SUPPLEMENTARY MATERIAL

PCR amplification and sequencing of *XPC* exons including splice donor and acceptor sites

In order to determine the DNA sequences at the splice donor and acceptor sites of all the 16 *XPC* exons, each exon and their respective splice donor and acceptor sites were PCR amplified using intronic primers flanking these sequences. A total of 17 primer pairs as shown in supplemental Table A were designed on the basis of the *XPC* genomic sequences we submitted (GenBank accession numbers AF261892-261901). The PCR reactions for the amplification of *XPC* exons 2 to 15 were performed using the Advantage cDNA PCR kit (Clontech, CA) as per the vendor's protocol. The PCR was conducted under the following conditions: 94 ^{o} C for 60 s, then 35 cycles of amplification (94 ^{o} C for 20 s and 66 ^{o} C for 3 min), ending with 66 ^{o} C for 3 min. The PCR amplification of *XPC* exon 1 was conducted using the Advantage-GC genomic PCR kit (Clontech, CA). The reaction was performed in a 50 ul volume containing 100-200 ng DNA, 100 ng of each primer, 0.5 M GC-melt solution and all the reagents indicated in the manufacturer's instructions. PCR was conducted under the following conditions: 94 ^{o} C for 60 s, then 35 cycles of amplification (94 ^{o} C for 30 s and 65 ^{o} C for 3 min), ending with 65 ^{o} C for 3 min. The PCR products were resolved on the agarose gel and bands of interest were sliced out and DNA was extracted from the slice using QIAquick gel extraction procedure (Qiagen, Santa Clara, CA). The sequence analysis was performed manually or by cycle sequencing employing dideoxy termination chemistry and an ABI 373A automated DNA sequencer (Applied Biosystems, CA).

Characterization of *XPC* genomic structure

The primers listed in supplemental Table B were used to amplify the *XPC* introns. They were designed on the basis of the *XPC* cDNA and sequences we submitted (GenBank accession numbers AF261892-261901) or the GenBank reference sequence (accession number AC090645). Some of the primers utilized were previously reported (1). A total of 16 primer pairs (either in exon or in intron) were employed to PCR amplify the entire *XPC* genome, which

spans ~ 33 kb using the conditions described in the supplemental material. The PCR amplification of *XPC* intron 1 was conducted in two different fragments. The 5' end was amplified using Advantage-GC genomic PCR kit (Clontech, CA) and primer pair S100 and S113 that resulted in a ~ 2.7 kb fragment. The PCR was conducted under the following conditions: 94 ^{o} C for 60 s, then 35 cycles of amplification (94 ^{o} C for 30 s and 64 ^{o} C for 3 min), ending with 64 ^{o} C for 3 min. The 3' end was amplified using primer pairs S116 and S87 and Advantage cDNA PCR kit (Clontech, CA) as per the manufacturer's protocols. The PCR conditions were same as we used for the amplification of 5' end of intron 1 and the size of the PCR product was 3.3 kb. Intron 2, intron 3, intron 13 and 14 were also amplified using Advantage cDNA PCR kit (Clontech, CA) as described above and annealing/extention temperatures were optimized for each primer pair. The primer pair, Exon2F and Exon3R results in a ~ 2.5 kb fragment (58 ^{o} C); A1 and S4 results in a ~ 2 kb fragment (66 ^{o} C); F1 and F2 results in a ~ 2 kb fragment (69 ^{o} C). Intron 6, intron 10 and intron 11 region of *XPC* gene was amplified using GeneAmp XL PCR kit (Applied Biosystems, CA) as per the vendor's protocol. The PCR was conducted under the following conditions: 94 ^{o} C for 60 s, then 35 cycles of amplification (94 ^{o} C for 30 s and 66 ^{o} C for 6 min), ending with 72 ^{o} C for 10 min. The gene specific primer pair, Exon6F2 and Exon7R results in a ~ 5 kb fragment (66 ^{o} C); J1 and D2 results in a ~ 3.5 kb fragment (66 ^{o} C); E1 and E2 results in a ~ 1 kb fragment (66 ^{o} C). For the amplification of remainder of the *XPC* genome, the PCR reaction was performed in a volume of 50 µl containing 1x PCR buffer, 1.5 mM MgCl2, 200 µM of each dNTPs, 1U Taq DNA polymerase (Invitrogen), 100 ng of each primer, and 50-200 ng DNA. The PCR was conducted under the following conditions: 94^{o}C for 60 s, then 35 cycles of amplification (94^{o}C for 15 s, optimized annealing temperature for 30 s, and 72^{o}C for 150 s), ending with 72^{o}C for 10 min. The primer pair Exon4F and exon 5.1R results in a ~ 1 kb fragment (55 ^{o} C); Exon5.1F and Exon5.2R results in a ~ 1.6 kb fragment (55.6 ^{o} C); Exon5.2F and Exon6R result in a ~ 0.7 kb fragment (55 ^{o} C); Exon7F and Exon8R result in a ~ 1 kb fragment (55 ^{o} C); M1 and M2b results in a ~ 2 kb fragment (56 ^{o} C); N1 and N2b results in a ~ 2.4 kb fragment (55.7 ^{o} C); N1c and N2 results in a ~ 1.65 kb fragment (54 ^{o} C); Exon12F and E2

results in a ~ 0.7 kb fragment (56 ^{o}C). The PCR products were purified and sequenced as described above.

Real time quantitative RT-PCR (QRT-PCR) for *XPC* Mrna

XPC mRNA isoforms were quantitated using isoform-specific real time QRT-PCR. One microgram of total RNA was converted to cDNA in a 20 µL reaction using Superscript II (Invitrogen) following the manufacturer's protocol. Both oligo (dT) (25 ng/µL) and random hexamer (2.5 ng/µL) were used as primers for RT. cDNA synthesis reactions were incubated at 25EC for 10 min followed by incubation at 37EC for 120 min. One µL (corresponding to 50 ng RNA or approximately 5,000 cells) of RT reaction was used as template for each real time quantitative PCR reaction. Real time QRT-PCR assays were carried out on a Bio-Rad iCycler iQ system (Bio-Rad, Hercules, CA, USA) using intercalation of SYBR Green as the fluorescence reporter. Reactions were carried out using the SYBR Green kit from PE Applied Biosystems following the manufacturer's protocol using 300 pM primers, 3 mM $MgCl_2$, 10 nM FAM, and 1 µL template (RT reaction or control plasmid) per 25 µL PCR reaction. Two step PCR (denaturation at 95EC for 15 s and annealing/extension at 60EC for 1 min) was done in the BioRad iCycler iQ with data collection and analysis during the combined annealing and extension step. Standard curves were created for each run using 10-fold serial dilutions of cloned *XPC* cDNA for the appropriate isoform. Melt curve analyses were performed on all PCR reactions to rule out nonspecific amplification.

Isoform-specific real time QRT-PCR assays were designed to discriminate between RNAs including or excluding specific exons (supplemental Table C and supplemental Figure 1). Primer sequences are based on the GenBank sequence for full length *XPC* mRNA (accession number D21089) and the exon junction information in Table 1. One primer for each assay spans a specific exon/exon junction, with only 2-4 nucleotides at the primer 3' end matching the 5' end of the downstream exon. Non-specific priming on other isoforms was prevented by at least a two out of three base pair mismatch at the primer 3' end. The second primer in each pair was

chosen using Vector NTI (Informax Inc.) to avoid primer dimer formation. All primer pairs were extensively validated prior to use. PCR reactions were carried out on cloned cDNAs for each *XPC* isoform and selected reverse transcription reactions and amplicon size were determined by electrophoresis on 12% Poly(NAT) gels (Elchrom Scientific). Primers were also tested for isoform specificity using cloned cDNAs. All primer pairs showed at least a 20 cycle difference between cDNA clones for exon inclusion and exon exclusion, indicating approximately a 10^6-fold specificity.

cDNA standards for real time QRT-PCR were generated by end-point PCR with the primers shown in supplementary Table C and subcloned with the TOPO TA cloning kit (Invitrogen) as per the vendor's protocol. The sequence of all clones was verified by cycle sequencing. The concentration of each standard was determined using picogreen (Molecular Probes) and also by real time quantitative polymerase chain reaction using a primer pair for the ampicillin resistance gene (supplementary Table C). This approach allows data from all assays to be expressed in terms of fg of the *XPC* clone, pXPC-3 (5) [this clone contains the partial *XPC* cDNA sequence GenBank accession number NM_004628 which does not contain the first 272 nucleotides of *XPC* cDNA GenBank accession number D21089] thus facilitating comparison between assays. For reference purposes, there are approximately 150 single stranded DNA molecules per fg of this clone.

Reference List

1. Legerski,R. and Peterson,C. (1992) Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C [published erratum appears in 1992 Dec 10;360(6404):610] *Nature*, **359**, 70-73.