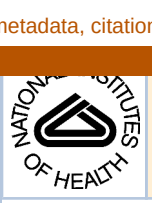

Published in final edited form as:

DNA Repair (Amst). 2005 August 15; 4(9): 1038–1046.

Chromosomal aberrations induced by double strand DNA breaks

Tamas Varga and **Peter D. Aplan**

Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20889, USA

Abstract

It has been suggested that introduction of double-strand DNA breaks into mammalian chromosomes can lead to gross chromosomal rearrangements through improper DNA repair. To study this phenomenon, we employed a model system in which a double-strand DNA break (DSB) can be produced in human cells *in vivo* at a predetermined location. The ensuing chromosomal changes flanking the breakage site can then be cloned and characterized. In this system, the recognition site for the I-*Sce*I endonuclease, whose 18 bp recognition sequence is not normally found in the human genome, is placed between a strong constitutive promoter and the Herpes simplex virus thymidine kinase (*HSV-tk*) gene, which serves as a negative selectable marker. We found that the most common mutation following aberrant DSB repair was an interstitial deletion; these deletions typically showed features of non-homologous end joining (NHEJ), such as microhomologies and insertions of direct or inverted repeat sequences. We also detected more complex rearrangements, including large insertions from adjacent or distant genomic regions. The insertion events that involved distant genomic regions typically represented transcribed sequences, and included both L1 LINE elements and sequences known to be involved in genomic rearrangements. This type of aberrant repair could potentially lead to gene inactivation via deletion of coding or regulatory sequences, or production of oncogenic fusion genes via insertion of coding sequences.

Keywords

chromosomal rearrangement; double strand DNA break; non-homologous end joining; insertion; I-*Sce*I

1. Introduction

DNA double strand breaks (DSBs) appear in the genome of human cells on a regular basis [1]. These lesions can lead to cell death if left unrepaired, therefore it is crucial that cells repair broken chromosomes. Double strand break repair can occur via either homology-dependent or non-homologous mechanisms [2,3]. Two forms of homology-directed repair have been identified. An intact copy of the sequence can be copied into the DSB from a sister chromatid or homologous chromosome [4]; this repair process is conservative. There also exists a non-conservative type of homologous recombination repair known as single-strand annealing (SSA), in which DSBs are processed to single-stranded tails, and repair occurs via annealing of the single-stranded tail at a nearby direct repeat sequence [5].

The second major form of DSB repair is non-homologous end joining (NHEJ), which does not require the presence of a homologous donor sequence. Short segments (1–4 nucleotides) of overlapping nucleotides (microhomologies) are often present at the NHEJ repair site [6],

Address correspondence to: Dr. Peter Aplan, NCI/NIH/Genetics Branch, National Naval Medical Center, Building 8 Room 5101, 8901 Rockville Pike, Bethesda, MD 20889-5105, Tel: 301-435-5005, FAX: 301-496-0047, Email: aplanp@mail.nih.gov.

suggesting that a short DNA overlap may be important for the religation of broken DNA ends. Since NHEJ is often accompanied by deletion or addition of nucleotides, NHEJ potentially compromises genetic information [7,8]. Despite the propensity for introducing small genetic changes upon repair, NHEJ is thought to act as a caretaker of genome, since cells deficient in NHEJ often display genetic instability [9,10]. To explain the presence of DNA end joining seen in systems lacking components of the NHEJ pathway, an alternative, highly error-prone NHEJ pathway has been proposed [11].

It has been suggested that improper DSB repair can lead to gross chromosomal rearrangements (GCRs), such as translocations, deletions and inversions [12]. Indeed, many of the chromosomal translocation breakpoints seen in patients with hematopoietic malignancies display hallmarks of NHEJ, suggesting that erroneous DNA repair may play a role in malignant transformation [10,13,14]. Several model systems have been employed to investigate the link between DSB repair and GCRs. Following insertion of a complementary pair of mutant neomycin phosphotransferase genes (neo^r) into mouse ES cells [15,16], production of a specific DSB within the mutant neo^r genes led to reconstruction of a functional neo^r gene via homologous recombination. Although this approach allows one to detect homologous recombination, it does not identify repair processes such as NHEJ, which do not produce a functional neo^r gene.

Loss-of-function reporter systems have also been used to gain information about mutations accompanying DSB repair in rodent cells. In these studies, a DSB was introduced either into the coding region of a Herpes simplex virus thymidine kinase (Tk) gene [17] or an intron of the phosphoribosyl transferase (*APRT*) gene [3,18]. Nonconservative repair processes abrogated the activity of the reporter genes and enabled cells to survive in selection media. In another experimental system [19], a DSB was introduced in an intron of the endogenous Tk gene of a human lymphoblastoid cell line, and aberrant DNA repair events were recovered by selecting for cells that had lost Tk activity. However, this assay favors the recovery of large scale changes, such as long deletions, and does not provide a means of assessing the frequency of short deletions.

We report here a loss-of-function reporter assay in which the outcome of the non-homologous repair of one single DSB can be studied in human cells. This approach provides an opportunity to analyze both the short deletions and the large scale changes that accompany the repair of a single, induced DSB.

2. Results

2.1 Design of the model system

We modified an existing experimental system (13, 14) to introduce a single DSB and identify cells that have sustained mutations accompanying the repair process. We generated a plasmid (pEF1αTk) that expresses the Herpes simplex virus thymidine kinase (*HSV-tk*) gene under the control of the *EF1α* promoter (fig. 1A), with the recognition sequence for the restriction endonuclease I-*Sce*I interposed between the *EF1α* promoter and the *HSV-tk* gene. The I-*Sce*I endonuclease, derived from the yeast *Sacharomyces cerevisiae*, has an 18 bp recognition sequence which is not present in the human genome [20]. The pEF1αTk vector can be introduced into mammalian cells, and those cells that have stably integrated the construct can be selected with G418. These cells will express *HSV-tk* and will therefore be sensitive to gancyclovir (GCV).

We electroporated the human monocytic cell line U937 [21] with the pEF1αTk plasmid and isolated clone F5, which contains a single copy of the construct (fig. 1). Inverse PCR demonstrated that the construct had been integrated into chromosome 7 between nucleotides

112329914-112329932 (all chromosomal coordinates refer to the July 2003 freeze, based on Human Genome Build 34 from NCBI).

The F5 cells were then transfected with I-*Sce*I expression vectors. For some experiments, we used an episomal vector (pCEP4-I*sce*I). As opposed to conventional plasmid vectors, in which only a fraction of cells that are successfully transfected will integrate the plasmid vector and supply ongoing antibiotic (eg., hygromycin) resistance, episomal vectors can supply ongoing antibiotic resistance in the absence of integration. This allows selection of relatively large numbers of unique, successfully transfected clones, selected solely on the basis of expression of the antibiotic resistance marker. In other experiments, we used previously described non-episomal I-SceI vectors (pPGK3XnlsI-*Sce*I and pCBASce). Since I-*Sce*I cleaves DNA between the *EF1α* promoter and the *HSV-tk* gene; this cleavage might result in chromosomal aberrations due to an imperfect repair that separates the *EF1α* promoter from *HSV-tk* coding sequences. A combination of negative (GCV) and positive selection (G418) enabled us to recover cells that sustained different forms of genomic rearrangements (fig. 1B).

2.2 Expression of I-SceI in F5 cells

F5 cells were transfected with the episomal expression vector pCEP4-I*sce*I, which contains a hygromycin resistance cassette (hygro^r) and expresses the I-*Sce*I enzyme under the control of a CMV promoter. Following transfection, the cells were grown in bulk, without any selection for 5 days, and were subsequently plated into 96 well plates at dilutions of 3, 9, or 27 cells per well. These plates were then selected with hygromycin alone (to identify cells that had been successfully transfected) or hygromycin and GCV. Although approximately 2.5% of the transfected cells were hygromycin resistant, only 0.4 % were both hygromycin and GCV resistant, indicating that approximately 84 % of the cells ($\frac{2.5 - 0.4}{2.5}$) that were successfully transfected with pCEP4-I*sce*I continued to express a functional *HSV-tk* gene. This finding suggested that minimal or no genetic changes had occurred in most of the cells that were transfected with the I-*Sce*I expression vector.

Using a series of PCR assays (fig. 1A), we found that half of the 22 clones selected on the basis of hygromycin resistance alone (termed 22-1, 2, 3 etc.) showed altered DNA sequences flanking the I-*Sce*I cleavage site, indicating that the DNA had been cleaved and imperfectly repaired. Two clones had deletions of *EF1α* or *HSV-tk*, as evidenced by the inability to amplify any of the segments indicated in fig. 1A. Sequence analysis of PCR amplified fragments demonstrated that 4 clones displayed a short deletion (9, 7, 3 and 1 nucleotides), 1 clone displayed a larger deletion (55 bp), and 4 clones had more complex rearrangements due to DNA insertions at the I-*Sce*I site. In these 4 clones the nucleotide sequence of the inserted DNA segments matched sequences from either the transfected episomal vector (1), or from distant regions of the genome (3) (table. 1).

As indicated above, half of the 22 clones that had been successfully transfected with the episomal I-*Sce*I expression vector contained “germline” PCR fragments and did not show clear evidence for sequence changes at the I-*Sce*I site upon sequence analysis. However, the chromatogram of several clones with “germline” PCR fragments exhibited an increase in background signal beginning precisely at the I-*Sce*I site (not shown), suggesting that these chromatograms represented mixed clones, some of which had mutations at the I-*Sce*I site. These results suggest that I-*Sce*I mediated DSB occurred in at least half of the cells that were successfully transfected.

2.3 Improved recovery of cells with more extensive rearrangements

As shown above, 50% of cells transfected with the I-*Sce*I expression vector exhibited mutations around the cleavage site. Given that malignant cells often display GCRs, we wanted to

determine if we could increase the recovery of clones containing GCRs by modifying our detection scheme. We utilized the negative selection provided by expression of *HSV-tk*, and transfected cells with pCBASce, a non-episomal I-*Sce*I expression vector.

Following transfection with the pCBASce vector, approximately 0.2 % of F5 cells were GCV resistant, and had retained G418 resistance, indicating that GCV resistance was not acquired through loss of the entire integrated chromosome. We anticipated that clones isolated in this fashion might have more extensive mutations, as clones with short (i.e., <20–30 bp) deletions would continue to express *Hsv-tk*, and therefore be eliminated by GCV selection. After 3 weeks of selection in GCV and G418, the transfected cells were single-cell cloned to obtain pure colonies; 58 GCV and G418 resistant colonies (termed as 6-1,2,3 etc.) were assayed for evidence of GCRs.

2.4 Screening strategy for analyzing GCRs generated by one DSB

We used Southern blot hybridization to distinguish clones which sustained an interstitial deletion from those that sustained more complex rearrangements, such as large insertions or translocations. To distinguish between these two events, we hybridized *Hin*dIII-digested genomic DNA from $GCV^r/G418^r$ clones with a chromosome 7 specific probe that hybridized upstream of the *EF1*α integration site (termed probe TV7), and a neo^r gene specific probe (fig. 1A). Since *Hin*dIII does not cleave within the pEF1αTk construct, both probes should identify identical genomic DNA fragments if the clone had sustained an interstitial deletion. However, if a translocation or other GCR separated the 3′ and 5′ portion of the integrated construct, the two probes should hybridize to distinct genomic DNA fragments. Fig. 2 shows that most of the clones sustained an interstitial deletion as evidenced by fragments of identical sizes in the duplicate hybridizations. In some cases, one hybridization signal was lost, suggesting that a large deletion, presumably triggered by I-*Sce*I cleavage between the hybridization target sequences, extended beyond the target sequence of the probe.

To investigate the mechanism of aberrant repair following I-*Sce*I cleavage, we analyzed the nucleotide sequence flanking the I-*Sce*I cleavage site in GCV resistant clones. We used a series of conventional and inverse PCR reactions to assess different regions of the integrated pEF1αTk vector (fig. 1A). Of the 58 clones isolated, four were shown to be a mixture of two independent clones, four clones were represented twice, and one clone was represented three times. Six clones showed no mutation of the region surrounding the I-*Sce*I site. Of the 50 independent clones which exhibited mutations around the DSB site, 48 were shown by Southern blot and/or PCR, to have sustained an interstitial deletion of DNA flanking the I-*Sce*I site. Two clones lost the entire *EF1*α promoter and at least 1 kb of adjacent chromosome seven region, but retained the neo^r gene downstream of the I-*Sce*I site. To verify that the deletions were generated as a result of erroneous DSB repair, as well as identify any additional mutations (such as insertions, inversions, duplications) accompanying the deletions, we determined the breakpoint sequences of 38 independent mutant clones. In one case, due to a large insertion of vector sequences at the I-*Sce*I site, we did not determine both ends of the deleted segment. Altogether, the nucleotide sequence changes from 37 mutant clones were completely characterized.

As shown in fig. 3., I-*Sce*I mediated cleavage of the integrated pEF1αTk construct occurs at position 1352, between the 3′ end of the *EF1*α promoter (1304) and the first ATG codon of the *HSV-tk* gene (1359). Fine analysis of the deletions demonstrated that the pEF1αTk construct sustained asymmetric deletions following I-*Sce*I cleavage. Upstream of the I-*Sce*I cleavage site, towards the *EF1*α promoter, the mean size of the deleted segment was 442 bp. However, these deletions are unevenly distributed. A majority of the clones (19) sustained small deletions, shorter than 10 bp, whereas nine clones had deletions of 11–200 bp and nine clones had much longer deletions that averaged 1752 bp. The length of the deleted segments downstream of the

I-*Sce*I induced DSB, extending into the *HSV-tk* gene, showed a more uniform distribution. In contrast to the deletions involving *EF1α*, no deletion was shorter than 100 bp or longer than 1700 bp. Sixteen clones showed short regions of microhomology (1–6 bp) at the breakpoint junction (fig. 3), and a short palindrome sequence (4–8 bp) was found at the breakpoint of 14 clones, nine of which also showed microhomologies.

In addition to the deletions described above, some clones contained additional rearrangements in the region cleaved by I-*Sce*I. The predominant type of additional event was a nucleotide insertion; 18 clones contained an insertion in the gap created by the deletion. These clones could be divided into several groups according to the characteristics of the inserted DNA segment.

Six clones showed insertions of only a few nucleotides, and it was not possible to determine whether these insertions were templated or not. A group of nine clones had short insertions of 11–24 bp. In these clones, the core of the inserted nucleotides was copied from either side of the breakpoint junction, creating either a direct (8 clones) or inverted (1 clone) repeat with the flanking sequences (fig. 3). Two clones contained complex rearrangements with insertions derived from the pEF1αTK vector. A single clone contained I-*Sce*I expression vector sequences inserted at the breakpoint junction.

2.5 Modification of the experimental system

Following characterization of the clones described above, we modified the experimental system in an effort to identify GCV resistant clones that exhibited GCRs. We used different expression vectors (pCEP4-I*sce*I vs. pCBASce), varied the transfection conditions (shorter period of bulk culture) and, in some experiments omitted G418 from the selection medium. We recovered 76 additional clones (termed 5-1, 2 3, etc. and 29-1, 2, 3, etc.) from these experiments, and detected primarily interstitial deletions similar to those described above. We again found unequal deletions around the I-*Sce*I cleavage site, the presence of short direct repeat sequences, and rare insertion events.

Five clones exhibited large insertions that matched gene segments from distant chromosomal regions; two additional clones sustained complex rearrangements involving DNA insertions from chromosome 7 regions adjacent to the pEF1αTk integration site (table 1). The Southern blot hybridization pattern of clone 5–22 could not be explained by a simple deletion (fig. 2) and was consistent with a chromosomal translocation, as unique *Hin*dIII fragments hybridized to probes Neo and TV7. PCR reactions revealed that this clone had retained the *EF1α* promoter, and sustained a short deletion of *HSV-tk* sequences. Inverse PCR reactions showed that *EF1α* promoter sequences immediately adjacent to the I-*Sce*I site had been joined to a sequence of at least 982 bp derived from chromosome 15 α satellite centromeric repeat sequences, suggesting that this clone had sustained a chromosomal translocation. Of the clones that were completely characterized, 56% had interstitial deletions, 24 % had deletions accompanied by small (<30 bp) insertions, and 20% deletions accompanied by large (>30 bp) insertions.

3. Discussion

Imperfect repair of DSBs is thought to play an important role in generating GCRs associated with malignant transformation. To examine how DSBs might lead to GCRs, we investigated the outcome of DSB repair in the human monocytic cell line U937. To obtain a detailed analysis of repair events following a single, specific DSB, we induced DSBs with the I-*Sce*I endonuclease. The simplest mechanism for repair of an I-*Sce*I induced DSB is a perfect religation event, which would recreate an I-*Sce*I recognition sequence that would remain susceptible to I-*Sce*I cleavage. However, if the cleaved recognition sequence were repaired imperfectly, no further I-*Sce*I cleavage would be possible, since the I-*Sce*I recognition site

would have been destroyed. We found that half (11/22) of the clones successfully transfected with an episomal I-*Sce*I expression vector showed clear evidence of I-*Sce*I cleavage, suggesting that I-*Sce*I mediated cleavage of genomic targets was reasonably efficient.

To enrich for clones which may have sustained more extensive mutations (such as inversions, insertions, translocations, or large deletions) following repair of an I-*Sce*I-induced DSB, we made use of the negative selection provided by the *HSV-tk* gene. Similar to prior reports that used rodent cells[3,18], we found that the most common mutation in the human U937 sub-clones that survived the GCV selection was an interstitial deletion, often accompanied by additional events (insertions, inversions, duplications). Many of the clones showed hallmarks of NHEJ, such as insertion of direct or inverted repeats or microhomologies at the breakpoints.

We recovered a total of fourteen clones that contained large segments of foreign DNA inserted at the I-*Sce*I site, and a single clone (clone 5–22) with a rearrangement consistent with a chromosomal translocation. Four clones contained sequences from the I-*Sce*I expression vectors, similar to previous reports [22], and two clones contained chromosome seven sequences derived from regions near the pEF1αTK integration site. Of interest, eight clones contained DNA insertions that matched sequences from distant regions of the genome (table 1); seven of the eight insertion events involved known transcribed regions, suggesting two possibilities as to the source of these sequences. The inserted segments could have been derived from double-stranded DNA fragments, or DNA/RNA hetero-duplexes, as suggested for the vector capture events. In this scenario, RNA molecules may have served as a source of DNA “patches” via reverse transcription, as seen in endonuclease-independent LINE L1 retrotranspositions [23–25]. A second possibility is that distant genomic regions served as a template for these sequences. These segments could have been copied into the gap initiated by I-*Sce*I cleavage or, alternatively, could have been excised from the genome and inserted at the breakpoint. Two of these eight insertions were likely the result of a three-way ligation, since the inserted DNA segments were derived from two distinct regions of the genome. In both instances, there were no sequence motifs shared between the two inserted sequences. Additionally, it is interesting to note that at least one inserted DNA segment (*D4Z4*) is associated with a known *in-vivo* genomic instability phenomenon. This sequence is found in tandem repeat on 4q35.2, a region that is thought to have been derived from an interchromosomal exchange with the nearly identical subtelomeric 10q26 [26], and often participates in intrachromosomal recombinations that lead to repeat contractions [27].

Although the mutations in this study took place within an artificial DNA construct, this type of faulty repair could easily lead to inactivation of tumor suppressor genes or the production of oncogenic fusion genes. For instance, a short interstitial deletion of the *BRCA1* gene, which shows nucleotide sequence changes similar to those described here, leads to inactivation of the gene [28], and insertions of *AF5Q31* sequences within the *MLL* locus can lead to oncogenic *MLL-AF5Q31* fusions [29]. The experimental system described here provides a platform useful for elaborating the mechanisms by which repair of DSBs by NHEJ may lead to fine and gross chromosomal aberrations.

4. Materials and Methods

4.1 Vector construction

pEF1α Tk was generated by digesting pEF1αNUPHOX (a pcDNA3 based vector containing a human *EF1α* promoter segment) with BamHI and NotI to remove the NUPHOX insert and to provide a vector with an EF1α promoter. The *HSV-tk* gene was amplified from the pTkNeo vector using primers Tkreverse (AATGCGGCCGCAGTTAGCCTCCCCCATCTC) and TkFwBamHI (AAGGGATCCTAGGGATAACAGGGTAATGGCTTCGTACCCCTGCCAT), with the

latter primer containing an I-*Sce*I recognition site immediately 5′ of the *HSV-tk* gene. The PCR product was digested with NotI and BamHI, purified, and ligated into the *EF1α* vector backbone. The pCEP4-I*sce*I plasmid was generated by cloning the 3Xnls-I-SceI fragment from the pPGK3XnlsI-*Sce*I vector [30], into the HindIII (blunted) and BamHI sites of the episomal expression vector pCEP4 (Invitrogen, Carlsbad, CA).

4.2 Transfections

The U937 human monocytic leukemia cell line (maintained in RPMI 1640, supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin and 100 μgml Streptomycin) was electroporated with ScaI linearized pEF1αTk using the Gene Pulser II electroporation system (Bio-Rad Laboratories, Inc., Hercules, CA), with pulses of 0.4 kV and 975 μF, and stable $G418^R$ clones were isolated.

Transfections with vectors pCEP4-I*sce*I, pPGK3XnlsI-*Sce*I or pCBASce [4]were performed using DMRIE-C reagent (Invitrogen, Carlsbad, CA). . On the fourth or fifth day following transfection, cells were expanded in bulk, or were plated into 96 or 24 well plates at 3 cells/well, 9 cells/well and 27 cells/well, or at 250 cells/well, 750 cells/well, 2250 cells/well and 6750 cells/well density, respectively. Selection agents GCV (from InvivoGen, San Diego, CA, CAS n°: 82410-32-0), G418 (from Invitrogen, Carlsbad, CA, CAS n°: 1405-41-0) and Hygromycin B (from Invitrogen, Carlsbad, CA, CAS n°: 31282-04-9) were added to the wells. The following selection conditions were applied: 40 μ M GCV, or 40 μM GCV and 400 μg/ml G418 in transfections done with pCBASce; 200 μg/ml Hygromycin, 200 μg/ml Hygromycin and 20 μM GCV, or 40 μM GCV in transfections done with pCEP4-I*sce*I.

4.3 Conventional and inverse PCR

The PCR reactions indicated in fig. 1A were carried out using 100 ng of genomic DNA as template in the presence of 0.2 μM primers with either Takara LA Taq (PCR #2, 3) (Takara Mirus Bio, Madison, WI) or with PCR Supermix (reactions 1, 4 and 5) (Invitrogen, Carlsbad, CA), according to the manufaturers' recommendations. For inverse PCR reactions, 1 μg of genomic DNA was digested with either *Cfo*I, *Hin*dIII or *Alu*I in 100 μl. The samples were extracted with phenol-chloroform, ethanol precipitated, and self-ligated in 250 μl using T4 Ligase (Promega, Madison, WI). Fragments containing *EF1α* sequences were amplified in Supermix (Invitrogen) using primers designed for a nested PCR. Details of primer sequences and amplification protocols are available upon request.

4.4 Southern Blot analysis

Duplicate samples containing 10 μg of genomic DNA from GCV resistant clones were digested with *Hin*dIII, size-fractioned on 0.8% agarose gels, and transferred to a nitrocellulose membrane and hybridized to a chromosome 7 specific fragment (probe TV7; nuc 112330867-112331229) or a NcoI-SmaI $G418^r$ fragment isolated from pPRC/CMV (Invitrogen, Carlsbad, CA). The fragments were labeled with ^{32}P using Ready-To-Go DNA labelling beads (Amersham Biosciences) and hybridized as previously described [31]. Final washing conditions were 0.1% SDS/0.1x SSC at 52°C for both probes.

4.5 Sequence Analysis

Nucleotide sequences were determined using an Applied Biosystems 3730 and compared to the human genome assembly (University of Santa Cruz July 2003 freeze, based on NCBI Human Genome Build 34) [32].

Acknowledgements

The authors thank Maria Jasin for providing the I-*Sce*I expression vector, pCBASce and Greg Donoho for providing the I-*Sce*I expression vector, pPGK3XnlsI-*Sce*I. We also thank Ilan Kirsch and Michael Kuehl for helpful and stimulating discussions.

References

1. Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci U S A 2003;100:12871–12876. [PubMed: 14566050]
2. Liang F, Han M, Romanienko PJ, Jasin M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A 1998;95:5172–5177. [PubMed: 9560248]
3. Sargent RG, Brenneman MA, Wilson JH. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. Mol Cell Biol 1997;17:267–277. [PubMed: 8972207]
4. Richardson C, Moynahan ME, Jasin M. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. Genes Dev 1998;12:3831–3842. [PubMed: 9869637]
5. Lin FL, Sperle K, Sternberg N. Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. Mol Cell Biol 1990;10:103–112. [PubMed: 2294396]
6. Roth DB, Wilson JH. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol Cell Biol 1986;6:4295–4304. [PubMed: 3025650]
7. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 2003;4:712–720. [PubMed: 14506474]
8. Gu H, Forster I, Rajewsky K. Sequence homologies, N sequence insertion and JH gene utilization in VHDJH joining: implications for the joining mechanism and the ontogenetic timing of Ly1 B cell and B-CLL progenitor generation. Embo J 1990;9:2133–2140. [PubMed: 2113468]
9. Smith J, Riballo E, Kysela B, Baldeyron C, Manolis K, Masson C, Lieber MR, Papadopoulo D, Jeggo P. Impact of DNA ligase IV on the fidelity of end joining in human cells. Nucleic Acids Res 2003;31:2157–2167. [PubMed: 12682366]
10. Ferguson DO, Alt FW. DNA double strand break repair and chromosomal translocation: lessons from animal models. Oncogene 2001;20:5572–5579. [PubMed: 11607810]
11. Lee GS, Neiditch MB, Salus SS, Roth DB. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. Cell 2004;117:171–184. [PubMed: 15084256]
12. Morgan WF, Corcoran J, Hartmann A, Kaplan MI, Limoli CL, Ponnaiya B. DNA double-strand breaks, chromosomal rearrangements, and genomic instability. Mutat Res 1998;404:125–128. [PubMed: 9729329]
13. Zucman-Rossi J, Legoix P, Victor JM, Lopez B, Thomas G. Chromosome translocation based on illegitimate recombination in human tumors. Proc Natl Acad Sci U S A 1998;95:11786–11791. [PubMed: 9751743]
14. Elliott B, Jasin M. Double-strand breaks and translocations in cancer. Cell Mol Life Sci 2002;59:373–385. [PubMed: 11915950]
15. Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. Nature 2000;405:697–700. [PubMed: 10864328]
16. Richardson C, Jasin M. Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. Mol Cell Biol 2000;20:9068–9075. [PubMed: 11074004]
17. Lukacsovich T, Yang D, Waldman AS. Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. Nucleic Acids Res 1994;22:5649–5657. [PubMed: 7838718]
18. Phillips JW, Morgan WF. Illegitimate recombination induced by DNA double-strand breaks in a mammalian chromosome. Mol Cell Biol 1994;14:5794–5803. [PubMed: 8065314]

19. Honma M, Izumi M, Sakuraba M, Tadokoro S, Sakamoto H, Wang W, Yatagai F, Hayashi M. Deletion, rearrangement, and gene conversion; genetic consequences of chromosomal double-strand breaks in human cells. Environ Mol Mutagen 2003;42:288–298. [PubMed: 14673874]

20. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet 1996;12:224–228. [PubMed: 8928227]

21. Sundstrom C, Nilsson K. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). Int J Cancer 1976;17:565–577. [PubMed: 178611]

22. Lin Y, Waldman AS. Capture of DNA sequences at double-strand breaks in mammalian chromosomes. Genetics 2001;158:1665–1674. [PubMed: 11514454]

23. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 2002;110:327–338. [PubMed: 12176320]

24. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 2002;31:159–165. [PubMed: 12006980]

25. Eickbush TH. Repair by retrotransposition. Nat Genet 2002;31:126–127. [PubMed: 12006979]

26. van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, van der Maarel SM, de Jong PJ, Hewitt JE. Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics 2002;79:210–217. [PubMed: 11829491]

27. Lemmers RJ, Van Overveld PG, Sandkuijl LA, Vrieling H, Padberg GW, Frants RR, van der Maarel SM. Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. Am J Hum Genet 2004;75:44–53. [PubMed: 15154112]

28. Hardouin A, Baumann J, Roussel G, Quillien V, Dugast C, Berthet P. A new mutation in the BRCA1 gene (g.5196–5201del6, 5195–5202ins12), a 6 bp deletion replaced by the duplication of a 12 bp adjacent upstream intronic sequence. Hum Mutat 2001;17:154. [PubMed: 11180603]

29. Deveney R, Chervinsky DS, Jani-Sait SN, Grossi M, Aplan PD. Insertion of MLL sequences into chromosome band 5q31 results in an MLL-AF5Q31 fusion and is a rare but recurrent abnormality associated with infant leukemia. Genes Chromosomes Cancer 2003;37:326–331. [PubMed: 12759932]

30. Donoho G, Jasin M, Berg P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol 1998;18:4070–4078. [PubMed: 9632791]

31. Aplan PD, Chervinsky DS, Stanulla M, Burhans WC. Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors. Blood 1996;87:2649–2658. [PubMed: 8639880]

32. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res 2002;12:656–664. [PubMed: 11932250]

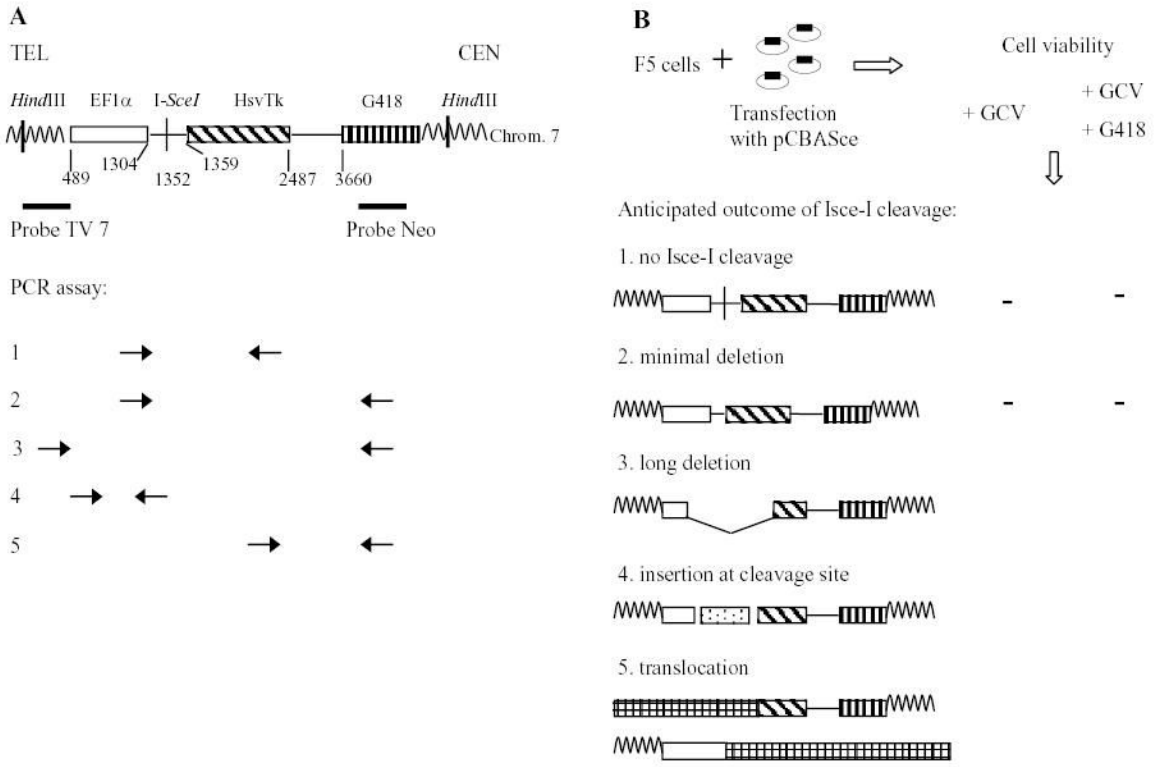

Fig. 1. Outline of reporter system

(A) Integration of the pEF1αTk construct into chromosome seven. Telomere and centromere are indicated; the numbers refer to pEF1αTk vector sequences. The *EF1α* promoter (open box), I-*Sce*I recognition sequence, *HSV-tk* cDNA (cross-hatched box), and G418^R cassette (vertically striped box) are shown. Locations of primers (PCR reactions #1–5) and probes used to analyze mutations around the I-*Sce*I site are indicated.. **(B)** Expression of I-*Sce*I cleaves genomic DNA at the I-*Sce*I site. Cells sustaining different forms of mutation (indicated in 1–5) upon DSB repair can be recovered using a combination of positive and negative selection.

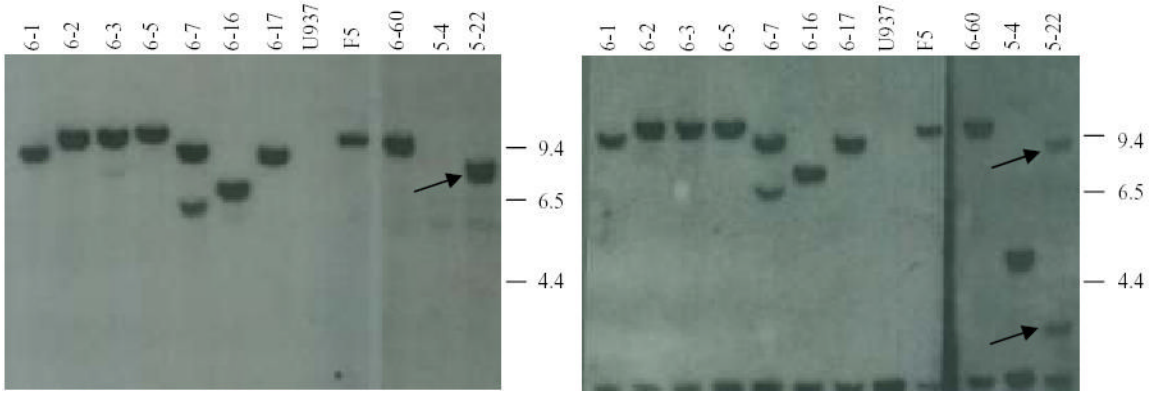

Fig. 2. Analysis of GCV resistant clones

*Hind*III digested genomic DNA from individual clones hybridized to a neo gene specific (left panel) or a chromosome 7 (TV 7) specific probe (right panel). In most cases, the two hybridization signals are of the same size, indicating that repair of the DSB was accompanied by an interstitial deletion. In clone 5-4, the deletion extended beyond the neo^R gene (verified by inverse PCR sequence analysis). In clone 5-22, the chromosome seven specific probe shows two signals of even intensity, neither of whose size corresponds to that of the neo specific signal (arrows). The endogenous chromosome seven signal is seen at 2.3 kb in all lanes. Size standards are in kb.

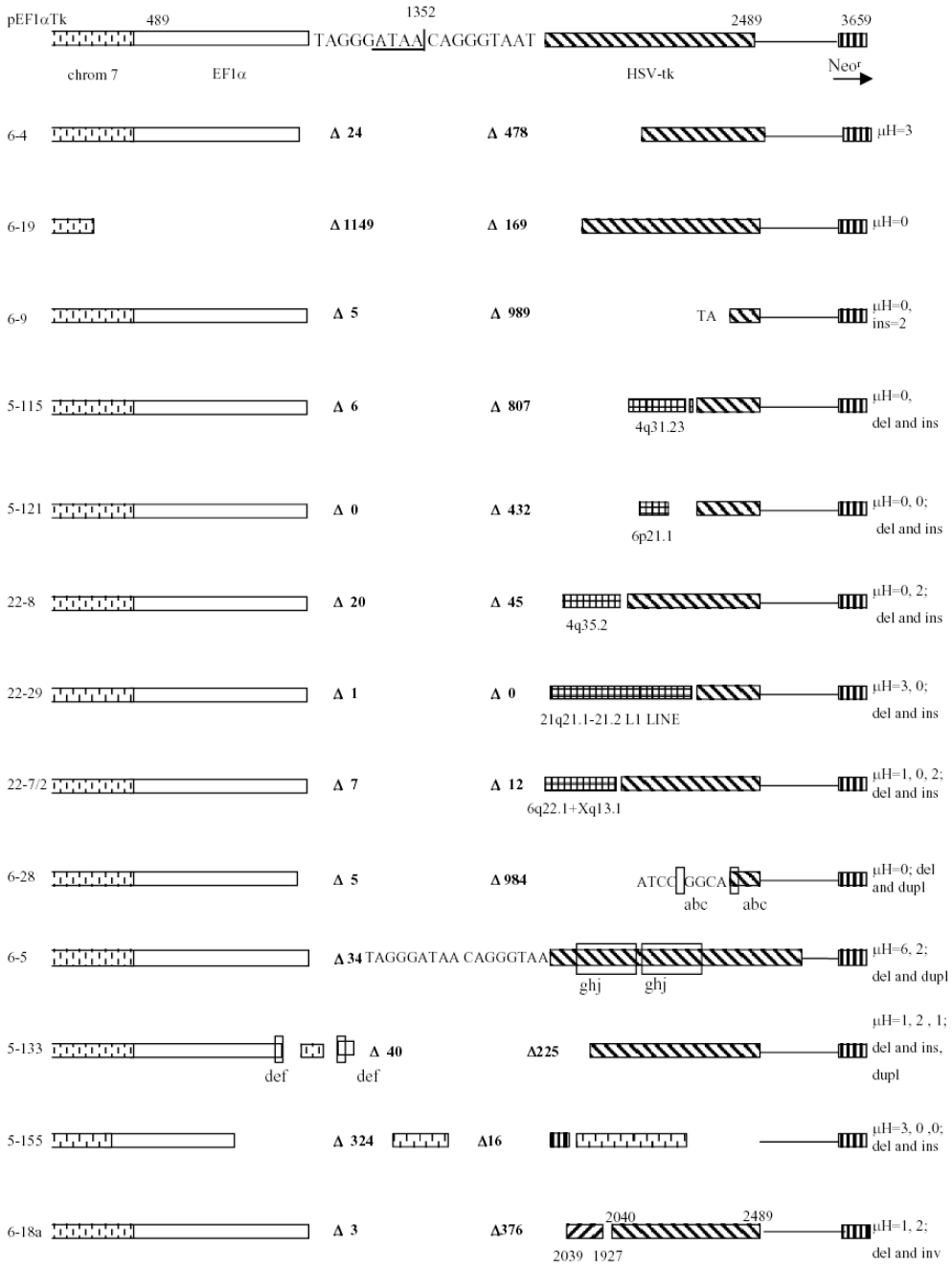

Fig. 3. Diagram of genetic changes around the DSB site

Representative examples of different types of rearrangements are shown. Numbers on vector DNA represent nucleotide position; sequences downstream of neo^r gene are not shown. The staggered I-*Sce*I cleavage site is underlined. Clone 5-115, 5-121, 22-8, 22-29 and 227/2 contain an insertion that matches distant regions of the human genome; clone 5-115 also contain an insertion of 14 bp segment of low complexity. Clone 5-133, 6-5 and 6-28 sustained more complex rearrangements involving a primary deletion and duplication leading to the formation of direct repeats. Clone 5-155 contains two insertions from adjacent chromosome 7 sequences; clone 6-18a contains an inversion of a segment of the *HSV-tk* gene. Δ refers to nucleotide deletion from *EF1α* or *HSV-tk*; μH indicates microhomology at each junction; del, ins, inv and

dupl refer to deletions, insertions, inversions or duplications, respectively. Open rectangles represent duplications, and letters (abc, def, and ghi) indicate the orientation of repeats. Figure is not drawn to scale.

Table 1

Features of non vector derived DNA insertion events.

clone	match for inserted DNA	description of inserted DNA	features of breakpoint junction
22-7/2	6q22.1 and Xq13.1	mRNA (gb *BU660518.1*) (nuc. 114,938,877-114,939,020) and histone deacetylase 8 (*HDAC8*) gene (nuc. 69,839,437-69,839,474	del. of 7 and 12 bp μH of 1, 1 and 0 bp
22-8	4q35.2	*D4Z4* repeat, ORF inside (nuc. 191,697,731-191,697,966)	del. of 20 and 45 bp μH of 0 and 2 bp
22-29	21q21.1- q21.2	*L1 LINE* and flanking unique sequence (nuc. 22,512,921-22,513,550)	staggered cut of I-SceI site is filled in, del. of 1 and 0 bp, μH of 3 and 0 bp
5-115	4q31.23	intronic sequence within the human mineralocorticoid receptor gene (nuc. 149631931-149632077)	extra 14 bp CT-rich sequence added, forms direct repeat with breakpoint junction, del. of 6 and 807 bp, μH of 0 bp
5-121	6p21.1	3′ UTR of translocating chain-associating membrane protein 2 (*TRAM2)* gene (nuc. 52364732-52364778)	del. of 0 and 432 bp μH of 0 bp
29-6	10q26.13	intronic region of arginyl-tRNA-protein transferase (*ATE1*) gene (nuc. 123208457 123208689)	staggered cut of I-SceI site is filled in, μH of 4 bp
29-45	12p11.21 and 12p13	intronic region of *DDX12*, a duplicated putative helicase gene (nuc. 31140689-31140834, 9469691- 9469836 and 9349729-9349874)	del. of 17 and 0 bp μH of 1 and 1 bp
29-1A	4q35.2 and 10p13	two distinct segments of non coding DNA (nuc. 188535445 188535618 and 16305232-16305415)	del. of 18 and 1 bp, μH of 1, 1 and 0 bp