Genetics: Early Online, published on November 14, 2016 as 10.1534/genetics.116.188680

Title: The Nucleotide Excision Repair pathway limits L1 retrotransposition.

Geraldine Servant[*], Vincent A. Streva[*, 1], Rebecca S. Derbes[*], Madushani I. Wijetunge[*], Marc Neeland[*], Travis B. White[*, 2], Victoria P. Belancio[†], and Astrid M. Roy-Engel[*^] and Prescott L. Deininger[*^].

* Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane Cancer Center, 1430 Tulane Ave., New Orleans, LA 70112.

† Tulane University, Department of Structural and Cellular Biology, School of Medicine, Tulane Cancer Center and Tulane Center for Aging, 1430 Tulane Ave., New Orleans, LA 70112.

^ Both authors share *equal senior authorship*.

1 Current Address: Division of Infectious Diseases, Boston Children's Hospital and Harvard Medical School, 300 Longwood Ave., Boston, MA 02115

2 Current Address: Developmental Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065

Running title: (35 characters including spaces)

NER limits L1 retrotransposition

Keywords:

L1 retrotransposon, Nucleotide Excision Repair, Target-Primed Reverse Transcription, DNA damage, genome stability

Corresponding author:

Prescott Deininger, Ph.D.

Tulane Cancer Center, SL66

Tulane University Health Sciences Center

1430 Tulane Ave.

New Orleans, LA 70112

pdeinin@tulane.edu

Phone: (504) 988-6385

Fax: (504) 988-5516

Abstract (230 words out of the 250-word limit)

Long interspersed elements 1 (L1) are active mobile elements that constitute almost 17% of the human genome. They amplify through a “copy-and-paste” mechanism termed retrotransposition, and *de novo* insertions related to these elements have been reported to cause 0.2% of genetic diseases. Our previous data demonstrated that the endonuclease complex ERCC1-XPF, which cleaves a 3’ DNA flap structure, limits L1 retrotransposition. Although the ERCC1-XPF endonuclease participates in several different DNA repair pathways, such as single strand annealing, or in telomere maintenance, its recruitment to DNA lesions is best characterized in the nucleotide excision repair (NER) pathway. To determine if the NER pathway prevents the insertion of retroelements in the genome, we monitored the retrotransposition efficiencies of engineered L1 elements in NER-deficient cells and in their complemented versions. Core proteins of the NER pathway, XPD and XPA, and the lesion binding protein, XPC, are involved in limiting L1 retrotransposition. In addition, sequence analysis of recovered *de novo* L1 inserts and their genomic locations in NER-deficient cells demonstrated the presence of abnormally large duplications at the site of insertion, suggesting that NER proteins may also play a role in the normal L1 insertion process. Here we propose new functions for the NER pathway in the maintenance of genome integrity: limitation of insertional mutations caused by retrotransposons and the prevention of potentially mutagenic large genomic duplications at the site of retrotransposon insertions events.

INTRODUCTION

Retrotransposons, including Long INterspersed Element 1 (L1), constitute about a third of the human genome (LANDER *et al.* 2001; DE KONING *et al.* 2011). D*e novo* insertions of retrotransposons have been reported as the cause of over 90 genetic diseases, indicating that these elements continue to amplify in the human genome (OSTERTAG and KAZAZIAN 2001; XING *et al.* 2007; BELANCIO *et al.* 2008a; HANCKS and KAZAZIAN 2012; HANCKS and KAZAZIAN 2016). Retrotransposons amplify throughout the genome using a "copy-and-paste" mechanism, termed retrotransposition, based on the reverse transcription of an RNA intermediate (BOEKE *et al.* 1985). The L1-encoded proteins, ORF1p and ORF2p, are responsible for the amplification of L1 elements in the genome (MATHIAS *et al.* 1991; FENG *et al.* 1996; MORAN *et al.* 1996). Reverse transcription of L1, a non-LTR retrotransposon, occurs in the nucleus, through a proposed process called target-primed reverse transcription (TPRT) ((LUAN *et al.* 1993; FENG *et al.* 1996; LUAN and EICKBUSH 1996) diagrammed in Figure 1A). In the TPRT model, the L1 ORF2p-encoded endonuclease cleaves between the T and the A of a consensus sequence in the DNA (5'-TTTT/AA-3') freeing a 3' T-rich DNA end that primes the reverse transcription from the polyA tail of the L1 mRNA. A 3' flap DNA structure is thought to be generated resulting from the elongation of L1 cDNA at the insertion site (Figure 1A and (FENG *et al.* 1996; LUAN and EICKBUSH 1996; BOEKE 1997; CHRISTENSEN *et al.* 2006)). The factors involved in the completion of the insertion process are unknown but a second nick must be made in order for second-strand DNA synthesis to occur.

We have previously reported that the enzymatic complex ERCC1-XPF, a 3' flap endonuclease, utilized in various DNA repair pathways, limits L1 retrotransposition (GASIOR *et al.* 2008). ERCC1-XPF is a structure-specific endonuclease that nicks double-stranded DNA 5' of a DNA lesion (DE LAAT *et al.* 1998). Therefore, we proposed that ERCC1-XPF cleaves the

predicted flap structure formed by the elongating cDNA during L1 insertion (Figure 1A and (GASIOR *et al.* 2008)). ERCC1-XPF has a global function in cellular DNA damage repair, notably in the removal of the DNA flap structures during single-strand annealing (SSA) repair of DSBs (SARGENT *et al.* 2000; AHMAD *et al.* 2008; AL-MINAWI *et al.* 2008). ERCC1-XPF is also a component of the telomeric TRF2 complex, involved in the protection of telomeres (ZHU *et al.* 2003). However, ERCC1-XPF function is best characterized in the cleavage of the damaged DNA strand in the nucleotide excision repair (NER) pathway, an important DNA repair system that removes a wide variety of lesions, including ultra-violet (UV) light-induced cyclobutane pyrimidine dimers (CPDs) and pyrimidine-(6,4)-pyrimidone photoproducts (6-4 PPs), as well as bulky chemical DNA adducts and intrastrand crosslinks (HOEIJMAKERS 2001; REARDON and SANCAR 2005; SUGASAWA *et al.* 2009). Therefore different pathways could be involved in the recruitment of the ERCC1-XPF complex to limit L1 retrotransposition. Although a DNA flap structure has not been identified as a lesion recognized by the NER pathway, the NER lesion binding protein XPCp can recognize a flap structure *in vitro* (SUGASAWA *et al.* 2001; SUGASAWA *et al.* 2002). In the present study, we thus investigated the role of the NER pathway in the regulation of L1 retrotransposition.

Rather than recognizing specific base modifications, NER senses structural DNA distortion and non-hydrogen-bonded bases, caused by DNA lesions (SUGASAWA *et al.* 2009; CLEMENT *et al.* 2010). In humans, defects in NER are associated with several autosomal recessive disorders, such as Xeroderma pigmentosum (XP) and Cockayne syndrome (CS) (LEHMANN *et al.* 2011; LAUGEL 2013). Patients suffering from these diseases exhibit an extreme sensitivity to sun exposure, neurological disorders, and a predisposition to cancers. Complementation studies have revealed nine factors (XPA through XPG, CSA and CSB) involved in NER activity (for review, see (CLEAVER *et al.* 2009; NOUSPIKEL 2009)). NER consists of two subpathways that have two

distinct mechanisms of lesion recognition but share a common central repair pathway: global genome repair (GGR) and transcription-coupled repair (TCR) (Figure 1B and (VERMEULEN and FOUSTERI 2013; PUUMALAINEN *et al.* 2016)). TCR only repairs DNA lesions in the transcribed strand of active genes (MELLON *et al.* 1987). In TCR, the CSA and CSB proteins are recruited to the site of a stalled RNA Polymerase II (RNAPII) to initiate the repair (TROELSTRA *et al.* 1992; HENNING *et al.* 1995; GROISMAN *et al.* 2003; VAN DEN BOOM *et al.* 2004). GGR is the most active branch that detects base lesions in the remainder of the genome through the XPC complex, consisting of the XPC protein, as well as hRAD53 and centrin2 proteins (SUGASAWA *et al.* 1998; RIEDL *et al.* 2003). After the recognition step, both pathways converge to use a common set of proteins involved in the removal of the damaged strand (SCHÄRER 2011). Two structure-specific endonucleases, the ERCC1-XPF complex and XPG excise the damaged strand on the 5' and 3' sides of the lesion, respectively, followed by DNA replication and ligation to fill the gap (O'DONOVAN *et al.* 1994; SIJBERS *et al.* 1996).

Our data show that multiple proteins of the GGR pathway are required to limit the mobility of the non-LTR retrotransposon L1. Additionally, new L1 inserts, recovered in NER-deficient cells, show abnormally large duplications, suggesting a potential role of NER proteins in the L1 insertion process. Therefore we hypothesize that the GGR pathway can recognize the elongating L1 cDNA in a TPRT model of retrotransposition and excise it, inhibiting the damage caused by *de novo* L1 inserts.

MATERIALS AND METHODS

Cell lines and culture conditions

HeLa cells (ATCC CCL2) were grown in eMEM supplemented with 10% Fetal Bovine Serum, 0.1 mM non-essential amino acids (Life Technologies) and 1 mM sodium pyruvate (Life Technologies) at 37° in a 5% carbon dioxide environment. The following cell lines were obtained from the Coriell Cells Repository: XPC-SV40 transformed fibroblast (GM15983), XPD- SV40 transformed fibroblast (GM08207), the stably complemented version of XPD- cell line (XPD+) (GM15877) and the isogenic pair of XPA- SV40 transformed fibroblast (GM04312) and the stably complemented version (XPA+) (GM15876). XPC-, XPD- and XPA- cell lines were grown in eMEM supplemented with 10% Fetal Bovine Serum, 0.1mM non-essential amino acids (Life Technologies) at 37° in a 5% carbon dioxide environment. XPD+ cell line was grown in the same conditions with the presence of 500μg/mL G418 (Life Technologies). XPA+ cell line was grown in the DMEM supplemented with 10% Fetal Bovine Serum (Life Technologies).

Plasmids

TAM102/L1.3 contains the CMV promoter upstream of L1.3 element, deleted for L1 5'UTR, and the *mblastI* indicator cassette cloned in pCEP4 (MORRISH *et al.* 2002).

TAMD702A/L1.3 derives from TAM102/L1.3 and contains the reverse transcriptase deficient mutant of the L1.3 element and the *mblastI* indicator cassette cloned in pCEP4 (MORRISH *et al.* 2002).

TAMH230A/L1.3 derives from TAM102/L1.3 and contains the endonuclease deficient mutant of the L1.3 element and the *mblastI* indicator cassette cloned in pCEP4 (MORRISH *et al.* 2002).

The synL1_neo vector used for the recovery of *de novo* L1 inserts was previously described (GASIOR *et al.* 2007).

pBS-L1PA1$_{CH}$_blast rescue vector is pBS-L1PA1$_{CH}$ plasmid (WAGSTAFF *et al.* 2011) tagged with Blast rescue cassette for the recovery of *de novo* L1 inserts.

pCMV6-XPC (ORIGENE) consists of XPC cDNA driven by CMV promoter.

pCMV6-XPCΔ, control vector, was constructed by deleting 342bp, between *Bgl*II (nt 753) and *Bam*HI (nt 1095) sites, in the XPC cDNA to create a defective deletion product.

pIRES2-EGFP (Clontech) contains a neomycin resistance gene expressed from a SV40 promoter that allows for the selection of transfected cells that is used as a toxicity control. The vector also contains a multi-cloning site downstream a CMV promoter and upstream of an IRES and eGFP marker.

pCMV-Bsd (Life Technologies) contains a blasticidin resistance gene.

Both plasmids above were used in parallel in the retrotransposition assays as a combined control for transfection and growth variations between the cell lines used.

All plasmid DNA were purified by alkaline lysis and twice purified by cesium chloride buoyant density centrifugation. DNA quality was also evaluated by the visual assessment of ethidium bromide stained agarose gel electrophoresed aliquots.

Retrotransposition assays

For cells with lower transfection efficiencies such as XPA-, XPA+, XPD-, XPD+ and XPC- fibroblasts, L1 retrotransposition assay were performed as described in (MORAN *et al.* 1996) using the L1 episomal vectors. Briefly, $5x10^6$ cells were seeded in T75 flasks. Cells were transfected the next day at about 90% confluence using Lipofectamine 2000 (Life Technologies) following the manufacturer's protocol. For L1 retrotransposition assays, cells were transfected

with 1 or 3µg of L1.3 or L1.3-RT (-) construct tagged with the *mblast* retrotransposition cassette (TAM102/L1.3 or TAMD702A/L1.3) in T75 flasks. 3 µg of XPC expression vector were used to transiently complement XPC-deficient cells. Cells were selected for the presence of the L1 plasmid in selective medium containing 200 µg/ml hygromycin (Life Technologies) for 5 days. The selection was removed and cells were grown for 7 days in growth medium. After the growing period, hygro^R cells were trypsinized and reseeded in 6-well plates in triplicate at serial dilutions, from 10^6 to 10^3 cells, and selected for transposition events in medium containing 10 µg/mL blasticidin (InvivoGen). After 15 days, cells were fixed and stained with crystal violet solution (0.2% crystal violet in 5% acetic acid and 2.5% isopropanol). The number of blast^R or neo^R colonies was counted in each well and the relative number of colonies per 10^6 transfected cells was determined.

L1 toxicity and colony formation assay

L1 toxicity and colony formation assays were performed using the L1 episomal vectors. Briefly, $5x10^6$ cells were seeded in T75 flasks. Cells were transfected the next day at about 90% confluence using Lipofectamine 2000 (Life Technologies) following the manufacturer's protocol. Cells were transfected with 3 µg of L1.3, L1.3-RT (-), or L1.3-EN (-) construct tagged with the *mblast* retrotransposition cassette (TAM102/L1.3, TAMD702A/L1.3, or TAMH230A/L1.3). 3 µg of XPC expression vector were used to transiently complement XPC-deficient cells. Cells were selected for the presence of the L1 plasmid in selective medium containing 200 µg/ml hygromycin (Life Technologies) for 14 days. The cells were then fixed and stained with crystal violet solution (0.2% crystal violet in 5% acetic acid and 2.5% isopropanol). The number of hygro^R colonies was counted in each flask.

RT-qPCR

$5x10^6$ cells were seeded in T75 flasks. Cells were transfected the next day at about 90% confluence using Lipofectamine 2000 (Life Technologies) following the manufacturer's protocol. Cells were transfected with 3µg of L1.3 construct tagged with the *mblast* retrotransposition cassette (TAM102/L1.3). To allow for the expression of the vector, total RNA were extracted the next day from each flask, using TRIzol Reagent (Life Technologies). We then carried out chloroform extraction and isopropanol precipitation. RNA was suspended in 80 µL DEPC-treated water. The cDNA was synthetized using the Reverse Transcription System (Promega), following the manufacturer's protocol. Briefly 1 µg total RNA was denatured at 75° for 5 min. Reverse transcription reaction was primed with $Oligo(dT)_{15}$ primers and incubated at 42° for 1 h in a thermocycler (BioRad, C1000 Touch). The enzyme was then heat-inactivated at 99° for 5 min. To avoid contamination with L1 DNA from the genome or the expression vector, and to make sure that only L1 transcript is quantified, we analysed the spliced *mblast* reporter cassette in L1 RNA using the TAQMAN qPCR approach. The PCR reaction (SsoFast Probes Supermix, BioRad) was performed in a qPCR thermocycler (BioRad, CFX96) using 1µL of cDNA and the blasticidin primers at 250 mM final concentration each: Primer F (5' GCAGATCGAGAAGCACCTGT 3'), Primer R (5' TGGTGTCAATGTATATCATTTTACTGG 3') and the 5'-FAM-labeled, ZEN internal quencher, Iowa-Black Fluorescent Quencher (IDT) probe at 150mM final concentration (5' /56-FAM/AGGTTGCCA/ZEN/GCTGCCGCA/3IABkFQ/ 3') as follows: initial denaturation at 95°, 45 cycles of 95° for 10 s, 60° for 20 s. The probe was designed to span the splice junction of the mature RNA so that it will not anneal to the unspliced DNA. The fluorescence was read during the annealing/elongation step (60°; 20 s). For normalization, Pre-Developed TaqMan Assay Reagents for Human beta-actin (cat# 4326315E, Life Technologies) were used. Reactions were

done in triplicate. The values of the Cq for each PCR reaction were determined using the BioRad CFX Manager software (BioRad).

Recovery of *de novo* L1 inserts

De novo L1 insert recovery was performed as previously described with slight modifications, as noted below (MORRISH *et al.* 2002; EL-SAWY *et al.* 2005). Briefly, 3 x 10^6 XPA-, XPD-, XPC- or XPD+ cells were transfected with 3µg synL1_neo rescue vector (GASIOR *et al.* 2007) or pBS-L1PA1$_{CH}$_blast rescue vector using Lipofectamine 2000 reagent (Life Technologies). Cells were selected with 200 µg/mL hygromycin for 5 days, and then put under 400 µg/mL of G418 or 10 µg/mL of Blasticidin for 14 days to allow for colony formation. Neo^R and blast^R cells were harvested by trypsinization and genomic DNA was extracted using a Qiagen DNeasy Blood and Tissue kit. Genomic DNA was digested with 100U *Hin*dIII (NEB) overnight at 37°. The following day, digested genomic DNA was self-ligated using 1200 U T4 DNA ligase (NEB) in a volume of 1 mL overnight at room temperature. DNA was purified and concentrated using centrifugal filters (Amicon Ultra, 0.5 mL, 50K, Millipore). Purified DNA was transformed by electroporation into competent DH5α *E. coli* (Life Technologies). Individual kanamycin- or blasticidin-resistant colonies were grown and plasmid DNA was harvested using SV Wizard miniprep kit (Promega). The 5' end of the *de novo* L1 insert was sequenced using primers specific to the L1 rescue plasmid and primer walking until the 5' end of the insert was recovered as described in (MORRISH *et al.* 2002). Because sequencing through a long adenosine tract at the 3' end of the L1 inserts is not effective, the 3' flanking genomic region was sequenced by ligation mediated PCR based on (YUANXIN *et al.* 2003; STREVA *et al.* 2015). Briefly, a pool of five to six L1 rescue vectors was digested with *Stu*I (NEB) to relax supercoils, and then sheared by sonication using a Bioruptor (Diagenode, high, 30 seconds on, 90 seconds

off, for 12 minutes). Sheared plasmid DNA was primer extended using an oligo specific to the 3' end of the L1 rescue plasmid (3'_rescue_1: 5' ATATATGAGTAACCTGAGGC 3' or 3'_rescue_1_secondpA: 5' GTGGGCATTCTGTCTTGTTC 3'). Duplexed T-linkers were ligated using 10U T4 DNA ligase and PCR was performed using the primers: linker specific (5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3') and 3'_rescue_1 (or 3'_rescue_1_secondpA) primers as follows: initial denaturation at 94°, 20 cycles of 94° for 30s, 60° for 1 min, 72° for 1 min, and a final extension for 10 min at 72°. PCR reactions were run on a 1% agarose gel and a light smear between 400-700 nt was gel extracted with the Qiaquick gel extraction kit (Qiagen). One μL of gel extracted DNA was subject to an additional 15 cycles of PCR amplification as described above using linker specific and nested 3' rescue vector primers (3'_rescue_2: 5' TGAGTAACCTGAGGCTATGCTG 3' or 3'_rescue_2_secondpA: 5' TTCTGTCTTGTTCCGGTTCTTAAT 3') primers. The nested PCR product was run on a 1% agarose gel and the resulting smear was gel extracted and cloned into TOPO-TA (Life Technologies). Cloned PCR products were Sanger sequenced using M13 forward and reverse primers to determine 3' end junctions. Samples were sent for sequencing to Elim Biopharmaceuticals, Inc, Hayward, California. Lasergene 10 SeqBuilder software was utilized for sequence analysis. Flanking regions were mapped on human reference genome hg19 (build 37) using Blat tool (https://genome.ucsc.edu/cgi-bin/hgBlat). We verified the presence of the 5' and 3' flanking sequences within a L1 rescue plasmid with a combination of PCR between the two regions (Supplemental Table S3 for primers and supplemental Figure S8B for results). Sequences from rescues are included in Supplemental File 1.

Immunoblot analysis

To evaluate expression of NER proteins in the cells, XPA-, XPA+, XPC-, XPD-, XPD+ cells were haverested in 300 µl of lysis buffer (50 mM Tris, pH7.2, 150 mM NaCl, 0.5% Triton X-100, 10 mM EDTA, 0.5% SDS). After 10 min of sonication (Bioruptor, Diagenode, manufacturer's recommended settings), lysates were clarified by centrifugation for 15 min at 4° at 13,000 rpm and the protein concentration was determined by Bradford assay (Biorad). 40 µg of proteins were loaded on 4-12% bis-tris polyacrylamide gel and 3-8% tris-acetate polyacrylamide gel (Life Technologies). Proteins were transferred to a nitrocellulose membrane using iBlot gel transfer system from Life Technologies (manufacturer's settings). Membrane was blocked for 1h at room temperature in PBS (pH7.4), 0.1% Tween 20 (Sigma), 5% skim milk powder (OXOID) and then incubated overnight at 4° with an anti-XPC monoclonal antibody (D-10, sc-74410, Santa Cruz Biotechnology), anti-XPD monoclonal antibody (ab54676, abcam) or anti-XPA polyclonal antibody (H-273, sc-853, Santa Cruz Biotechnology) diluted at 1:1,000 in PBS, 0.1% Tween 20, 3% non-fat dry milk. Membrane was then incubated for 1 hour at room temperature with the secondary donkey anti-rabbit HRP-conjugated antibody (sc-2317, Santa Cruz Biotechnology) diluted at 1:100,000 in PBS, 0.1% Tween 20, 3% non-fat milk. Signals were detected using Super Signal West Femto Chemiluminescent Substrate (Pierce).

To determine XPC protein expression, 24 hours after transfection with 3 µg of XPC expression vector or control vector, XPC-deficient cells and untransfected HeLa cells were harvested in 300 µl of lysis buffer (50 mM Tris, pH7.2, 150 mM NaCl, 0.5% Triton X-100, 10 mM EDTA, 0.5% SDS). After 10 min of sonication (Bioruptor, Diagenode, manufacturer's recommended settings), lysates were clarified by centrifugation for 15 min at 4° at 13,000 rpm and the protein concentration was determined by Bradford assay (Biorad). 20 µg of proteins for transfected cells and 40 µg of proteins for HeLa cells were loaded on 3-8% tris-acetate polyacrylamide gel (Life Technologies). Proteins were transferred to a nitrocellulose membrane

using iBlot gel transfer system from Life Technologies (manufacturer's settings). Membrane was blocked for 1h at room temperature in PBS (pH7.4), 0.1% Tween 20 (Sigma), 5% skim milk powder (OXOID) and then incubated overnight at 4° with an anti-XPC polyclonal antibody (H-300, sc-30156, Santa Cruz Biotechnology) diluted at 1:1,000 in PBS, 0.1% Tween 20, 3% non-fat dry milk. Membrane was then incubated for 1 hour at room temperature with the secondary donkey anti-rabbit HRP-conjugated antibody (sc-2317, Santa Cruz Biotechnology) diluted at 1:100,000 in PBS, 0.1% Tween 20, 3% non-fat milk. Signals were detected using Super Signal West Femto Chemiluminescent Substrate (Pierce).

UV sensitivity assay

The protocol was adapted from (EMMERT *et al.* 2000). Briefly $5x10^5$ cells were seeded in 6-cm plates and grown in growth medium for 24 hours. The growth medium was removed and the cells were irradiated in the presence of 1 mL of 1X phosphate buffer saline (PBS) with a bactericidal UVC lamp (254 nm, 1.57 J/m^2/s) at 0, 3, 6, 9 and 12 J/m^2 UVC dose. The PBS was removed and replaced with growth medium. After four days, cells were counted with a hemocytometer to determine cell survival. Cell survival was calculated as the percent of live cells in the irradiated sample relative to the untreated sample. To determine the efficiency of the XPC transient complementation, $3x10^6$ XPC-deficient cells were transfected with 3 μg of XPC expression vector or the control vector in T75 flasks using Lipofectamine 2000 (Life Technologies) and were reseeded in 6-cm plates the next day in the conditions mentioned above. The UV assay was then performed as mentioned above.

Cell cycle analysis

Cells from a confluent T75 plates were harverested and fixed for over 24 hrs with ice cold 70% ethanol at -20°C. After fixation cells were washed once with PBS and incubated in PI/Triton X-100 staining solution (20 μg/ml propidium iodide, 0.1% (v/v) Triton X-100 and 0.2 mg/mL RNAse A in PBS) for at least 2 hours. Flow cytometry of the stained cells was carried out on a Becton Dickinson LSRII using DiVA Software (Louisiana Cancer Research Consortium (LCRC) FACS Core) and 50,000 events were collected. The data was analyzed using Modfit Software (Verity Software House).

RESULTS

Two essential proteins of the NER pathway, XPA and XPD, limit L1 retrotransposition.

To determine whether the NER pathway controls L1 retrotransposition and is responsible for the role of the ERCC1-XPF complex in preventing L1 mobility, we investigated the effect of the deficiency of XPA and XPD, two crucial factors of the NER pathway, on L1 retrotransposition. XPA recruits ERCC1-XPF to the site of damage by directly binding the complex (Figure 1B and (VOLKER *et al.* 2001; SAIJO *et al.* 2004)). The helicase activity of the XPD protein forms part of the core structure of the transcriptional/DNA repair complex TFIIH that opens the chromatin around the lesion during the NER process (Figure 1B and (COIN *et al.* 1998)). We performed L1 retrotransposition assays with marked L1 elements (MORAN *et al.* 1996) in XPA- and XPD-deficient cell lines, and their stably complemented partners, XPA+ and XPD+ cells. We transfected the cells with the episomal pCEP4 vector carrying a L1.3 element (TAM102/L1.3) tagged at its 3' end with the blasticidin (*mblast*) retrotransposition cassette (MORRISH *et al.* 2002). The principle of the assay (Figures 2A and B) is based on the introduction of an intron into the reporter gene in a manner that the reporter gene becomes functional after transcription of the L1 element, splicing of L1 mRNA, target-primed reverse transcription. Thus the formation of blasticidin-resistant colonies, due to the expression of the reporter gene, indicates that retrotransposition occurred. The results of the L1 retrotransposition assays in XPA-, XPA+, XPD- and XPD+ cells showed that stable complementation of NER-deficiency greatly decreased the number of $blast^R$ colonies (Figures 2C to F). These results are not due to differences in L1 expression in the cells, because the quantification of L1 mRNA by RT-qPCR showed equal amount of RNA in all the cells, deficient or proficient for the NER pathway (Supplemental Figure S1). In addition, immunoblot analysis verified that XPA and XPD protein expression in

the complemented XPA- and XPD-deficient cell lines was restored (Supplemental Figure S2A). A UV sensitivity assay (LEVY *et al.* 1995) validated that the repair of UV-caused damage is not efficient in XPA- and XPD-deficient cells (low cell survival at high UV dose exposure), and becomes efficient again in their complemented versions (Supplemental Figures S2B and S2C). The increase in L1 retrotransposition in NER-deficient cells does not seem to be explained by the growth rate of the different cell lines, because the cell cycle analysis by FACS revealed a slightly longer cycle for NER-deficient cells, as a few more cells accumulated in G2 phase (Supplemental Figure S3). A longer cell cycle does not seem to be a significant factor of a retrotransposition rate increase in our studies, because previous publications showed that, although transiting the cell cycle does not seem to be required for retrotransposition (KUBO *et al.* 2006) any arrest in cell cycle seemed to lead to a decrease in L1 retrotransposition (KUBO *et al.* 2006; SHI *et al.* 2007; XIE *et al.* 2013).

As a control, we tested for the possibility that NER-deficiency would promote the retrotransposition of L1 elements defective in the endonuclease (EN) as has been seen in other DNA repair defects (MORRISH *et al.* 2002). No blast^R colonies were observed performing retrotransposition assays in NER-deficient cells, using L1 constructs individually defective in either endonuclease or reverse transcriptase (TAMH230A102/L1.3, EN(-) or TAMD702A/L1.3, RT(-)) (Supplemental Figure S4). In parallel, to rule out any bias engendered by various transfection efficiency, growth rates and L1-caused toxicity, we verified that L1 expression did not alter colony formation in a manner that could cause the retrotransposition rate differences in any of the cells used in the experiment (Supplemental Figure S5).

The GGR DNA lesion binding protein, XPC, also contributes to the limitation of L1 retrotransposition.

NER can recognize and repair a wide variety of lesions that are not related in their chemical structures, such as bulky chemical DNA adducts and UV-induced lesions (GILLET and SCHÄRER 2006). Some reports suggest that it is the helical distortion that attracts the NER repair proteins (CLEMENT *et al.* 2010; NAEGELI and SUGASAWA 2011). Whether the DNA lesion occurs in a transcriptionally active or inactive region in the genome will influence the NER sensor used for the repair. Because L1 insertion occurs throughout the genome (DEININGER and BATZER 2002; MORRISH *et al.* 2002) and not just in transcribing genes, we evaluated the GGR DNA lesion binding protein XPC for its ability to limit L1 retrotransposition. We monitored L1 retrotransposition events in XPC-deficient cells and in transiently complemented (XPC+) versions of the same cells.

Monitoring XPC protein (XPCp) expression level and repair efficiency of the UV-induced DNA damage, we determined that the transient transfection of XPC-deficient cells resulted in a partial complementation of the deficiency (Figure 3A and B). In fact, the cell survival rate was under 10% for moderate UV dose (6 J/m^2), while the cell survival rate of stably complemented XPA+ and XPD+ cell lines and HeLa cells were about 50% for the same UV dose (Supplemental Figures S2B and S2C). However the cell survival of transiently complemented XPC- cells (XPC+) using a vector expressing XPC cDNA (pCMV6-XPC), which express a large amount of XPCp (Figure 3A), was significantly higher than the cell survival of the XPC-deficient cells (XPC-) transfected with the vector control (pCMV6-XPCΔ) ($P<0.0001$, Chi-square test for trend) (Figure 3B). Showing the same trend as in XPA- and XPD-deficient cells, L1 retrotransposition increased three fold in the XPC-deficient cells (XPC-) in comparison to the partially complemented cells (XPC+) (Figure 3C). These results indicate that the lesion binding protein in the GGR pathway, XPCp, is likely involved in the regulation of L1 retrotransposition.

The deficiency in NER proteins does not have an impact on the size of *de novo* L1 inserts, but *de novo* L1 inserts present abnormally large target-site duplications in NER-deficient cell lines.

Because complementation of NER-deficiency leads to a decrease in the apparent retrotransposition rate, suggesting that NER proteins inhibit retrotransposition events, we wished to investigate whether mutations in NER proteins impact the characteristics of *de novo* L1 inserts. We extracted the genomic DNA of XPC-, XPA- and XPD-deficient cells and in the stably complemented partner, XPD+, expressing the marker of retrotransposition. We then sequenced *de novo* inserts and the flanking regions to evaluate the features of the inserts, such as the length of the insert, the length of the polyA tail, the size of the target site duplication (TSD), and the sequence of the L1 endonuclease cleavage site. The principle of the method is presented in Supplemental Figure S6A. The previously published method (MORRISH *et al.* 2002; YUANXIN *et al.* 2003; EL-SAWY *et al.* 2005; GASIOR *et al.* 2007) was slightly modified to facilitate analysis of the 3' flanking sequences of the L1 elements recovered as plasmids (details in the Materials and Methods). Because the stably complemented cell lines were resistant to neomycin, we could not use the previously described synL1_neo vector (EL-SAWY *et al.* 2005). As a result, we designed a new rescue vector, pBS-L1PA1$_{CH}$_blast, using a blasticidin retrotransposition cassette (Supplemental Figure S6B). Except for the insert size that depends on the length of the reporter cassette, the characteristics of *de novo* L1 inserts obtained with the new L1PA1$_{CH}$_blast rescue vector in HeLa cells were similar to those previously published in (GILBERT *et al.* 2002; SYMER *et al.* 2002; GILBERT *et al.* 2005) (Table S1 and S2).

We recovered 29 *de novo* L1 inserts in XPD-deficient cells (XPD-), nine in XPA-deficient cells (XPA-) and 26 in XPC-deficient cells (XPC-) using the synL1_neo vector and 17 in stably complemented XPD-deficient cells (XPD+) using L1PA1$_{CH}$_blast rescue vector (Tables 1, 2, 3

and 4). We particularly focused more on recovering L1 inserts in the XPD-deficient cell lines in order to allow a statistically significant analysis. The recovered inserts were dispersed randomly throughout the genome with no noticeable preference for a specific site or orientation on chromosomes. No full-length *de novo* L1 insertions were recovered in NER-deficient cells or in the stably complemented cell line, XPD+. The inserts were all 5' truncated and the size of the inserts varied from 1.8 to 2.8 kb in the complemented cells and from 2.3 to 4.3 kb in NER-deficient cells (the minimal recoverable size of the insert is 1.8 kb using L1PA1$_{CH}$_blast rescue vector for the assay or 2.3 kb using synL1_neo vector as the presence of the selection marker, blasticidin or neomycin respectively, and the bacterial origin of replication at the 3' end of the L1 construct are required for recovery). In comparison with data obtained in HeLa cells, we detected no significant difference in the median length of L1 inserts in NER-deficient and proficient cells lines (P-value Mann-Whitney U-test > 0.05, Supplemental Figure S7A and S7B). Therefore, NER proteins do not seem to influence the size of the *de novo* L1 inserts.

The majority of *de novo* L1 inserts occurred at typical L1 endonuclease cleavage sites (Tables 1, 2, 3 and 4 and (MORRISH *et al.* 2002)). However, one out of the nine (11%) recovered inserts (clone #AM7, Table 4) in XPA-deficient cells, nine out of 26 (35%) in XPC- cells (clones #CM3, #CM7, #CM8, #CM9, #CM15, #CM16, #CM18, #CM20, and #CM21, Table 3) and three out of 28 (11%) in XPD- cells (clones #DM8, #DM15, and #DM20, Table 1) occurred at an atypical cleavage site and also did not contain the polyA tail at the 3' end of the L1 sequence, a hallmark of an L1 insertion event. Inserts similar to these have been previously described as the result of endonuclease-independent insertion events in DNA repair-deficient cells (MORRISH *et al.* 2002). However in the present study, the frequency of these events is significantly higher in XPC-deficient cells than in the other NER-deficient cells and HeLa cells (GILBERT *et al.* 2005) and we did not observe any retrotransposition events in NER-deficient cells transfected with an

L1 construct defective in the endonuclease activity of ORF2p (Supplemental Figure S4). Therefore, endonuclease-independent insertion of L1 in NER-deficient cells does not seem to explain these events and these atypical inserts may be the results of recombination events between L1 and genomic sequences occurring at the insertion site during the insertion process.

Genomic deletions at the L1 insertion site were observed in three out of 26 (12%) in XPC-deficient cells (clones #CM24 to #CM26, Table 3), and two out of nine (22%) in XPA-deficient cells (clones #AM8, and #AM9, Table 4). These observations are similar in frequency to small deletions found from analysis of polymorphic genomic L1 inserts (KONKEL *et al.* 2007) although our deletions were slightly longer in general. A significantly higher rate of deletions was observed in the complemented XPD+ cells (eight out of 17, 47%). For nearly half of these events, deletions were generated by homologous recombination between the sequences of the L1 element in the genome at the insertion site and the elongating *de novo* L1 (see TPRT process, Figure 1A). This type of event has been previously reported and described in HeLa cells although not at this high rate (GILBERT *et al.* 2005).

One area in which NER-deficient L1 inserts differed from those in NER-proficient cells is the size of the target size duplication. In NER-deficient cells, TSDs ranged from 5 bp to 5.3 kb in size with a median length of 1495 bp, while in XPD+ cells TSDs ranged from 2 bp to 1.6 kb in length with a median of 14 bp (Tables 1-4). TSD length in NER-deficient cells was significantly larger than those observed in either cells stably complemented to correct the NER deficiency or in NER-proficient HeLa cells (P-value Mann-Whitney U-test < 0.05, Figure 4). Previous studies have also reported that the typical size of a TSD from L1 inserts recovered from normal culture cells and from L1 insert naturally occurring in the genome is usually 5-30 bp (MORRISH *et al.* 2002; SYMER *et al.* 2002; GILBERT *et al.* 2005). Events generating large TSDs (over 100 bp) have been observed in HeLa cells but the frequency of these events never exceeded 10% of the

total recovered events (Table S1 and S2 and (GILBERT *et al.* 2002; SYMER *et al.* 2002; GILBERT *et al.* 2005)). In the NER-deficient cell lines tested in the present study, four out of seven (57%) *de novo* inserts on XPA- cells, 21 out of 23 (91.3%) in XPC- cells, and 22 out of 28 (78.6%) in XPD- cells presented TSDs over 100 bp in length (Figure 4 and Supplemental Figure S8) and of these, 18 (38%) had TSDs of over 1 kb in length. Only three out of nine (33%) events presented TSDs over 100 bp in the stably complemented cells (XPD+), a difference that is statistically significant (P-value Fisher exact test < 0.05). It seems possible that the slightly elevated levels of longer TSDs in the complemented cells relative to previous studies in HeLa, for instance, may be due to the complementation not necessarily bringing the NER pathway to maximum efficiency. Therefore, our findings seem to imply that the expression of NER proteins is required to prevent large genomic rearrangements at the L1 insertion site.

DISCUSSION

L1 elements have a tremendous capacity to cause damage to the human genome, potentially causing disease, cell death and aging (BELGNAOUI *et al.* 2006; GASIOR *et al.* 2006; WALLACE *et al.* 2010; BELANCIO *et al.* 2014). Because L1 retrotransposition is a threat to genome integrity, it is not surprising that numerous cellular factors control the mobility of these elements at different levels. For example, L1 expression is maintained at a low level by methylation (ALVES *et al.* 1996), premature polyadenylation and alternative RNA splicing (BELANCIO *et al.* 2008b). In addition, at a post-transcriptional level, members of human cytidine deaminase APOBEC3 family, involved in the response against viral infection, inhibit the activity of L1 retrotransposons (BOGERD *et al.* 2006; KINOMOTO *et al.* 2007; LOVSIN and PETERLIN 2009; RICHARDSON *et al.* 2014).

Our data demonstrate that the entire global NER pathway represents an additional mechanism regulating L1 activity. The ERCC1-XPF complex, previously identified as an L1 regulator, is involved in many DNA repair pathways and is more essential for life than the other NER components. All three NER proteins tested in this study (XPAp, XPDp and XPCp, Figures 2 and 3) show a similar ability to suppress L1 retrotransposition as previously reported for the ERCC1-XPF complex (GASIOR *et al.* 2008). This suppression seems likely to be the result of an altered ability to complete the retrotransposition process successfully, rather than an ability to make longer inserts in NER deficiency given that we detected no significant differences in insert length (Figure S7).

It seems likely that the NER pathway is primarily responsible for the inhibition of L1 retrotransposition by ERCC1-XPF, as the magnitude of the effect of XPA, XPD and ERCC1 deficiencies on L1 retrotransposition were very similar. This does not, however, rule out the possibility that some of the other sensors and pathways that recruit ERCC1-XPF to DNA lesions

may also be involved, to a lesser extent, in the control of L1 insertion. In particular, TCR, the other NER subpathway, may also contribute to the sensing of L1 insertion events. However, because TCR only protects the portion of the genome that is being actively transcribed in that specific cell type (MELLON *et al.* 1987), the TCR subpathway could only have a more limited effect on L1 insertion rates. Because the TCR pathway is strand specific, its primary influence would primarily be limited to influencing the orientation preference of the insertion within genes rather than on the global rate of retrotransposition events.

Previously, the Garfinkel laboratory reported that Ssl2 and Rad3, two DNA helicases of the TFIIH complex, homologs of human XPB and XPD proteins, respectively, can inhibit Ty1 retrotransposition in *Saccharomyces cerevisiae* (LEE *et al.* 1998). Ssl2 and Rad3 helicases are proposed to induce the degradation of Ty1 cDNA in the cytoplasm by destabilizing its structure. However, Ty1 and L1 elements have very different integration mechanisms. The former is a virus-like retrotransposon, similar to retroviruses, and its reverse transcription occurs in a pseudo-viral particle in the cytoplasm. The cDNA is imported into the nucleus and is integrated in the genome. In contrast, L1 mRNA is reverse transcribed inside the nucleus (TPRT), directly at the site of the insertion of the new copy. Therefore, it is not expected that these two kinds of elements would induce a similar cellular response. Intriguingly in both cases, it seems likely that the NER components are involved in limiting the insertion by limiting the synthesis/stability of the cDNA intermediate.

In addition to the role of the NER factors in limiting the insertion of a new copy of L1, our L1 insert recovery data suggest that the NER pathway may also be involved in the less well-defined last step of the L1 TPRT process, as a high number of extremely large TSDs was observed in NER-deficient cell lines. These results suggest that the NER factors present at the site of L1 insertion may influence the formation or selection of the second nick necessary for the

completion of the TPRT process at a location close to the cleavage site of the L1 endonuclease (Figure 5). Although speculative, in NER-deficient cells, the location of the second nick is highly variable and might depend on a more random event that can occur at a location very far away from the primary cleavage site, thereby generating extremely large TSDs. The distant location of the second nick can be the result of the direct influence of the chromosomal region on L1 insertion process such as exposed DNA ends generated during replication or DNA repair processes, serving as primers for the synthesis of the second L1 cDNA strand. An alternative possibility is that the predicted cDNA flap intermediate may persist in ERCC1-XPF deficient cells, giving the insertion process a longer time to reach a naturally occurring nick at a more distant location through the strand displacement activities of helicases recruited to the site of L1 insertion by DNA repair machinery.

In addition to the long TSDs in these L1 insertions, the only other unusual feature is an unusually high number of events that do not look like they went through the normal TPRT process on the polyA tail. Although most of the inserts have the typically long A tails expected of de novo L1 insertions, these 'tailless' inserts seem to indicate some other undefined aberrant processes during the insertion.

The extremely large TSDs found in NER-deficient L1 insertions fall within the size range of genomic CNVs (EICHLER 2001). This large size makes them potentially more deleterious as insertions. However it also, has the consequence of duplicating regions that may have functional significance, such as exons. Thus, they represent a unique source of this type of genomic duplication. It is also worth considering, however, that large segments of homology near one another also have a propensity to recombine with one another (STANKIEWICZ and LUPSKI 2002). This would be similar to the formation of solo LTRs following retroviral insertion (BELSHAW *et al.* 2007). Thus, this type of event might be expected to have a shortened lifespan in the genome

and some L1 insertions may be eliminated by recombination between their long TSDs. Our data suggest that individuals with defects in NER, such as xeroderma pigmentosum patients, are likely to be subject to higher levels of DNA damage caused by L1 elements. Other studies reported that NER capacity varies between individuals (SPITZ *et al.* 2001; TOMESCU *et al.* 2001; HOU *et al.* 2002) in response to different stimuli such as circadian rhythm (GADDAMEEDHI *et al.* 2011) or are specifically repressed in some cancers (XU *et al.* 2011). Under those conditions and in combination with a deregulation of L1 expression, NER deficiency may increase the potential impact of L1 retrotransposition on the human genome, favoring somatic genetic instability and the generation of age-related diseases, such as cancer.

FUNDING

This research was supported by National Institutes of Health (NIH) P20 P20GM103518/ P20RR020152 (PLD, VPB and AMR-E), R01GM079709A (AMR-E), R01GM45668 and EPSCOR/BORSF grant (PLD), and NIH/NIA 5K01AG030074 (VPB). VPB is also supported by The Ellison Medical Foundation New Scholar in Aging award, AG-NS-0447-08, and the Kay Yow foundation. This paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

ACKNOWLEDGEMENTS

The lentivirus production and purification was performed at the Vector Core at the LSU Health Sciences Center in New Orleans.

GS and BJW were involved in the design and with support from RSD and MN performed most of the experimental assays and data collection. MN performed the XPA experiments and complementation experiments. SLG performed the siRNA experimental assays and analysis. VAS performed experiments to process *de novo* L1 insertion events and performed their analysis with GS. GS, RSD, and VPB performed immunoblot analysis. GS and BJW contributed to the data analysis and editing of the manuscript. PLD and AMR-E performed the overall conception and study design, some data collection, analysis, interpretation and data writing of the paper and decision to submit it for publication.

REFERENCES

Ahmad, A., A. R. Robinson, A. Duensing, E. van Drunen, H. B. Beverloo *et al.*, 2008 ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol Cell Biol 28**:** 5082-5092.

Al-Minawi, A. Z., N. Saleh-Gohari and T. Helleday, 2008 The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic Acids Res 36**:** 1-9.

Alves, G., A. Tatro and T. Fanning, 1996 Differential methylation of human LINE-1 retrotransposons in malignant cells. Gene 176**:** 39-44.

Belancio, V. P., D. E. Blask, P. Deininger, S. M. Hill and S. M. Jazwinski, 2014 The aging clock and circadian control of metabolism and genome stability. Front Genet 5**:** 455.

Belancio, V. P., D. J. Hedges and P. Deininger, 2008a Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18**:** 343-358.

Belancio, V. P., A. M. Roy-Engel and P. Deininger, 2008b The impact of multiple splice sites in human L1 elements. Gene 411**:** 38-45.

Belgnaoui, S. M., R. G. Gosden, O. J. Semmes and A. Haoudi, 2006 Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. Cancer Cell Int 6**:** 13.

Belshaw, R., J. Watson, A. Katzourakis, A. Howe, J. Woolven-Allen *et al.*, 2007 Rate of recombinational deletion among human endogenous retroviruses. J Virol 81**:** 9437-9442.

Boeke, J. D., 1997 LINEs and Alus--the polyA connection. Nat Genet 16**:** 6-7.

Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink, 1985 Ty elements transpose through an RNA intermediate. Cell 40**:** 491-500.

Bogerd, H. P., H. L. Wiegand, B. P. Doehle, K. K. Lueders and B. R. Cullen, 2006 APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res 34**:** 89-95.

Christensen, S. M., J. Ye and T. H. Eickbush, 2006 RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc Natl Acad Sci U S A 103**:** 17602-17607.

Cleaver, J. E., E. T. Lam and I. Revet, 2009 Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat Rev Genet 10**:** 756-768.

Clement, F. C., U. Camenisch, J. Fei, N. Kaczmarek, N. Mathieu *et al.*, 2010 Dynamic two-stage mechanism of versatile DNA damage recognition by xeroderma pigmentosum group C protein. Mutat Res 685**:** 21-28.

Coin, F., J. C. Marinoni, C. Rodolfo, S. Fribourg, A. M. Pedrini *et al.*, 1998 Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. Nat Genet 20**:** 184-188.

de Koning, A. P., W. Gu, T. A. Castoe, M. A. Batzer and D. D. Pollock, 2011 Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7**:** e1002384.

de Laat, W. L., E. Appeldoorn, N. G. Jaspers and J. H. Hoeijmakers, 1998 DNA structural elements required for ERCC1-XPF endonuclease activity. J Biol Chem 273**:** 7835-7842.

Deininger, P. L., and M. A. Batzer, 2002 Mammalian retroelements. Genome Res 12**:** 1455-1465.

Eichler, E. E., 2001 Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17**:** 661-669.

El-Sawy, M., S. P. Kale, C. Dugan, T. Q. Nguyen, V. Belancio *et al.*, 2005 Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. J Mol Biol 354**:** 246-257.

Emmert, S., N. Kobayashi, S. G. Khan and K. H. Kraemer, 2000 The xeroderma pigmentosum group C gene leads to selective repair of cyclobutane pyrimidine dimers rather than 6-4 photoproducts. Proc Natl Acad Sci U S A 97**:** 2151-2156.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke, 1996 Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87**:** 905-916.

Gaddameedhi, S., C. P. Selby, W. K. Kaufmann, R. C. Smart and A. Sancar, 2011 Control of skin cancer by the circadian rhythm. Proc Natl Acad Sci U S A 108**:** 18790-18795.

Gasior, S. L., G. Preston, D. J. Hedges, N. Gilbert, J. V. Moran *et al.*, 2007 Characterization of pre-insertion loci of de novo L1 insertions. Gene 390**:** 190-198.

Gasior, S. L., A. M. Roy-Engel and P. L. Deininger, 2008 ERCC1/XPF limits L1 retrotransposition. DNA Repair (Amst) 7**:** 983-989.

Gasior, S. L., T. P. Wakeman, B. Xu and P. L. Deininger, 2006 The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357**:** 1383-1393.

Gilbert, N., S. Lutz, T. A. Morrish and J. V. Moran, 2005 Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol 25**:** 7780-7795.

Gilbert, N., S. Lutz-Prigge and J. V. Moran, 2002 Genomic deletions created upon LINE-1 retrotransposition. Cell 110**:** 315-325.

Gillet, L. C., and O. D. Schärer, 2006 Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem Rev 106**:** 253-276.

Groisman, R., J. Polanowska, I. Kuraoka, J. Sawada, M. Saijo *et al.*, 2003 The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. Cell 113**:** 357-367.

Hancks, D. C., and H. H. Kazazian, 2012 Active human retrotransposons: variation and disease. Curr Opin Genet Dev 22**:** 191-203.

Hancks, D. C., and H. H. Kazazian, 2016 Roles for retrotransposon insertions in human disease. Mob DNA 7**:** 9.

Henning, K. A., L. Li, N. Iyer, L. D. McDaniel, M. S. Reagan *et al.*, 1995 The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. Cell 82**:** 555-564.

Hoeijmakers, J. H., 2001 Genome maintenance mechanisms for preventing cancer. Nature 411**:** 366-374.

Hou, S. M., S. Falt, S. Angelini, K. Yang, F. Nyberg *et al.*, 2002 The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis 23**:** 599-603.

Kinomoto, M., T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima *et al.*, 2007 All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res 35**:** 2955-2964.

Konkel, M. K., J. Wang, P. Liang and M. A. Batzer, 2007 Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. Gene 390**:** 28-38.

Kubo, S., M. C. Seleme, H. S. Soifer, J. L. Perez, J. V. Moran *et al.*, 2006 L1 retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci U S A 103**:** 8036-8041.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature 409**:** 860-921.

Laugel, V., 2013 Cockayne syndrome: the expanding clinical and mutational spectrum. Mech Ageing Dev 134**:** 161-170.

Lee, B. S., C. P. Lichtenstein, B. Faiola, L. A. Rinckel, W. Wysock *et al.*, 1998 Posttranslational inhibition of Ty1 retrotransposition by nucleotide excision repair/transcription factor TFIIH subunits Ssl2p and Rad3p. Genetics 148**:** 1743-1761.

Lehmann, A. R., D. McGibbon and M. Stefanini, 2011 Xeroderma pigmentosum. Orphanet J Rare Dis 6**:** 70.

Levy, D. D., M. Saijo, K. Tanaka and K. H. Kraemer, 1995 Expression of a transfected DNA repair gene (XPA) in xeroderma pigmentosum group A cells restores normal DNA repair and mutagenesis of UV-treated plasmids. Carcinogenesis 16**:** 1557-1563.

Li, L., E. S. Bales, C. A. Peterson and R. J. Legerski, 1993 Characterization of molecular defects in xeroderma pigmentosum group C. Nat Genet 5**:** 413-417.

Lovsin, N., and B. M. Peterlin, 2009 APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. Ann N Y Acad Sci 1178**:** 268-275.

Luan, D. D., and T. H. Eickbush, 1996 Downstream 28S gene sequences on the RNA template affect the choice of primer and the accuracy of initiation by the R2 reverse transcriptase. Mol Cell Biol 16**:** 4726-4734.

Luan, D. D., M. H. Korman, J. L. Jakubczak and T. H. Eickbush, 1993 Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72**:** 595-605.

Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke and A. Gabriel, 1991 Reverse transcriptase encoded by a human transposable element. Science 254**:** 1808-1810.

Mellon, I., G. Spivak and P. C. Hanawalt, 1987 Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell 51**:** 241-249.

Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke *et al.*, 1996 High frequency retrotransposition in cultured mammalian cells. Cell 87**:** 917-927.

Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato *et al.*, 2002 DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31**:** 159-165.

Naegeli, H., and K. Sugasawa, 2011 The xeroderma pigmentosum pathway: decision tree analysis of DNA quality. DNA Repair (Amst) 10**:** 673-683.

Nouspikel, T., 2009 DNA repair in mammalian cells: So DNA repair really is that important? Cell Mol Life Sci 66**:** 965-967.

O'Donovan, A., A. A. Davies, J. G. Moggs, S. C. West and R. D. Wood, 1994 XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. Nature 371**:** 432-435.

Ostertag, E. M., and H. H. Kazazian, Jr., 2001 Biology of mammalian L1 retrotransposons. Annu Rev Genet 35**:** 501-538.

Puumalainen, M. R., P. Rüthemann, J. H. Min and H. Naegeli, 2016 Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. Cell Mol Life Sci 73**:** 547-566.

Reardon, J. T., and A. Sancar, 2005 Nucleotide excision repair. Prog Nucleic Acid Res Mol Biol 79**:** 183-235.

Richardson, S. R., I. Narvaiza, R. A. Planegger, M. D. Weitzman and J. V. Moran, 2014 APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife 3**:** e02008.

Riedl, T., F. Hanaoka and J. M. Egly, 2003 The comings and goings of nucleotide excision repair factors on damaged DNA. EMBO J 22**:** 5293-5303.

Saijo, M., T. Matsuda, I. Kuraoka and K. Tanaka, 2004 Inhibition of nucleotide excision repair by anti-XPA monoclonal antibodies which interfere with binding to RPA, ERCC1, and TFIIH. Biochem Biophys Res Commun 321**:** 815-822.

Sargent, R. G., J. L. Meservy, B. D. Perkins, A. E. Kilburn, Z. Intody *et al.*, 2000 Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells. Nucleic Acids Res 28**:** 3771-3778.

Schärer, O. D., 2011 Multistep damage recognition, pathway coordination and connections to transcription, damage signaling, chromatin structure, cancer and aging: current perspectives on the nucleotide excision repair pathway. DNA Repair (Amst) 10**:** 667.

Shi, X., A. Seluanov and V. Gorbunova, 2007 Cell divisions are required for L1 retrotransposition. Mol Cell Biol 27**:** 1264-1270.

Sijbers, A. M., P. J. van der Spek, H. Odijk, J. van den Berg, M. van Duin *et al.*, 1996 Mutational analysis of the human nucleotide excision repair gene ERCC1. Nucleic Acids Res 24**:** 3370-3380.

Spitz, M. R., X. Wu, Y. Wang, L. E. Wang, S. Shete *et al.*, 2001 Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. Cancer Res 61**:** 1354-1357.

Stankiewicz, P., and J. R. Lupski, 2002 Genome architecture, rearrangements and genomic disorders. Trends Genet 18**:** 74-82.

Streva, V. A., V. E. Jordan, S. Linker, D. J. Hedges, M. A. Batzer *et al.*, 2015 Sequencing, identification and mapping of primed L1 elements (SIMPLE) reveals significant variation in full length L1 elements between individuals. BMC Genomics 16**:** 220.

Sugasawa, K., J. Akagi, R. Nishi, S. Iwai and F. Hanaoka, 2009 Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. Mol Cell 36**:** 642-653.

Sugasawa, K., J. M. Ng, C. Masutani, S. Iwai, P. J. van der Spek *et al.*, 1998 Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell 2**:** 223-232.

Sugasawa, K., T. Okamoto, Y. Shimizu, C. Masutani, S. Iwai *et al.*, 2001 A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes Dev 15**:** 507-521.

Sugasawa, K., Y. Shimizu, S. Iwai and F. Hanaoka, 2002 A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex. DNA Repair (Amst) 1**:** 95-107.

Symer, D. E., C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost *et al.*, 2002 Human l1 retrotransposition is associated with genetic instability in vivo. Cell 110**:** 327-338.

Takayama, K., E. P. Salazar, A. Lehmann, M. Stefanini, L. H. Thompson *et al.*, 1995 Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D. Cancer Res 55**:** 5656-5663.

Tomescu, D., G. Kavanagh, T. Ha, H. Campbell and D. W. Melton, 2001 Nucleotide excision repair gene XPD polymorphisms and genetic predisposition to melanoma. Carcinogenesis 22**:** 403-408.

Troelstra, C., A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma *et al.*, 1992 ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell 71**:** 939-953.

van den Boom, V., E. Citterio, D. Hoogstraten, A. Zotter, J. M. Egly *et al.*, 2004 DNA damage stabilizes interaction of CSB with the transcription elongation machinery. J Cell Biol 166**:** 27-36.

Vermeulen, W., and M. Fousteri, 2013 Mammalian transcription-coupled excision repair. Cold Spring Harb Perspect Biol 5**:** a012625.

Volker, M., M. J. Mone, P. Karmakar, A. van Hoffen, W. Schul *et al.*, 2001 Sequential assembly of the nucleotide excision repair factors in vivo. Mol Cell 8**:** 213-224.

Wagstaff, B. J., M. Barnerssoi and A. M. Roy-Engel, 2011 Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. PLoS One 6**:** e19672.

Wallace, N. A., V. P. Belancio, Z. Faber and P. Deininger, 2010 Feedback inhibition of L1 and alu retrotransposition through altered double strand break repair kinetics. Mob DNA 1**:** 22.

Xie, Y., L. Mates, Z. Ivics, Z. Izsvák, S. L. Martin *et al.*, 2013 Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. Mob DNA 4**:** 10.

Xing, J., D. J. Witherspoon, D. A. Ray, M. A. Batzer and L. B. Jorde, 2007 Mobile DNA elements in primate and human evolution. Am J Phys Anthropol Suppl 45**:** 2-19.

Xu, X. S., L. Wang, J. Abrams and G. Wang, 2011 Histone deacetylases (HDACs) in XPC gene silencing and bladder cancer. J Hematol Oncol 4**:** 17.

Yuanxin, Y., A. Chengcai, L. Li, G. Jiayu, T. Guihong *et al.*, 2003 T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. Nucleic Acids Res 31**:** e68.

Zhu, X. D., L. Niedernhofer, B. Kuster, M. Mann, J. H. Hoeijmakers *et al.*, 2003 ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Mol Cell 12**:** 1489-1498.

TABLES

Table 1. Characteristics of *de novo* L1 inserts recovered in XPD-deficient cells

Clone #	Insert size	Chromosome	Orientation	Position (5' -> 3' end)	TSD length	TSD verification	pA length	Cleavage site
DM.1	2663	chr5	+	137495110 - 137495108	2	S	41nt	TCTT/a
DM.2	2883	Chr18	-	57193322 - 57193327	5	S	55nt	TTTA/a
DM.3	3147	chr12	+	112775935 - 112775928	7	S	45nt	CTTT/a
DM.4	2780	Chr7	-	61968906 - 61968919	13	S	81nt	TTTC/a
DM.5	2827	chr12	-	98026422 - 98026438	16	S	30nt	TTTT/a
DM.6	3929	Chr8	+	128728905 - 128728851	54	S	58nt	TCTT/a
DM.7	3989	chr17	-	13596243 - 13596460	217	S	83nt	TTTT/a
DM.8	2804	chr17	-	50628899 - 50629306	407	S/P	-	CTAG/a
DM.9	3626	chr1	-	121484567 - 121484978	411	S	73nt	TTCT/g
DM.10	3686	chr3	-	174149054 - 174149550	496	S/P	84nt	TCTT/a
DM.11	2550	chr7	+	144760424 - 144759911	513	S/P	34nt	TTTT/g
DM.12	3008	Chr3	+	29453614 - 29453069	545	S/P	25nt	TTTT/a
DM.13	3167	chr11	+	105190706 -105190152	554	S/P	46nt	ACTT/g
DM.14	2559	ChrX	-	10708604 - 10709170	571	S/P	77nt	TTCT/a
DM.15	3719	chr15	+	30302496 - 30301796	700	S/P	-	CCCA/g
DM.16	3103	Chr8	-	58341572 - 58342341	769	S/P	97nt	TTTC/a
DM.17	3411	Chr14	-	77523239 - 77524010	771	S/P	69nt	TTTT/a
DM.18	3610	chr9	+	5534460 - 5533597	863		37nt	TTTT/a
DM.19	2928	ChrX	+	33775626 - 33774740	886	S/P	33nt	TTTA/g
DM.20	2582	chr7	-	130375857 - 130376914	1057	S/P	-	TGGT/c
DM.21	2380	chr9	+	40823585 - 40822515	1070	S/P	56nt	CTTT/a
DM.22	2894	chrX	+	122778988 - 122777801	1187		39nt	TTTC/a
DM.23	2536	chr5	+	64034619 - 64033086	1533	P	19nt	TTTT/c
DM.24	3917	chr11	+	14727287 - 14725642	1645		57nt	TTTC/a
DM.25	3857	Chr1	-	120251234 - 120249499	1735	S/P	65nt	GTTT/g
DM.26	3991	chr3	+	29455524 - 29453069	2455	P	22nt	TTTT/a
DM.27	3940	Chr15	-	28869801 - 28873325	3361	P	87nt	TATT/a
DM.28	3996	Chr13	-	44857642 - 44861667	4025	P	57nt	TTTG/a
DM.29	2655	Chr3	+	29458420 - 29453069	5896		30nt	TTTT/a

S: sequence; P: PCR

Table 2. Characteristics of *de novo* L1 inserts recovered in (XPD+) complemented XPD-deficient cells

Clone #	Insert size	Chromosome	Orientation	Position (5' -> 3' end)	TSD length	Deletion length	pA length	Cleavage site
DP.1	1956	chr3	+	166704010 - 166704008	3		31nt	TTTT/a
DP.2	1989	chr8	+	121836950 - 121836947	4		31nt	TCTT/a
DP.3	2318	chr20	+	52652085 - 52652079	7		38nt	TTTA/a
DP.4	2270	chr3	-	156453738 - 156453742	4		28nt	CTTT/a
DP.5	2269	chr7	+	90201740 - 90201726	15		45nt	TTTC/a
DP.6	2038	chr16	-	53268517 - 53268532	16		33nt	TTTT/a
DP.7	1866	chr8	+	30763655 - 30763645	11		46nt	TTAT/a
DP.8	2013	chr2	-	24215408 - 24216927	1520		35nt	TTTT/a
DP.9	1977	chr17	+	35458775 - 35457164	1612		40nt	TCTT/a
DP.10	2704	chr12	-	18901143 - 18901142	-	-	52nt	TCTT/a
DP.11	1866	ch15	-	87439350 - 87439347		2	52nt	TTTA/a
DP.12	2791	chr6	+	99948429 - 99948435		5	19nt	TTTT/a
DP.13	1900	chr14	-	51873941 - 51873933		7	25nt	TTTG/a
DP.14	2062	chr1	-	94735344 - 94735334		9	34nt	TCTT/a
DP.15	2380	chr4	-	30539318 - 30539303		14	25nt	TTTT/a
DP.16	2002	chr2	-	59684354 - 59684338		15	110nt	TTCT/a
DP.17	2741	chr6	-	248880 - 248410		469	55nt	TTTT/a

Table 3. Characteristics of *de novo* L1 inserts recovered in XPC-deficient cells

Clone #	Insert size	Chromosome	Orientation	Position (5' -> 3' end)	TSD length	Deletion length	pA length	Cleavage site
CM.1	2491	chr15	-	52326493 - 52326495	2		25nt	TTTT/a
CM.2	2736	chr4	+	58109300 - 58109280	20		29nt	TTTA/a
CM.3	2421	chr4	-	81232879 - 81233154	275		-	ATTT/a
CM.4	3167	chr5	+	41311414 - 41311137	277		60nt	TCTT/g
CM.5	2886	chr15	+	33641486 - 33640983	503		15nt	TTTT/a
CM.6	2661	chr8	+	23490649 - 23489121	516		18nt	TGAA/a
CM.7	2946	chr2	+	182583931 - 182583379	552		-	AATA/a
CM.8	2886	chr6	-	85235896 - 85236510	614		-	AATT/t
CM.9	2567	chr5	-	65162148 - 65162763	615		-	AATG/a
CM.10	2609	chr16	-	9557849 - 9558491	642		10nt	TTTT/a
CM.11	3386	chr8	+	26232474 - 26231790	684		25nt	TCTT/a
CM.12	2677	chr9	+	66811576 - 66810877	699		15nt	TTTT/g
CM.13	3293	chr17	+	49728154 - 49727312	842		6nt	GATT/g
CM.14	3112	chr11	-	114392300 - 114393187	887		23nt	TCTT/a
CM.15	2894	chr3	+	34830281 - 34829315	966		-	AGTT/c
CM.16	2604	chr1	-	211947488 - 211948623	1135		-	CTGC/a
CM.17	2684	chr8	+	127633155 - 127631920	1235		84nt	TTCT/a
CM.18	3044	chr3	+	5007476 - 5005443	2033		-	ATTT/t
CM.19	3096	chr17	+	48833262 - 48831219	2043		21nt	TTTC/a
CM.20	2718	chr20	-	12381555 - 12384581	3026		-	ACCT/g
CM.21	3035	chr18	+	20321307 - 20317622	3685		-	TAAG/c
CM.22	2890	chr2	-	137563056 - 137567601	4545		31nt	TTTT/a
CM.23	4006	Chr19	-	27732520 - 27739285	6765		38nt	TTTT/g
CM.24	2700	chr18	-	22016206 - 22016201		5	21nt	CTTT/a
CM.25	4306	chr14	-	67575620 - 67575614		6	20nt	TTAA/g
CM.26	4192	chr9	+	115010933 - 115010942		9	13nt	TTTT/a

Table 4. Characteristics of *de novo* L1 inserts recovered in XPA-deficient cells

Clone #	Insert size	Chromosome	Orientation	Position (5' -> 3' end)	TSD length	Deletion length	pA length	Cleavage site
AM.1	2490	chr2	-	36487725 - 36487738	13		35nt	TTTC/a
AM.2	4230	chr3	-	47259213 - 47259253	40		79nt	TTTC/a
AM.3	2678	chr11	+	85736502 - 85736455	47		42nt	TCTT/a
AM.4	3409	chr15	-	28728325 - 28728567	242		25nt	TTTC/a
AM.5	2265	chr9	-	135357719 - 135359341	1622		5nt	TTTT/g
AM.6	2712	chr2	+	89248581 - 89246458	2123		-	ACTG/a
AM.7	2756	chr12	-	19866410 - 19866397		13	9nt	TTTT/a
AM.8	2306	chr3	+	36006942 - 36010967		4025	50nt	TGTT/a

FIGURE LEGENDS

Figure 1: A similar 3'flap DNA structure is generated during L1 insertion process (TPRT) and the NER pathway.

A. Schematic of the first steps of L1-TPRT reaction. 1) ORF2 endonuclease recognizes a consensus sequence 5'-TTTTAA-3' and cleaves the DNA between the T and A nucleotides resulting in a T-rich free 3' end. 2) This end is allowed to base-pair with the polyA tail of the L1 mRNA (orange) and prime the reverse transcription of the mRNA. In this model, the proposed structure formed by the elongating cDNA is a 3' flap intermediate. This 3' flap intermediate, a known substrate for the structure-specific endonuclease ERCC1-XPF, is proposed to be cleaved, limiting L1 retrotransposition and leading to the restoration of the original DNA sequence.

B. Schematic model of the first steps of the human NER pathway. The NER pathway consists of two subpathways: Transcription Coupled Repair (TCR) and the Global Genome Repair (GGR). Choice of pathway is determined by DNA lesion recognition. In the TCR subpathway, the base lesion in actively transcribed regions of the genome induces the arrest of transcription elongation by RNA Polymerase II (RNAPII). CSA and CSB proteins are recruited to the site of stalled RNAPII and initiate the repair process. In the non-transcribed genomic regions, NER repair occurs through the GGR subpathway. A wide variety of DNA base damage is detected by the DNA lesion-binding protein XPC through the structural distortion of the DNA helix. The next steps of the repair are identical in both subpathways. Once the base lesion is recognized and signaled, the general transcription factor TFIIH, a complex of ten components including the helicases XPD and XPB, is recruited to the damage site and responsible for DNA unwinding around the lesion. XPA-RPA proteins stabilize the opened chromatin structure and recruit the endonuclease ERCC1-XPF to the damage site. XPG endonuclease is recruited with the complex TFIIH and seems to be required for the DNA unwinding. XPG incises the damaged

strand at the 3' of the lesion and ERCC1-XPF excises 5' of the damage. Proteins in bold are the factors evaluated in the present study.

Figure 2: XPD and XPA proteins limit L1 retrotransposition.

A. Schematic of the L1-retrotransposition assay. The L1 vector is a pCEP4 episomal vector carrying a hygromycin resistance ($hygro^R$) gene, for the selection of transfected cells, and a full length L1 element tagged at the 3' end with a retrotransposition cassette. The retrotransposition cassette consists of a reporter gene such as blasticidin resistance ($blast^R$) or a neomycin resistance (neo^R) gene interrupted with an intron. The reporter gene is in the reverse orientation in comparison to the L1 element and its transcription is driven by its own promoter, whereas the intron is in the sense orientation (direction of transcription is indicated by arrows). The splice donor and acceptor of the intron are indicated as SD and SA. The resistance gene becomes functional after L1 transcription, splicing, and TPRT of a new L1 copy into the genome. Therefore, only when retrotransposition of the cassette occurs, cells can grow under selection conditions ($blast^R$ or neo^R colonies).

B. Schematic of the protocol of the L1-retrotransposition assay used with the $hygro^R$ pCEP4 episomal vectors. The day after transfection, cells containing plasmid are selected with hygromycin for three days. Then selection medium is removed and replaced by growth medium for recovery. Ten days post-transfection, cells were reseeded in 6-well plates at serial dilution, from 10^6 to 10^3 cells, in triplicate and grown under the appropriate selection. Two weeks later, cells were fixed and stained in crystal violet solution and the number of $blast^R$ or neo^R colonies in each well is counted.

C. Relative retrotransposition rates of a $blast^R$-tagged L1 reporter element (TAM102/L1.3) were determined in an XPA-deficient cell line (XPA-) and a commercially available stably

complemented XPA-deficient cell line (XPA+). The results were normalized relative to XPA+, which was arbitrarily set to 1.0. L1 retrotransposition assay was performed three times independently. Bars represent the average and standard error of the mean from the three independent experiments. Statistical significance is indicated by * P-value: $1x10^{-15}$ (two-tailed, two-sample T-test).

D. Representative example of blast^R colony formation resulting from L1 retrotransposition assay performed in XPA+ and XPA- cells. The L1 retrotransposition rate for each cell line is indicated below.

E. Relative number of blast^R colonies resulting from L1 retrotransposition assay performed in XPD+ and XPD- cell lines. The results were normalized relative to the XPD+, which was arbitrarily set to 1.0. L1 retrotransposition assay was performed three times independently. Bars represent the average with standard error of the mean from the three independent experiments. Statistical significance is indicated by * P-value: $6.5x10^{-6}$ (two-tailed, two-sample T-test).

F. Representative example of blast^R colony formation resulting from L1 retrotransposition assay performed in XPD+ and XPD- cells. The L1 retrotransposition rate for each cell line is indicated below.

Figure 3: XPC limits L1 retrotransposition.

A. Expression of XPC protein (XPCp) in HeLa cells and in XPC-deficient cells, transfected with either wildtype XPC (pCMV6-XPC) (XPC+) or mutant XPC (pCMV6-XPCΔ) (XPC-) expression vector was analyzed by immunoblotting 24h post transfection. The extra bands observed in XPC+ and XPC- lanes can be the truncated form of XPC protein, present in XPC-deficient cell line and caused by two deletions in the XPC gene sequence, generating a frameshift

and a truncated protein (LI *et al.* 1993). The truncated protein might be stabilized by the overexpression of the wildtype or mutant XPC protein.

B. Determination of the nucleotide excision repair competency of the transient complementation of the XPC-deficient cells with a wildtype XPC (pCMV6-XPC) or mutant XPC (pCMV6-XPCΔ) expressing vector by using a UV sensitivity assay. Top panel shows the schematic of the UV sensitivity assay (details in Materials and Methods). XPC- cells and its transiently complemented version, XPC+ cells, were exposed to 0, 3, 6, 9 or 12 J/m^2 UVC dose. Cell survival was determined four days post UV exposure. Data represent the logarithm of the percentage of cell survival plotted against the UV dose. The assay was performed at least three times independently for each condition, a representative is shown.

C. Relative number of $blast^R$ colonies resulting from L1 retrotransposition assay performed in XPC-deficient cell line and its transiently complemented version. XPC- cells were co-transfected with XPC cDNA (XPC+) or control (XPC-) vectors and the tagged L1 vector (TAM102/L1.3). The L1 retrotransposition assay was performed as described for Figure 2. The results were normalized to the XPC+ results, which was arbitrarily set to 1.0. Data correspond to the relative number of $blast^R$ colonies determined in each condition. L1 retrotransposition assays have been performed independently at least three times. Data represent the average with standard error of the mean. Statistical significance is indicated by *P-value: 0.0001 (two-tailed, two-sample T-test).

Figure 4: Recovered *de novo* L1 inserts in NER-deficient cells have large target site duplications (TSDs).

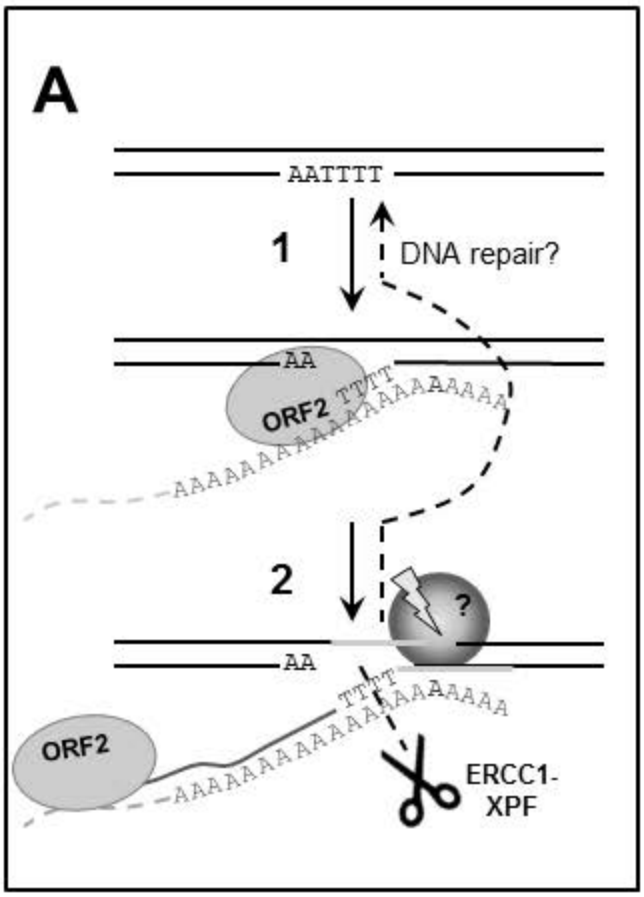
Dot plot representation of the size of TSDs of *de novo* L1 inserts recovered from XPD-deficient cells (XPD-), stably complemented XPD-deficient cells (XPD+) and HeLa cells. The

boxes represent the interquartile range (the range between the first and third quartiles) of the distribution of TSD size for each cell line. The line in the middle of the box represents the median. The count (n) of recovered inserts for each cell line is indicated above the name of the cell line. Statistical significance is indicated by * P-value: 0.032, **** P-value: $3.13*10^{-7}$ and ns: non-significant (P-value: 0.38) (Mann-Whitney test).

Figure 5: Model of limitation of damage caused by L1 TPRT by GGR-NER pathway.

The 3' flap DNA structure is generated by L1 elongating cDNA during TPRT process. The XPC complex recognizes the structure and recruits the other proteins of the NER pathway to the L1 insertion site. ERCC1-XPF endonuclease cleaves the elongating L1 cDNA, inhibiting the insertion of a new L1 element in the genome and leading to the restoration of the original DNA sequence. To a lesser extent, the NER pathway seems to be involved in generating a second nick during the L1 insertion process, in close proximity to the first cleavage generated by the L1 endonuclease. Therefore, in WT cells, the *de novo* insert is flanked by small (less than 100 bp) TSDs, whereas in NER-deficient cells, a more distal unrelated nick may be used to complete the retroelement insertion resulting in large TSDs.

Figure 1


A
AATTTT
1
DNA repair?
AA
ORF2
TTTT
AAAAAAAAAAAAAAAAAAAA
2
?
AA
TTTT
AAAAAAAAAAAAAAAAAAAA
ORF2
ERCC1-
XPF

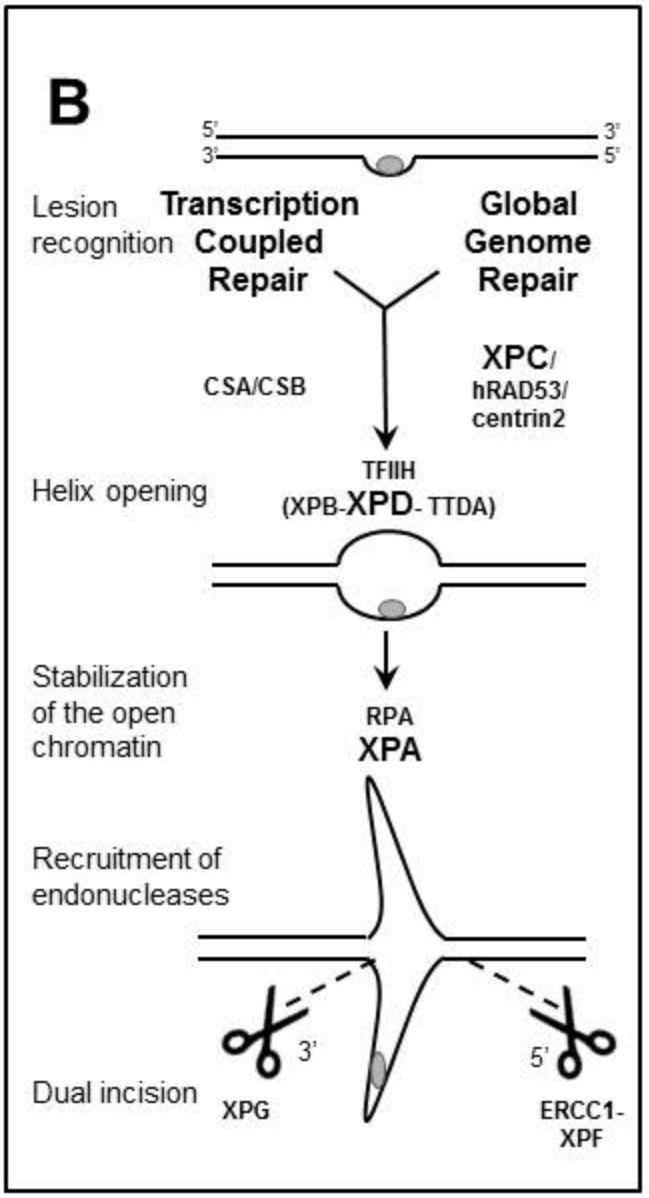

B
5'
3'
3'
5'
Lesion
recognition
Transcription
Coupled
Repair
Global
Genome
Repair
CSA/CSB
XPC/
hRAD53/
centrin2
Helix opening
TFIIH
(XPB-XPD- TTDA)
Stabilization
of the open
chromatin
RPA
XPA
Recruitment of
endonucleases
Dual incision
3'
5'
XPG
ERCC1-
XPF

Figure 2

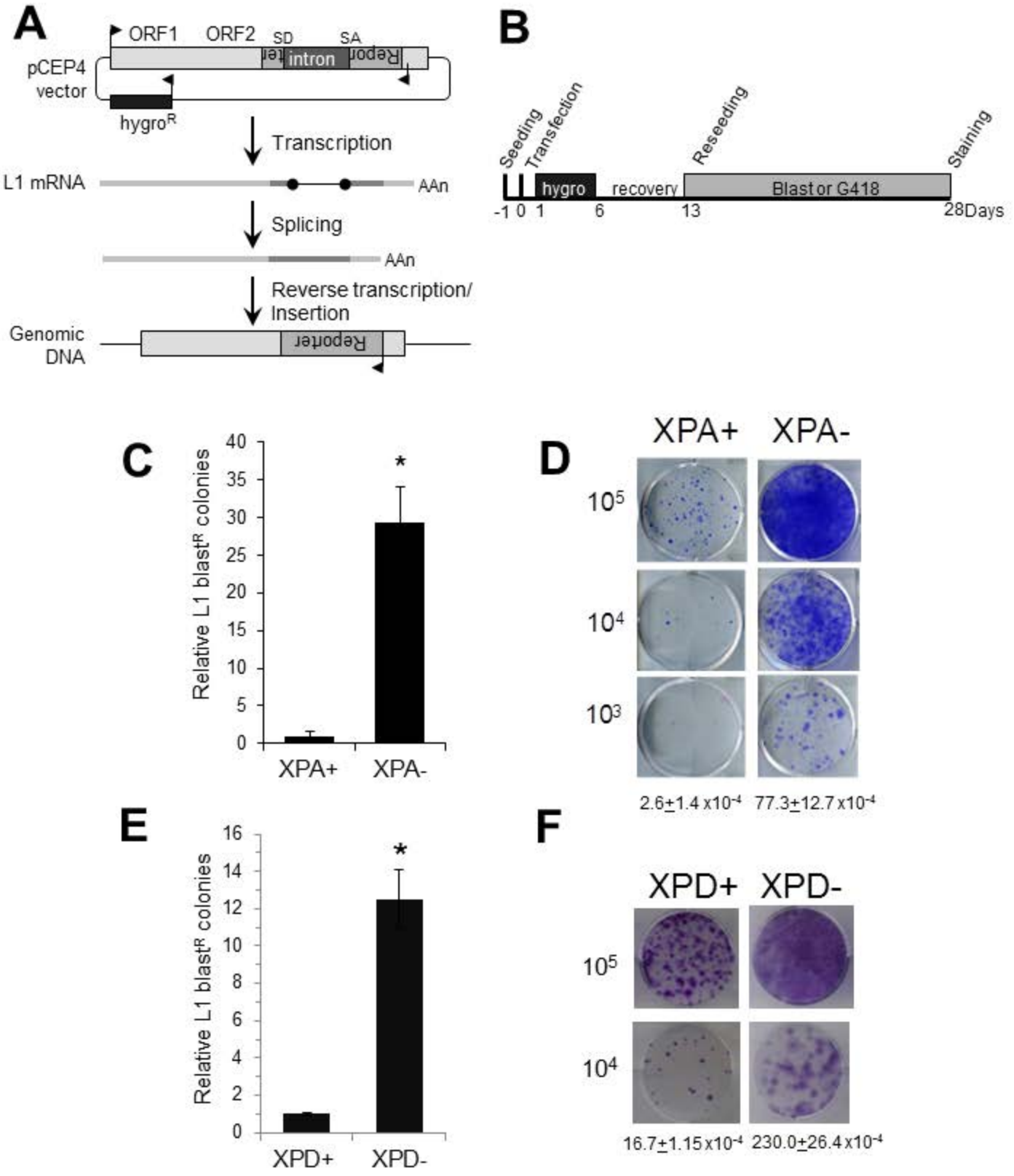

A
ORF1
ORF2
SD
SA
intron
pCEP4
vector
hygroR
Transcription
L1 mRNA
AAn
Splicing
AAn
Reverse transcription/
Insertion
Genomic
DNA
B
Seeding
Transfection
Reseeding
Staining
hygro
recovery
Blastor G418
-1 0 1
6
13
28Days
C
Relative L1 blastR colonies
40
35
30
25
20
15
10
5
0
*
XPA+
XPA-
D
XPA+
XPA-
10^5
10^4
10^3
2.6±1.4 x10^-4
77.3±12.7 x10^-4
E
Relative L1 blastR colonies
16
14
12
10
8
6
4
2
0
*
XPD+
XPD-
F
XPD+
XPD-
10^5
10^4
16.7±1.15 x10^-4
230.0±26.4 x10^-4

Figure 3

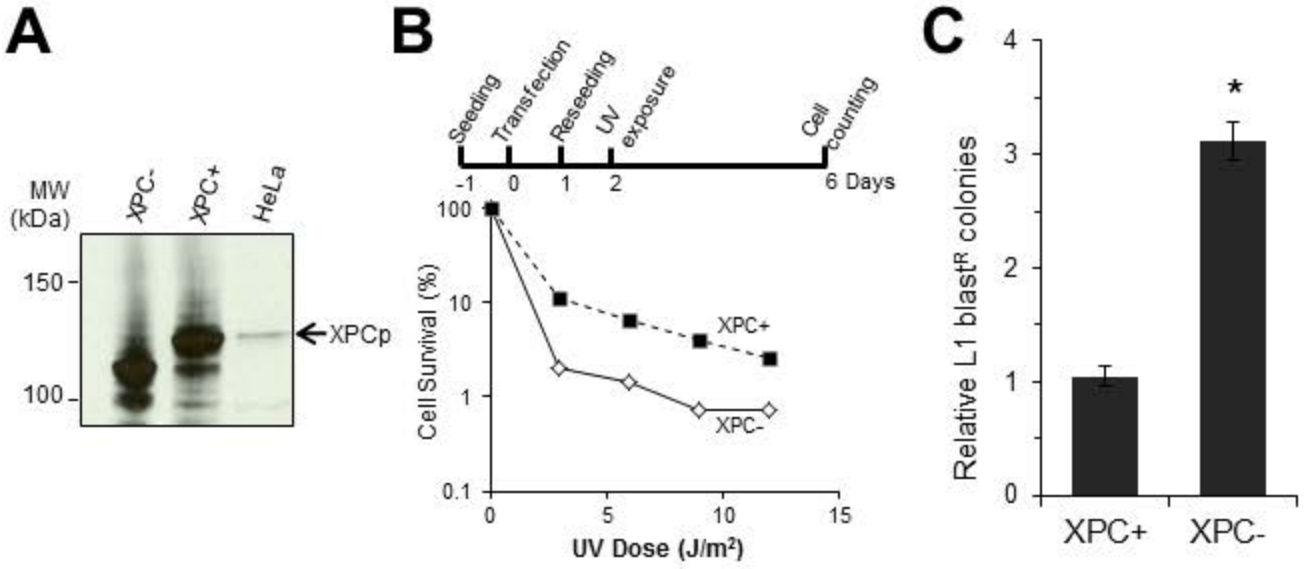

A
MW
(kDa)
XPC-
XPC+
HeLa
150
100
XPCp
B
Seeding
Transfection
Reseeding
UV
exposure
Cell
counting
-1
0
1
2
6 Days
100
10
1
0.1
Cell Survival (%)
XPC+
XPC-
0
5
10
15
UV Dose (J/m²)
C
Relative L1 blastR colonies
4
3
2
1
0
*
XPC+
XPC-

Figure 4

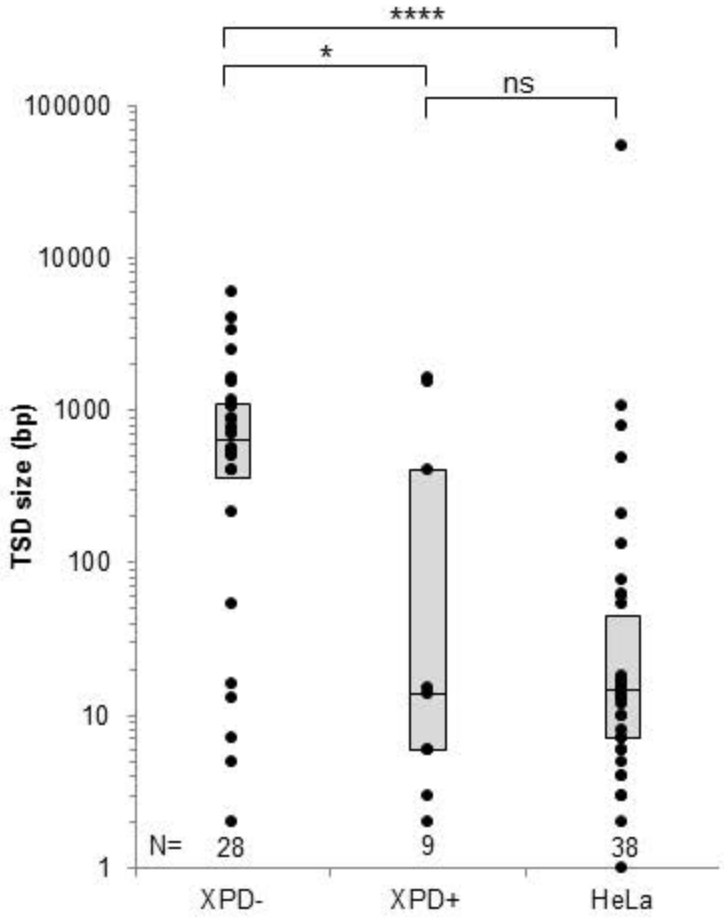
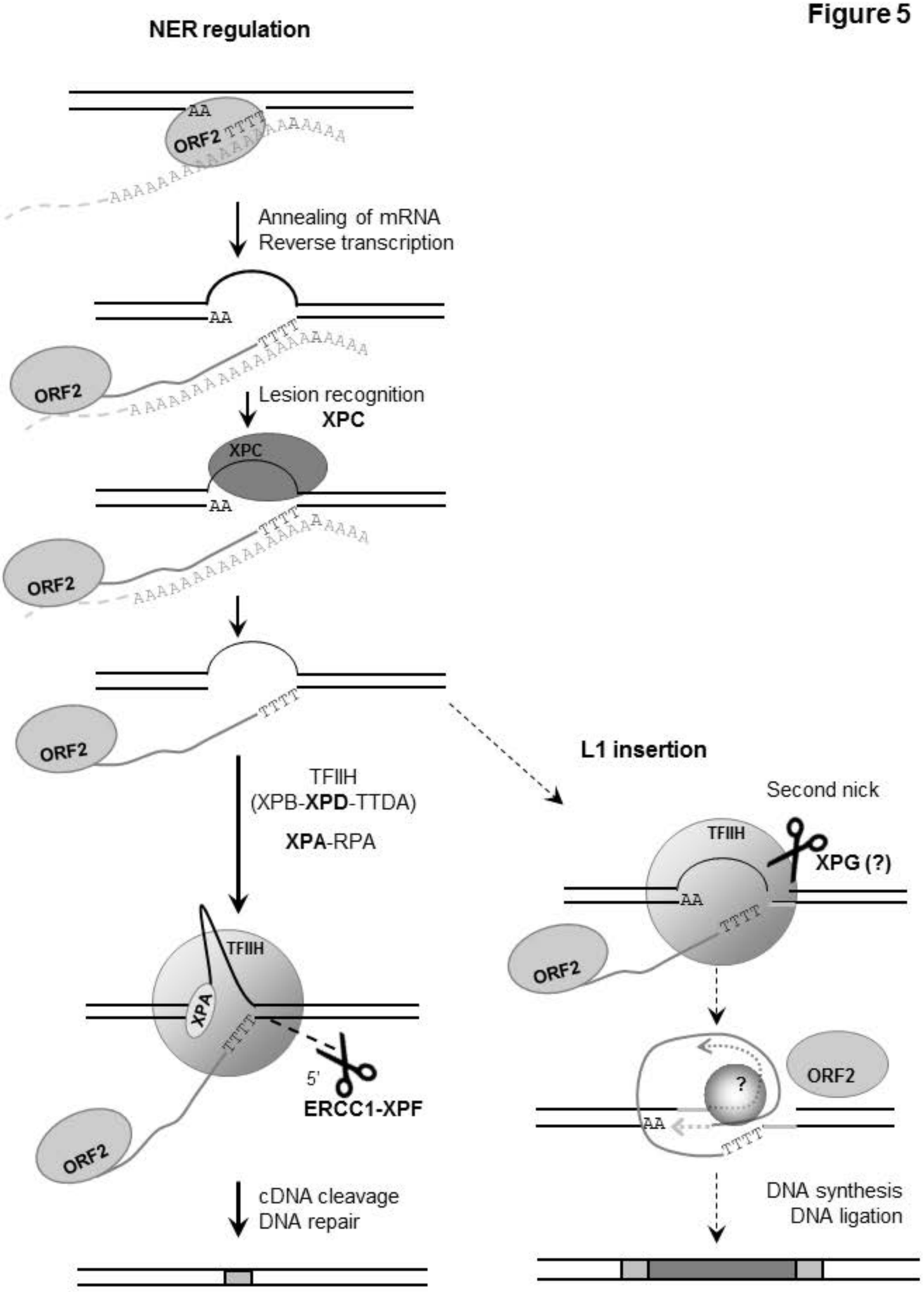


Figure 5


NER regulation
AA
ORF2
TTTT
AAAAAAAAAAAAAAAAAAAA
Annealing of mRNA
Reverse transcription
Lesion recognition
XPC
TFIIH
(XPB-XPD-TTDA)
XPA-RPA
XPA
5'
ERCC1-XPF
cDNA cleavage
DNA repair
L1 insertion
Second nick
XPG (?)
?
DNA synthesis
DNA ligation