visible in these overexpression experiments. The relative quantification of all RAD proteins and PolQ is summarized for the shown blots on the right side of **Fig. 4B**.

Next, we evaluated the frequency of GFP-positive cells under these experimental conditions. As shown in **Fig. 5**, overexpression of each of the 4 RAD proteins or PolQ alone had no significant effect on the number of GFP-positive cells, when comparing to the targeting sgRNA alone (t, n=16; t was set artificially to 100). The number of green cells was similar or less than in the non-targeting control experiments (nt, n=16). When RAD50 and RAD51 were used together with non-targeting or targeting sgRNAs, the reversion rate was again similar to the nt- and t-experiments, respectively. RAD52 and RAD54 transfections displayed no differences between nt- and t-sgRNAs. Transient transfection of PolQ with targeting sgRNAs significantly increased the reversion of RCL^{mut*} cells to GFP-positive cells by a factor of 1.5-fold (p-value < 0,05), indicating that Pol θ alone is slightly enhancing DNA repair processes after DNA damage induction via the Doxycycline-induced I-SceI enzyme. To our suprise, the combination of RAD50/PolQ and RAD51/PolQ in combination with the targeting sgRNA resulted in a 3 to 8-fold increase of GFP-positive cells (p-value is < 0,001), pointing to a putative role of these two RAD proteins in combination with PolQ for the A->G transition of the defective splice acceptor site.

Vice versa, all 4 RAD proteins were knocked-down by an siRNA-mediated approach and analyzed in the presence of overexpressed PolQ (see **Fig. 6**). Here, we observed an effect when endogenous RAD proteins were knocked-down together with targeting sgRNAs when compared to the nt- or the nt/RAD-experiments. Knocking-down the endogenous RAD50 and RAD51 protein had still a 2-fold higher number of GFP-positive cells when compared to the nt-control, while a knock-down of RAD52 or RAD54 dropped to the level of the nt-control in a significant fashion (p-value < 0,001). Only the knock-down of RAD52 resulted in a very low number of GFP-positive cells (in the range of the non-targeting experiment). Thus, an overexpression of RAD50 or RAD51 enhanced, but the absence of RAD52 abolished the proposed RNA-mediated DNA repair pathway.

4. Discussion

Here, we present our efforts to establish a mammalian test system that allows to analyze DSB repair processes in the presence of edited pre-mRNA molecules. Noteworthy, the process of establishing such a test system took quite a long time, as we first tried to use all necessary components on single plasmids in combination with transient transfection experiments, until we learned that no significant data could be obtained this way. Only if we stably integrated all components into the genome - apart from sgRNA expressing constructs - then robust and reproducible data could be obtained. The final cellular test system is composed of a (1) stably integrated reporter gene that produces an unsplicable pre-mRNA which cannot be translated into an intact

EGFP protein. Importantly, this defective reporter gene was linked via a P2A autocleavage domain to a Blasticidin resistance gene which allowed the selection of cells that had reverted a single nucleotide at the 3-end of the intronic DNA sequence. The second component is a (2) recently published dead Cas13b protein variant that was fused in frame via a linker sequence to a mutated ADAR protein with high activity and specificity (ADAR2$_{DD}$(E488Q/T375G; 32), The third component is an (3) inducible I-SceI endonuclease that is concomitantly expressed with the dCas13b::ADAR RNA editor in a Doxycycline-inducible fashion. The cell line with both stably integrated Sleeping Beauty vector constructs (EGFP reporter gene/BSD; I-SceI/dCas13b::ADAR) was named RCL^{mut*}, while a corresponding control cell line was named PCL (see **Fig. 1A,B**, Suppl. **Fig. S2**).

We have chosen this complicated system to reassure that the necessary mutation in the genomic DNA of our reporter gene could only derive from the RNA-edited pre-mRNA molecules in order to prove, validate and quantify the concept of RNA-mediated DNA repair (RmDR) in mammalian cells.

Upon transfection of vectors coding for either non-targeting or targeting sgRNAs, we were able to demonstrate the reversion of the "defective reporter gene" into an "active reporter gene" in the RCL^{mut*} cells (see **Fig. 1B**). The frequency of genetic reversion was quite low, however, reflects only the tip of the iceberg because the used dPspCas13b::ADAR2$_{DD}$(E488Q/T375G) fusion protein was shown to exhibit only an activity of 28% to revert a specific nucleotide (32). Further, the RNA-editing ADAR portion is reverting single nucleotides scattered in a region of ± 25 nucleotides, and thus, represents only a good but not perfect system when RNA editing is experimentally required. Moreover, editing the pre-mRNA at the specific A-nucleotide (into Inosine) will ultimately lead to a pre-mRNA transcript that becomes spliceable, and thus, will lead to a transient GFP-positivity. Only the few remaining non-spliced pre-mRNAs could be used for the RmDR process. We had chosen this system because the dCas13b protein exhibits virtually no binding to DNA, and thus, allowed us to perform experiments where the sgRNA/dCas13 protein complex binds specifically only to the unspliced pre-mRNA molecules produced from the defective reporter gene. Thus, the reproducible obtained 40 to 50 GFP-positive cells per million investigated cells represent a tiny but significant cell population which could be selected by using Blasticidin selection after the medium change. This experimental scheme helped us to differentiate between RNA-editing alone (transiently GFP-positive cells after 24h) and the expected reversion of the intronic mutation in the genomic DNA (stably GFP-positive cells after 48h) which could be then further selected to purity before analyzing these revertants at the genomic DNA level by direct DNA sequencing (see **Fig. 3**).

The data that we obtained in repetitive experiments using the RCL^{mut*} cell line were quite promising, as we could see a significant increase of reporter gene revertants when comparing the results obtained with non-targeting and targeting sgRNAs (~10-fold, see **Fig. 1C**). This sounds not very much, but one has to take into the account that even in the non-targeting sgRNA experiments, the I-SceI endoculease is hydrolyzing the intron which could be then error-prone repaired by the endogenous NHEJ DNA repair system. That may also cause some changes in the nucleotide sequences and probably a low amount of genetic revertants (~10^{-6}).

We also thought to enhance the process of RmDR and tested several RAD proteins (RAD50, RAD51, RAD52 and RAD54) as well as PolQ. RAD50 has been described as part of the MRN complex (MRE11, RAD50,NBS1) that has DNA melting capacity and is recruiting RNA Polymerase II (RNAPII) to DNA double stand breaks and to cause the onset of a "damage-induced long non-coding RNA (dilncRNA) transcript by RNAPII

(38). This process is even enhanced by the single-strand binding protein RPA and activates usually the ATM protein. The dilncRNA is making hybrids with the single stranded DNA end and is substantial part of the subsequent DNA repair process. In our case we hoped that the ADAR-edited reporter mRNA may substitute for this dilncRNA, and is therefore involved in the reversion of our mutant splice accpetor site. In our experiments, overexpression of RAD50 alone increased the reversion rate by a factor of 1.5 in the targeting sgRNA experiment. The combination of RAD50 and PolQ enhanced the reversion rate by a factor of about 3-fold. PolQ is known to reverse transcribe single stranded RNA into single stranded DNA (29), which might explain the observed enhancing effect. In that case, a ssDNA is probably enhancing the genetic reversion of the single nucleotide at the end of the artificial intron.

RAD51 is the human orthologue of the bacterial RecA protein, and known to bind to DNA. RAD51 is of general importance for DNA repair processes, because RAD51 is able to form long helical filaments attached to single-stranded DNA, and allows a DNA strand replacement with a base-pairing single-stranded DNA strand. In our experiments, overexpression of RAD51 alone increased the reversion rate by a factor of 1.2 in the targeting sgRNA experiment. The combination of RAD51 and PolQ enhanced the reversion rate by a factor of about 8-fold, indicating that this combination could enhance DNA repair processes via a ssDNA-intermediate as a template. Recently, RAD51 has been described to bind to RNA in order to facilitate "R-loop formation". R-loops are structures consisting of an RNA-DNA duplex and an unpaired DNA strand. They can form during transcription upon nascent RNA threadback invasion into the DNA duplex to displace the non-template strand (39). R-loop formation occurs also in telomer maintenance, e.g. when RAD51 binds to the lncRNA TERRA during R-loop formation, which is a necessary step for telomer maintenance (40). Also here, only the combination of RAD51 and PolQ, which exhibits RAD51 binding domains (41), increased the genetic reversion rate, which is a hint for a secondary pathway, which is using a reverse transcribed mRNA to increase the nucleotide transition from A to G by about 8-fold.

The recombination protein Rad52 promotes translesion synthesis and mutagenesis by non-recombinogenic functions, which include the recruitment of the PCNA ubiquitylation complex Rad6/Rad18 to chromatin. RAD52 is implicated in the repair of Methyl methanesulfonate- and UV-induced ssDNA lesions through Rad51/Rad57-dependent and independent non-recombinogenic functions (42). RAD52 has been described to bind to ssRNA, ssDNA, RNA/DNA hybrids and dsDNA, all features which might help to facilitate RmDR. In addition, RAD52 is mediating an inverse strand exchange between dsDNA and ssRNA (43) and is able to promote the interaction of RNA with DNA. But in our experiments, a slight overexpression of RAD52 had no effect on the reversion rate, neither alone nor in combination with PolQ. However, when we applied an siRNA-mediated knock-down of RAD52, the reversion rate dropped completely to levels of the non-targeting control. This indicated that RAD52 seems to have an impact for RmDR, although an overexpression of RAD52 did not increase the mutation rate.

RAD54 is a motor protein that is required for HR events and DNA repair. Upon DNA damage situation, an acetylated K515 of RAD51 seems to recruits BRD9 which in turn facilitates the interaction with RAD51, also essential for the HR process (44). The process of HR is initiated at the site of DNA breaks or gaps and involves a search for homologous sequences promoted by Rad51 and auxiliary proteins followed by the subsequent invasion of broken DNA ends into the homologous duplex DNA that then serves as a template for repair. Rad54 activities contribute to the late phase of HR, especially the branch migration of Holliday junctions (45). In our experiments, over-

expression of RAD54 had no effect on the reversion rate, neither alone nor in combination with PolQ. The siRNA-mediated knock-down of RAD54 (±PolQ) decreased slightly the reversion rate, but overall RAD54 does not seem to have an impact on this DSB repair pathway.

Thus, the concomitant overexpression of RAD50/PolQ or RAD51/PolQ had a positive effect in RmDR, while the combined knock-down of RAD52 and PolQ seem to completely abolish this activity. RAD54 had neither an effect after overexpression nor after knock-down.

In conclusion, here we present data that support the assumption that RmDR is a real existing, new mechanism that uses transcribed nuclear RNA *in cis* for potential DNA repair processes. The tested pre-mRNA alone was capable of causing the A to G (via Inosine in the preRNA) nucleotide transition, but if cells are able to reverse transcribe the RNA into a single stranded cDNA, this repair process is enhanced by 3 to 8-fold. As a matter of fact, large RNA-Seq studies have revealed that 90% of the mammalian genome is constantly transcribed into hnRNA (46), but only a fraction is coding for protein, and thus, processed to mRNA or other important RNA molecules. Most of the transcribed hnRNA molecules represent simply copies of genomic DNA, which may help to guarantee genetic integrity by the here proposed RmDR pathway. Assuming that RNA has this new function, it may support on one hand the genetic integrity of organisms, but could also be the reason for bi-allelic mutations in cancer cells and the onset of SNP formation. RNA transcription by RNAPII displays an error rate of ~10^{-4}, and thus, both sides of this medal are present in mammalian cells. To this end, the proposed RmDR mechanism is contributing new facets to cellular biology.

5. Funding

This study was supported by grants MA 1876/12-1 from the DFG, and 2020.070 from the Wilhelm Sander foundation to R.M.

6. Disclosure of interest

The authors declare that they have no competing interests.

7. Data availability

All experimental data or any additional information required to reanalyze the data reported in this paper are available from the Lead Contact. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

8. References

1. C. Biernaux, M. Loos, A Sels, G. Huez, P. Stryckmans, Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals, *Blood* **1995**, *86*, 3118.
2. F.M. Uckun, K. Herman-Hatten, M.L. Crotty, et al, Clinical significance of MLL-AF4 fusion transcript expression in the absence of a cytogenetically detectable t(4;11)(q21;q23) chromosomal translocation, *Blood* **1998**, *92*, 810.
3. G. Marcucci, M.P. Strout, C.D. Bloomfield, M.A. Caligiuri, Detection of unique ALL1 (MLL) fusion transcripts in normal human bone marrow and blood: distinct origin of normal versus leukemic ALL1 fusion transcripts, *Cancer Res* **1998**, *58*, 790.

4. C. Caldas, C.W. So, A. MacGregor, et al, Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene, *Gene* **1998**, *208*, 167.

5. S. Bose, M. Deininger, J. Gora-Tybor, J.M. Goldman, J.V. Melo, The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease, *Blood*, **1998**, *92*, 3362.

6. M. Eguchi-Ishimae, M. Eguchi, E. Ishii, et al, Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals, *Blood* **2001**, *97*, 737.

7. A.S. Quina, P. Gameiro, M. Sá da Costa, M. Telhada, L. Parreira, PML-RARA fusion transcripts in irradiated and normal hematopoietic cells, *Genes Chromosomes Cancer* **2000**, *29*, 266.

8. B. Maes, V. Vanhentenrijk, I. Wlodarska, et al, The NPM-ALK and the ATIC-ALK fusion genes can be detected in non-neoplastic cells, *Am J Pathol* **2001**, *158*, 2185.

9. M. Beylot-Barry, A. Groppi, B. Vergier, K. Pulford, J.P. Merlio, Characterization of t(2;5) reciprocal transcripts and genomic breakpoints in CD30+ cutaneous lymphoproliferations, *Blood* **1998**, *91*, 4668.

10. H. Li, J. Wang, G. Mor, J. Sklar, A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells, *Science* **2008**, *321*, 1357.

11. H. Li, J. Wang, X. Ma, J. Sklar, Gene fusions and RNA trans-splicing in normal and neoplastic human cells, *Cell Cycle* **2009**, *8*, 218

12. D.S. Rickman, D. Pflueger, B. Moss, et al, SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer, *Cancer Res* **2009**, *69*, 2734.

13. E. Kowarz, J. Merkens, M. Karas, T. Dingermann, R. Marschalek, Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer, *Am J Blood Res* **2011**, *1*, 1.

14. E. Kowarz, T. Dingermann, R. Marschalek, Do non-genomically encoded fusion transcripts cause recurrent chromosomal translocations?, *Cancers* **2012**, *4*, 1036.

15. S. Lianoglou, V. Garg, J.L. Yang, C.S. Leslie, C. Mayr, Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression, *Genes Dev* **2013**, *27*, 2380.

16. K. Kamieniarz-Gdula, N.J. Proudfoot, Transcriptional Control by Premature Termination: A Forgotten Mechanism, *Trends Genet* **2019**, *35*, 553.

17. D.A. Jackson, A.B. Hassan, R.J. Errington, P.R. Cook, Visualization of focal sites of transcription within human nuclei, *EMBO J* **1993**, *2*, 1059.

18. F.J. Iborra, A. Pombo, D.A. Jackson, P.R. Cook, Active RNA polymerases are localized within discrete transcription "factories' in human nuclei, *J Cell Sci* **1996**, *109*, 1427.

19. C.S. Osborne, L. Chakalova, K.E. Brown, et al, Active genes dynamically colocalize to shared sites of ongoing transcription, *Nat Genet* **2004**, *36*, 1065.

20. H. Neves, C. Ramos, M.G. da Silva, A. Parreira, L. Parreira, The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation, *Blood* **1999**, *93*, 1197.

21. C.S.Osborne, L. Chakalova, J.A. Mitchell, et al, Myc dynamically and preferentially relocates to a transcription factory occupied by Igh, PLoS Biol 2007, 5, e192.

22. P. Streb, E. Kowarz, T. Benz, J. Reis, R. Marschalek, How chromosomal translocation arise to cause cancer: the role of RNA-directed DNA repair, *IScience* submitted Nov 2022.

23. S. Francia, F. Michelini, A. Saxena, et al, Site-specific DICER and DROSHA RNA products control the DNA-damage response, *Nature* **2012**, *488*, 231.

24. W. Wei, Z. Ba, M. Gao, et al, A role for small RNAs in DNA double-strand break repair, *Cell* **2012**, *149*, 101.

25. F. Michelini, S. Pitchiaya, V. Vitelli, et al, Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks, *Nat Cell Biol* **2017**, *19*, 1400.

26. L.K. Derr, J.N. Strathern, A role for reverse transcripts in gene conversion, *Nature* **1993**, *361*, 170.

27. C. Ohle, R. Tesorero, G. Schermann, N. Dobrev, I. Sinning, T. Fischer, Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair, *Cell* **2016**, *167*, 1001.

28. F. Storici, K. Bebenek, T.A. Kunkel, D.A. Gordenin, M.A. Resnick, RNA-templated DNA repair. *Nature* **2007**, *447*, 338.

29. G. Chandramouly, J. Zhao, S. McDevitt, et al, Polθ reverse transcribes RNA and promotes RNA-templated DNA repair, *Sci Adv* **2021**, *7*, eabf1771.

30. Y. Shen, P. Nandi, M.B. Taylor, et al, RNA-driven genetic changes in bacteria and in human cells, *Mutat Res* **2011**, *717*, 91.

31. E. Kowarz, D. Löscher, R. Marschalek, Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines, *Biotechnol J* **2015**, *10*, 647.

32. D.B.T. Cox, J.S. Gootenberg, O.O. Abudayyeh, et al, RNA editing with CRISPR-Cas13, *Science* **2017**, *358*, 1019.

33. A. Plessis, A. Perrin, J.E. Haber, B. Dujon, Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus, *Genetics* **1992**, *130*, 451.

34. R. Leman, H. Tubeuf, S. Raad, et al, Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants, *BMC Genomics* **2020**, *21*, 86.

35 A. Morano, T. Angrisano, G. Russo, et al, Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene, *Nucleic Acids Res* **2014**, *42*, 804.

36. S.H. Chan, A.M. Yu, M. McVey, Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila, *PLoS Genet* **2010**, *6*, e1001005.

37. A.M. Yu, M. McVey, Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions, *Nucleic Acids Res* **2010**, *38*, 5706.

38. S. Sharma, R. Anand, X. Zhang, et al, MRE11-RAD50-NBS1 Complex Is Sufficient to Promote Transcription by RNA Polymerase II at Double-Strand Breaks by Melting DNA Ends, *Cell Rep* **2021**, *34*, 108565.

39. B.P. Belotserkovskii, S. Tornaletti, A.D. D'Souza, P.C. Hanawalt, R-loop generation during transcription: Formation, processing and cellular outcomes, *DNA Repair* **2018**, *71*, 69.

40. M. Feretzaki, M. Pospisilova, R. Valador Fernandes, T. Lunardi, L. Krejci, J. Lingner, RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops, *Nature* **2020**, *587*, 303.

41. A. Schrempf, J. Slyskova, J.I. Loizou, Targeting the DNA Repair Enzyme Polymerase θ in Cancer Therapy, *Trends Cancer* **2021**, *7*, 98.

42. M.I. Cano-Linares, A. Yáñez-Vilches, N. García-Rodríguez, et al, Non-recombinogenic roles for Rad52 in translesion synthesis during DNA damage tolerance, *EMBO Rep* **2021**, *22*, e50410.

43. O.M. Mazina, H. Keskin, K. Hanamshet, F. Storici, A.V. Mazin, Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair, *Mol Cell* **2017**, *67*, 19.

44. Q. Zhou, J. Huang, C. Zhang, et al, The bromodomain containing protein BRD-9 orchestrates RAD51-RAD54 complex formation and regulates homologous recombination-mediated repair, *Nat Commun* **2020**, 11, 2639.

45. A.V. Mazin, O.M. Mazina, D.V. Bugreev, M.J. Rossi, Rad54, the motor of homologous recombination, *DNA Repair* **2010**, *9*, 286.

46. ENCODE Project Consortium, E. Birney, J.A. Stamatoyannopoulos, et al, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* **2007**, *447*, 799.

Figure 1

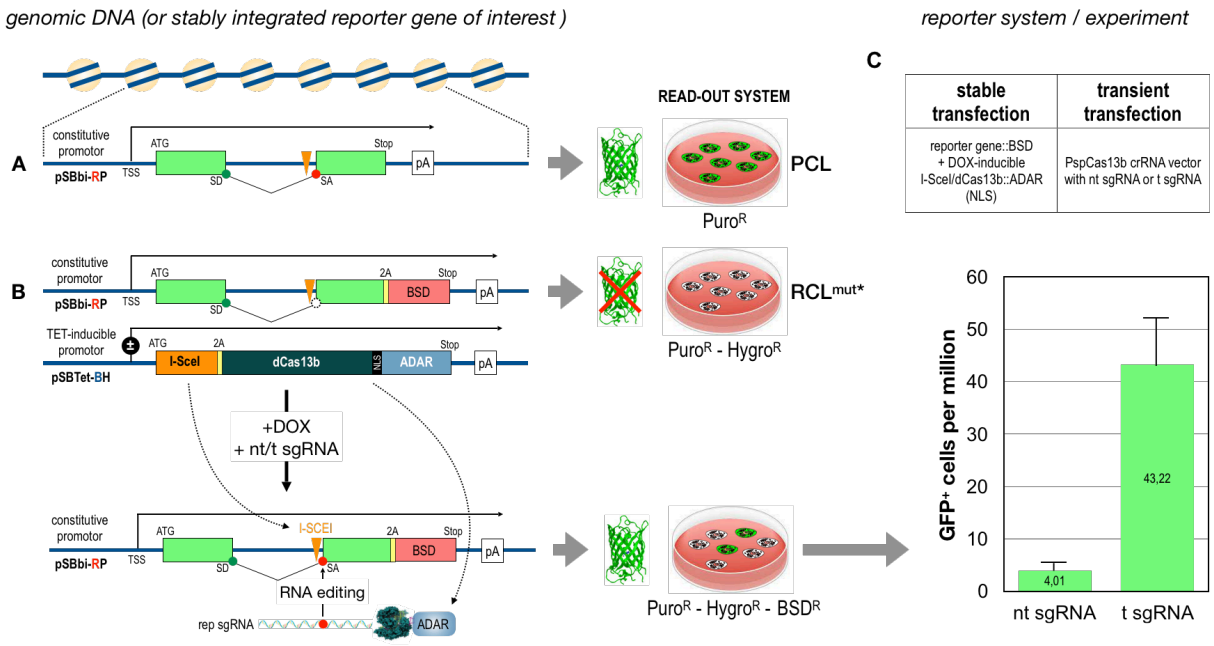


Figure 1: Experimental design of the reporter cell lines and results

A. Design of the reporter cell line PCL. The EGFP reporter gene was interrupted by a short synthetic intron (80 bp) that derived from 2 different mouse genes (5'-portion of the *Lonp1* intron 17 (30 bp) was fused to an inverted I-SceI recognition site (18 bp) and the 3'-portion of *Safb* intron 11, 32 bp). This construct was tested in in HEK293T cells to validate splicing of this intron in stably transfected PCL cells (see also Supplementary **Fig. S2**). The vector backbone produces constitutively the RFP protein and the Puromycin resistance protein (pSBbi-RP). **B.** Design of the reporter cell line RCL^{mut*}. This cell line contained a single point mutation in the 3'-splice acceptor signal sequence to inhibit splicing. In addition, the EGFP reading frame was fused via a P2A peptide with the Blasticidin resistance gene. The vector backbone produces constitutively the RFP protein and the Puromycin resistance protein (pSBbi-RP). This cell line was transfected a second time with another Sleeping Beauty vector, pSBTet-BH, that encodes an inducible cassette with the restriction enzyme I-SceI separated by an P2A sequence from the fusion protein dCas13b::ADAR. The vector backbone produces constitutively the BFP protein and the Hygromycin resistance protein (pSBTet-BH). The resulting RCL^{mut*} cell line was used for transient transfection experiments with the PspCas13b crRNA vector that produces constitutively from an U6 promoter either the non-targeting or the targeting sgRNAs. As shown in the lower part, only the targeting sgRNA will promote the RNA-editing process of the inactive reporter gene mRNA. This will result in a transient GFP-positivity (up to 24h). I-SceI-induced hydrolysis at the artificial intron will then induce a DDR situation where the RNA-edited mRNA could contribute to RmDR which then results in a stable GFP positivity (> 48h). The revertant cells were further selected by adding Blasticidin to the medium for additional 2-3 weeks. **C.** Overview of the transfections with the reporter cell line RCLmut*, with stably integrated Sleeping Beauty vectors, and transiently transfected PspCas13b crRNA vectors expressing either the non-targeting or the targeting sgRNAs. Below is the result of 5 individual experiments where the mean GFP positive cells per million cells (±SD) are given.

Figure 2

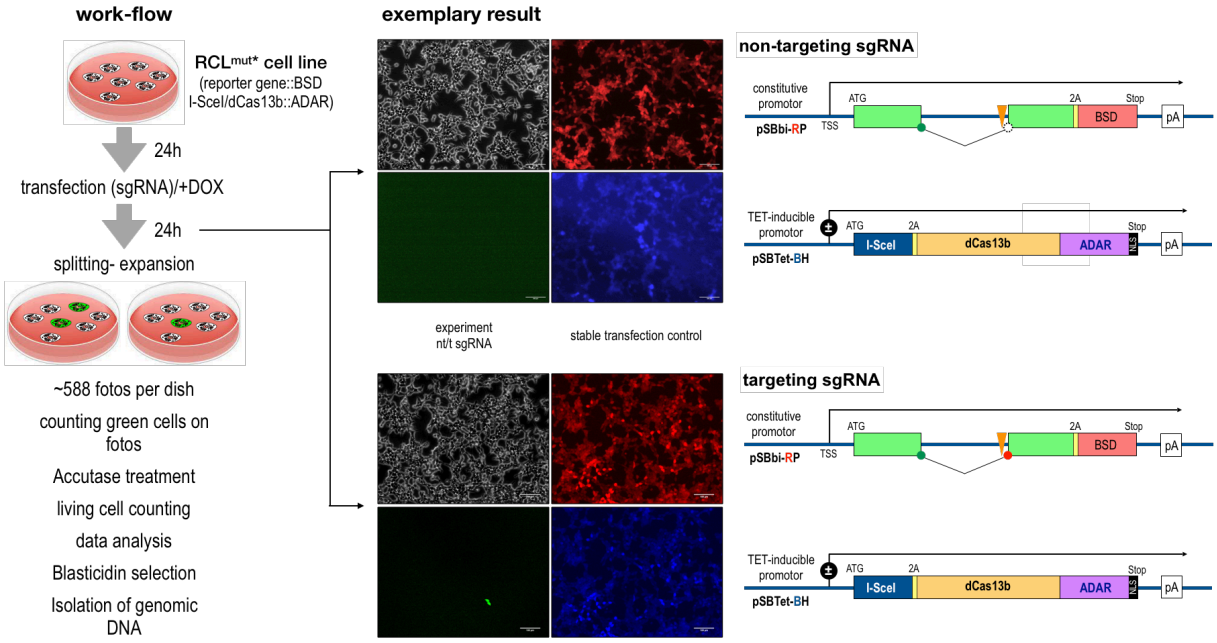


Figure 2: Work flow and results

This figure describes the work flow that has been established to determine the genetic reversion of the inactive reporter gene into an active reporter gene. The RCL^{mut}* cell line was transiently transfected with corresponding PspCas13b crRNA vectors and medium containing 1 µg/ml Doxycycline. After 24h, cells were splitted into 2 wells and expanded for another 24 h. From each well, 588 photographs were taken and GFP-positive cells were counted; afterwards, cells were detached by an Accutase treatment and total cell numbers were quickly investigated before putting the cells back to fresh medium without any Doxycycline, but containing Blasticidin for selection. After 2-3 weeks, all cells in culture become green and DNA was isolated. Middle panels: single picture series from 1 of the 588 pictures taken per well with phase contrast as well as green, red and blue channel. Red and blue channel shows the expression of RFP and GFP from the two different vector backbones, while the green channel displays a single GFP-positive revertant cell. Right panel: genetic situation in the non-targeting control RCLmut* cells, and the targeting RCLmut*cells, where in single cells the inactive reporter gene becomes back-mutated to an active reporter gene; such cells become Blasticidin resistant.

Figure 3

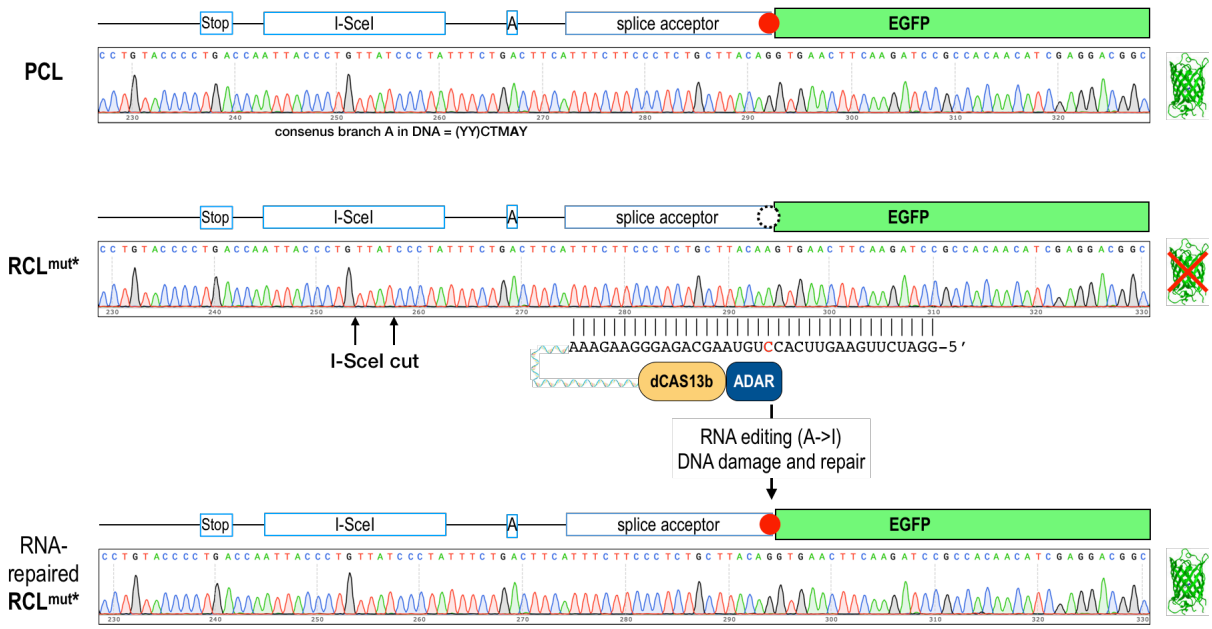
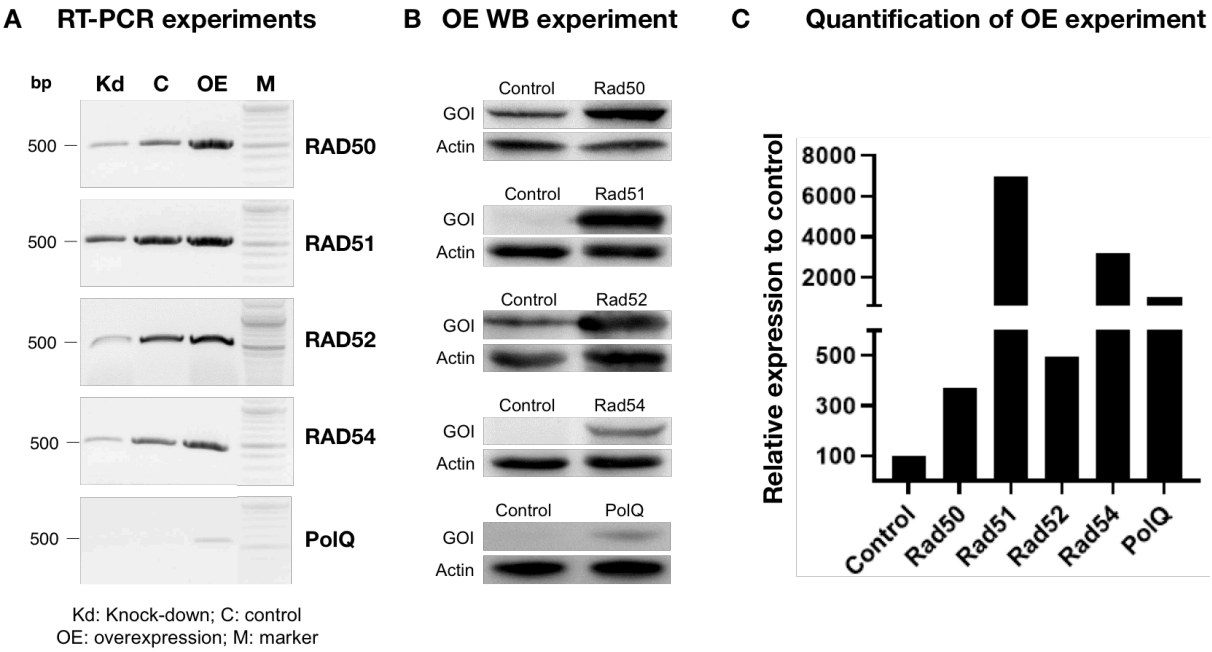


Figure 3: Sequence analysis of PCL and RCL^{mut*} cells
The figure displays the direct sequence of 500ng genomic DNA isolated from PCL and RCL^{mut*} cells, prior and after transient transfection with a vector expressing the targeting sgRNA. Upper part: the active reporter gene that is correctly spliced which results in GFP-positive cells (see also Supplemental **Fig. S2**). Middle part: situation in RCL^{mut*} cells. The defective splice acceptor was inhibiting splicing, and thus, all cells remain GFP-negative and are Blasticidin sensitive. Lower part: after transient transfection with appropriate sgRNA, the guided dCas13b::ADAR RNA-Editor is converting the A-nucleotide into an Inosine. The edited mRNA is now able to splice correctly and could be used in DNA damage situations for an RmDR process. This will lead to the genetic conversion of the intronic DNA sequence of the reporter gene, resulting in stable GFP-positivity (correct splicing) and Blasticidin resistance. The sequence shown in this Figure was deriving from a single GFP-positive cell clone after Blasticidin selection for 2 weeks. We tested always the genomic DNA, with no prior PCR amplification, to reassure that the sequences display the actual genomic situation in a "quantitative fashion", because mixed cell clones would give rise to double peaks which were never observed in these experiments.

Figure 4



Kd: Knock-down; C: control
OE: overexpression; M: marker

Figure 4: overexpression and knock-down of RAD and PolQ gene

A. RT-PCR experiments of the knock-down (Kd) experiment of all 4 RAD genes, endogenous transcription of all 5 genes (c) and overexpression of all 5 genes in RCLmut* cells. Based on the transcriptional data, PolQ was poorly expressed in the reporter cells. **B.** Western blot experiment with endogenous expression and after overexpression of all 4 RAD proteins and PolQ. Here, the endogenous steady-state expression of RAD51 and RAD54 seems to be very low, eventually explained by a high turn-over of these proteins. The RCL^{mut*} cells displayed undetectable amounts of PolQ while overexpression of plasmid encoded RAD proteins and PolQ was well visible. **C.** Quantification of the overexpressed protein relative to the actin protein signal and the endogenous signal.

Figure 5

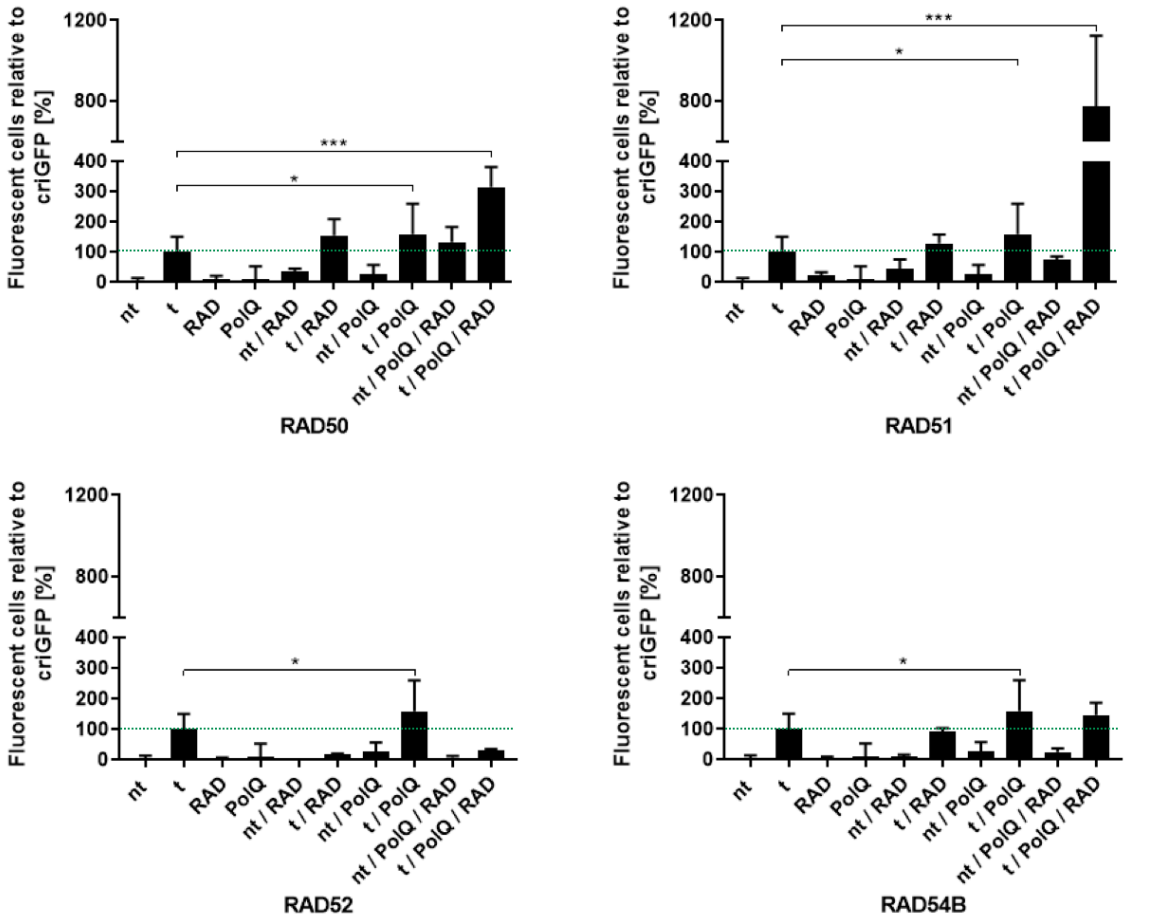


Figure 5: Results of overexpressed RAD proteins ± PolQ on the amont of GFP-positive cells

The 4 panels are the mean results of at least 4 (RAD, nt/t-RAD, nn/t PolQ/RAD) and a maximum of 16 experiments (nt/t, PolQ, nt/PolQ). Here, we investigated the increase of GFP-positive cells in the presence of either RAD50, RAD51, RAD52 or RAD54 in conjunction with PolQ. Neither the expression of RAD protein nor PolQ alone (without targeting sgRNA) did significantly influence the number of GFP-positive cells when compared to the non-targeting control. First significant effects were observed with PolQ in the presence of targeting sgRNAs (p-value < 0,05). Similar result - but not significant - were obtained with RAD50 and RAD51 again with targeting sgRNAs. The transient co-transfection of RAD50 and RAD51 with PolQ and targeting sgRNA enhanced significantly the number of GFP-positive cells (p-value < 0,001).

Figure 6

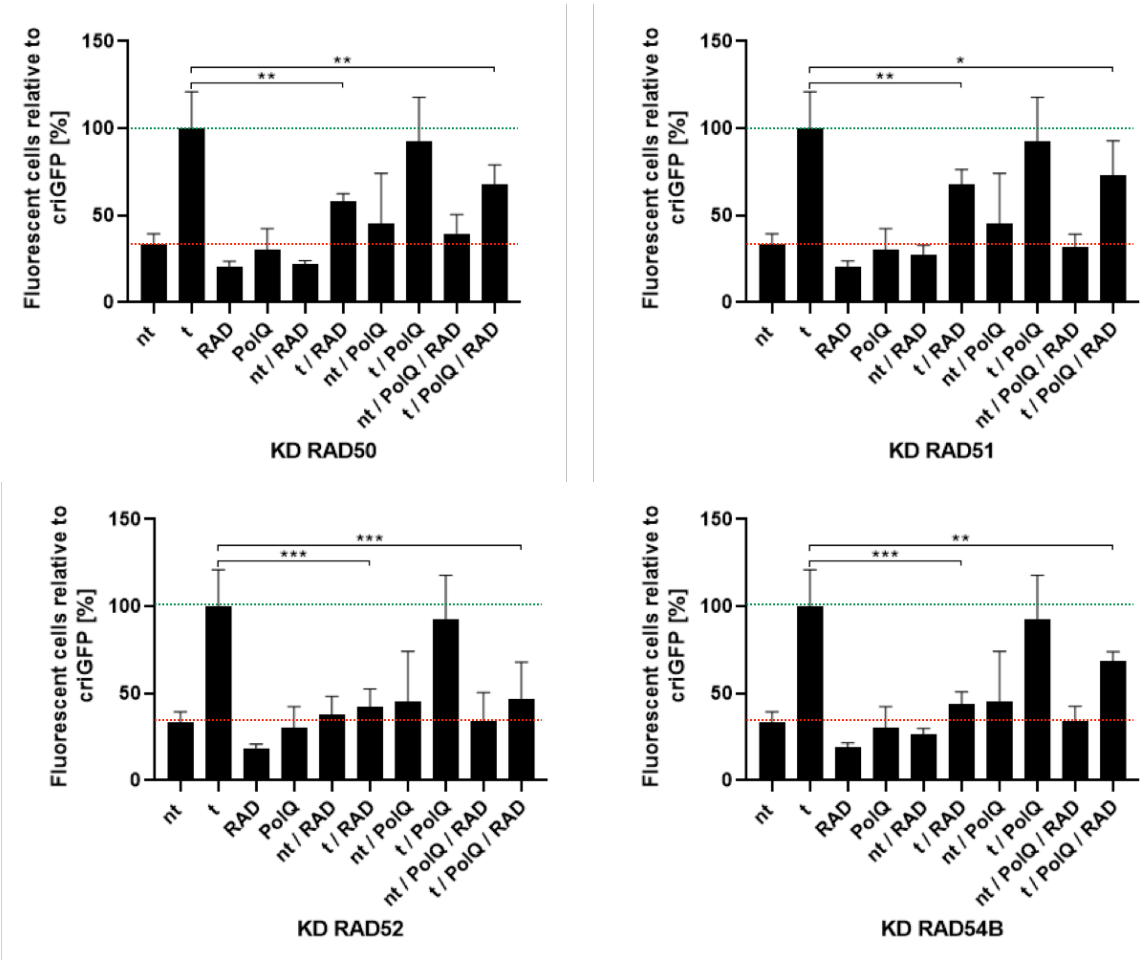
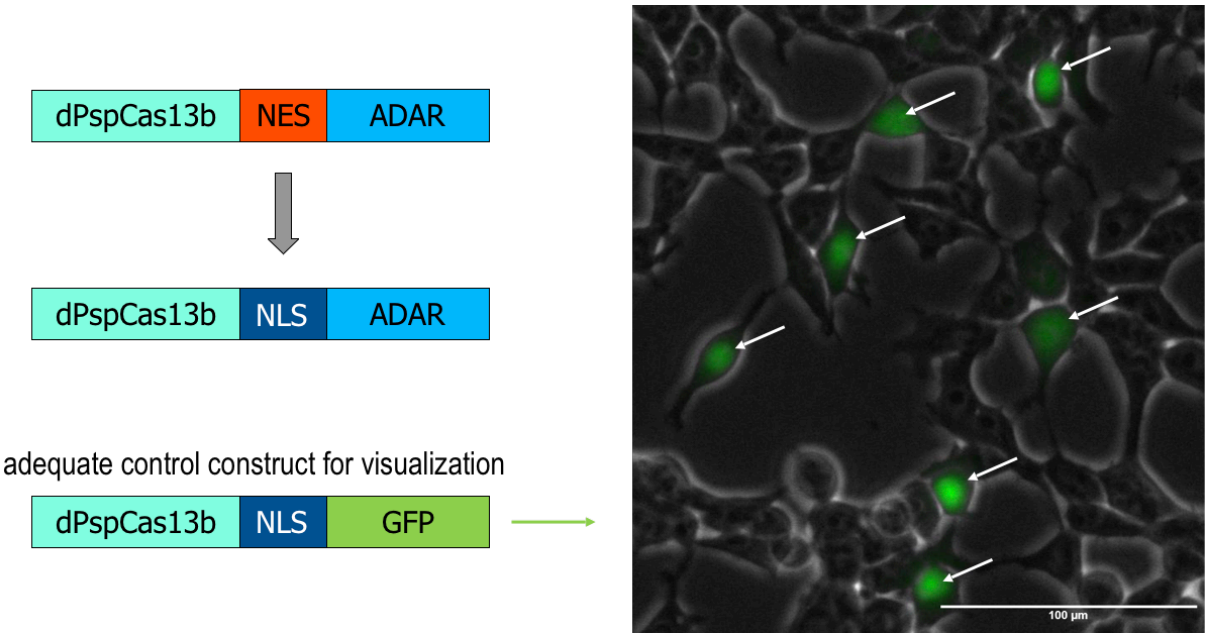


Figure 6: Results of knocking-down RAD proteins ± PolQ on the number of GFP-positive cells

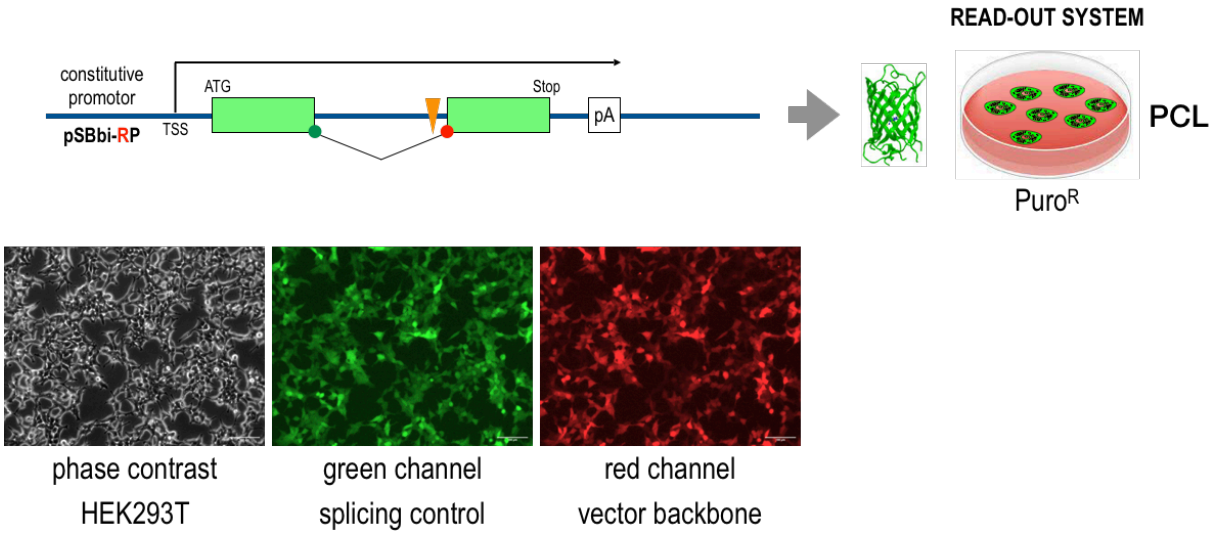
Again, the 4 panels are the mean results of at least 4 (RAD, nt/t-RAD, nt/t PolQ/RAD) and a maximum of 16 experiments (nt/t, PolQ, nt/PolQ). Here, we investigated the decrease of GFP-positive cells in the absence of either RAD50, RAD51, RAD52, RAD54 in combination with overexpressed PolQ. Neither the expression of RAD protein nor PolQ alone (without targeting sgRNA) did significantly influence the number of GFP-positive cells when compared to the non-targeting control. The only significant result was obtained when RAD52 was knocked-down, as the number of GFP-positive cells dropped to the negative control (nt). All other combinations gave significant reductions, but still GFP-positive cells were obtained higher than the nt controls.

Supplemental Figure S1



Supplemental Figure S1: Validation of the dCas13b-NLS::ADAR construct
The established dPspCas13b::ADAR2$_{DD}$(E488Q/T375G) fusion construct for RNA-editing (32) had an NES sequence at the Carboxy-terminus of the dCas13b portion. This sequence was exchanged with an NLS sequence, as we wanted to investigate RNA-editing in the cell's nucleus rather than the cytosol. For this purpose, we cloned the Cas13b-NLS::GFP fusion protein, by substituting the ADAR portion with a normal GFP reporter sequence. As shown on the right side, transient transfection into HEK293T cells display clearly the nuclear localization of the dCas13b-NLS::GFP fusion protein in transfected cells. From these data we predicted that the dCas13b-NLS::ADAR fusion will also be localized in the cell's nucleus.

Supplemental Figure S2



Supplemental Figure S2: Validation of correct splicing in the PCL reporter cell line
This validation experiment was necessary to confirm that the synthetic intron, that was artificially introduced into the EGFP gene, is correctly spliced-out and results in GFP positive cells. The reporter gene was cloned into the pSBbi-RP Sleeping Beauty vector backbone, which constitutively express the RFP and Puromycin resistance genes. A series of pictures was taken, displaying the phase contrast, as well as the green and red channel. These pictures revealed the correct splicing of the reporter gene in PCL cells. In addition, also the comparable intensities of fluorescent cells in both channels, indicated an equal expression of both the reporter gene (green) and the constitutive marker gene (red).