

# The Gene Targeting Approach of Small Fragment Homologous Replacement (SFHR) Alters the Expression Patterns of DNA Repair and Cell Cycle Control Genes

Silvia Pierandrei<sup>1,2</sup>, Andrea Luchetti<sup>3</sup>, Massimo Sanchez<sup>4</sup>, Giuseppe Novelli<sup>3</sup>, Federica Sangiuolo<sup>3</sup> and Marco Lucarelli<sup>1,2</sup>

Cellular responses and molecular mechanisms activated by exogenous DNA that invades cells are only partially understood. This limits the practical use of gene targeting strategies. Small fragment homologous replacement (SFHR) uses a small exogenous wild-type DNA fragment to restore the endogenous wild-type sequence; unfortunately, this mechanism has a low frequency of correction. In this study, we used a mouse embryonic fibroblast cell line with a stably integrated mutated gene for enhanced green fluorescence protein. The restoration of a wild-type sequence can be detected by flow cytometry analysis. We quantitatively analyzed the expression of 84 DNA repair genes and 84 cell cycle control genes. Peculiar temporal gene expression patterns were observed for both pathways. Different DNA repair pathways, not only homologous recombination, as well as the three main cell cycle checkpoints appeared to mediate the cellular response. Eighteen genes were selected as highly significant target/effectors of SFHR. We identified a wide interconnection between SFHR, DNA repair, and cell cycle control. Our results increase the knowledge of the molecular mechanisms involved in cell invasion by exogenous DNA and SFHR. Specific molecular targets of both the cell cycle and DNA repair machineries were selected for manipulation to enhance the practical application of SFHR.

*Molecular Therapy—Nucleic Acids* (2016) 5, e304; doi:10.1038/mtna.2016.2; published online 5 April 2016

**Subject Category:** gene addition deletion and modification

## Introduction

Of the various gene therapy approaches that have been explored in recent years, gene targeting has received increasing attention as compared with traditional gene augmentation. The former approach is aimed at the direct correction of genetic mutations, stably restoring the wild-type gene function *in situ* and maintaining gene regulation under the native sequence context.<sup>1–3</sup> Consequently, the targeted gene continues to be regulated by its endogenous machinery, maintaining its physiological expression pattern.<sup>3</sup> By contrast, the latter approach is based on the delivery of additional copies of the wild-type therapeutic gene, which is expressed under the control of strong exogenous promoters.<sup>4–6</sup> Gene targeting has drawn new interest due to its concrete applicability to stem cells, a primary aim of gene therapy.<sup>7</sup> In addition, approaches aimed at introducing site-specific modification by site-directed zinc-finger nucleases (ZFNs),<sup>8,9</sup> transcription activator-like effector nucleases (TALENs),<sup>8,9</sup> and clustered regulatory interspaced short palindromic repeats associated with Cas-9 (CRISPR-Cas 9)<sup>9,10</sup> promise a new revolution in gene targeting. These approaches are aimed at the generation of a site-specific DNA double-strand break and consequent site-specific error-prone nonhomologous end joining (NHEJ) or homology-directed repair. The techniques which are used to disrupt or to restore gene function have potential for therapeutic applications. However, considerable advances

in the comprehension of their mechanisms, particularly those involved in the downstream response to the induced damage, are needed before these approaches move from bench to bedside.<sup>11</sup>

Within the gene targeting possibilities, the small fragment homologous replacement (SFHR) is characterized by the use of small exogenous wild-type DNA fragments (SDFs, up to 1 kb). SFHR involves an interaction between the wild-type SDF sequence, probably used as a template, and the mutated endogenous sequence, which ultimately is corrected through a still largely undefined mechanism.<sup>12–14</sup> Generally, the correction is thought to involve homologous recombination (HR) and/or other DNA repair pathways that may function between two specific homologous DNA sequences.<sup>15</sup> In a previous study performed in mouse embryonic fibroblasts (MEF) using an enhanced green fluorescent protein (EGFP)-based mammalian reporter assay system,<sup>16</sup> we demonstrated interconnections between epigenetics, DNA repair, and cell cycle control pathways in the response to cell invasion by exogenous DNA.

SFHR has already been used to modify different types of genomic mutations *in vitro* and *in vivo* in both human and mouse primary, immortalized and stem cells, as well as in animal models, demonstrating its potential for the treatment of several disease-associated genes. These genes include cystic fibrosis transmembrane conductance regulator (CFTR, responsible for cystic fibrosis),<sup>15,17–21</sup> dystrophin (*DMD*

<sup>1</sup>Department of Cellular Biotechnologies and Hematology, Sapienza University of Rome, Rome, Italy; <sup>2</sup>Pasteur Institute – Cenci Bolognietti Foundation, Sapienza University of Rome, Rome, Italy; <sup>3</sup>Department of Biomedicine and Prevention, Tor Vergata University, Rome, Italy; <sup>4</sup>Department of Cell Biology and Neurosciences, Istituto Superiore di Sanità, Rome, Italy. Correspondence: Marco Lucarelli, Dipartimento di Biotechnologie Cellulari ed Ematologia, Sapienza Università di Roma, c/o Azienda Policlinico Umberto I, Edificio 27, III piano, Viale Regina Elena, 324, 00161 Roma, Italia. E-mail: [marco.lucarelli@uniroma1.it](mailto:marco.lucarelli@uniroma1.it)

**Keywords:** arrayed gene expression; cell cycle control; DNA repair; gene targeting; SFHR

Received 10 November 2015; accepted 12 December 2015; published online 5 April 2016. doi:10.1038/mtna.2016.2

(Duchenne muscular dystrophy), responsible for muscular dystrophies),<sup>22–24</sup> survival motor neuron (*SMN*, responsible for spinal muscular atrophy),<sup>25,26</sup> hypoxanthine phosphoribosyltransferase 1 (*HPRT1*, responsible for Lesch–Nyhan syndrome),<sup>27</sup> DNA-dependent protein kinase catalytic subunit (*DNA-PKcs*, responsible for severe combined immune deficiency),<sup>28</sup> and  $\beta$ -globin (*HBB*, responsible for  $\beta$ -thalassemia and sickle cell disease).<sup>29–31</sup>

The SFHR approach has both advantages and disadvantages. Its primary advantages are the easy production and use of an SDF without the need for complex vectors, the permanent homologous modification of the genomic sequence of interest, the inheritable and physiologically expressed modification, and the potential to correct any genetic disease (also dominant negative). In addition, the size of the gene to be corrected does not constitute a limiting factor because the targeting sequence corrects only the altered DNA portion. The greatest drawbacks of the SFHR approach are the low and variable correction efficiency, ranging from 0.01 to 5%, as well as the general limited reproducibility and the possibility of undesirable random integration of the SDF.<sup>2,15–17,23,32–35</sup> However, TALENs have recently been demonstrated to enhance SFHR efficiency.<sup>36</sup>

Many variables seem to influence SFHR, including the type and differentiation status of the target cells, the degree of chromatin condensation and the phases of the cell cycle, as well as the DNA damage repair machinery and the level of recombination activity.<sup>16</sup> In addition, the accessibility of the SDF to the target locus, the methodology of SDF delivery, the level of nuclear uptake, and the half-life of the SDF influence the result.

A serious obstacle to overcoming of main drawbacks of SFHR and thus its consequent improvement is the lack of a complete understanding of the cellular response(s) and molecular mechanism(s) triggered by exogenous DNA. Many steps involved in the correction process remain unknown.<sup>37</sup> Furthermore, even main pathways already known to be possibly involved, such as epigenetics, DNA repair, and cell cycle control, have scarcely been studied in relation to SFHR at both general and gene-specific levels. Nevertheless, before its application to clinical medicine, SFHR must become efficient, safe, and reproducible.

The aim of this work was to contribute to the clarification of molecular mechanisms underlying the cell invasion by exogenous DNA and its processing and to a better understanding of the mechanisms underlying the SFHR gene therapy approach. In particular, we focused on the primary pathways involved in host defense, genome integrity maintenance, and cell cycle control, after treatment with an exogenous SDF. Greater knowledge of the SFHR mechanism might lead experimental interventions to enhance its efficiency and safety for use in clinical protocols, particularly in light of new ZNFs, TALENs, and CRISPR-Cas 9 tools.

## Results

### SFHR-mediated modification efficiency and cellular growth

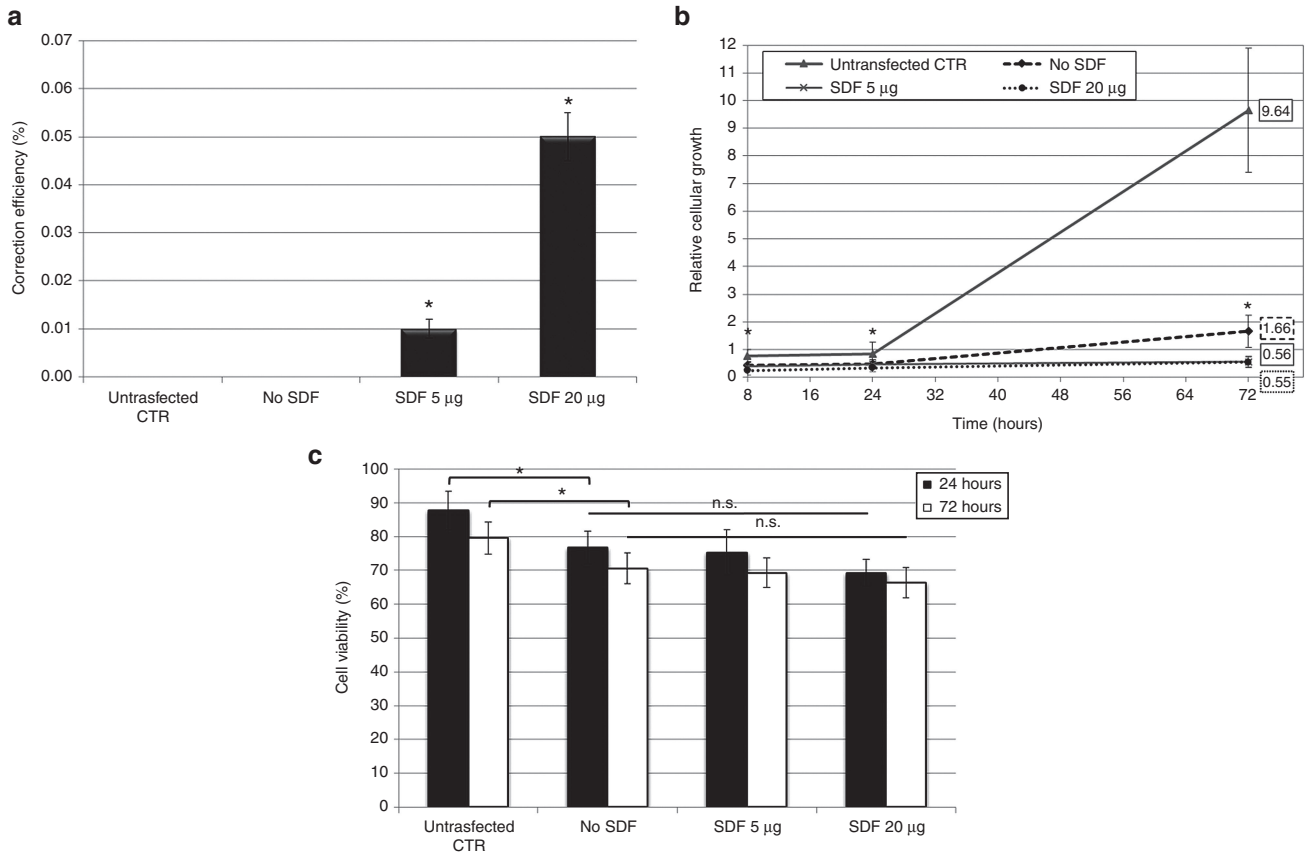
These experiments were conducted on unsynchronized MEF-mutEGFP (MEF with an integrated mutated EGFP

gene) transfected with different amounts of SDF-PCR-WT (double-stranded PCR-amplified wild-type SDF, 5 and 20  $\mu$ g). After transfection (or plating for untransfected controls), the cells were analyzed by flow cytometry to determine the modification efficiency and counted, by both microscope and flow cytometry, to assess the cellular growth and viability (Figure 1). As expected, and already demonstrated by us,<sup>16</sup> gene modification efficiency (Figure 1a) was enhanced when cells were treated with a high dose (20  $\mu$ g) of SDF-PCR-WT, reaching 0.05% modification (Student's *t*-test,  $P < 0.001$ ) compared with 0.01% obtained with the low dose (5  $\mu$ g) (Student's *t*-test,  $P < 0.001$ ). Transfection had a negative effect on growth (Figure 1b). Even the control cells transfected without the SDF showed growth levels that were significantly lower than those of the nontransfected control cells (Student's *t*-test,  $P < 0.05$ ), particularly after 72 hours from transfection. This effect was further accentuated when the cells underwent transfection with the SDF (Student's *t*-test,  $P < 0.05$ ). This effect did not appear to be dose dependent, because growth values were similar after administration of different amounts of the SDF. Overall cell viability of adherent cells resulted reduced, for the combined effect of plating and SDF transfection, from 22% (in the untransfected control at 24 hours) up to 33% (in cells transfected with the high dose of SDF at 72 hours) (Figure 1c). The quote of this effect depending on the combined effect of transfection and SDF seems to mainly depend on transfection and not to be SDF dose dependent. In fact, the control cells transfected without the SDF showed a viability reduced of 11% (at 24 hours) and 10% (at 72 hours) in respect to untransfected control (both Student's *t*-test,  $P < 0.05$ ) but similar to cells transfected either with low or high SDF dose at both 24 and 72 hours (analysis of variance (ANOVA), nonsignificant (n.s.)).

### Effect of SFHR on DNA repair genes

After RNA extraction, the quantitative expression of 84 genes involved in the response to several types of DNA damage was investigated in MEF-mutEGFP using quantitative real-time PCR (qRT-PCR) arrays. These genes were classified as follows: 18 related to HR, 7 to NHEJ, 12 to mismatch repair, 19 to base excision repair, 27 to nucleotide excision repair, and 1 with an interconnected and regulatory role within several repair pathways (Supplementary Figure S1). The basal expression levels of DNA repair genes in untreated MEF-mutEGFP were heterogeneous (Supplementary Figure S3), with some highly expressed and several weakly expressed genes, resulting in changes depending on the experimental time point (8, 24, or 72 hours).

Tables 1 and 2 (columns 1–24) summarize the results of the overall pattern of expression change by the preliminary statistical analysis (see Materials and Methods). The number of genes upregulated, downregulated, and unchanged at 8, 24, and 72 hours was analyzed in the following six comparisons: cells transfected with 5 or 20  $\mu$ g of the SDF with respect to the untransfected control cells, cells transfected with 20  $\mu$ g of the SDF with respect to cells transfected with 5  $\mu$ g of the SDF, cells transfected with no SDF with respect to untransfected control cells, and cells transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF. Every  $\chi^2$  was significant, with  $P < 0.001$ . We also analyzed



**Figure 1 Correction efficiency and cellular growth after MEF-mutEGFP were transfected with different amounts of SDF-PCR-WT.** (a) Correction efficiency. Student's *t*-test,  $*P < 0.001$  with respect to control. (b) Relative cellular growth. The values of relative cellular growth with respect to the number of cells plated in control, 72 hours after transfection, are indicated in the corresponding boxes. Student's *t*-test,  $*P < 0.05$  for all treatments with respect to control. (c) Cell viability by flow cytometry analysis. Student's *t*-test,  $*P < 0.05$  for cells transfected with no SDF in respect to untransfected control; n.s. = analysis of variance for all transfected conditions (with no SDF, as well as with low or high SDF dose) not significant. The two experimental times tested are indicated. Untransfected CTR = cells that did not undergo transfection; No SDF = cells that underwent transfection without the SDF; SDF 5  $\mu$ g = cells transfected with the low SDF dose; SDF 20  $\mu$ g = cells transfected with the high SDF dose. Error bars indicate SD. CTR, control; MEF-mutEGFP, mouse embryonic fibroblasts with an integrated mutated EGFP gene; SDF, small DNA fragment; SDF-PCR-WT = double-stranded PCR-amplified wild-type SDF.

the SDF dose effect testing the following four comparisons: genes upregulated with respect to the untransfected control cells ( $\chi^2$ ,  $P < 0.005$ ), genes upregulated with respect to cells transfected with no SDF ( $\chi^2$ ,  $P < 0.05$ ), genes downregulated with respect to the untransfected control cells ( $\chi^2$ ,  $P < 0.001$ ), and genes downregulated with respect to cells transfected with no SDF ( $\chi^2$ ,  $P < 0.01$ ). Overall, all differences in the proportion of both upregulated and downregulated genes in the experimental conditions tested were statistically significant. The preliminary statistical analysis was also performed on the average fold change of upregulated and downregulated genes in the following comparisons: cells transfected with 0, 5, or 20  $\mu$ g of the SDF with respect to untransfected control and cells transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF. Statistically significant dose-dependent differences in the average fold change of upregulated genes were evident at both 8 and 24 hours (ANOVA and Bonferroni's posttest, respectively,  $P < 0.05$  and  $P < 0.001$ ). Also paired data analysis was performed in the following conditions: cells transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF and cells

transfected with 20  $\mu$ g of the SDF with respect to cells transfected with 5  $\mu$ g of the SDF. In this case, statistically significant dose-dependent differences in the average fold change were observed for upregulated genes at 8, 24, and 72 hours (Student's *t*-test for paired data, respectively,  $P < 0.05$ ,  $P < 0.001$ ,  $P < 0.001$ ), as well as for downregulated genes at 24 and 72 hours (Student's *t*-test for paired data, respectively,  $P < 0.001$  and  $P < 0.05$ ). In the text below, we do not repeat the significance level of each comparison.

First, to evaluate the overall effect of both the SDF and the nucleofection protocol, we compared cells transfected with the SDF at low (5  $\mu$ g) and high (20  $\mu$ g) doses with untransfected controls at the corresponding experimental time points. At the low SDF dose (5  $\mu$ g, **Tables 1 and 2**, columns 3 and 11; **Supplementary Figure S4a**) and at the early experimental time point (8 hours after nucleofection), 56 (66.7%) DNA repair genes were upregulated, with only 2 (2.4%) genes downregulated. At the intermediate experimental time point (24 hours after treatment), 42 (50.0%) over-expressed genes remained, with 1 (1.2%) downregulated gene. Finally (72 hours after treatment), only 4 (4.8%) genes

**Table 1** Summary of results regarding the proportion and fold change of DNA repair genes modulated by SFHR (upregulated DNA repair genes)

		DNA repair, upregulated															
		1		2		3		4		5		6		7		8	
		No SDF		In respect to untransfected control:				20 versus 5 µg									
Hours		N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD				
8		51 (60.7%)	1.90±0.75	56 (66.7%)	1.98±0.96	58 (69.0%)	2.37±1.71	27 (32.1%)	1.68±0.34								
24		29 (34.5%)	1.36±0.11	42 (50.0%)	1.55±0.30	70 (83.3%)	2.23±1.10	64 (76.2%)	1.75±0.48								
72		4 (4.8%)	1.44±0.18	4 (4.8%)	1.50±0.19	21 (25.0%)	1.56±0.27	43 (51.2%)	1.77±0.45								
				17		18		19		20							
				In respect to cells transfected without SDF:													
		No SDF		5 µg SDF		20 µg SDF											
Hours		N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD										
8		11 (13.1%)	1.83±0.89	29 (34.5%)	1.85±1.10												
24		37 (44.0%)	1.39±0.26	73 (86.9%)	2.00±1.00												
72		4 (4.8%)	1.38±0.17	45 (53.6%)	1.65±0.59												

The number, percentage, and average fold change of modulated genes after transfection with 0 µg of the SDF (No SDF), low SDF (5 µg), or high SDF (20 µg) dose as compared with untransfected control, as well as with controls transfected without SDF, at different experimental time points (8, 24, and 72 hours) are shown. A direct comparison of the SDF dose effect (20 versus 5 µg) is also shown. See text for statistical analysis and explanations. FC, fold change; SDF, small DNA fragment; SFHR, small fragment homologous replacement.

**Table 2** Summary of results regarding the proportion and fold change of DNA repair genes modulated by SFHR (downregulated DNA repair genes)

		DNA repair, downregulated															
		9		10		11		12		13		14		15		16	
		No SDF		In respect to untransfected control:				20 versus 5 µg									
Hours		N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD				
8		3 (3.6%)	0.68±0.12	2 (2.4%)	0.54±0.23	5 (6.0%)	0.64±0.17	16 (19.0%)	0.61±0.12								
24		9 (10.7%)	0.76±0.05	1 (1.2%)	0.75±0.00	0 (0.0%)	ND	2 (2.4%)	0.73±0.01								
72		38 (45.2%)	0.65±0.10	50 (59.5%)	0.61±0.13	35 (41.7%)	0.68±0.12	5 (6.0%)	0.65±0.11								
				21		22		23		24							
				In respect to cells transfected without SDF:													
		No SDF		5 µg SDF		20 µg SDF											
Hours		N (%)	Mean FC ± SD	N (%)	Mean FC ± SD	N (%)	Mean FC ± SD										
8		10 (11.9%)	0.70±0.01	17 (20.2%)	0.68±0.12												
24		7 (8.3%)	0.76±0.01	1 (1.2%)	0.78±0.00												
72		14 (16.7%)	0.64±0.05	15 (17.9%)	0.67±0.10												

The number, percentage, and average fold change of modulated genes after transfection with 0 µg of the SDF (No SDF), low SDF (5 µg), or high SDF (20 µg) dose as compared with untransfected control, as well as with controls transfected without SDF, at different experimental time points (8, 24, and 72 hours) are shown. A direct comparison of the SDF dose effect (20 versus 5 µg) is also shown. See text for statistical analysis and explanations. FC, fold change; ND, not determined; SDF, small DNA fragment; SFHR, small fragment homologous replacement.

were upregulated, with 50 (59.5%) genes reaching a lower expression level than (and 30 genes with expression similar to) those of the untransfected control.

At the high SDF dose (20 µg, **Tables 1** and **2**, columns 5 and 13; **Supplementary Figure S4b**) and at the early experimental time point (8 hours after nucleofection), 58 (69.0%) DNA repair genes were upregulated, with only 5 (6.0%) genes downregulated. At the intermediate experimental time point (24 hours after treatment), the number of overexpressed genes increased to 70 (83.3%), with no downregulated genes. Finally (72 hours after treatment), the number of upregulated genes decreased to 21 (25.0%), with 35 (41.7%) genes reaching a lower expression level than (and 28 genes with expression similar to) those of the untransfected control cells.

After an early phase of upregulation, the DNA repair genes appeared to be progressively downregulated, with most of them reaching expression levels lower than those of the control cells (set to 1) during the late phase. The proportion of early upregulated genes seems to be similar at low (5 µg) and high (20 µg) SDF doses. However, the early phase of upregulation seems to be shorter after treatment with a low SDF dose than with a high SDF dose. In the latter condition, at 24 hours after treatment, the number of upregulated genes continued to grow, whereas in the former condition, the number of upregulated genes began to decrease. At 72 hours, the number of upregulated genes remained considerably higher at the high SDF dose than at the low SDF dose. In addition, a higher number of downregulated genes were observed in the latter condition. Furthermore, the treatment



with the high SDF dose appeared to produce a greater effect on the fold change of upregulated genes than the treatment with the low SDF dose, primarily at 8 and 24 hours (Table 1, column 6 with respect to 4). The overall dose effect that produced the prolonged and quantitatively greater change to the expression of DNA repair genes (primarily those upregulated genes) at the high dose of SDF could also be clearly seen by directly comparing the cells treated with 20  $\mu\text{g}$  of the SDF with those treated with 5  $\mu\text{g}$  of the SDF (in this case, the latter experimental condition was set to 1) (Tables 1 and 2, columns 7 and 15 for gene number and columns 8 and 16 for fold change; Supplementary Figure S4c).

To investigate a possible effect of the nucleofection protocol on the expression of DNA repair genes, we analyzed cells that underwent the transfection protocol without the SDF (treated with the same nucleofection protocol and buffer with no SDF inside) (Tables 1 and 2, columns 1 and 9 for gene number and columns 2 and 10 for fold change; Supplementary Figure S4d). The temporal expression pattern after nucleofection without the SDF was very similar to the general temporal expression pattern, with early upregulation followed by progressive downregulation. However, this effect was reduced in both the proportion of involved genes and their fold change, particularly for upregulated genes, compared with that induced by the SDF (also at the low dose) and the nucleofection protocol. In addition, the genes modulated by the transfection protocol appear to differ, at least partially, from those modulated by the SDF.

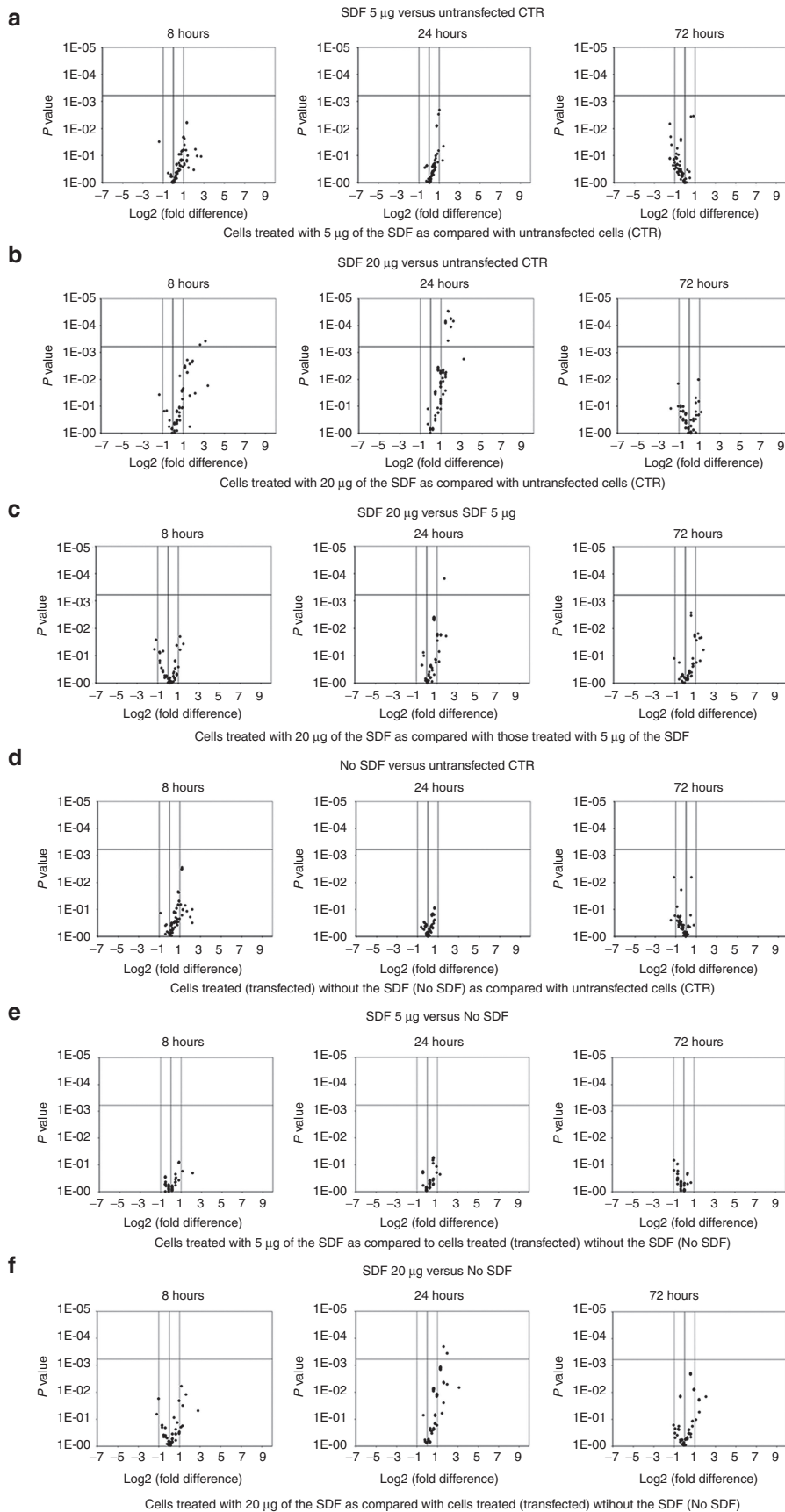
To subtract the effect of the nucleofection protocol, which allows the specific effect of SDF to be distinguished, cells nucleofected with or without the SDF were compared at either the low (Tables 1 and 2, columns 17 and 21 for gene number and columns 18 and 22 for fold change; Supplementary Figure S4e) or high SDF dose (Tables 1 and 2, column 19 and 23 for gene number and columns 20 and 24 for fold change; Supplementary Figure S4f). After subtracting the effect of the nucleofection protocol, the specific upregulatory effect of 5  $\mu\text{g}$  of the SDF alone appeared to be quantitatively reduced (in terms of both gene number and fold change) with respect to the additive effects of 5  $\mu\text{g}$  of the SDF and the nucleofection protocol (see Table 1, columns 3 and 4; Supplementary Figure S4a). The temporal pattern specifically induced by the SDF was quite similar to that induced by the cumulative effect of the SDF and the nucleofection protocol at 24 and 72 hours. However, the large number of upregulated genes at 8 hours observed after treatment with the SDF and nucleofection seemed to depend more on the nucleofection protocol. By contrast, the specific effect exerted by 20  $\mu\text{g}$  of the SDF after subtracting the effect of the nucleofection protocol appeared to induce a quantitatively greater upregulation response in terms of gene number at both 24 and 72 hours, specifically depending on the SDF itself (compared with Table 1, columns 5 and 6; Supplementary Figure S4b). In particular, the longer persistence (up to 72 hours) of a greater number of upregulated genes appeared to be specifically induced by the high SDF dose. After subtracting the effect of the nucleofection protocol, a dose-specific effect of the SDF was evident at every experimental time point, particularly at 24 hours (Table 1, columns 19 and 20 as compared with columns 17 and 18).

After the overall analysis reported above, a more statistically stringent analysis was performed (see Materials and Methods). Fourteen DNA repair genes had statistically significant expression differences in at least one experimental time point. The statistically significant increase in the expression of these genes was evident (at corrected  $P < 0.0006$ ) only after nucleofection with 20  $\mu\text{g}$  of the SDF (Figure 2b,c,f); the increase occurred for 2 genes at both 8 and 24 hours and for the remaining 12 genes at 24 hours. No statistically significant effects (at corrected  $P < 0.0006$ ) were detectable after nucleofection with 5  $\mu\text{g}$  of the SDF (Figure 2a,e) or after transfection with no SDF (Figure 2d). Each of these genes showed a quantitative expression value greater than 2. These results suggest a dose effect and a specific temporal expression pattern.

### Effect of SFHR on cell cycle regulatory genes

Using qRT-PCR arrays, we studied the quantitative expression of 84 genes involved in cell cycle control in MEF-mutEGFP. These genes were classified as follows: 29 generally related to checkpoint and arrest, 9 specifically to G1/G1-S transition, 10 specifically to S and replication phases, 10 to M phase, 19 to a general positive regulation, and 7 to a general negative regulation of cell cycle (Supplementary Figure S2). For cell cycle regulatory genes, the overall results are described in detail below and summarized in Tables 3 and 4 (columns 25–48). To evaluate the overall gene expression pattern, we performed the same six different comparisons listed above for DNA repair genes (see also Materials and Methods). For cell cycle control genes, every  $\chi^2$  was significant at  $P < 0.001$ . We also analyzed the SDF dose effect testing the additional four comparisons with the following results: genes upregulated with respect to the untransfected control ( $\chi^2$ , n.s.), genes upregulated with respect to cells transfected with no SDF ( $\chi^2$ ,  $P < 0.001$ ), genes downregulated with respect to the untransfected control ( $\chi^2$ , n.s.), and downregulated with respect to cells transfected with no SDF ( $\chi^2$ ,  $P < 0.001$ ). For cell cycle control genes, the great majority of differences in the proportion of both upregulated and downregulated genes in the experimental conditions tested were also statistically significant. Additionally, for cell cycle control genes, the preliminary statistical analysis was extended to the average fold change of upregulated and downregulated genes to test the same four above-listed comparisons. For cell cycle control genes, a statistically significant dose-dependent difference in the average fold change of upregulated genes was evident at 24 hours (ANOVA Bonferroni's posttest,  $P < 0.05$ ). Moreover, for both up- and downregulated cell cycle control genes, the three different pairs of experimental conditions listed above were tested. In this case, a statistically significant dose-dependent difference in the average fold change was evident at 24 hours for both upregulated genes (Student's *t*-test for paired data,  $P < 0.005$ ) and downregulated genes (Student's *t*-test for paired data,  $P < 0.05$ ). In the text below, we do not repeat the significance level of each comparison for these genes.

For cell cycle genes, the basal expression levels in untreated MEF-mutEGFP were heterogeneous (Supplementary Figure S5), with some highly expressed and several weakly expressed genes, as well as some genes with



**Figure 2** Volcano plots of the studied DNA repair genes in MEF-mutEGFP at 8, 24, and 72 hours after treatment, according to different experimental conditions (from a to f). The spots above the horizontal line represent those genes with a statistically significant expression differences with respect to controls (Student's *t*-test after Bonferroni's correction for multiple comparisons,  $P < 0.0006$ ). The spots on the left and right of the vertical lines represent genes with expression levels that are twofold lower or higher than controls, respectively. CTR, untransfected control; MEF-mutEGFP, mouse embryonic fibroblasts with an integrated mutated EGFP gene; SDF, small DNA fragment.

variable expression depending on the experimental time point (8, 24, or 72 hours). Also in this case, to first evaluate the overall effect of both the SDF and the nucleofection protocol, we compared the cells transfected by the SDF at low (5  $\mu$ g) and high (20  $\mu$ g) doses with untransfected controls at the corresponding experimental time point. Applying the same thresholds as those for the DNA repair gene analysis at the low SDF dose (5  $\mu$ g; **Tables 3** and **4**, columns 27 and 35; **Supplementary Figure S6a**) and the early experimental time point (8 hours after nucleofection), 59 (70.2%) cell cycle genes were upregulated and 6 (7.1%) genes were downregulated. At the intermediate experimental time point (24 hours after treatment), 41 (48.8%) genes were upregulated and 8 (9.5%) were downregulated. Finally, at 72 hours after treatment, 21 (25.0%) genes were upregulated and 25 (29.8%) genes were downregulated, with 38 displaying expression levels similar to those of the untransfected control.

At the high SDF dose (20  $\mu$ g; **Tables 3** and **4**, columns 29 and 37; **Supplementary Figure S6b**) and early experimental time point (8 hours after nucleofection), 70 (83.3%) cell cycle genes were upregulated, with only 6 (7.1%) downregulated genes. At the intermediate experimental time points (24 hours after treatment), the number of overexpressed genes decreased to 52 (61.9%), with 6 (7.1%) downregulated genes. Finally, at 72 hours after treatment, the number of upregulated genes decreased to 25 (29.8%), with 27 (32.1%) genes reaching a lower expression level than (and 32 genes having expression levels similar to) those of the untransfected control.

Also for cell cycle genes, there is an early phase of expression stimulation followed by a progressive phase of downregulation; in the late phase (72 hours), most of the cell cycle genes reached expression levels lower than those of the control (set to 1). As for the DNA repair genes, the treatment with the high SDF dose appeared to produce a higher proportion of upregulated genes and a quantitatively greater effect on their fold change, particularly at 24 hours, than the treatment with the low SDF dose (**Table 3**, column 30 with respect to 28). However, in contrast to DNA repair genes, the treatment with either the low or high SDF dose appeared to induce a similar temporal pattern, with consistent upregulation still at 24 hours, although with no further increase (at this experimental time point) induced from the high SDF dose.

The overall high SDF dose effect that resulted in the quantitatively greater changes to the cell cycle genes could also be clearly observed via direct comparison between the cells treated with 20  $\mu$ g of the SDF and those treated with 5  $\mu$ g (experimental condition in this case set to 1) (**Tables 3** and **4**, columns 31 and 39 for gene number and 32 and 40 for fold change; **Supplementary Figure S6c**).

As for DNA repair genes, to evaluate a possible effect of the nucleofection protocol on cell cycle gene expression, we analyzed cells that underwent the transfection protocol without the SDF (**Tables 3** and **4**, columns 25 and 33 for gene number and columns 26 and 34 for fold change; **Supplementary Figure S6d**). The temporal expression pattern after nucleofection without the SDF was similar to that observed for cells treated with the SDF. From the point of view of the number of upregulated genes and their average fold change, this experimental condition treated with no SDF was very similar to that

treated with 5  $\mu$ g of the SDF, whereas quantitative differences in upregulation could be detected when compared with the experimental condition with the high SDF dose. In this case, the genes modulated by the transfection protocol were also partially different from those modulated by the SDF.

To subtract the effect of the nucleofection protocol, which allows the specific effect of the SDF to be distinguished, we directly compared cells nucleofected with or without the SDF (set to 1), at either the low (**Tables 3** and **4**, columns 41 and 45 for gene number and columns 42 and 46 for fold change; **Supplementary Figure S6e**) or high dose (**Tables 3** and **4**, column 43 and 47 for gene number and columns 44 and 48 for fold change; **Supplementary Figure S6f**). After subtracting the effect of the nucleofection protocol, the specific effect of 5  $\mu$ g of the SDF alone appeared to be quantitatively reduced, in terms of both the proportion of upregulated genes and their average fold change, with respect to the additive effects of 5  $\mu$ g of the SDF and the nucleofection protocol (as compared with **Table 3**, columns 27 and 28; **Supplementary Figure S6a**). In this case, the large number of upregulated genes at 8 hours, and even at 24 hours, evidenced after treatment with the SDF and nucleofection seemed to depend more on the nucleofection protocol than on a specific effect of the SDF itself. Furthermore, longer persistence (up to 72 hours) of upregulated genes was demonstrated. The upregulation specifically induced by the high SDF dose (20  $\mu$ g) after subtracting the effect of the nucleofection protocol is quite similar to the cumulative effect of the SDF and the nucleofection protocol (see **Table 3**, columns 29 and 30; **Supplementary Figure S6b**) at 24 and 72 hours (even with a reduction in the average fold change at 24 hours). The longer persistence (up to 72 hours) of upregulated genes appeared to be specifically induced by the high SDF dose. However, the predominant effect also at high SDF dose at early experimental time (8 hours) seemed to be that of the nucleofection protocol (at least on gene number) rather than that of the SDF itself. Also for cell cycle genes, after subtracting the effect of the nucleofection protocol, a dose-specific effect of the SDF could be observed at every experimental time point (**Table 3**, columns 43 and 44 as compared with columns 41 and 42).

After the overall analysis, a more statistically stringent analysis of cell cycle gene expression was performed (see Materials and Methods). Fifteen cell cycle genes had statistically significant expression differences at, at least, one experimental time point. The statistically significant increase in the expression of these genes was evident (using the corrected  $P < 0.0006$  level) only after nucleofection with 20  $\mu$ g of the SDF compared with the untransfected control (**Figure 3b**); for two genes, the increase occurred at both 8 and 24 hours, for five genes only at 8 hours, and for eight genes only at 24 hours. No statistically significant effects (at corrected  $P < 0.0006$ ) were observed under the other experimental conditions. Each of these genes showed a quantitative expression value greater than 2. These results also suggest a dose effect and a specific temporal expression pattern.

#### Selection of specific gene targets affected by SFHR within the DNA repair and cell cycle genes

To select the best specific gene targets with expression modulation induced by SFHR within both the DNA repair and cell

**Table 3** Summary of results regarding the proportion and fold change of cell cycle genes modulated by SFHR (upregulated cell cycle genes)

Hours	Cell cycle, upregulated																
	25		26		27		28		29		30		31		32		
	No SDF		5 $\mu$ g SDF		20 $\mu$ g SDF		20 versus 5 $\mu$ g		No SDF		5 $\mu$ g SDF		20 $\mu$ g SDF		20 versus 5 $\mu$ g		
	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	
8	57 (67.9%)	1.94 $\pm$ 0.67	59 (70.2%)	2.12 $\pm$ 1.10	70 (83.3%)	3.33 $\pm$ 6.86	39 (46.4%)	2.05 $\pm$ 1.31									
24	39 (46.4%)	1.52 $\pm$ 0.30	41 (48.8%)	1.53 $\pm$ 0.28	52 (61.9%)	3.49 $\pm$ 4.57	57 (67.9%)	2.26 $\pm$ 1.94									
72	10 (11.9%)	1.48 $\pm$ 0.30	21 (25.0%)	1.54 $\pm$ 0.30	25 (29.8%)	1.41 $\pm$ 0.22	25 (29.8%)	1.54 $\pm$ 0.41									
			41	42	43	44											
			In respect to cells transfected without SDF:														
			5 $\mu$ g SDF		20 $\mu$ g SDF												
			N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD											
8			16 (19.0%)	1.59 $\pm$ 0.97	29 (34.5%)	3.45 $\pm$ 7.06											
24			7 (8.3%)	1.51 $\pm$ 0.33	52 (61.9%)	2.66 $\pm$ 4.18											
72			22 (26.2%)	1.49 $\pm$ 0.28	29 (34.5%)	1.53 $\pm$ 0.30											

The number, percentage, and average fold change of modulated genes after transfection with 0  $\mu$ g of the SDF (No SDF), low SDF (5  $\mu$ g), or high SDF (20  $\mu$ g) dose as compared with untransfected control, as well as with controls transfected without SDF, at different experimental time points (8, 24, and 72 hours) are shown. A direct comparison of the SDF dose effect (20 versus 5  $\mu$ g) is also shown. See text for statistical analysis and explanations. FC, fold change; SDF, small DNA fragment; SFHR, small fragment homologous replacement.

**Table 4** Summary of results regarding the proportion and fold change of cell cycle genes modulated by SFHR (downregulated cell cycle genes)

Hours	Cell cycle, downregulated																
	33		34		35		36		37		38		39		40		
	No SDF		5 $\mu$ g SDF		20 $\mu$ g SDF		20 versus 5 $\mu$ g		No SDF		5 $\mu$ g SDF		20 $\mu$ g SDF		20 versus 5 $\mu$ g		
	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD	
8	6 (7.1%)	0.63 $\pm$ 0.15	6 (7.1%)	0.67 $\pm$ 0.09	6 (7.1%)	0.59 $\pm$ 0.04	10 (11.9%)	0.65 $\pm$ 0.05									
24	2 (2.4%)	0.77 $\pm$ 0.03	8 (9.5%)	0.60 $\pm$ 0.12	6 (7.1%)	0.56 $\pm$ 0.13	7 (8.3%)	0.66 $\pm$ 0.09									
72	30 (35.7%)	0.66 $\pm$ 0.12	25 (29.8%)	0.62 $\pm$ 0.13	27 (32.1%)	0.60 $\pm$ 0.14	16 (19.0%)	0.54 $\pm$ 0.16									
			45	46	47	48											
			In respect to cells transfected without SDF:														
			5 $\mu$ g SDF		20 $\mu$ g SDF												
			N (%)	Mean FC $\pm$ SD	N (%)	Mean FC $\pm$ SD											
8			20 (23.8%)	0.77 $\pm$ 0.05	9 (10.7%)	0.63 $\pm$ 0.08											
24			20 (23.8%)	0.65 $\pm$ 0.10	13 (15.5%)	0.65 $\pm$ 0.16											
72			4 (4.8%)	0.66 $\pm$ 0.01	15 (17.9%)	0.62 $\pm$ 0.13											

The number, percentage, and average fold change of modulated genes after transfection with 0  $\mu$ g of the SDF (No SDF), low SDF (5  $\mu$ g), or high SDF (20  $\mu$ g) dose as compared with untransfected control, as well as with controls transfected without SDF, at different experimental time points (8, 24, and 72 hours) are shown. A direct comparison of the SDF dose effect (20 versus 5  $\mu$ g) is also shown. See text for statistical analysis and explanations. Notably, the high values of SD reported in Tables 1–4 originate from the fact that the average fold change was calculated for different genes and not for different experimental replicates. Thus, in this case, the SD value reflects the heterogeneous response of the genes, indicating biological variability and not experimental variability. For example, in column 44, the treatment at 8 hours with 20  $\mu$ g of the SDF referring to cells transfected without the SDF displayed an average fold change of 3.45, with an SD of 7.06. This value was found because within the corresponding 29 upregulated genes, some have very high and others have very low fold changes (essentially, the fold changes of these 29 genes ranged from 1.31 to 39.64). FC, fold change; SDF, small DNA fragment; SFHR, small fragment homologous replacement.

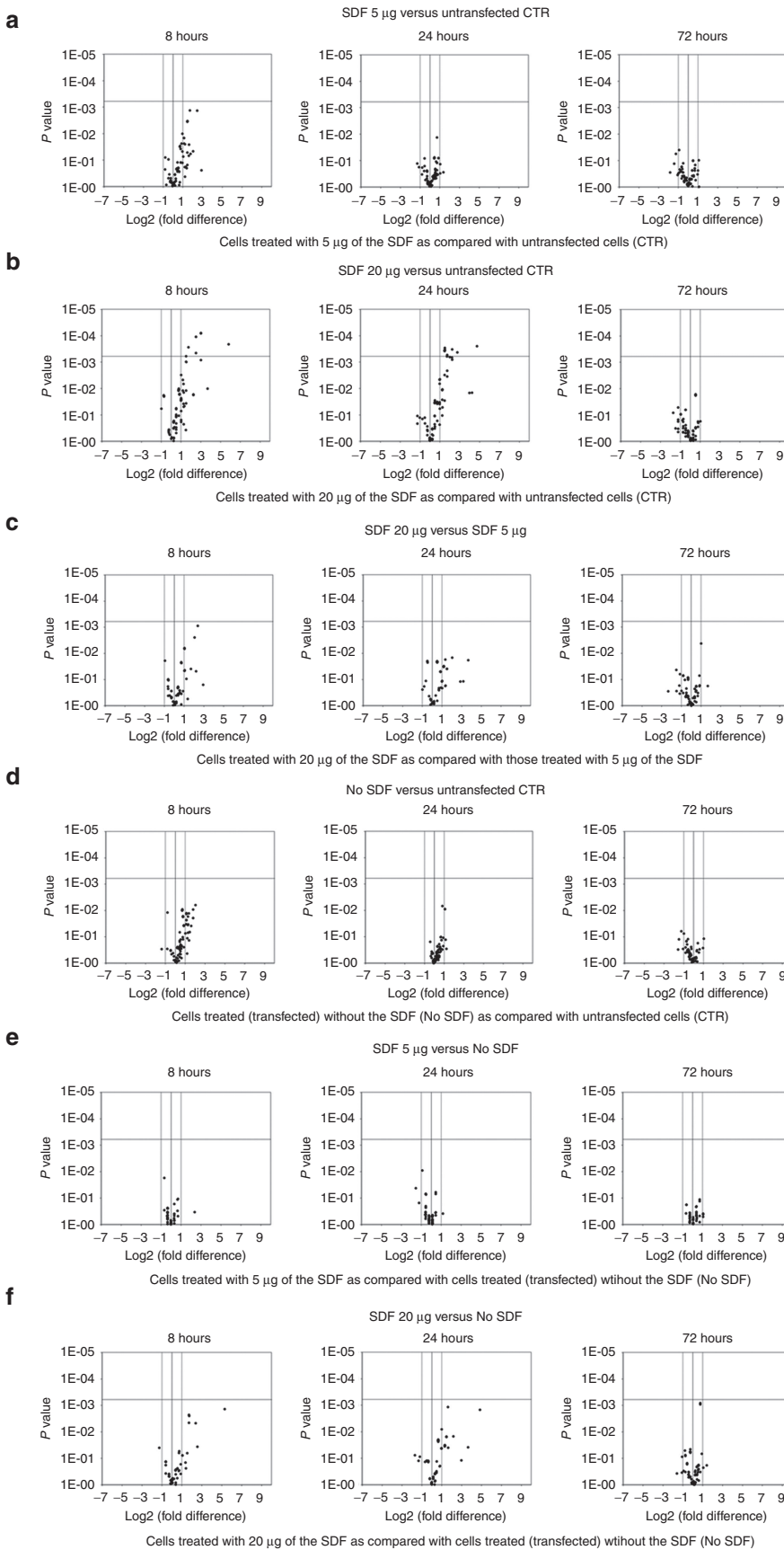
cycle genes, we used a combination of statistical and biological criteria. Because a gene was selected, the following four conditions had to be simultaneously verified:

1. statistically significant expression difference compared with the control for at least one experimental condition (Student's *t*-test after Bonferroni's correction,  $P < 0.0006$ );
2. statistically significant dose effect with clear differences between the experimental conditions of cells transfected with 0, 5, and 20  $\mu$ g of the SDF for at least one experimental condition (in this case, Student's *t*-test,  $P < 0.05$ );

3. a fold change greater than 3 with respect to the control, for at least one experimental condition;
4. a temporal pattern reflecting the overall temporal pattern of all genes taken together with early (8 hours) and/or intermediate (24 hours) upregulation, followed by a progressive late (72 hours) return to baseline and/or downregulation.

According to these criteria, the following 18 genes were selected. Five genes more specific for the DNA repair pathway (**Figure 4**): *Neil2*, *Parp3*, *Pms2*, *Rad51l3*, and *Trex1*; seven genes more specific for the cell cycle pathway (**Figure 5**): *Cdkn1a*, *Chek1*, *Ddit3*, *Gpr132*, *Mdm2*, *Prm1*, and *Slnf1*;





**Figure 3** Volcano plots of the studied cell cycle genes in MEF-mutEGFP at 8, 24, and 72 hours after treatment, according to different experimental conditions (from a to f). The spots above the horizontal line represent those genes with statistically significant expression differences with respect to controls (Student's *t*-test after Bonferroni's correction for multiple comparisons,  $P < 0.0006$ ). The spots on the left and right of the vertical lines represent genes with expression levels twofold lower or higher than controls, respectively. CTR, untransfected control; MEF-mutEGFP, mouse embryonic fibroblasts with an integrated mutated EGFP gene; SDF, small DNA fragment.

six genes involved in both pathways (**Figure 6**): *Atm*, *Brca1*, *Brca2*, *Ppm1d*, *Rad9*, and *Sesn2*. Brief descriptions of the selected genes, along with links to databases where full information may be found, are provided as a supplementary information under the heading **Brief descriptions and database links for the selected genes**.

## Discussion

Despite the evidence regarding the interconnection of SFHR and many cellular pathways,<sup>16</sup> the interacting molecular mechanisms have not yet been fully characterized. Throughout the cell cycle, the DNA is constantly monitored for damage and replication errors by a complex regulatory network that controls DNA duplication and the cell cycle at multiple levels and that functionally cooperates with DNA repair pathways.<sup>38</sup>

In this work, we focused on the relationship between SFHR and these two main cellular pathways at the single-gene level.

The quantitative analysis of DNA repair gene expression indicated a peculiar temporal gene expression pattern depending on the combined effects of the nucleofection protocol and the exogenous DNA. The nucleofection protocol induced early (8 hours) upregulation of DNA repair genes, which progressively returned to basal expression levels and were finally (72 hours) downregulated in most cases. The specific effect of the SDF was quantitatively additive in a dose-dependent manner with respect to that of nucleofection. At the low SDF dose, most of the DNA repair genes remained upregulated at the intermediate experimental time point (24 hours) but later (72 hours) returned to basal levels or were downregulated. At the high SDF dose, the effect was greater from both the quantitative and the persistence points of view: at the intermediate time point (24 hours), there was a further increase in the number of upregulated genes, and at the late time point (72 hours), most of the genes still appeared upregulated.

At present, electroporation is one of the most efficient and widely used methods to transfer exogenous DNA into cells.<sup>39,40</sup> Recent developments of methodologies able to introduce genetic material directly into the nuclei of several cell types have led to further improvement of DNA transfer methods and protocol.<sup>41–45</sup> Although these methods are very efficient, common drawbacks are their invasiveness and their variable efficiency depending on the cell type. Our results demonstrated the relevant effect exerted by the technique; this effect is additive with respect to that specifically exerted by the transfected DNA on the expression of genes involved in DNA repair and cell cycle control. In effect, the growth of cells is nearly arrested by the nucleofection with a quote of adherent cells not viable, mainly depending on nucleofection and not on the dose of the SDF. In the remaining quote of viable cells, the gene expression is modulated by SDF in a dose-dependent manner. This is therefore another variable to be considered when using gene transfer approaches based on electroporation methods.

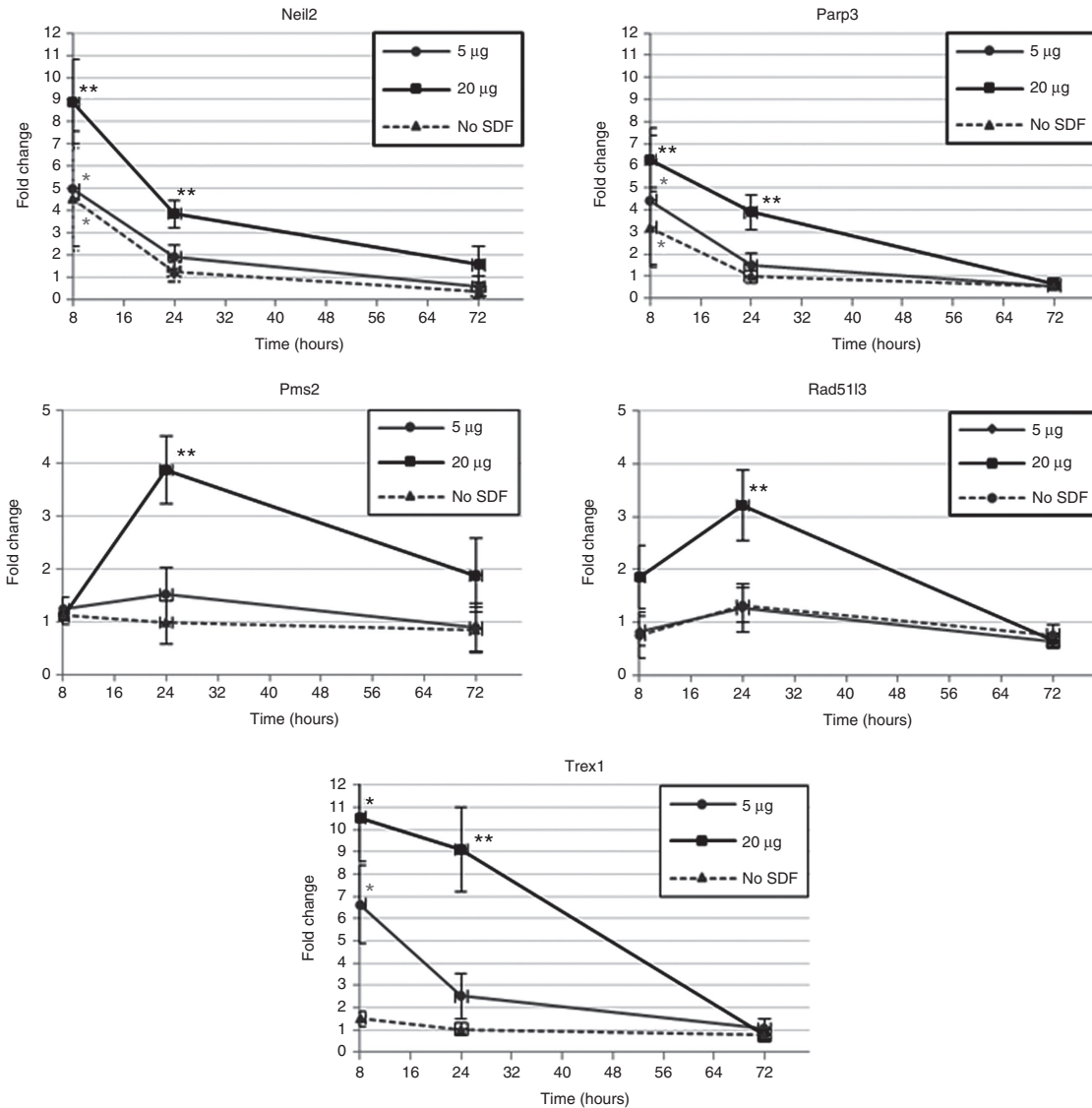
In contrast to current opinions, the DNA repair genes significantly affected by the SFHR belong to different DNA repair pathways as mismatch repair, base excision repair, and HR,

and not only to the HR pathway. The early upregulation of the DNA repair genes observed both as a general pattern and for the 14 selected genes may constitute an adaptive response aimed at the preservation of DNA integrity after exogenous DNA invasion, which may also constitute a molecular basis for DNA modification. However, the subsequent general and gene-specific late downregulation of these genes may be responsible for the reduced modification efficiency, possibly as a subsequent adaptive response aimed at limiting the undesirable effects of exogenous DNA on genomic DNA once the repair attempt has been performed.

In our previous work,<sup>16</sup> no random integration of the SDF was detectable after SFHR. However, the possibility that such integration can occur below the level of detection could not be dismissed. In this work, the quantitative expression analysis of seven NHEJ genes did not show any statistically significant increase in their expression. This is confirmation that the basal activity of the error-prone NHEJ subpathway at least does not increase with this approach.

A dramatic effect of cell cycle arrest, which was delayed until the final experimental time point (72 hours), was induced by both the nucleofection protocol and the SDF. Accordingly, the quantitative study of cell cycle gene expression revealed a wide effect of SFHR on cell cycle regulation. The involvement of cell cycle regulation genes is consistent with our previous results that showed a greater tendency for the SDF to access the target locus during G2/M phase.<sup>16</sup> As for DNA repair genes, the quantitative effect on the number of genes involved and on their level of expression was higher and more prolonged when the SDF was included; this effect was dose dependent with respect to the effect induced by the nucleofection protocol alone. The effect seemed to be exerted on each of the three main cellular checkpoints, as the cell cycle genes with expression modulated by SFHR were functionally linked to both G1/S and G2/M transitions, as well as to the intra-S-phase checkpoint. The effect of cell cycle arrest correlates well with the fact that the vast majority of the upregulated genes are well-recognized negative regulators of checkpoints and of the cell cycle. These results are in excellent agreement with those obtained for DNA repair genes, as cellular checkpoint activation and cell cycle arrest are mandatory for DNA damage repair. Furthermore, the enhanced expression of negative regulators of the cell cycle may constitute part of the molecular basis for the DNA modification. However, similar to observations for DNA repair, the induced modulation of cell cycle genes might be insufficient, from a quantitative and/or qualitative point of view, to warrant higher levels of modification. The marked SDF dose effect on the expression of both DNA repair and cell cycle genes, correlates well with the enhanced modification efficiency reached at the high SDF dose.

Because the DNA repair events occur in a cellular context, for a more realistic description of DNA lesion processing and of the cell cycle progression an integrated vision of DNA repair pathways, checkpoints and cell cycle phases must be considered. Achieving functional separation between sensors and effectors of DNA repair and the cell cycle is not easy and probably not useful. Sensors and effectors of DNA repair extensively influence reciprocal responses with dynamic and highly regulated mechanisms. By contrast, several proteins are multifunctional,

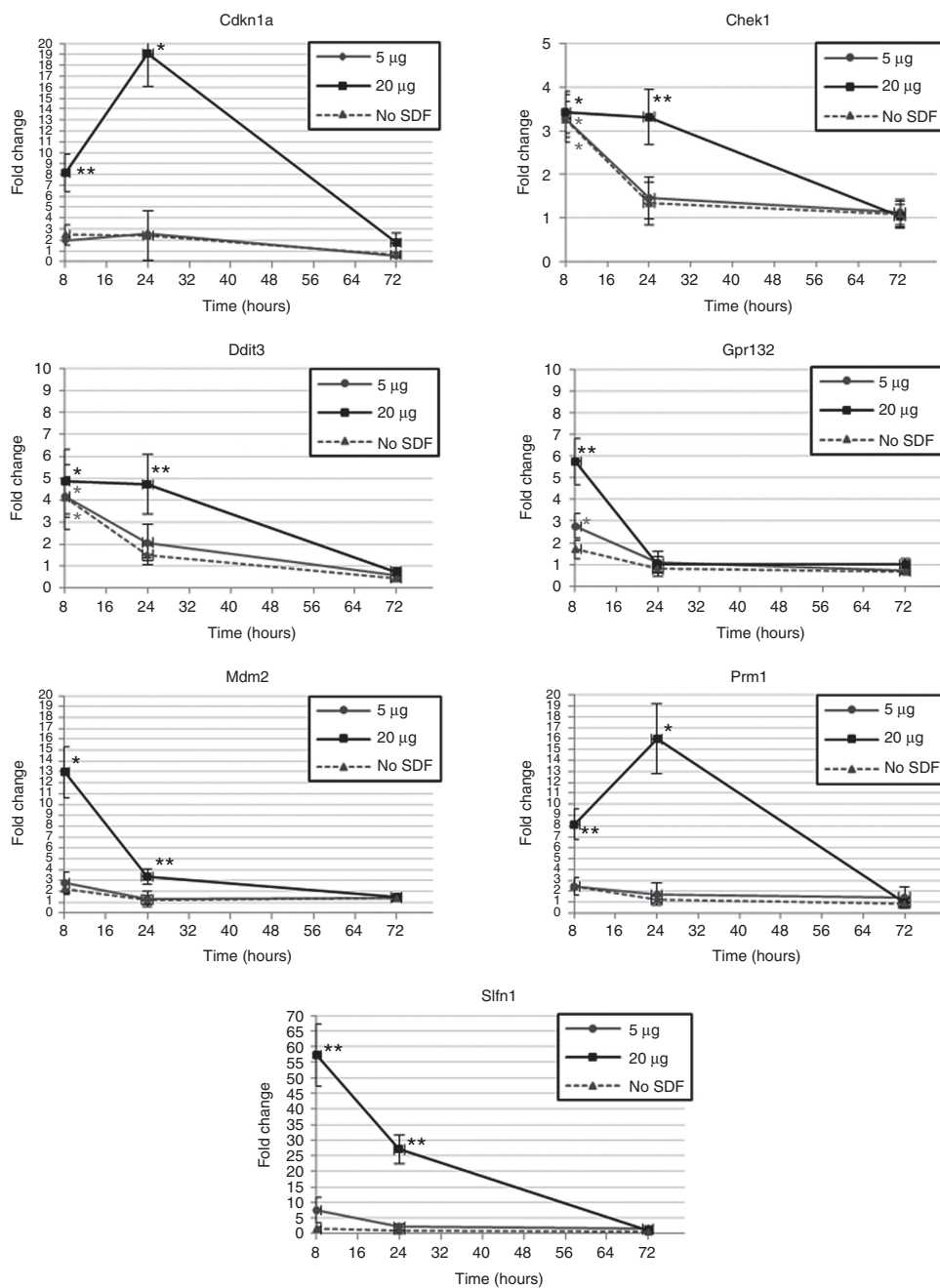


**Figure 4** Temporal pattern after SFHR modification of selected genes within the DNA repair pathway (see text for selection criteria). \*\*Student's *t*-test with Bonferroni's correction,  $P < 0.0006$ ; \*Student's *t*-test,  $P < 0.05$ . SDF, small DNA fragment; SFHR, small fragment homologous replacement.

with direct involvement in both pathways. Exogenous DNA interferes with, and is treated by, this complicated network. The experimental evidence obtained by our studies confirms the involvement and interconnection of the biochemical pathways analyzed in the cellular response to cell invasion by exogenous DNA. In particular, of the studied genes, 18 were selected as the best targets/effectors of SFHR. These genes often have a dual (sometimes multiple) role in both DNA repair and cell cycle regulation pathways. This dual role further stresses the interconnection between SFHR, DNA repair effectors, and cellular checkpoint regulators and clearly indicates that only an integrated study approach may highlight the network. The epigenetic changes and chromatin remodeling, as well as the host defense responses involved in the maintenance of genome structural integrity, which continuously occur during the different phases of the cell cycle, are likely to influence the targeting

mechanism. These actions may represent essential key factors for SFHR efficiency.

Our results could point a way to better yields of SDF-induced gene repair. In addition to constitute the molecular basis of the correction, the selected genes may represent ideal candidates for manipulations aimed to the enhancement of SFHR efficiency. A further gene selection aimed to a practical application of SFHR may be based on specific characteristics of genes, for example, their functions, highest expression levels, and specific temporal patterns. From the point of view of the level of expression, *Sifn1* and *Cdkn1a* are involved in the cell cycle checkpoint and arrest by the negative regulation of the G1/S transition with an increase of, respectively, 58 times (at 8 hours) and 19 times (at 24 hours). As well, *Prm1* is involved in mitotic chromosome condensation (M phase), replacing histones during chromatin remodeling, with an increase of 16-fold at 24 hours. For these



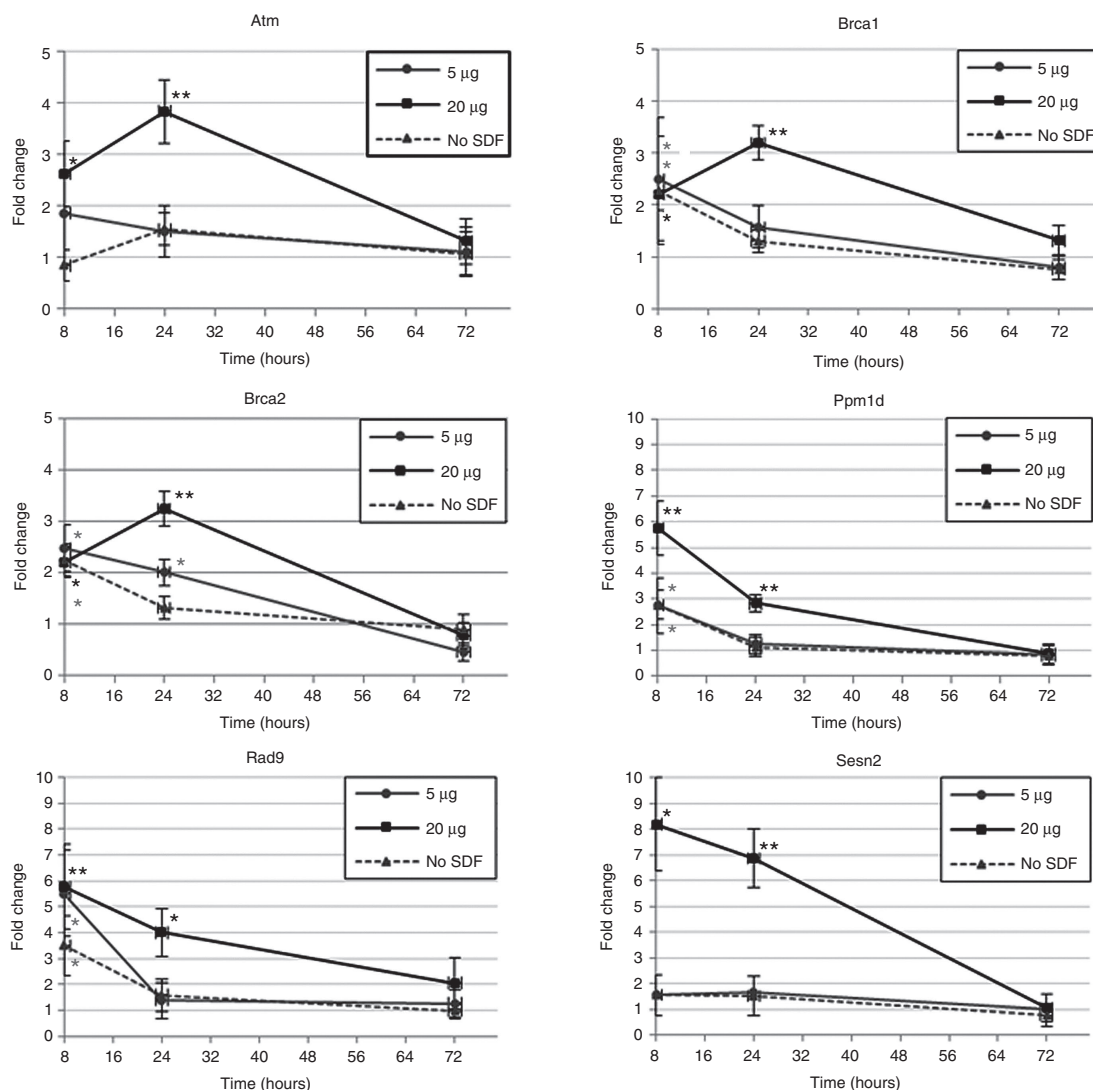
**Figure 5** Temporal pattern after SFHR modification of selected genes within the cell cycle pathway (see text for selection criteria). \*\*Student's  $t$ -test with Bonferroni's correction,  $P < 0.0006$ ; \*Student's  $t$ -test,  $P < 0.05$ . SDF, small DNA fragment; SFHR, small fragment homologous replacement.

genes, a favorable role in SFHR could be hypothesized. From the functions point of view, also for other two of the selected genes, *Trex1* and *Rad9*, a favorable role in enhancing the SFHR modification efficiency could be hypothesized based on the facts that *Trex1* has a role in mismatch repair and in a proof-reading function for DNA polymerase and that *Rad9* has a role in mismatch repair and HR, as well as in the cell cycle arrest, acting at the G2/M checkpoint. However, both genes have a 3' to >5' exonuclease activity on double-stranded DNA.<sup>46,47</sup> Consequently, a contrasting hypothesis is that their exonuclease activity might degrade the SDF, with consequent unfavorable

effects on SFHR efficiency. These are examples of interesting targets to be down- or upregulated to enhance SFHR efficiency and to distinguish between contrasting hypotheses.

Taken together, these results contribute to the comprehension of the molecular mechanisms underlying cell invasion by exogenous DNA and its genomic effect. Adding to our previous findings that DNA methylation is involved in the response to exogenous DNA invasion,<sup>16</sup> in this work, we determined that the two main pathways of the cell cycle and DNA repair also appear to mediate the cellular response to this invasion. The interplay between these pathways and their specific





**Figure 6** Temporal pattern after SFHR modification of selected genes with a role within both the DNA repair and the cell cycle pathways (see text for selection criteria). \*\*Student's *t*-test with Bonferroni's correction,  $P < 0.0006$ ; \*Student's *t*-test,  $P < 0.05$ . SDF, small DNA fragment; SFHR, small fragment homologous replacement.

temporal patterns appear to influence the modification efficiency of SFHR. The selection of specific molecular targets to manipulate provides suggestions for increasing gene repair efficiency to achieve higher modification for practical SFHR applications in *ex vivo* therapeutic approaches. These results extend our knowledge of the molecular mechanisms that respond to exogenous DNA and are consequently involved in gene targeting. A more complete understanding of these mechanisms is also relevant for future therapeutic applications of protocols based on ZFN, TALEN, and CRISPR-Cas 9 approaches.

## Materials and methods

### Cellular system and modification protocol

**MEF-mutEGFP cellular system.** A cellular system composed of MEF with a stably integrated mutated EGFP gene

was used as previously set up and described by our group.<sup>16</sup> Briefly, MEF were isolated initially from a knockout SMA1 mice (generated by prof. Arthur Burghes, Ohio State University)<sup>48</sup> and subsequently immortalized with the SV40 virus.<sup>16</sup> The wild-type sequence of the EGFP gene (cloned between the pCMV promoter and SV40-pA) was mutated using a QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) to insert a nonsense mutation at codon 70 (CAG > TAG), thus creating a premature stop and a truncated nonfluorescent EGFP protein. Stable MEF clones integrating a variable number of mutated EGFP genes were produced. The C2 MEF-mutEGFP clone, integrating 13 copies of the mutated EGFP (mutEGFP) gene sequence, was chosen for the experiments described in this work. Cells were cultured, at 37 °C under 5% CO<sub>2</sub>, in Dulbecco's modified Eagle medium with 10% fetal bovine serum, 1% L-glutamine, 1% penicillin/streptomycin, 120 ng/µl G418, 1% nonessential amino acids, 20 mmol/l HEPES (all from Euroclone, Milan,

Italy), 0.01 mmol/l 2-β-mercaptoethanol (Gibco, Life Technologies, Foster City, CA), 200 ng/μl of hygromycin (Sigma-Aldrich, Milan, Italy), according to a previously optimized protocol.<sup>16</sup> For each experimental condition,  $1.7 \times 10^6$  MEF-mutEGFP were used. Each experimental condition was assayed at least in triplicate.

**SDF features and nucleofection.** An SDF with previously optimized features and nucleofection parameters already set up was used.<sup>16</sup> Briefly, a 876 bp double-stranded DNA SDF homologous to EGFP wild-type sequence (named SDF-PCR-WT) was obtained by amplifying the wild-type EGFP region cloned in the pCR-2.1 vector (Invitrogen, CA) using the following primers: forward, 5'-ACTCATCAATGTATCT-TATCAT-3'; reverse, 5'-AGGTCTATATAAGCAGAGCT-3'. The PCR product was purified from a 1% agarose gel using a QIAquick Gel Extraction Kit (Qiagen, Manchester, UK). The sequence of the SDF was checked by DNA sequencing and its quantity was determined using a spectrophotometer (ND-1000, Nanodrop, Thermo Scientific Europe, Monza, Italy). In particular,  $3 \times 10^6$  SDFs/cell (corresponding to 5 μg/ $1.7 \times 10^6$  cells) or  $12 \times 10^6$  SDFs/cell (corresponding to 20 μg/ $1.7 \times 10^6$  cells) were used in experiments using a real-time expression array (qRT-PCR array) to test the effects of low and high SDF doses, respectively. Cells were suspended in 100 μl of nucleofection buffer MEF-2, and electroporation was performed using the Amaxa Nucleofection System (Lonza, Cologne, Germany) with the nucleofection program T-20 for the transfection of  $1.7 \times 10^6$  cells with the two different amounts of the SDF. Simultaneously,  $1.7 \times 10^6$  cells were electroporated under the same experimental conditions without adding SDF. Then, the cells were incubated at 37 °C under 5% CO<sub>2</sub>. Untransfected control cells were also plated.

**Cellular growth and modification.** For both growth and modification evaluations, every experimental condition was tested at least in triplicate (three independent biological samples from three independent experiments). Cellular growth was evaluated by counting trypsinized cells for every experimental condition at 8, 24, and 72 hours after transfection (or after plating for untransfected controls). For microscope count, 10 μl of cell suspension was added to 10 μl of 2× Trypan blue to exclude dead cells. Ten microliters of this mix was placed in a glass slide to perform cell counts. Cellular growth is expressed as the ratio of the number of total living cells to the number of cells that initially underwent transfection (or plating for untransfected controls). Cellular viability was evaluated by flow cytometry by using BD FACSAria (BD Biosciences, Erembodegem, Belgium). Dead cells were excluded by adding the nucleic acid dye TO-PRO-3 Iodide (20 nmol/l; Invitrogen) to a suspension of  $10^6$  cells. Cellular viability was evaluated at 24 and 72 hours after transfection (or plating for untransfected controls) and expressed as the percentage of living cells. The targeted modification rate was evaluated at 72 hours after transfection (or plating for untransfected controls) by flow cytometry analysis. The data from  $3 \times 10^5$  living cells were analyzed using BD FACSDiva Software version 6.1.3 (BD Biosciences), to obtain the percentage of EGFP-positive cells. To gate EGFP-positive cells, parental wild-type

EGFP cells<sup>16</sup> were used. The remaining trypsinized cells were pelleted and used to perform RNA extraction.

**RNA extraction, DNase treatment, and retrotranscription.** Total RNA was extracted according to the RNeasy Mini Kit protocol (Qiagen) and dosed using a spectrophotometer (ND-1000, Nanodrop) to determine the concentration and the A260/A280 ratio. The quality of the extracted RNA was evaluated by agarose denaturing gel electrophoresis. DNase treatment was performed to remove residual amounts of contaminating genomic DNA according to the following protocol: 6 μg of total RNA was incubated with 2.2 units of DNase I (New England Biolabs, Ipswich, MA) at 37 °C for 10' in a final volume of 20 μl. Subsequently, the sample was treated with 1× ethylenediaminetetraacetic acid (50 mmol/l, pH = 8) at 75 °C for 10' to deactivate the DNase I. A further DNase treatment step was performed by treating 2.4 μg of RNA (300 ng/μl) with Buffer GE (5×) (SABioscience, Qiagen) by incubation at 42 °C for 5'. After both DNase treatments, reverse transcription was performed using a RT<sup>2</sup> First Strand Kit (SABioscience, Qiagen) according to the manufacturer's protocol: the reverse transcription mix was prepared in a final volume of 10 μl and subsequently added to 10 μl of DNase-treated RNA preparation. This mix was incubated at 42 °C for 15' and then stopped immediately by incubating at 95 °C for 5'. Then, the reaction was placed on ice for 5', and 91 μl of RNase-free water was added to proceed with the qRT-PCR array protocol.

**Quantitative expression study using qRT-PCR arrays of 84 DNA repair genes and 84 cell cycle genes.** For the quantitative expression analysis, every experimental condition was tested at least in triplicate (three independent biological samples from three independent experiments) at 8, 24, and 72 hours after transfection (or plating for untransfected controls). The quantitative analysis of the expression of 84 genes involved in different DNA repair (**Supplementary Figure S1**) and cell cycle (**Supplementary Figure S2**) pathways was conducted using the RT<sup>2</sup> Profiler PCR Array (SABioscience, Qiagen). Each array contains primer assays for 84 pathway genes and for 5 housekeeping genes that enable data normalization. In addition, several controls are included: one genomic DNA control, three reverse transcription controls, and three positive PCR controls. The genomic DNA control specifically detects genomic DNA contamination with a high level of sensitivity. The reverse transcription control tests the efficiency of the reverse transcription reaction performed using the RT<sup>2</sup> First Strand Kit, detecting a template synthesized from the kit's built-in external RNA control. The positive PCR controls consist of a pre-dispensed artificial DNA sequence and an accompanying detection assay to determine PCR efficiency. The controls provided in replicates are also used to test for interwell and intraplate consistency. The analyzed genes belonging to DNA repair and cell cycle pathways are listed in the legends of **Supplementary Figures S1 and S2**, respectively.

Real-time PCR was performed using RT<sup>2</sup> Profiler PCR Arrays in combination with RT<sup>2</sup> SYBR Green Mastermix (SABioscience, Qiagen). The PCR array master mix was composed of 1,350 μl of 2× RT<sup>2</sup> SYBR Green Master Mix,

102  $\mu$ l of cDNA synthesis reaction, and 1,248  $\mu$ l of RNase-free water in a final volume of 2,700  $\mu$ l. Twenty-five microliters of the PCR array master mix was dispensed into each well of the RT<sup>2</sup> Profiler PCR Array. The plates were run on an Applied Biosystems 7500 Fast real-time PCR system (Applied Biosystems, Foster City, CA) according to the following cycle: 95 °C for 10', followed from 40 cycles at 95 °C for 15," and 60 °C for 1'. As control for primer specificity, the derivative melting curves of the PCR products generated at the end of the amplification cycle were used.

Once the threshold cycles ( $C_T$ ) of the individual genes were acquired, the analysis was performed using the provided Excel analysis template (SABioscience, Qiagen).  $\Delta C_T$  values were calculated using the difference between the  $C_T$  values of the individual genes of interest and the average  $C_T$  values of the housekeeping genes. All the tested housekeeping genes remained stable in both the controls and the experiments (with a difference  $< 1.5 C_T$ ); thus, the average  $C_T$  values of the housekeeping genes were calculated considering all values. The  $\Delta\Delta C_T$  was calculated by the difference between the  $\Delta C_T$  of the gene in the experiment and the  $\Delta C_T$  of the gene in the control. The fold change was calculated as  $2^{(-\Delta\Delta C_T)}$ . Average values were calculated using the results of at least three independent experiments.

**Statistical analysis.** The SPSS (SPSS, Chicago, IL) package was used for statistical analysis. For the analysis of modification efficiency and growth, a two-tailed Student's *t*-test and one-way ANOVA were performed. For the analysis of the overall pattern of the number of modulated genes,  $\chi^2$  and contingency tables were used. For the overall analysis of the average fold change of modulated genes, a two-way ANOVA with Bonferroni's posttest was used. For the comparison of the average fold change of modulated genes between specific pairs of experimental conditions, Student's *t*-test for paired samples was used.

Preliminarily, we studied the overall pattern of expression change, analyzing both the number of genes with an altered expression pattern and their average fold change. After this preliminary analysis, a more stringent statistical analysis was performed (see below).

As preliminary analysis, we compared the expression fold change of experiments with the respective controls to select three different classes of expression modulation. We considered genes with a fold change greater than 1.2 with respect to the control to be upregulated, genes with a fold change less than 0.8 with respect to the control to be downregulated, and genes with a fold change from 0.8 to 1.2 with respect to the control to have expression levels similar to those of the control. We performed a statistical analysis of the number of upregulated, downregulated, and unchanged (three expression classes) genes at 8, 24, and 72 hours (three experimental times) using 3×3 contingency tables (three rows for the experimental times and three columns for the expression classes). The following six comparisons were tested using a corresponding number of contingency tables: cells transfected with 5 or 20  $\mu$ g of the SDF with respect to the untransfected control cells, cells transfected with 20  $\mu$ g of the SDF with respect to cells transfected with 5  $\mu$ g of the SDF, cells transfected with no SDF with respect to untransfected control cells, and cells

transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF. In addition, we analyzed the SDF dose effect using 3×3 contingency tables (three rows for control, 5 and 20  $\mu$ g of the SDF, and three columns for the three experimental times) testing the following four comparisons: genes upregulated with respect to the untransfected control cells, genes upregulated with respect to cells transfected with no SDF, genes downregulated with respect to the untransfected control cells, and genes downregulated with respect to cells transfected with no SDF.

The preliminary statistical analysis was also performed on the average fold change of upregulated and downregulated genes by a two-way ANOVA (treatments and timing) and Bonferroni's posttest. The following comparisons were tested for both up- and downregulated genes (overall four comparisons): cells transfected with 0, 5, or 20  $\mu$ g of the SDF with respect to untransfected control and cells transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF. For both up- and downregulated genes, three pairs of experimental conditions were tested by Student's *t*-test for paired data: cells transfected with 5 or 20  $\mu$ g of the SDF with respect to cells transfected with no SDF and cells transfected with 20  $\mu$ g of the SDF with respect to cells transfected with 5  $\mu$ g of the SDF.

After the overall analysis reported above, which considered all genes grouped into arbitrary classes of expression level, a more statistically stringent analysis was performed to select the genes with statistically significant expression modifications with respect to the control. In this case, a two-tailed Student's *t*-test with Bonferroni's correction for multiple tests was used, with a corrected level of  $P < 0.0006$  considered statistically significant. Only for these preselected genes, a two-tailed Student's *t*-test was further used to statistically evaluate both the expression differences with respect to the control at any experimental time point and a dose effect according to the quantity of the SDF administered. In this case,  $P < 0.05$  was considered statistically significant.

## Supplementary material

**Figure S1.** DNA repair RT<sup>2</sup> Profiler PCR Array.

**Figure S2.** Cell cycle RT<sup>2</sup> Profiler PCR Array.

**Figure S3.** Basal expression of DNA repair genes in untransfected MEF-mutEGFP at 8 h (black bars), 24 h (green bars) and 72 h (red bars) after plating (simultaneously to nucleofected experimental lines).

**Figure S4.** Overall analysis of the fold change of the studied genes of the DNA repair pathway in MEF-mutEGFP at 8 h, 24 h and 72 h after treatment, according to different experimental conditions (from a to f).

**Figure S5.** Basal expression of cell cycle genes in untransfected MEF-mutEGFP at 8 h (purple bars), 24 h (yellow bars) and 72 h (light blue bars) after plating (simultaneously with nucleofected experimental lines).

**Figure S6.** Overall analysis of the fold change of the studied genes of the cell cycle pathway in MEF-mutEGFP at 8 h, 24 h and 72 h after treatment, according to different experimental conditions (from a to f).

**Brief descriptions and database links for the selected genes**



**Acknowledgments** This work was supported by the Istituto Pasteur Fondazione Cenci Bolognetti (project 2013–2015). S.P. was partially supported by the Istituto Pasteur Fondazione Cenci Bolognetti. The authors declare no conflict of interests. The work was done in Rome, Italy.

- Sanguolo, F and Novelli, G (2004). Sequence-specific modification of mouse genomic DNA mediated by gene targeting techniques. *Cytogenet Genome Res* **105**: 435–441.
- Gruenert, DC, Bruscia, E, Novelli, G, Colosimo, A, Dallapiccola, B, Sanguolo, F et al. (2003). Sequence-specific modification of genomic DNA by small DNA fragments. *J Clin Invest* **112**: 637–641.
- Capecchi, MR (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**: 507–512.
- Fischer, A and Cavazzana-Calvo, M (2008). Gene therapy of inherited diseases. *Lancet* **371**: 2044–2047.
- Sorrell, DA and Kolb, AF (2005). Targeted modification of mammalian genomes. *Biotechnol Adv* **23**: 431–469.
- Vasquez, KM, Marburger, K, Intody, Z and Wilson, JH (2001). Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci USA* **98**: 8403–8410.
- Tenzen, T, Zembowicz, F and Cowan, CA (2010). Genome modification in human embryonic stem cells. *J Cell Physiol* **222**: 278–281.
- Gaj, T, Gersbach, CA and Barbas, CF 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**: 397–405.
- Gupta, RM and Musunuru, K (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**: 4154–4161.
- Doudna, JA and Charpentier, E (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096.
- Xiao-Jie, L, Hui-Ying, X, Zun-Ping, K, Jin-Lian, C and Li-Juan, J (2015). CRISPR-Cas9: a new and promising player in gene therapy. *J Med Genet* **52**: 289–296.
- Gruenert, DC (1998). Gene correction with small DNA fragments. *Curr Res Mol Ther* **1**: 607–613.
- Gruenert, DC (1999). Opportunities and challenges in targeting genes for therapy. *Gene Ther* **6**: 1347–1348.
- Yáñez, RJ and Porter, AC (1998). Therapeutic gene targeting. *Gene Ther* **5**: 149–159.
- Goncz, KK and Gruenert, DC (2000). Site-directed alteration of genomic DNA by small-fragment homologous replacement. *Methods Mol Biol* **133**: 85–99.
- Luchetti, A, Filareto, A, Sanchez, M, Ferraguti, G, Lucarelli, M, Novelli, G et al. (2012). Small fragment homologous replacement: evaluation of factors influencing modification efficiency in an eukaryotic assay system. *PLoS One* **7**: e30851.
- Colosimo, A, Goncz, KK, Novelli, G, Dallapiccola, B and Gruenert, DC (2001). Targeted correction of a defective selectable marker gene in human epithelial cells by small DNA fragments. *Mol Ther* **3**: 178–185.
- Kunzelmann, K, Legendre, JY, Knoell, DL, Escobar, LC, Xu, Z and Gruenert, DC (1996). Gene targeting of CFTR DNA in CF epithelial cells. *Gene Ther* **3**: 859–867.
- Maurisse, R, Cheung, J, Widdicombe, J and Gruenert, DC (2006). Modification of the pig CFTR gene mediated by small fragment homologous replacement. *Ann N Y Acad Sci* **1082**: 120–123.
- Sanguolo, F, Bruscia, E, Serafino, A, Nardone, AM, Bonifazi, E, Lais, M et al. (2002). *In vitro* correction of cystic fibrosis epithelial cell lines by small fragment homologous replacement (SFHR) technique. *BMC Med Genet* **3**: 8.
- Sanguolo, F, Scaldaferrri, ML, Filareto, A, Spitalieri, P, Guerra, L, Favia, M et al. (2008). Cfr gene targeting in mouse embryonic stem cells mediated by small fragment homologous replacement (SFHR). *Front Biosci* **13**: 2989–2999.
- Kapsa, R, Quigley, A, Lynch, GS, Steeper, K, Kornberg, AJ, Gregorevic, P et al. (2001). *In vivo* and *in vitro* correction of the mdx dystrophin gene nonsense mutation by short-fragment homologous replacement. *Hum Gene Ther* **12**: 629–642.
- Kapsa, RM, Quigley, AF, Vadolos, J, Steeper, K, Ioannou, PA, Byrne, E et al. (2002). Targeted gene correction in the mdx mouse using short DNA fragments: towards application with bone marrow-derived cells for autologous remodeling of dystrophic muscle. *Gene Ther* **9**: 695–699.
- Todaro, M, Quigley, A, Kita, M, Chin, J, Lowes, K, Kornberg, AJ et al. (2007). Effective detection of corrected dystrophin loci in mdx mouse myogenic precursors. *Hum Mutat* **28**: 816–823.
- Sanguolo, F, Filareto, A, Spitalieri, P, Scaldaferrri, ML, Mango, R, Bruscia, E et al. (2005). *In vitro* restoration of functional SMN protein in human trophoblast cells affected by spinal muscular atrophy by small fragment homologous replacement. *Hum Gene Ther* **16**: 869–880.
- Spitalieri, P, Cortese, G, Pietropoli, A, Filareto, A, Dolci, S, Klinger, FG et al. (2009). Identification of multipotent cytotrophoblast cells from human first trimester chorionic villi. *Cloning Stem Cells* **11**: 535–556.
- Bedayat, B, Abdolmohamadi, A, Ye, L, Maurisse, R, Parsi, H, Schwarz, J et al. (2010). Sequence-specific correction of genomic hypoxanthine-guanine phosphoribosyl transferase mutations in lymphoblasts by small fragment homologous replacement. *Oligonucleotides* **20**: 7–16.
- Zayed, H, McIvor, RS, Wiest, DL and Blazar, BR (2006). *In vitro* functional correction of the mutation responsible for murine severe combined immune deficiency by small fragment homologous replacement. *Hum Gene Ther* **17**: 158–166.
- Colosimo, A, Guida, V, Antonucci, I, Bonfini, T, Stuppia, L and Dallapiccola, B (2007). Sequence-specific modification of a beta-thalassemia locus by small DNA fragments in human erythroid progenitor cells. *Haematologica* **92**: 129–130.
- Goncz, KK, Prokopyshyn, NL, Chow, BL, Davis, BR and Gruenert, DC (2002). Application of SFHR to gene therapy of monogenic disorders. *Gene Ther* **9**: 691–694.
- Goncz, KK, Prokopyshyn, NL, Abdolmohamadi, A, Bedayat, B, Maurisse, R, Davis, BR et al. (2006). Small fragment homologous replacement-mediated modification of genomic beta-globin sequences in human hematopoietic stem/progenitor cells. *Oligonucleotides* **16**: 213–224.
- Goncz, KK, Kunzelmann, K, Xu, Z and Gruenert, DC (1998). Targeted replacement of normal and mutant CFTR sequences in human airway epithelial cells using DNA fragments. *Hum Mol Genet* **7**: 1913–1919.
- Goncz, KK, Colosimo, A, Dallapiccola, B, Gagné, L, Hong, K, Novelli, G et al. (2001). Expression of DeltaF508 CFTR in normal mouse lung after site-specific modification of CFTR sequences by SFHR. *Gene Ther* **8**: 961–965.
- Thorpe, P, Stevenson, BJ and Porteous, DJ (2002). Optimising gene repair strategies in cell culture. *Gene Ther* **9**: 700–702.
- Luchetti, A, Malgieri, A and Sanguolo, F (2014). Small fragment homologous replacement (SFHR): sequence-specific modification of genomic DNA in eukaryotic cells by small DNA fragments. *Methods Mol Biol* **1114**: 85–101.
- Sargent, RG, Suzuki, S and Gruenert, DC (2014). Nuclease-mediated double-strand break (DSB) enhancement of small fragment homologous recombination (SFHR) gene modification in human-induced pluripotent stem cells (hiPSCs). *Methods Mol Biol* **1114**: 279–290.
- Sargent, RG, Kim, S and Gruenert, DC (2011). Oligo/polynucleotide-based gene modification: strategies and therapeutic potential. *Oligonucleotides* **21**: 55–75.
- Warmerdam, DO and Kanaar, R (2010). Dealing with DNA damage: relationships between checkpoint and repair pathways. *Mutat Res* **704**: 2–11.
- Mir, LM (2014). Electroporation-based gene therapy: recent evolution in the mechanism description and technology developments. *Methods Mol Biol* **1121**: 3–23.
- Yao, C, Guo, F, Li, C and Sun, C (2013). Gene transfer and drug delivery with electric pulse generators. *Curr Drug Metab* **14**: 319–323.
- Gresch, O, Engel, FB, Nestic, D, Tran, TT, England, HM, Hickman, ES et al. (2004). New non-viral method for gene transfer into primary cells. *Methods* **33**: 151–163.
- Lakshminath, U, Pelacho, B, Sudo, K, Linehan, JL, Coucouvanis, E, Kaufman, DS et al. (2004). Efficient transfection of embryonic and adult stem cells. *Stem Cells* **22**: 531–543.
- Lenz, P, Bacot, SM, Frazier-Jessen, MR and Feldman, GM (2003). Nucleoporation of dendritic cells: efficient gene transfer by electroporation into human monocyte-derived dendritic cells. *FEBS Lett* **538**: 149–154.
- Maasho, K, Marusina, A, Reynolds, NM, Coligan, JE and Borrego, F (2004). Efficient gene transfer into the human natural killer cell line, NK1, using the Amaxa nucleofection system. *J Immunol Methods* **284**: 133–140.
- Trompeter, HI, Weinhold, S, Thiel, C, Wernet, P and Uhrberg, M (2003). Rapid and highly efficient gene transfer into natural killer cells by nucleofection. *J Immunol Methods* **274**: 245–256.
- Lieberman, HB (2006). Rad9, an evolutionarily conserved gene with multiple functions for preserving genomic integrity. *J Cell Biochem* **97**: 690–697.
- Mazur, DJ and Perrino, FW (1999). Identification and expression of the TREX1 and TREX2 cDNA sequences encoding mammalian 3'→5' exonucleases. *J Biol Chem* **274**: 19655–19660.
- Conner, DA (2001). Mouse embryo fibroblast (MEF) feeder cell preparation. *Curr Protoc Mol Biol Unit* **23.2**.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (<http://www.nature.com/mtna>)

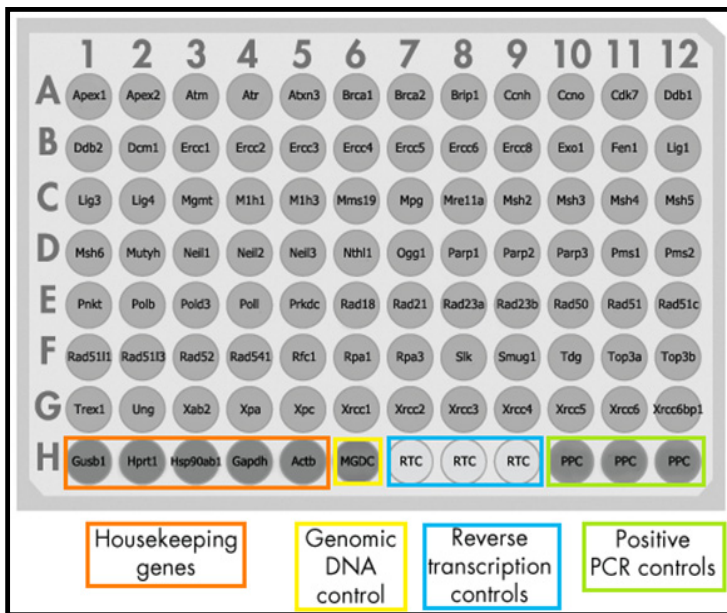


## SUPPLEMENTAL MATERIALS

### Abbreviations

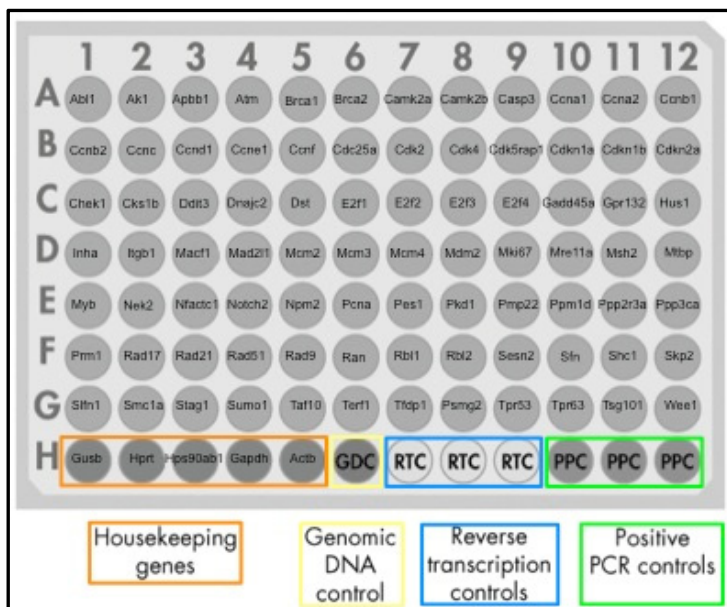
BER= base excision repair  
 NER= nucleotide excision repair  
 MMR= mismatch repair  
 HR= homologous recombination  
 NHEJ= non homologous end joining  
 SSB = single-strand break  
 DSB = double-strand break  
 ORF = open reading frame

### Materials and Methods



**Figure S1 - DNA repair RT<sup>2</sup> Profiler PCR Array.** (See text for explanations).

**BER:** Apex1, Apex2, Ccno, Lig3, Mpg, Mutyh, Neil1, Neil2, Neil3, Nth1, Ogg1, Parp1, Parp2, Parp3, Polb, Smug1, Tdg, Ung, Xrcc1 (19 genes). **NER:** Atr, Atxn3, Brip1, Ccnh, Cdk7, Ddb1, Ddb2, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Ercc6, Ercc8, Lig1, Mms19, Pnkp, Poll, Rad23a, Rad23b, Rfc1, Rpa1, Rpa3, Slk, Xab2, Xpa, Xpc, (27 genes). **MMR:** Exo1, Mlh1, Mlh3, Msh2, Msh3, Msh4, Msh5, Msh6, Pms1, Pms2, Pold3, Trex1 (12 genes). **HR:** Atm, Bracl, Bracl2, Dcm1, Mre11a, Rad18, Rad21, Rad50, Rad51, Rad51c, Rad51l1, Rad51l3, Rad52, Rad54l, Top3a, Top3b Xrcc2, Xrcc3, (18 genes). **NHEJ:** Fen1, Lig4, Prkdc, Xrcc4, Xrcc5, Xrcc6, Xrcc6bp1 (7 genes). **Other gene involved in DNA repair:** Mgmt (1 gene).

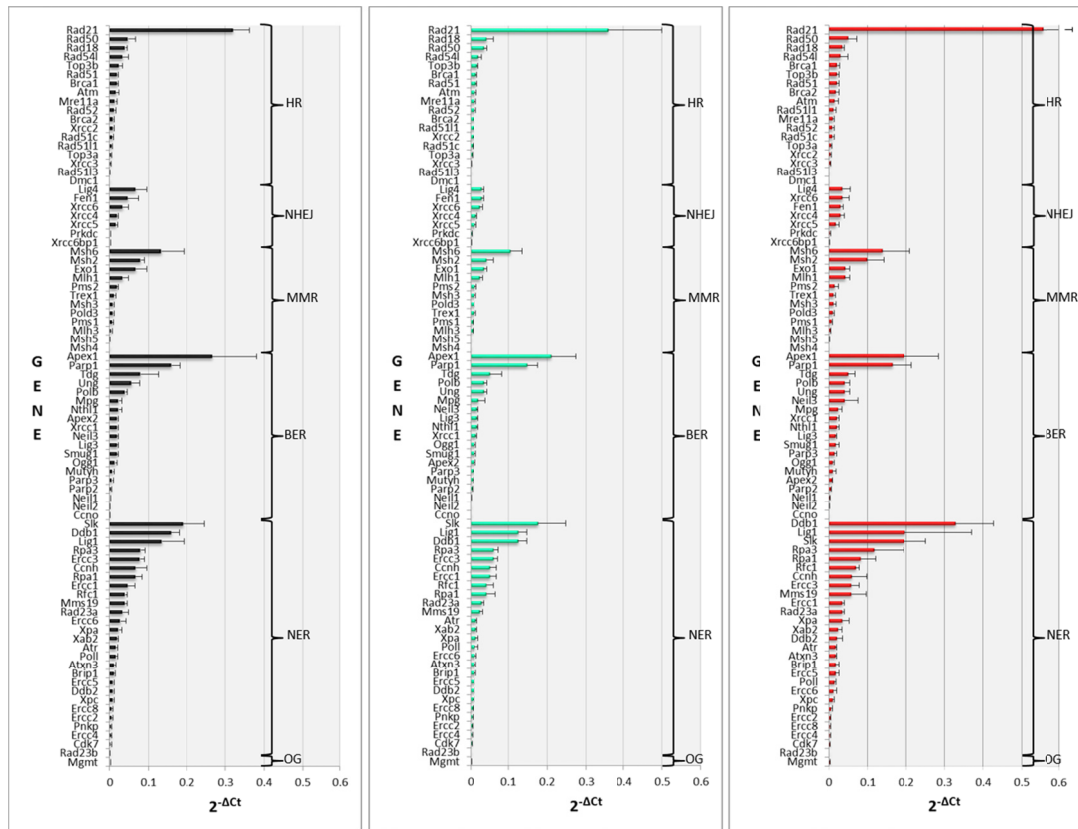


**Figure S2 - Cell cycle RT<sup>2</sup> Profiler PCR Array.** (See text for explanation).

**Checkpoint and Arrest:** Ak1, Apbb1, Bracl2, Casp3, Cdk5rap1, Cdkn1a, Cdkn1b, Cdkn2a, Chek1, Cks1b, Ddit3, Dst, Gadd45a, Hus1, Inha, Macf1, Mad2l1, Mdm2, Msh2, Notch2, Pkd1, Pmp22, Ppm1d, Rad9, Sesn2, Sfn, Sfn1, Smc1a, Tsg101 (29 genes). **Positive Regulation:** Abi1, Ccna1, Ccna2, Ccnb, Ccnb2, Ccnc, Ccnd1, Ccne1, Ccnf, Cdk4, E2f1, E2f2, E2f3, E2f4, Psmg2, Ran, Shc1, Skp2, Tfdp1 (19 genes). **Negative Regulation:** Atm, Bracl, Itgb1, Rbl1, Rbl2, Trp53, Trp63 (7 genes). **G1 and G1/S transition:** Camk2a, Camk2b, Gpr132, Mtbp, Myb, Nfatc1, Ppp2r3a, Ppp3ca, Taf10 (9 genes). **S and replication:** Dnajc2, Mcm2, Mcm3, Mcm4, Mki67, Mre11a, Pcna, Rad17, Rad51, Sumo1 (10 genes). **M phase:** Cdc25a, Cdk2, Nek2, Npm2, Pes1, Prm1, Rad21, Stag1, Terf1, Wee1 (10 genes).

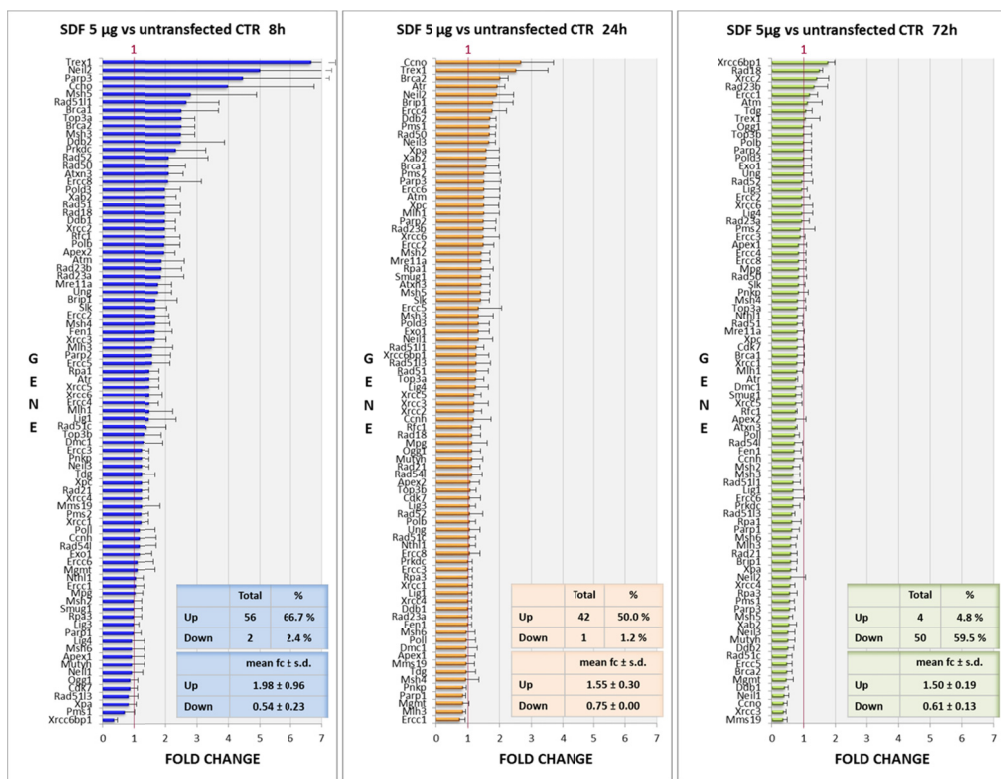
Mki67, Mre11a, Pcna, Rad17, Rad51, Sumo1 (10 genes). **M phase:** Cdc25a, Cdk2, Nek2, Npm2, Pes1, Prm1, Rad21, Stag1, Terf1, Wee1 (10 genes).

## Results

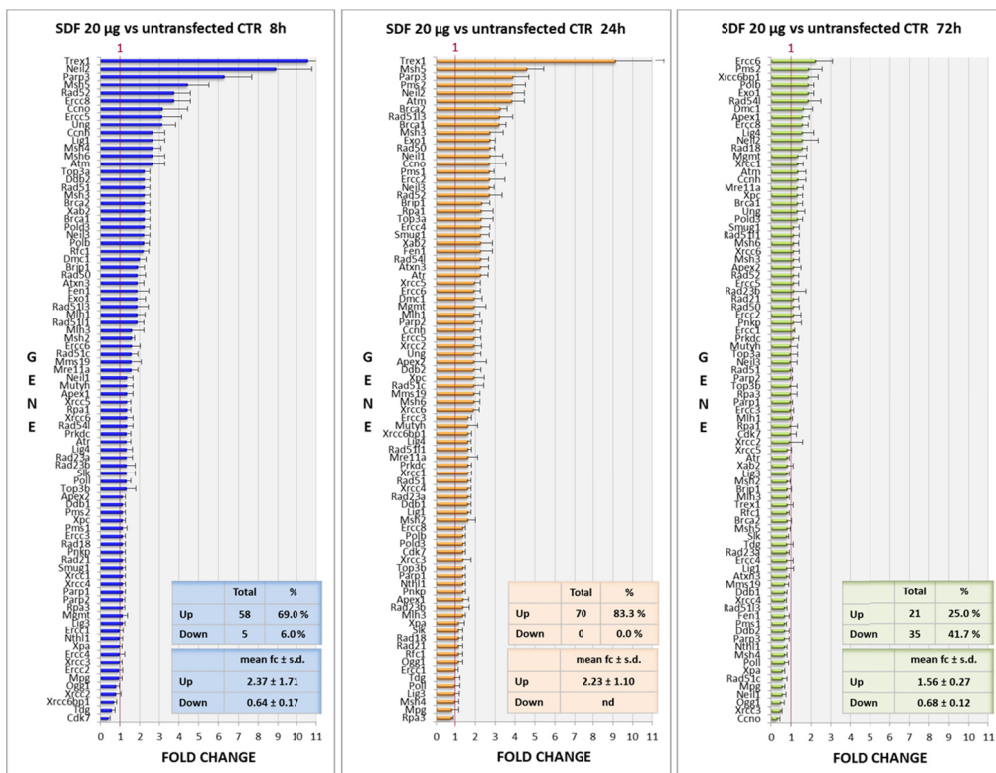


**Figure S3: Basal expression of DNA repair genes in untransfected MEF<sup>-</sup>-mutEGFP at 8 (black bars), 24 (green bars) and 72 hours (red bars) after plating (simultaneously to nucleofected experimental lines). On the right of each image, the grouping of the analyzed genes according to their primary role in the DNA repair pathway is indicated. The gene order is according to the decreasing expression level of each gene; consequently, although similar, the gene order is not identical between panels. Standard deviation (s.d.) bars are indicated.**

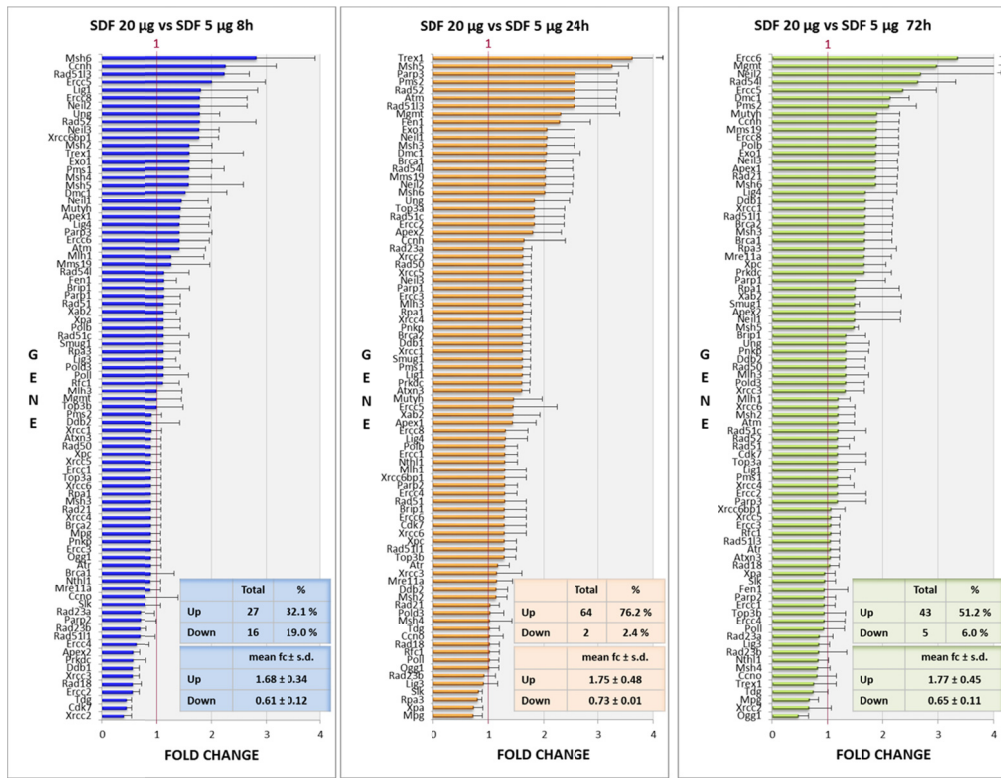
**Figure S4 - Overall analysis of the fold change of the studied genes of the DNA repair pathway in MEF-mutEGFP at 8, 24 and 72 hours (h) after treatment, according to different experimental conditions (from a to f). The average fold change of each gene is shown, with standard deviation (s.d.) bars. Inserts show both the number and percentage of up-regulated (Up) and down-regulated (Down) genes and their overall average fold change (fc) with s.d. (also shown in Table 1 and 2 in the main text).**



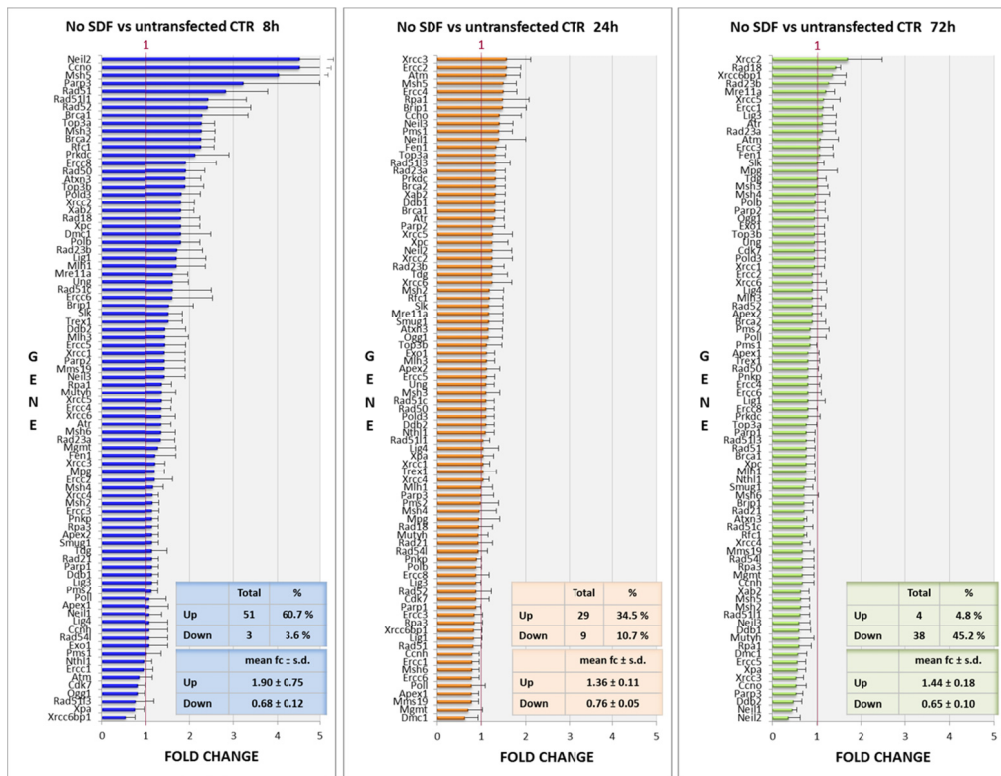
**Figure S4 - a)** cells treated with 5 µg of SDF as compared with untransfected cells (CTR, set to 1).



**Figure S4 - b)** cells treated with 20 µg of SDF as compared with untransfected cells (CTR, set to 1).

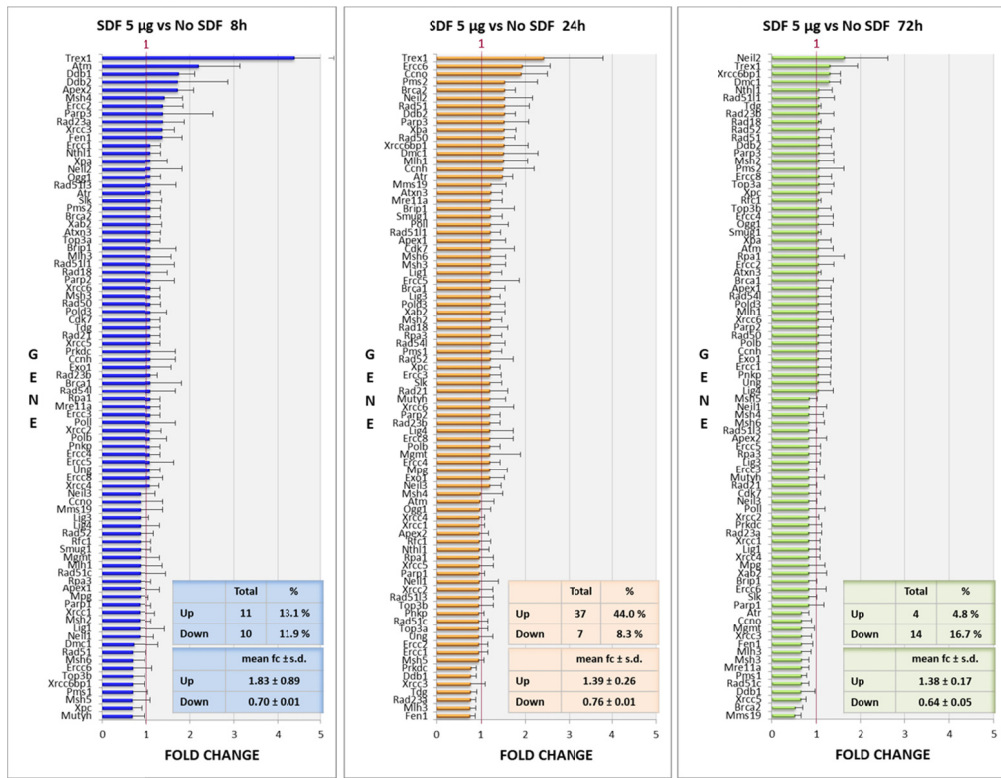


**Figure S4 - c)** cells treated with 20 µg of SDF as compared with those treated with 5 µg of the SDF (the latter set to 1).

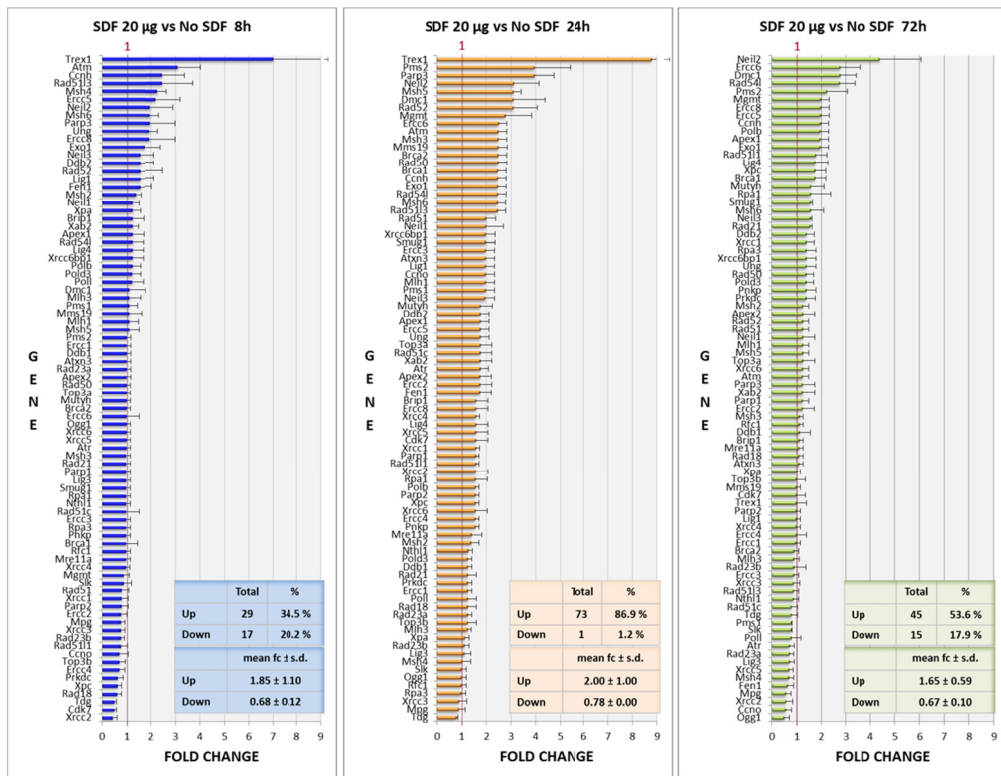


**Figure S4 - d)** cells treated (transfected) without the SDF (indicated in the figure as No SDF) as compared with untransfected cells (CTR, set to 1).

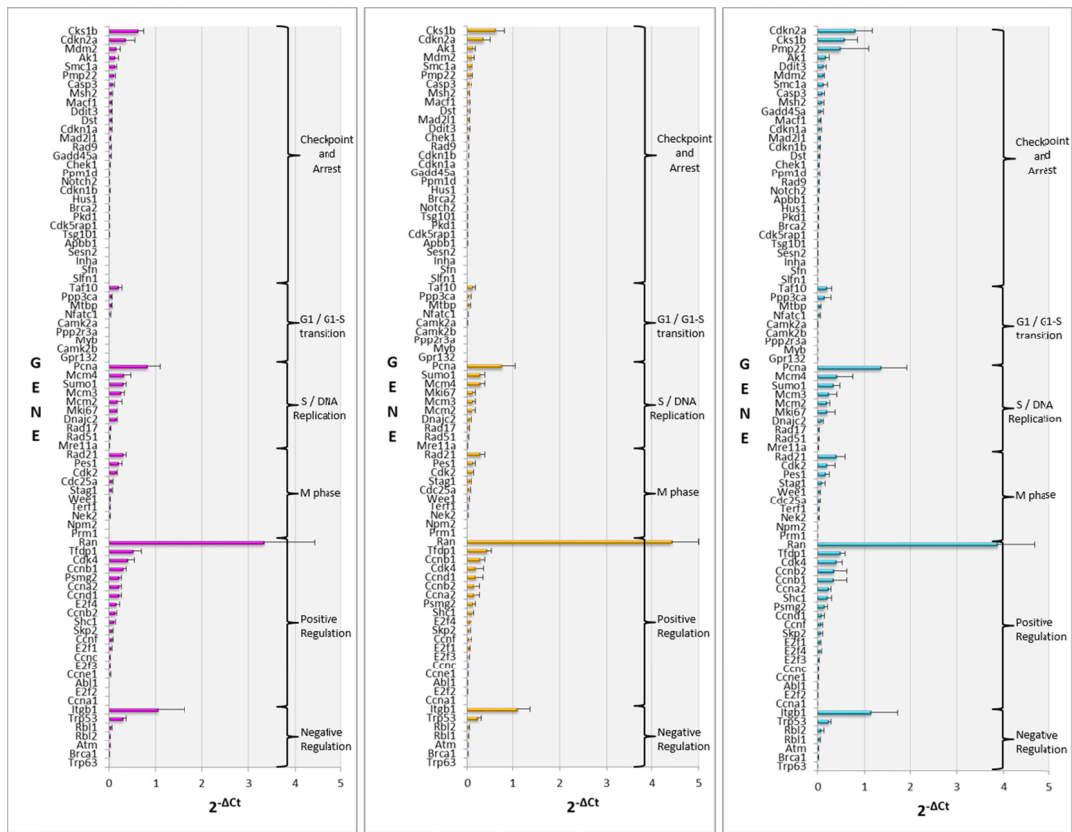




**Figure S4 - e** cells treated with 5 µg of the SDF as compared with cells treated (transfected) without the SDF (indicated in the figure as No SDF, set to 1).

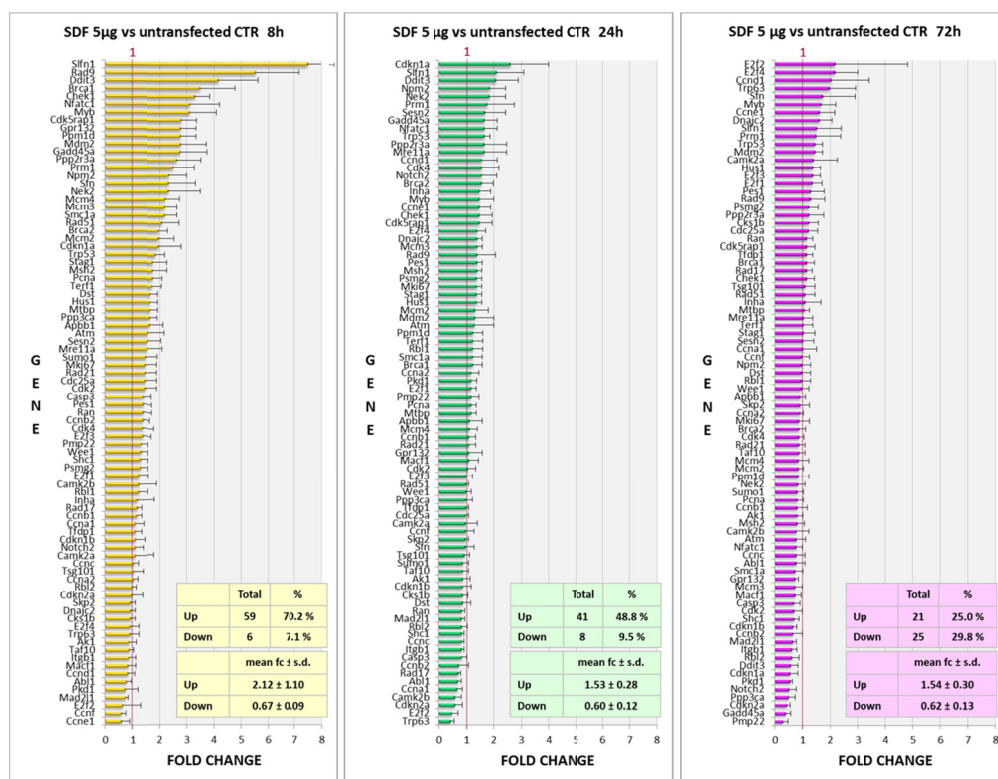


**Figure S4 - f** cells treated with 20 µg of the SDF as compared with cells treated (transfected) without the SDF (indicated in the figure as No SDF, set to 1).

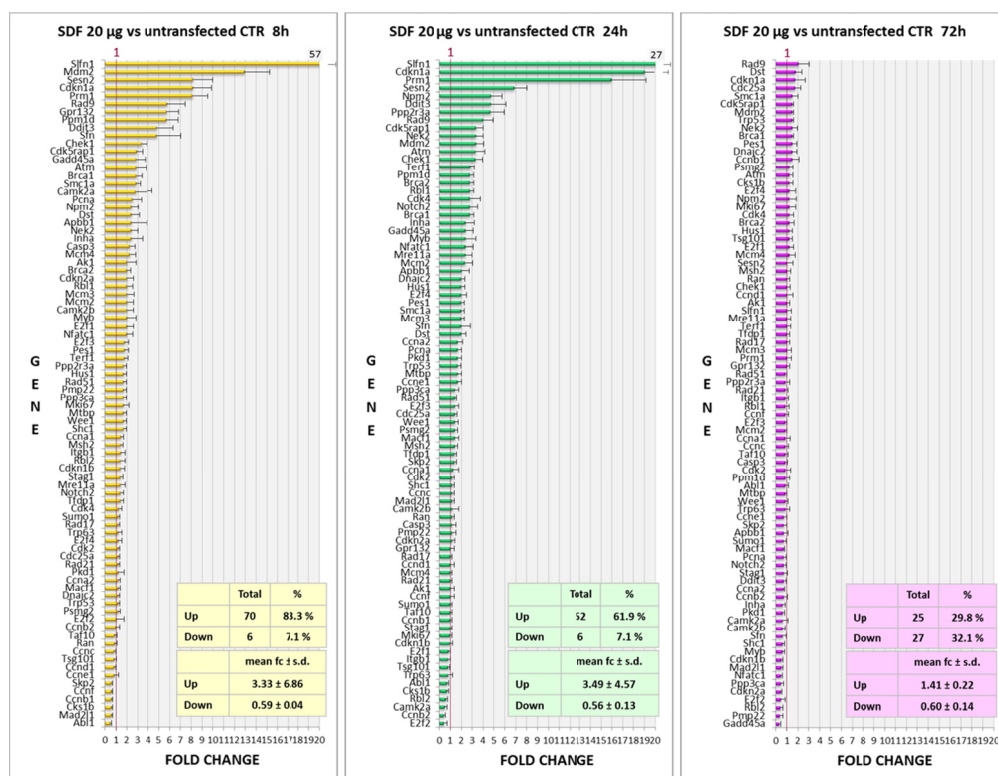


**Figure S5 - Basal expression of cell cycle genes in untransfected MEF-mutEGFP at 8 (purple bars), 24 (yellow bars) and 72 hours (light blue bars) after plating (simultaneously with nucleofected experimental lines). On the right of each image, the grouping of the analyzed genes according to their primary role in the cell cycle pathway is indicated. The gene order is according to the decreasing expression level of each gene; consequently, although similar, the gene order is not identical between panels. Standard deviation (s.d.) bars are indicated.**

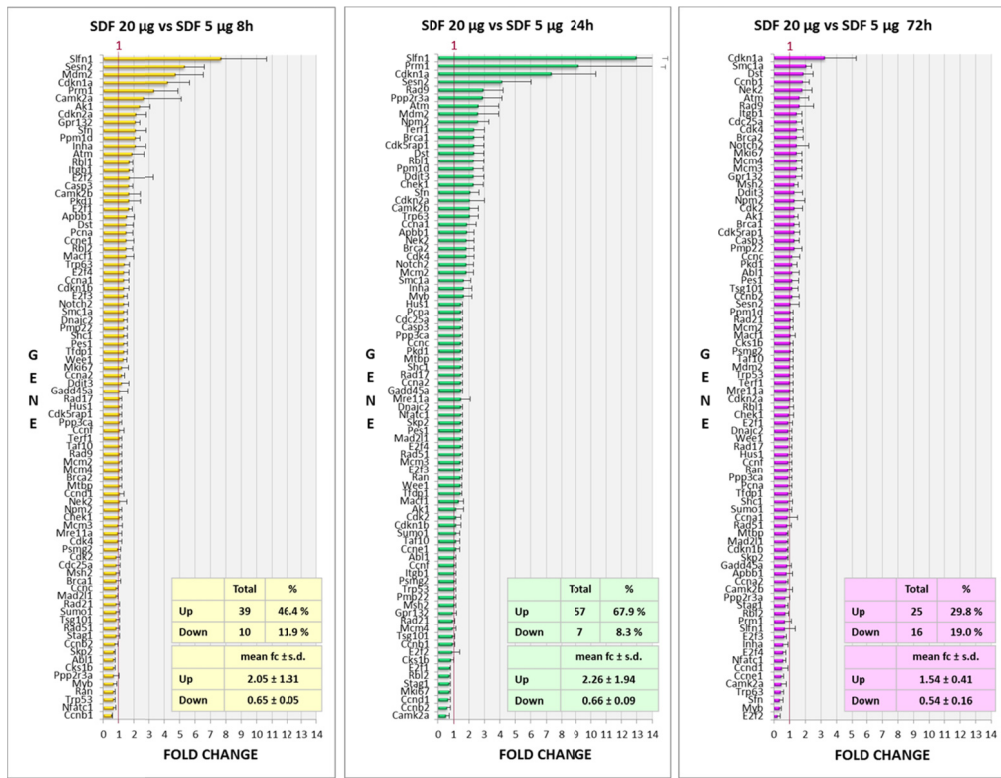
**Figure S6 - Overall analysis of the fold change of the studied genes of the cell cycle pathway in MEF-mutEGFP at 8, 24 and 72 hours (h) after treatment, according to different experimental conditions (from a to f). The average fold change of each gene is shown, with standard deviation (s.d.) bars. Inserts show both the number and percentage of up-regulated (Up) and down-regulated (Down) genes and their overall average fold change (fc) with s.d. (also shown in Table 3 and 4 in the main text).**



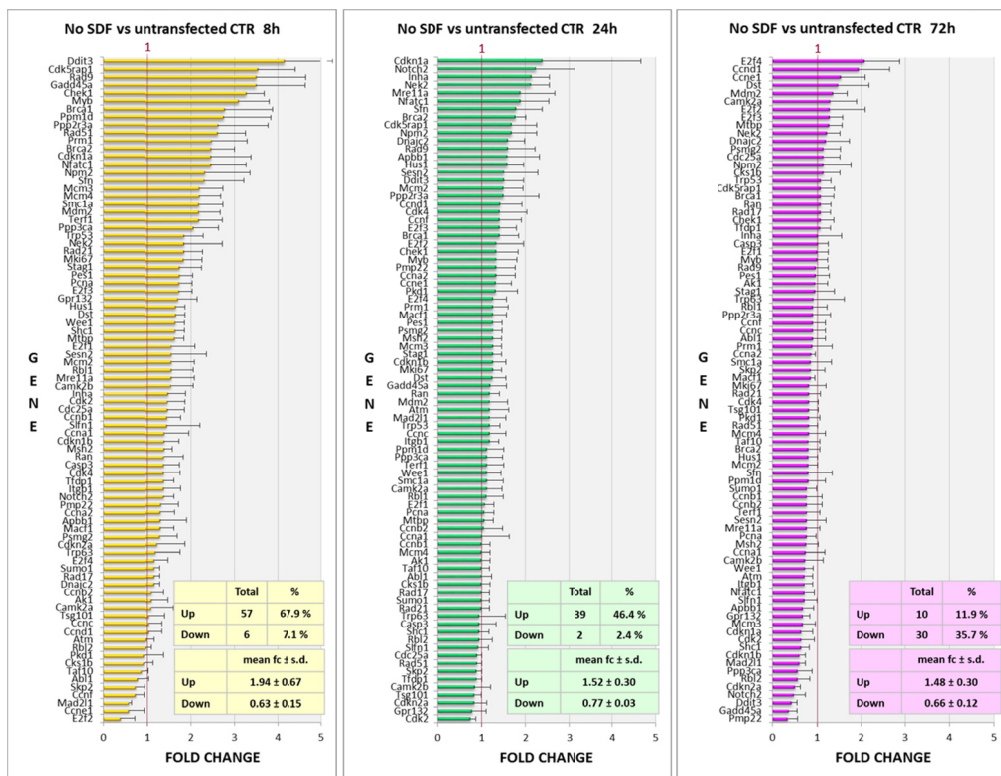
**Figure S6 - a)** cells treated with 5 µg of the SDF as compared with untransfected cells (CTR, set to 1).



**Figure S6 - b)** cells treated with 20 µg of the SDF as compared with untransfected cells (CTR, set to 1).

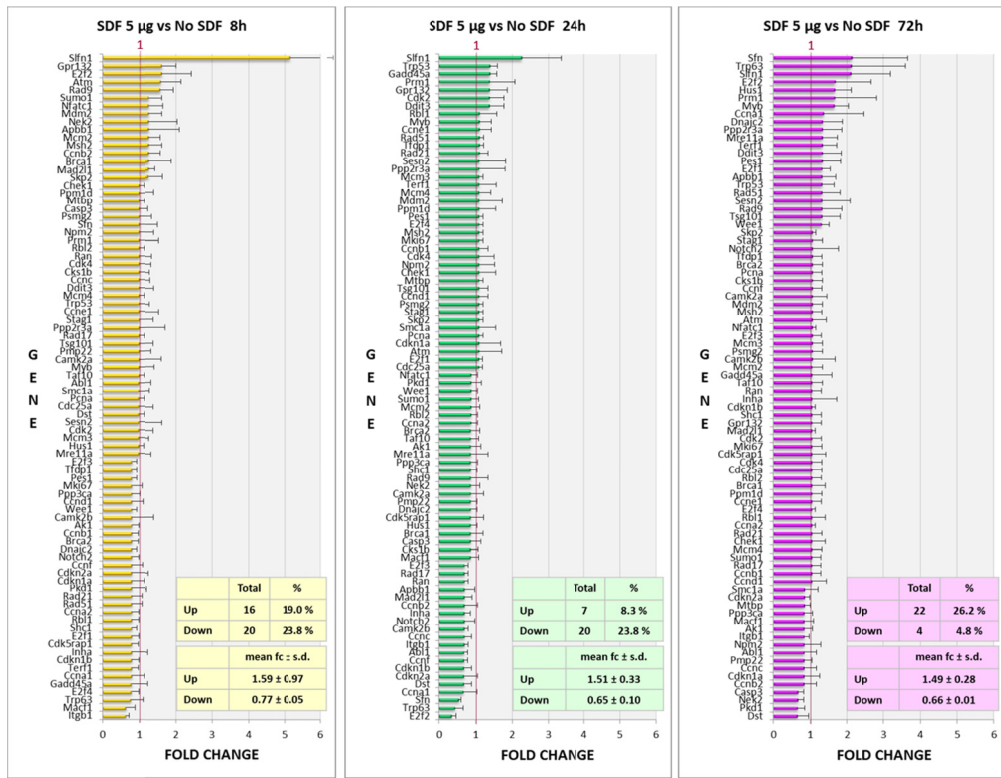


**Figure S6 - c)** cells treated with 20 µg of the SDF as compared with those treated with 5 µg of the SDF (the latter set to 1).

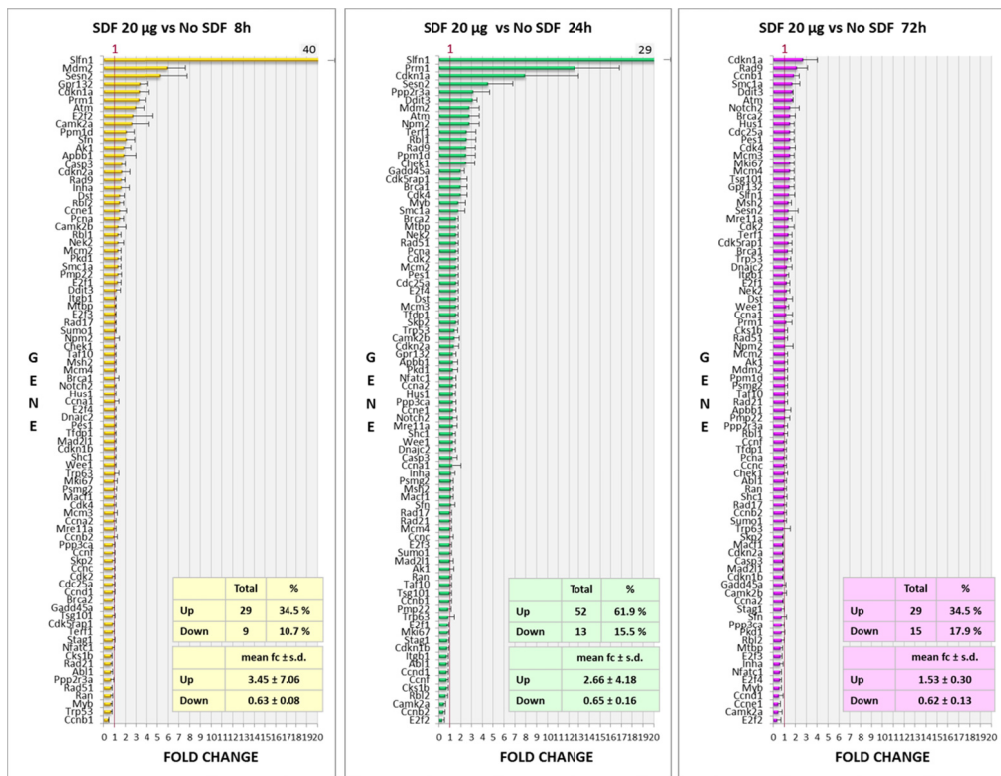


**Figure S6 - d)** cells treated (transfected) without the SDF (indicated in the figure as No SDF) as compared with untransfected cells (CTR, set to 1).





**Figure S6 - e)** cells treated with 5 µg of the SDF as compared with cells treated (transfected) without the SDF (indicated in the figure as No SDF, set to 1).



**Figure S6 - f)** cells treated with 20 µg of the SDF as compared with cells treated (transfected) without the SDF (indicated in the figure as No SDF, set to 1).

## **Brief descriptions and database links for the selected genes.**

The **Neil2** (*nei* endonuclease VIII-like 2 (*E. coli*)) gene encodes for a DNA glycosylase that initiates the first step in BER. This protein recognizes apurinic / apyrimidinic sites and is involved in the repair of lesions in DNA generated during transcription and/or replication, primarily by removing oxidative products of cytosine.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=18956](http://www.genenames.org/data/hgnc_data.php?hgnc_id=18956)

<http://www.informatics.jax.org/marker/MGI:2686058>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=NEIL2>

The **Parp3** (poly (ADP-ribose) polymerase family, member 3) gene encodes for a protein that belongs to a superfamily of ADP-ribosyl transferases (ARTs) that catalyzes the poly(ADP-ribosyl)ation of a limited number of acceptor proteins involved in chromatin architecture and DNA metabolism. This protein is involved in the detection/signaling pathway leading to the repair of both DNA SSBs by BER and DSBs by the NHEJ. It also links the DNA damage surveillance network to the mitotic fidelity checkpoint and is also involved in transcriptional silencing. This protein is preferentially localized to the daughter centriole throughout the cell cycle.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=273](http://www.genenames.org/data/hgnc_data.php?hgnc_id=273)

<http://www.informatics.jax.org/marker/MGI:1891258>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=PARP3>

The **Pms2** (postmeiotic segregation increased 2 (*S. cerevisiae*)) gene encodes for an endonuclease involved in MMR. This protein allows the correction of base-base mismatches and insertion-deletion loops resulting from DNA replication and recombination events.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=9122](http://www.genenames.org/data/hgnc_data.php?hgnc_id=9122)

<http://www.informatics.jax.org/marker/MGI:104288>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=PMS2>

The **Rad51B** (RAD51 homolog D (*S. cerevisiae*)) gene encodes for a protein involved in DSB repair by HR. In complexes formed with other members of the RAD51 family, this protein catalyzes homologous pairing between single- and double-stranded DNA playing a role in the early stage of recombination-based repair of DNA. It is also involved in the disruption of Holliday junctions.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=9823](http://www.genenames.org/data/hgnc_data.php?hgnc_id=9823)

<http://www.informatics.jax.org/marker/MGI:1261809>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=RAD51D>

The **Trex1** (three prime repair exonuclease 1) gene encodes for a protein with both DNA damage checkpoint activity (upstream ORF) and exonuclease activity (downstream ORF). This protein has a critical role in the MMR pathway and a proofreading function for DNA polymerase. It is the major 3' → 5' DNA exonuclease in mammalian cells, with a direct role in the degradation of genomic double-stranded DNA to minimize potential autoimmune activation by persistent self DNA during cell death.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=12269](http://www.genenames.org/data/hgnc_data.php?hgnc_id=12269)

<http://www.informatics.jax.org/marker/MGI:1328317>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=TREX1>

The **Cdkn1a** (cyclin-dependent kinase inhibitor 1A, also known as p21) gene encodes for a protein involved in checkpoint and cell cycle arrest by the negative regulation of the G1 / S transition. This protein binds and inhibits the activity of cyclin-CDK2 or cyclin-CDK1 complexes. Its expression is tightly controlled by p53. It is a negative upstream regulator of DNA methyl transferase 1 (Dnmt1).

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=1784](http://www.genenames.org/data/hgnc_data.php?hgnc_id=1784)

<http://www.informatics.jax.org/marker/MGI:104556>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=CDKN1A>

The **Chek1** (checkpoint kinase 1) gene encodes for a serine / threonine protein kinase that modulates signaling that prevents cell proliferation. This protein plays a role in the G1 / S and intra-S-phase checkpoints leading to cell cycle arrest. It phosphorylates histone H3, inducing chromatin-mediated transcriptional repression. Additionally, it acts as an integrator for Atm and Atr signaling and is activated by BRCA1.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=1925](http://www.genenames.org/data/hgnc_data.php?hgnc_id=1925)

<http://www.informatics.jax.org/marker/MGI:1202065>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=CHEK1>

The **Ddit3** (DNA-damage inducible transcript 3, also known as C / EBP homologous protein (CHOP)) gene encodes for a protein that is a basic leucine zipper transcription factor of the dimer forming C / EBP family. This protein acts primarily as dominant negative regulator of several other transcription factors but can also induce the transcription of downstream target genes. The promoters of its target genes showed no common sequence motifs, reflecting that Ddit3 forms heterodimers with several alternative transcription factors that bind to different motifs. This protein is a stress-inducible nuclear protein involved in the G1 / S checkpoint, in cell cycle arrest, and in the possible induction of apoptosis.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=2726](http://www.genenames.org/data/hgnc_data.php?hgnc_id=2726)

<http://www.informatics.jax.org/marker/MGI:109247>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=DDIT3>

The **Gpr132** (G protein-coupled receptor 132) gene encodes a member of the G protein-coupled receptor superfamily. This protein is able to induce cell cycle arrest during the G2 / M transition to delay mitosis.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=17482](http://www.genenames.org/data/hgnc_data.php?hgnc_id=17482)

<http://www.informatics.jax.org/marker/MGI:1890220>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=GPR132>

The **Mdm2** (p53 E3 ubiquitin protein ligase homolog (mouse), also known as murine double minute 2) gene encodes for a protein involved in bypassing the G1 checkpoint and in suppressing cell cycle arrest. This protein acts as a negative regulator of p53, achieving p53 repression by binding and blocking the N-terminal trans-activation domain of p53, by acting as an E3 ubiquitin ligase that targets both itself and p53 for degradation, and by inhibiting p53 transcriptional activation. Although the major function of Mdm2 is to suppress p53 activities, emerging research has identified p53-independent roles of Mdm2. The overall effect of Mdm2 overexpression is to induce genomic instability through inhibiting DSB repair and suppressing cell cycle arrest.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=6973](http://www.genenames.org/data/hgnc_data.php?hgnc_id=6973)

<http://www.informatics.jax.org/marker/MGI:96952>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=MDM2>

The **Prm1** (protamine 1) gene encodes for a basic chromosomal protein that substitute for histones in the chromatin of sperm during the haploid phase of spermatogenesis. This protein is also involved in mitotic chromosome condensation (M phase), replacing histones during chromatin remodeling.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=9447](http://www.genenames.org/data/hgnc_data.php?hgnc_id=9447)

<http://www.informatics.jax.org/marker/MGI:97765>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=PRM1>

The **Slfn1** (Schlafen family member 1) gene encodes for a protein that belongs to the Schlafen family of proteins. This protein is involved in the cell cycle checkpoint and cell cycle arrest by the negative regulation of the G1 / S transition.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=25500](http://www.genenames.org/data/hgnc_data.php?hgnc_id=25500)

<http://www.informatics.jax.org/marker/MGI:1313259>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=SLFN12>

The **Atm** (ataxia telangiectasia mutated) gene encodes for a serine / threonine protein kinase that belongs to the superfamily of phosphatidylinositol 3-kinase-related kinases (PIKKs). This protein is recruited and activated by DSBs. The ATM-mediated DNA damage response consists of both a rapid and a delayed



response. The rapid response, including the maintenance of a phosphorylated state of CDK2-cyclin, results in cell cycle arrest at the G1 / S checkpoint, which may allow the DSB repair. If DSBs cannot be repaired, subsequent phosphorylation events lead to stabilization and activation of p53 and to the transcription of several p53 downstream genes, causing long-term cell cycle arrest or even apoptosis.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=795](http://www.genenames.org/data/hgnc_data.php?hgnc_id=795)

<http://www.informatics.jax.org/marker/MGI:107202>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=ATM>

The **Brca1** (breast cancer 1, early onset) gene encodes for a multifunctional protein involved in several pathways that undergo functional modulation by ATM / ATR kinases. This protein activates the repair of DNA DSBs by HR in cooperation with, among others, Brca2 and Rad51. It is a negative regulator of cell growth, acting at the G2 / M checkpoint by activating Chk1 upon DNA damage. It may also induce apoptosis by inducing Ddit1 (Gadd45). The protein interacts with RNA polymerase II and histone deacetylase complexes, playing a role in chromatin remodeling and transcriptional regulation (for example of Cdkn1a (p21)). It also mediates ubiquitination.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=1100](http://www.genenames.org/data/hgnc_data.php?hgnc_id=1100)

<http://www.informatics.jax.org/marker/MGI:104537>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=BRCA1>

The **Brca2** (breast cancer 2, early onset) gene, despite having a very different structure with respect to Brca1, encodes for a multifunctional protein with at least some functions similar and / or interrelated to those of the Brca1 protein. The Brca2 protein is involved in DSB repair by the HR pathway; in particular, the localization of Rad51 to the DSBs requires the formation of a Brca1-Palb2-Brca2 complex. It is also involved in the cell cycle regulation at the G2 / M checkpoint. It has intrinsic histone acetyltransferase activity and is involved in chromatin remodeling and transcriptional control; peculiarly, this protein is involved in transcription-associated recombination.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=1101](http://www.genenames.org/data/hgnc_data.php?hgnc_id=1101)

<http://www.informatics.jax.org/marker/MGI:109337>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=BRCA2>

The **Ppm1d** (protein phosphatase, Mg<sup>2+</sup>/Mn<sup>2+</sup> dependent, 1D) gene is a member of the PP2C family of serine / threonine protein phosphatases. It negatively contributes to both DNA repair and growth inhibitory pathways activated in response to DNA damage in a p53-dependent manner. It appears to suppress NER, turn off DNA damage checkpoint responses, restore chromatin structure, and inhibit senescence and apoptosis.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=9277](http://www.genenames.org/data/hgnc_data.php?hgnc_id=9277)

<http://www.informatics.jax.org/marker/MGI:1858214>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=PPM1D>

The **Rad9** (Rad 9 homolog A (*S. pombe*)) gene encodes for a protein that plays a role in both DNA repair (HR and MMR by base-pair excision repair) and cell cycle checkpoint and arrest. It possesses a 3'→5' exonuclease activity required for cell cycle arrest at the G2 checkpoint in response to incompletely replicated or damaged DNA. It plays a role in telomere stability and homologous recombinational repair as a mechanism for promoting cell survival after ionizing radiation exposure as well as a role in locating Claspin to sites of DNA damage, facilitating its function during the Chk1-mediated checkpoint response.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=9827](http://www.genenames.org/data/hgnc_data.php?hgnc_id=9827)

<http://www.informatics.jax.org/marker/MGI:1328356>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=RAD9A>

The **Sesn2** (sestrin 2) gene encodes for a protein implicated in the link between DNA damage and growth. This protein protects cells against oxidative, genotoxic and energetic stresses and exerts its cytoprotective function by regenerating overoxidized peroxiredoxins, with a major role in the antioxidant defense of the cell. The functional interconnection between the response to DNA damage and cell cycle originates from the fact that the Sesn2 protein is a target of p53 and a negative regulator of mTOR (mammalian target of rapamycin), a positive regulator of cell growth that belongs to the phosphatidylinositol kinase-related kinase (PIKK) family.

[http://www.genenames.org/data/hgnc\\_data.php?hgnc\\_id=20746](http://www.genenames.org/data/hgnc_data.php?hgnc_id=20746)

<http://www.informatics.jax.org/marker/MGI:2651874>

<http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=SESN2>