SPANDIDOS PUBLICATIONS

Discrepancies between *in silico* and *in vitro* data in the functional analysis of a breast cancer-associated polymorphism in the *XRCC6/Ku70* gene

EMANUELA MARRAS[1], PETRA WILLEMS[2], VEERLE VANDERSICKEL[2], ISABELLA CERIANI[1], HUBERT THIERENS[2], ANNE VRAL[2] and GIANPAOLO PERLETTI[1]

[1]Department of Structural and Functional Biology, Laboratory of Toxicology and Pharmacology, Università degli Studi dell'Insubria, Busto Arsizio, Varese, Italy; [2]Department of Basic Medical Sciences, Ghent University, Ghent, Belgium

Received July 3, 2008; Accepted September 12, 2008

DOI: 10.3892/mmr_00000032

Abstract. Previous results from our research group have shown that the c.-1310 C→G single nucleotide polymorphism in the promoter region of the *XRCC6/Ku70* gene is significantly associated with breast cancer in a sample human patient population. In an attempt to attribute a functional meaning to this polymorphism, we performed a thorough analysis using a number of established *in silico* tools that strongly suggested that the c.-1310C→G transversion would activate a cryptic splicing acceptor located upstream of the canonical promoter of *Ku70*, but downstream of a putative alternative promoter (PAP) of the same gene. Experimental investigation of alternative transcripts, as well as of the activity of the PAP detected *in silico*, did not support the initial hypothesis of a functional role of the c.-1310C→G mutation in alternative splicing. Although a functional role of the SNP has yet to be determined, some evidence points to the linkage disequilibrium of the G variant of the polymorphism, with mutations located at critical sites within the promoter region of *Ku70*.

Introduction

Breast cancer is a major public health issue worldwide. Globally, it is by far the most frequent cancer in women, with an estimated 1.15 million new cases in 2002 (23% of all cancers), and more than half of these cases occurring in the industrialized countries of Europe and North America (1). In Europe in 2006, breast cancer accounted for 28.9% of all cancer cases in women, and was the leading cause of cancer-related death within the European Union (16.7%) (2).

In several independent studies also conducted by our group, an enhanced *in vitro* chromosomal radiosensitivity was observed in a significant number of breast cancer patients (3-7). Since the most detrimental form of radiation-induced DNA damage is the double-strand break (DSB), it is plausible that DSB-initiated chromosomal instability drives breast carcinogenesis. Furthermore, the key breast cancer susceptibility genes, BRCA1 and BRCA2, ATM and TP53, play important roles in DSB repair and chromosome stability (8,9). These genetic factors are present in only 6-11% of the general breast cancer patient population, indicating that other mutations in low penetrant genes or subtle defects arising from low penetrant variations in highly penetrant genes may also create a predisposition to breast cancer (9-12). Recently, several population-based case-control studies have shown a link between single nucleotide polymorphisms (SNPs) in DSB repair genes and breast cancer risk (8,9,12-17). Moreover, the breast is a selected microenvironment, subjected to endogenous oxidative stress through hormone exposure. Estrogen in particular has attracted considerable attention, as it may act as a complete carcinogen (8). During the oxidative metabolism of estrogen, reactive oxygen species (ROS) and DNA adducts are formed, leading to the oxidation and depurination of the DNA (18). This type of damage can further result in clustered sites of DNA damage, including DSBs. DSBs are also induced in proliferating cells when the replication fork encounters single-stranded DNA damage (19). Thus, estrogen plays a dual role as a growth-stimulating hormone and strand break-inducing agent, which may explain why DSB repair is so important in breast tumorigenesis.

The two main mechanisms of DSB repair are homologous recombination (HR) and non-homologous end-joining (NHEJ) (reviewed in ref. 19). Which pathway is activated in eukaryotic cells depends primarily on the cell cycle stage. The more accurate HR pathway is activated in S/G_2, when the sister chromatid provides a perfect copy of the sequence for alignment, while the more error-prone NHEJ pathway is the favoured mechanism in resting cells (G_0/G_1) (20). NHEJ is

Correspondence to: Dr Gianpaolo Perletti, Dipartimento di Biologia Strutturale e Funzionale, Università degli Studi dell'Insubria, Via A. Da Giussano 12, I-21052 Busto A, Italy
E-mail: gianpaolo.perletti@uninsubria.it

Abbreviations: *Ku70*, *XRCC6/Ku70* gene; LD, linkage disequilibrium; PAP, putative alternative promoter; SNP, single nucleotide polymorphism

Key words: XRCC6, *Ku70*, non-homologous end joining, splicing, alternative promoter

also the main mechanism for repairing radiation-induced DSBs in mammalian cells (21). Defects in this pathway result in enhanced chromosomal radiosensitivity, which is a hallmark for breast cancer. In NHEJ, broken DNA termini are first processed to make them compatible and then sealed by a ligation step, which often results in the loss of a few nucleotides at the broken ends. The key protein components of NHEJ (20) include the catalytic subunit of DNA protein kinase (DNA-PK_{CS}), the two regulatory subunits of the DNA-PK complex Ku70 and Ku80, DNA ligase IV with its cofactor XRCC4 (the X-ray cross complementing group 4 protein) and the nuclease artemis. The Ku70/Ku80 (Ku) heterodimer is the first protein to bind to the damaged ends of a DSB. When bound to DNA, Ku recruits and activates DNA-PK_{CS}. As these proteins play a prominent role in DNA DSB repair, they are vital for genome stability and act as tumour suppressors. Severe defects in these genes result in a high level of genetic instability and lead to cell death, triggered by cell cycle checkpoint surveillance. However, if these genes harbour small variations, such as SNPs, which result in mild defects that escape checkpoint surveillance, the resulting suboptimal DNA DSB-repair could allow the accumulation of DNA damage, possibly leading to chromosomal instability and tumorigenesis (12).

In our previous studies, we investigated the association between SNPs in NHEJ genes and breast cancer susceptibility. A positive association was found between the variant allele of the c.-1310 C→G SNP in *Ku70* and breast cancer, with a significant odds ratio (OR_{He} = 1.85; p=0.048) observed in sporadic breast cancer patients. In familial breast cancer patients the positive OR was not significant, which could indicate that other environmental factors, besides genetic aptitude, modify the association between the c.-1310 C→G SNP in *Ku70* and breast cancer (17). The possible effect of estrogen exposure was examined in an enlarged unselected patient population. A significant OR of 1.69 (p=0.017) was found for the total patient population. By dividing the patient population into smaller groups based on estrogen exposure parameters, high and significant ORs were observed for patients with a longer estrogen exposure (early age at first menarche, OR_{He+HV} = 1.84, p=0.008; late age at menopause, OR_{He+HV} = 1.96, p=0.029; long exposure interval, OR_{He+HV} = 1.81, p=0.035), while the ORs of patients with a short estrogen exposure were lower and non-significant (Willems *et al*, unpublished data). Besides being involved in breast cancer predisposition, the c.-1310 C→G SNP in *Ku70* has also been implicated in *in vivo* chromosomal radiosensitivity after radiation therapy in head and neck cancer patients. Carriers of the variant G allele show a 4-fold increased risk of developing severe dysphagia (Werbrouck *et al*, unpublished data).

These results suggest that the c.-1310 C→G SNP located in the promoter region of *Ku70* might influence DSB repair, and prompted us to investigate the possible functional role of this mutation at the molecular level using *in silico* and *in vitro* molecular techniques.

Materials and methods

Cell cultures. The EBV-immortalized lymphoblastic cell lines C002, C003 and C025 from healthy donors, as well as a panel of an additional 23 cell lines all homozygous for the 'C' variant of the c.-1310C→G SNP, were obtained from the cultured cell bank of the Department of Basic Medical Sciences, Ghent University, Ghent, Belgium. The EBV-immortalized GM12802 cell line, homozygous for the 'G' variant of c.-1310C→G SNP, was purchased from Coriell Cell Repositories, NJ, USA. This panel of lymphoblastoid cell lines was used to assess chromosomal radiosensitivity by cytogenetic tests, to amplify specific sequences within the promoter region of *XRCC6/ Ku70*, and to screen the transcript variants of *Ku70*. The COS-7 cell line and the HEK 293 cell line (Istituto Zooprofilattico, Brescia, Italy) were used to test promoter sequences with the DLR Luciferase Reporter Assay System (Promega, Italy). Lymphoblastoid cell lines were maintained in RPMI-1640 medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin and 10% FBS. COS-7 and HEK 293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin, 1 mM sodium pyruvate and 0.1 mM MEM Non-Essential Amino Acids Solution. All cell lines were grown at 37˚C in 5% CO_2 in a humidified incubator.

RT-PCR analysis. Total RNA was isolated from 10^7 lymphoblastoid cells using the SV Total RNA Isolation System (Promega) according to the manufacturer's instructions. A 2- to 3-μg aliquot of total RNA was used to generate single-stranded cDNA and to amplify transcript fragments using the RobusT II RT-PCR Kit (Finnzymes, Italy) under the following conditions: 48˚C for 30 min followed by 94˚C for 2 min (first-strand cDNA synthesis), 35 PCR amplification cycles (94˚C, 30 sec; 60˚C, 60 sec; 72˚C, 60 sec) and a final extension step at 72˚C for 7 min. To detect hypothetical transcript variants, primers were designed as described below. To amplify a normal transcript fragment of *XRCC6/Ku70* (NCBI accession no. NM_001469.3), the following primers were used: primer 2 (5'-ACACTGGATGCTCA-3'), a common reverse primer that annealed a specific sequence within *Ku70* exon 2, and forward primer 1 (5'-CATGCGTGGATTGTC-3') to give an amplified product of 264 base pairs. To detect fragments generated by alternative splicing, we designed forward primer 3 (5'-GCG CCACCCTCTG-3'), forward primer 4 (5'-ATGTGAAGC CAAATA-3') and forward primer 5 (5'-TCCTGGCCATGT GT-3') to be used with reverse primer 2. RT-PCR products were analyzed on 2.5% agarose gel, and the fragments of interest were analyzed by enzymatic restriction digestion and automated sequencing.

Construction of expression vectors. Human genomic DNA was purified from 10^7 immortalized lymphoblasts using the Jetquick Blood & Cell Culture DNA System (Genomed GmbH, Germany) according to the manufacturer's instructions. For the construction of reporter plasmids containing the canonical sequence promoter of human Ku70 (http://genomics.senescence.info/genes/seq.php?id=0117&type=promo), 2 μg genomic DNA was used as a template for amplification of a region of 748 bp spanning from position -1418 to -670, upstream of the translation start site of *XRCC6/Ku70* (5'-AGA TGGCCCAAACTTCAGACC-3'; 5'-GGAAGCGACCAACTT

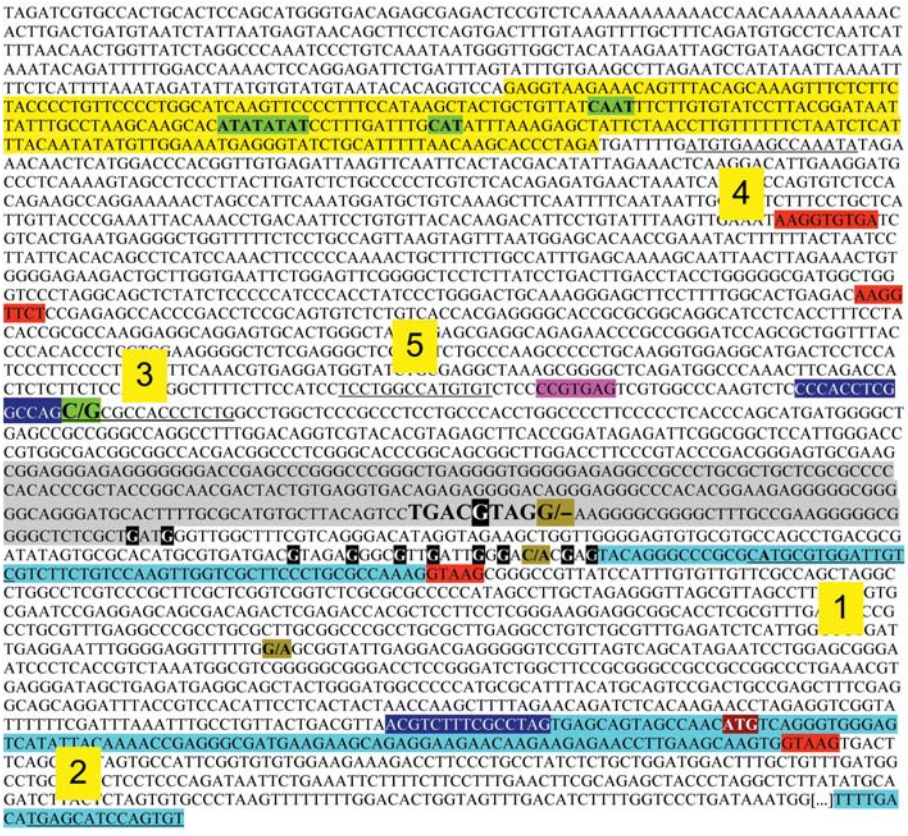

Figure 1. The promoter region of *XRCC6*/*Ku70* (chromosome 22, accession ref. no. NC_000022.9 NC_000022:40347241-40389998). The canonical *XRCC6*/*Ku70* promoter is represented in grey; the *Ku70* putative alternative promoter (PAP) in yellow. Putative TATA box, CAAT and CAT motifs within the PAP are shown in green. The 5'-UTR and the first exon of Ku70 are shown in light blue (ATG highlighted). Natural and putative splicing acceptor sites are shown in dark blue; the polymorphic nucleotide at position -1310 is highlighted in green, with a bold enlarged font. A putative splice acceptor branch site is shown in magenta. Functional and putative splicing donors are highlighted in red. The CREB (*C-fos*) transcription factor binding motif, altered by the G/- mutation in linkage disequilibrium (LD) with the c.-1310C→G SNP, is shown in a bold enlarged font. Transcription start sites (TSS) of published Ku70 mRNAs or ESTs are shown in black boxes. The TSS of the EST DB181025, likely generated by a putative alternative promoter (28), is highlighted within the CREB (*C-fos*) binding motif. Primer sequences used to analyze, by RT-PCR, the presence of alternative aberrant transcripts are underlined and numbered. Mutations in LD with the c.-1310C→G transversion, described in the NCBI LD Map, are highlighted in brown.

GGACA-3' forward and reverse primers, respectively). Similarly, a putative alternative promoter (PAP) region of 601 bp, spanning from position -2814 to -2213 upstream of the transcription start site of *XRCC6*/*Ku70,* was amplified (5'-AGG CCCAAATCCCTGTCAAATA-3' forward primer; 5'AGT TCATCTCTGTGAGACGAGGG-3' reverse primer). All sequences were generated using 0.02 U Fusion High-Fidelity DNA Polymerase (Finnzymes) under the following conditions: 98˚C for 1 min followed by 35 amplification cycles (98˚C, 30 sec; 62˚C, 1 min; 72˚C, 30 sec) and by a final extension step at 72˚C for 10 min. The fragments containing both the canonical promoter and the PAP were subcloned into the *Sma*I site of pBluescript II SK(-) (Stratagene, USA) and subsequently subcloned into the pGL3 basic vector (*Xho*I and *Hin*dIII sites) upstream of the firefly luciferase reporter gene, thus generating the pGL3/canonical_promoter vector and the pGL3/PAP vector. Correct cloning was confirmed by restriction enzyme mapping and sequencing.

Transient transfection and reporter gene assays. Transfection experiments were performed using COS-7 and HEK 293 cell lines seeded into 12-well plates at 70% confluence. Twenty-four hours later, 1 μg of pGL3/canonical_promoter plasmid, pGL3/PAP plasmid or pGL3 control vector was co-transfected with 0.05 μg pRL-TK vector into COS-7 and HEK 293 cells in triplicate using Lipofectamine 2000 Reagent (Invitrogen, Milan, Italy) according to the manufacturer's protocol. Cells were lysed with 200 μl passive lysis buffer (Promega) 48 h after transfection. Cell lysates were vortexed and briefly centrifuged to sediment cell debris. A 20-μl aliquot of cell lysate was then assayed for luciferase activity using a Lumat LB 9507 EG&G luminometer (Berthold Technologies GmbH, Germany) and the DLR Luciferase Reporter Assay System (Promega) according to the supplier's recommendations. A pGL3-control-promoter vector was included in all assays. All experiments were repeated at least three times. Data were expressed as M1/M2: RLUs *firefly luciferase*/RLUs *Renilla luciferase* relative luciferase activity, defined as luciferase activity normalized to the activity of Renilla luciferase, used to correct for variability caused by differences in cell viability or transfection efficiency.

Results and discussion

In an attempt to ascribe a functional meaning to the c.-1310C→G SNP located at position 40 346 645 on chromosome 22, we ascertained from both the published literature (22) and from online SNP databases (http://snp.ims.u-tokyo.ac.jp/cgi-bin/

Table I. Detection of splice acceptor sites in a 511-bp region surrounding the c.-1310C>G transversion.

Polymorphic variant: -1310C Total acceptor sites detected: 7						
Position	Sequence, acceptor site	Score	Relative position	Sequence, branch site	Score	Total splice site score
16	CACCCTCGTCGAAG|G	77.7	-20	TTTAC	72.8	150.5
152	GGCCCAAACTTCAG|A	72.7	-29	CTAAA	89.2	161.9
303	TCCCCCTCACCCAG|C	83.1	-25	CCCAC	67.8	150.9
330	AGCCGCCGGGCCAG|G	69.5	-33	CTCAC	92.1	161.6
342	AGGCCTTTGGACAG|G	77.2	-26	CTGAG	88.6	165.8
355	GGTCGTACACGTAG|A	67.8	-39	CTGAG	88.6	156.4
357	TCGTACACGTAGAG|C	66.1	-41	CTGAG	88.6	154.7

Polymorphic variant: c.-1310G Total acceptor sites detected: 8						
Position	Sequence	Score	Relative position	Sequence, branch site	Score	Total splice site score
16	CACCCTCGTCGAAG|G	77.7	-20	TTTAC	72.8	150.5
152	GGCCCAAACTTCAG|A	72.7	-29	CTAAA	89.2	161.9
242	**CCCACCTCGGCCAG|G**	**81.1**	**-32**	**GTGAG**	**67.0**	**148.1**
303	TCCCCCTCACCCAG|C	83.1	-25	CCCAC	67.8	150.9
330	AGCCGCCGGGCCAG|G	69.5	-33	CTCAC	92.1	161.6
342	AGGCCTTTGGACAG|G	77.2	-26	CTGAG	88.6	165.8
355	GGTCGTACACGTAG|A	67.8	-39	CTGAG	88.6	156.4
357	TCGTACACGTAGAG|C	66.1	-41	CTGAG	88.6	154.7

SnpInfo.cgi?SNP_ID=IMS-JST020860, linked from the NCBI SNP database: http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2267437) that this mutation was attributed to the promoter region of the *XRCC6*/*Ku70* (Ku70) gene (NM_001469.3).

Since the c.-1310C→G SNP was not located within an exon but rather within the promoter region of *Ku70*, we hypothesized that this mutation might be involved in the alteration of the transcription or splicing processes. Examination of the DNA sequence surrounding the nucleotide at position -1310 suggested that the C→G transversion, being preceded by a 'CAG' motif and by a significant number of upstream pirimidines, might be involved in the activation of a cryptic splice acceptor site (Fig. 1).

To verify this hypothesis, we analyzed a region of 511 base pairs surrounding the SNP by means of the Alex Dong Li's Splice Site Finder, version 0.5 (http://violin.genet.sickkids.on.ca/~ali/splicesitefinder.html), based on a modified algorithm according to the matrix system of Shapiro and Senapathy (23). The program did not identify a splicing site generated in the region adjacent to the common 'C' variant of the polymorphism (Table I). However, when the sequence harboring the 'G' variant was inserted, the program identified a new splice acceptor sequence, showing a total score of 81.1 and a putative associated branch site lying 35 bases upstream of the mutant nucleotide (Table I, Fig. 1).

This finding suggests that the c.-1310C→G transversion might activate a cryptic splicing acceptor within the 5'-end of the *Ku70* gene. To investigate the characteristics of this cryptic splicing acceptor in greater detail, we analyzed the site by information analysis. Using Schneider's Lab tools (Molecular Information Theory Group, Center for Cancer Research Nanobiology Program, National Cancer Institute, Frederick, MD, USA; http://www.lecb.ncifcrf.gov/~toms/delila.html), we generated sequence walkers to compare the information content (and hence the 'strength') of the first natural splice acceptor of *Ku70* and of the putative cryptic acceptor generated by the c.-1310C→G transversion (Fig. 2).

Analysis showed that the first natural splicing acceptor of *Ku70* was characterized by relatively low information content (3.1 bits), whereas the cryptic site containing the 'C' variant of the polymorphism scored 2.3 bits. In the presence of the c.-1310C→G mutation, the splice site score was increased to 4.3 bits, a value higher than the one shown by the natural competitor site (3.1 bits). Since, according to the information theory, the minimum quantity of information required for splicing was approximately 2.4 bits (24), and since splicing likelihood was shown to be correlated to the amount of information displayed by competing splicing sites (25), our data predicted that the c.-1310C→G mutation might be responsible for the generation of alternatively spliced and possibly aberrant transcripts of the *Ku70* gene.

This hypothesis was supported by the evidence shown in Fig. 3, which demonstrated that the c.-1310C→G transversion would generate a Holliday splicing configuration characterized by optimal Watson-Crick base-pairing between the immature RNA and the U5 and U1 snRNP-RNAs (26).

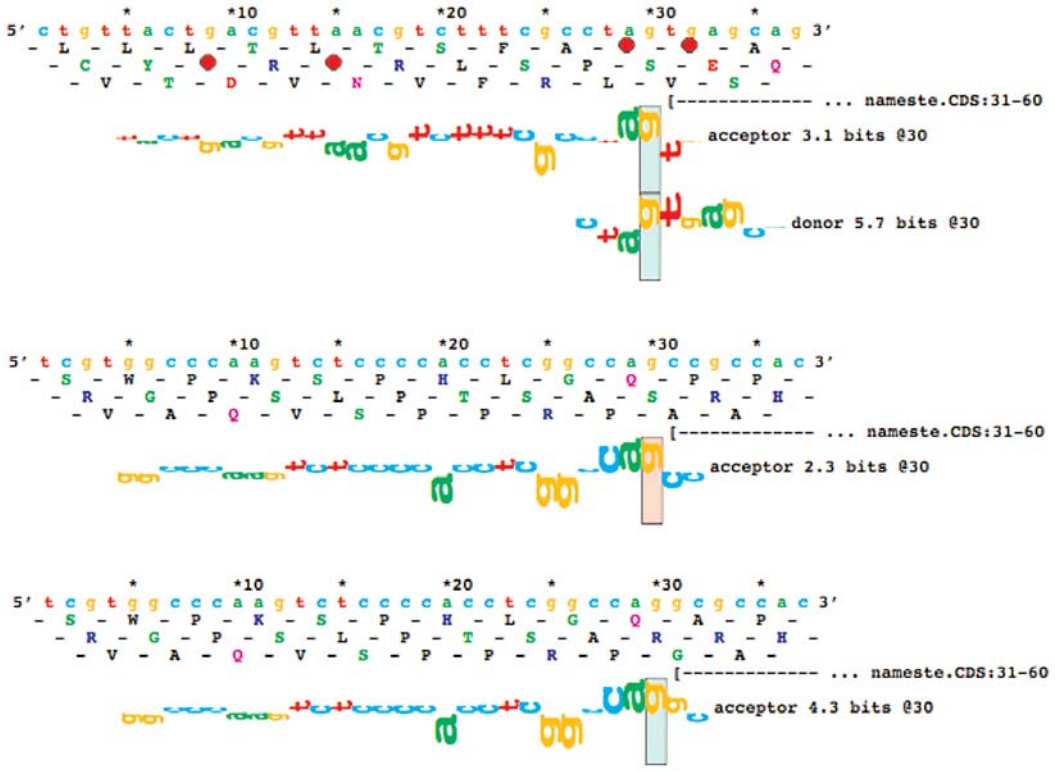

Figure 2. Lister map, sequence walkers and information content of the first natural splicing acceptor of *XRCC6/Ku70* (top). The information content of the putative cryptic splicing acceptor containing the 'C' variant of the SNP is shown in the middle sequence. Bottom, putative cryptic splicing acceptor activated by c.-1310C→G transversion. The boxed bases in the sequence walkers represent the last intronic guanine; the base immediately downstream is -1310C (middle sequence) or mutant -1310G (bottom sequence). Schneider's Lab Tools were used to produce the lister map and sequence walkers, and to calculate the information content of these splice sites.

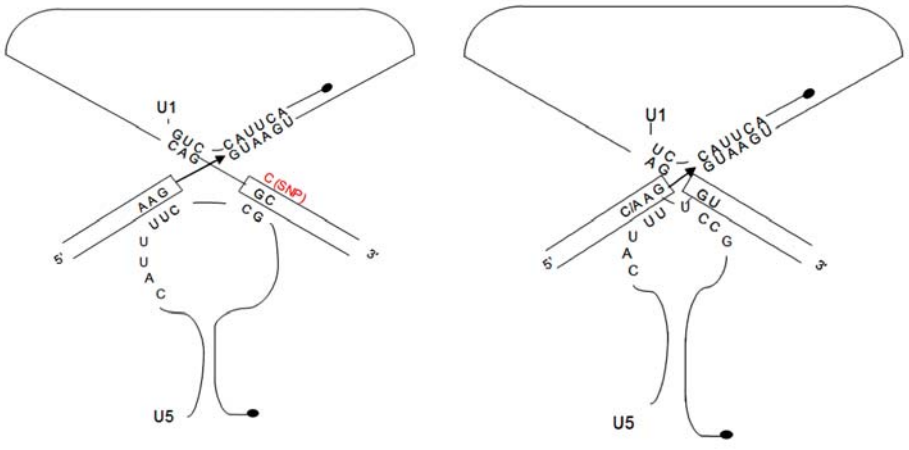

Figure 3. Schematic representation of the formation of a Holliday-like structure during assembly of the mammalian spliceosome. The c.-1310C→G transversion (left) generates a Holliday splicing configuration characterized by optimal Watson-Crick base-pairing between the immature RNA and the U5 and U1 snRNP-RNAs. The right panel shows base-pairing and interactions between RNA and U5/U1 snRNP-RNAs, reproduced according to the model described by Steitz (26).

Moreover, information analysis allowed for the identification of two putative donor sites, characterized by relatively high information content (7 bits; highlighted in Fig. 1).

To confirm these *in silico* findings, we performed RT-PCR analysis in cell lines homozygous for the 'C' or 'G' polymorphic variants in order to identify potential alternative fragments generated by aberrant splicing. Critical sequences were amplified using a common reverse primer (primer 2; for sequence see Materials and methods; highlighted in Fig. 1) annealing with a specific sequence within *Ku70* exon 2. Four different forward primers (see Materials and methods and Fig. 1), were designed to selectively detect aberrant fragments generated by alternative splicing in case the newly characterized SNP-activated splicing acceptor was functional. In particular, primers 1-2 were designed to amplify a normal 264-bp transcript fragment embracing the 5'-UTR and exon 1; primers 3-2, a 864-bp fragment embracing exon 1 and a newly formed exonic region downstream of the splicing acceptor

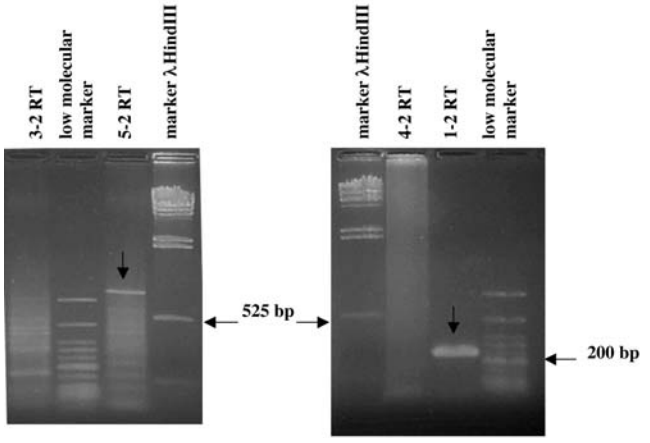

Figure 4. Agarose gel showing RT-PCR fragments obtained using different forward primers to detect alternative transcripts within the 5' region of the *XRCC6/Ku70* gene in the GM12802 cell line. The arrow in the lane showing the 5-2 fragment indicates the position of the 1-kb product of a putative transcript originating from an unknown promoter region upstream of the canonical promoter of *Ku70*; the arrow in the lane showing the 1-2 RT-PCR fragment indicates the position of a 'normal' fragment of 264 bp embracing the 5'-UTR and exon 1 of *XRCC6/Ku70*. In lane 3-2 and 4-2, aspecific amplicons are visible. These bands disappeared when the experiment was replicated.

generated by the c.-1310C→G transversion; primers 4-2, a 1922-bp fragment comprising exon 1 and a hypothetical 5'-UTR transcribed downstream of a putative promoter region upstream of the SNP (described in the following section) and primers 2-5, a 919-bp fragment comprising exon 1 and a putative intronic region located upstream of the hypothetic SNP-activated cryptic splice acceptor.

Whereas primers 1-2 allowed for the amplification of a 'normal' 264-bp sequence (Fig. 4), we were unable to detect in the GM12802 G/G homozygous cell line fragments suggesting the presence of alternative transcripts, indicating the activation of the cryptic splicing site at the site of the SNP. However, the presence of a 2-5 fragment of ~1,000 bp, detected by repeated RT-PCR reactions (Fig. 4), was suggestive of the presence of a transcript originating from an unknown promoter region located upstream of the canonical promoter of Ku70, described by Hosoi *et al* (27).

Analysis of a *Ku70* EST database indicated the presence of longer transcripts, extending well upstream of the 'normal' mRNA clones of Ku70. An example is represented by EST DB181025 (http://www.genome.ucsc.edu/cgi-bin/hgTracks?hgsid=107862218&hgt.dummyEnterButton.x=0&hgt.dummyEnterButton.y=0&clade=vertebrate&org=Human&db=hg18&position=DB181025&pix=620&hgsid=107862218). The transcription start site of this EST lies 166 bases upstream of known '+1' sites of Ku70 (Fig. 1). This specific transcript of *Ku70* was identified in the context of a research project aimed at isolating new 5'-extended ESTs that would likely be generated by PAPs (28). In fact, it is estimated that 52% of human genes may be subject to regulation by a PAP (28).

This is of interest because the presence of an additional promoter region upstream of the TATA-less Sp1 cluster, acknowledged to canonically drive the transcription of *Ku70*, would justify the presence and activity of an alternative SNP-generated splicing site upstream of the Sp1 cluster.

Promoter Score: 100.87 (Promoter Cutoff = 53.000000)
Significant Signals:

Name	TFD #	Strand	Location	Weight
ICSbf	S00413	+	18470	1.359000
AP-2	S01936	+	18491	1.108000
TFIID	S01540	+	18582	1.971000
Ig_dc	S00035	+	18593	50.000000
Oct	S01150	+	18593	8.606000
dc_box	S00015	+	18593	17.211000
IgHC.2	S00814	+	18595	3.442000
Oct-factors	S00050	+	18595	1.721000
Ig_cd	S00033	-	18602	8.606000
Oct-factors	S01029	-	18602	3.442000
OTF-2A	S02131	-	18603	7.376000

Promoter Score: 67.66 (Promoter Cutoff = 53.000000)
Significant Signals:

Name	TFD #	Strand	Location	Weight
SIF	S02021	-	13	1.161000
OBP	S02129	-	21	1.434000
JCV_repeat_seq	S01193	+	33	1.427000
SDR_RS	S01561	+	60	1.554000
PuF	S02016	+	63	1.082000
AP-2	S01936	-	69	1.091000
AP-2	S00346	-	71	1.672000
UCE.2	S00437	+	75	1.278000
AP-2	S00346	+	77	1.355000
Sp1	S00802	+	77	3.292000
Sp1	S00978	-	82	3.361000
GCF	S01964	+	93	2.361000
(Sp1)	S00857	+	168	4.589000
T-Ag	S00974	+	169	1.086000
Sp1	S00979	+	169	6.023000
Sp1	S00326	+	169	3.129000
Sp1	S00978	+	170	3.013000
JCV_repeated	S01193	+	170	1.427000
Sp1	S00781	+	171	3.191000
Sp1	S00802	-	175	3.061000
Sp1	S00801	-	176	3.119000
EARLY-SEQ1	S01081	-	177	5.795000
(Sp1)	S01187	-	177	6.819000
AP-2	S01936	-	177	1.091000
Sp1	S00956	-	178	3.129000
ATF	S01059	+	213	1.157000
ATF	S01940	+	213	3.721000
CREB	S00489	-	217	1.147000
E4F1	S01256	-	218	1.201000
ATF/CREB	S00534	-	218	1.138000
CREB	S00144	-	219	2.549000
Sp1	S00645	+	224	12.906000
Sp1	S00064	+	224	10.681000
Sp1	S01542	+	224	6.661000
GCF	S01964	+	227	2.361000
APRT	S00216	-	233	7.604000
T-Ag	S00974	+	249	1.086000
NF-kB	S01498	+	249	1.080000

Figure 5. Analysis of the region upstream of the *XRCC6/Ku70* gene, performed using the Promoter Scan program (http://www-bimas.cit.nih.gov/molbio/proscan/).

To identify a PAP upstream of the c.-1310C→G SNP, we submitted a sequence spanning 4,000 bp upstream of the canonical promoter of *Ku70* to the Promoter Scan program (http://www-bimas.cit.nih.gov/molbio/proscan/) of the Center for Information Technology, National Institutes of Health. The program detected the 'canonical' TATA-less *Ku70* promoter, and assigned it a score of 67.66. Interestingly, the program also detected a promoter region 1,297 bp upstream

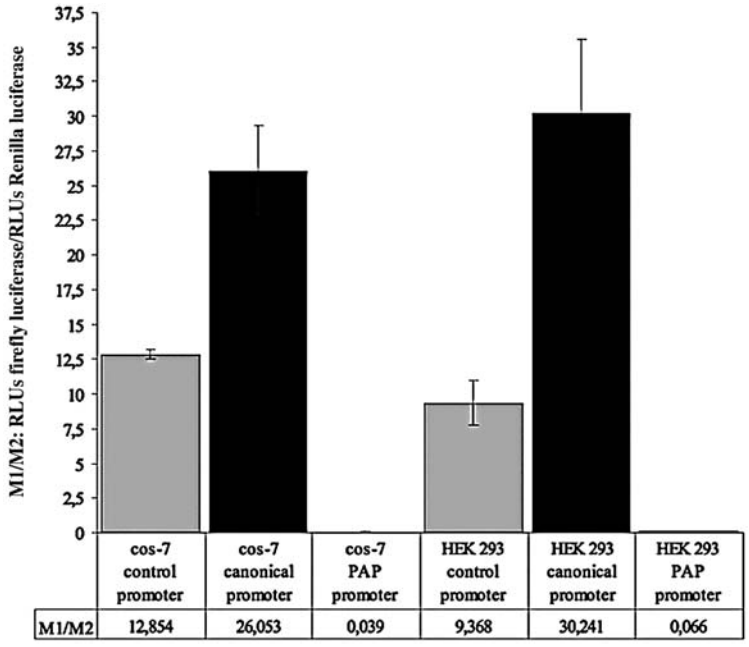

Figure 6. Analysis of firefly luciferase expression in COS-7 and HEK 293 cells, expressed as RLU/s *firefly luciferase*/RLU/s *Renilla luciferase* normalization ratio. The pGL3/canonical_promoter or pGL3/PAP plasmids, containing the *Ku70* canonical promoter or the PAP, respectively, or the pGL3 control vector, were co-transfected with the pRL-TK vector into COS-7 and HEK 293 cells; 48 h after transfection, cell lysates were assayed for luciferase activity using the Dual Luciferase Reporter Assay System as described in Materials and methods.

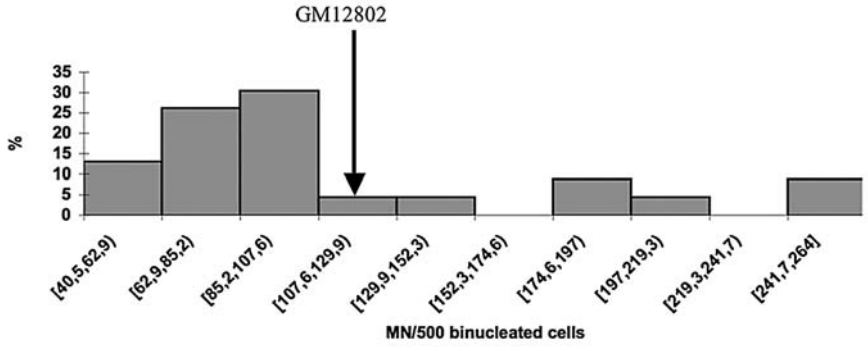

Figure 7. Frequency distribution of micronucleus formation (micronucleus-bearing cells/500 binucleated cells) in 23 EBV-immortalized control cell lines homozygous for the 'C' variant of the c.-1310C>G transversion in the *XRCC6*/*Ku70* gene. The micronucleus score of the GM12802 EBV-immortalized cell line, homozygous for the 'G' variant of the polymorphism, is indicated with an arrow.

of the previous one, to which the maximum score of 100 was assigned (Fig. 5). This was mostly due to the presence of a unique Ig_dc sequence, exclusively found in human gene promoters. Analysis of the newly detected PAP showed the presence of a putative TATA box preceded by a CAAT motif (~40 bp upstream of the TATA box) and followed (20 bp downstream) by a CAT motif (highlighted in Fig. 1).

To experimentally ascertain the activity of the newly found PAP, we amplified a fragment of 748 bp using PCR, embracing the canonical promoter of *Ku70* (27) and a 601-bp fragment corresponding to the PAP. Fragments were subsequently cloned upstream of a luciferase reporter gene within a pGL3 vector specifically designed for functional identification and characterization of promoter elements. Fig. 6 shows that, whereas the canonical *Ku70* promoter induced a strong expression of firefly luciferase, the putative Ku70 alternative promoter failed to act as a transcriptional activator of the reporter gene.

In summary, although we performed a very promising *in silico* analysis that suggested the existence of a relatively strong alternative splicing acceptor and pointed to the activity of a strong PAP of *XRCC6*/*Ku70*, experimental data ruled out the hypothesis of the involvement of the c.-1310C→G mutation in alternative/aberrant splicing of *Ku70* gene transcripts. Thus, care must be taken when translating data obtained *in silico* to experimental evidence.

Since Ku70 is a key protein involved in DNA DSB repair, we investigated whether the c.-1310C→G mutation would impair the function of Ku70 and in turn render the GM12802 G/G homozygous cell line more sensitive to DNA damaging agents. The cytokinesis-block micronucleus assay revealed a slight increase in radiosensitivity of GM12802 cells as compared to a panel of 23 homozygous (C/C) EBV-immortalized lymphoblastic cell lines (Fig. 7).

To explore an alternative role of the c.-1310C→G SNP in the regulation of Ku70 expression, we hypothesized that this

mutation might be in linkage disequilibrium (LD) with other mutations or SNPs involved in the regulation of gene expression. The existence of LD between the c.-1310C→G and other mutations in Ku70 has also been hypothesized by Tsai *et al (22)* and Bau *et al* (29).

Analysis of the LD maps published in the NIH-NCBI database (http://www.ncbi.nlm.nih.gov/SNP/GeneGt.cgi?geneID=2547) showed that the c.-1310C→G transversion is in strong LD with a number of SNPs located in the canonical promoter region of Ku70 (Fig. 1). In particular, the rs11342289 deletion (G/-) is placed between a CRE (TGACGT) and a SP1 site (GGGGCGGG, Fig. 1). Notably, the mutant guanine is part of a c-fos binding site (TGACGTAGG) (30), as shown in Fig. 1. The rare variant of this SNP (guanine deletion) could affect the generation of Ku70 transcripts.

The rs28384707 (C/A) SNP, in strong LD with the c.-1310C→G mutation, is located at the exact position where most Ku70 mRNA transcription start sites have to date been described (Fig. 1). This polymorphism may also affect the transcription efficiency of Ku70. Other SNPs in LD with the c.-1310C→G mutation (e.g., rs28384708, rs28384711), located in intronic regions, may also affect the expression of Ku70 in different ways. However, this does not exclude LD with additional, as yet unidentified, SNPs located within exonic regions. Research is underway to identify exonic SNPs throughout the Ku70 gene and to characterize the LD of these polymorphisms with the c.-1310C→G transversion or with other disease-associated mutations.

References

1. Parkin DM, Bray F, Ferlay J and Pisani P: Global cancer statistics, 2002. CA Cancer J Clin 55: 74-108, 2005.
2. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M and Boyle P: Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol 18: 581-592, 2007.
3. Baeyens A, Thierens H, Claes K, Poppe B, Messiaen L, De Ridder L and Vral A: Chromosomal radiosensitivity in breast cancer patients with a known or putative genetic predisposition. Br J Cancer 87: 1379-1385, 2002.
4. Baeyens A, van den Broecke R, Makar A, Thierens H, De Ridder L and Vral A: Chromosomal radiosensitivity in breast cancer patients: Influence of age of onset of the disease. Oncol Rep 13: 347-353, 2005.
5. Jones L, Scott D, Cowan R and Roberts S: Abnormal radiosensitivity of lymphocytes from breast-cancer patients with excessive normal tissue-damage after radiotherapy - chromosome-aberrations after low dose-rate irradiation. Int J Radiat Biol 67: 519-528, 1995.
6. Scott, D, Barber J, Levine E, Burrill W and Roberts S: Radiation-induced micronucleus induction in lymphocytes identifies a high frequency of radiosensitive cases among breast cancer patients: a test for predisposition? Br J Cancer 77: 614-620, 1998.
7. Scott, D, Barber J, Spreadborough A, Burrill W and Roberts S: Increased chromosomal radiosensitivity in breast cancer patients: a comparison of two assays. Int J Radiat Biol 75: 1-10, 1999.
8. Ralhan R, Kaur J, Kreienberg R and Wiesmüller L: Links between DNA double strand break repair and breast cancer: accumulating evidence from both familial and nonfamilial cases. Cancer Lett 248: 1-17, 2007.
9. Dapic V, Carvalho MA and Monteiro AN: Breast cancer susceptibility and the DNA damage response. Cancer Control 12: 127-136, 2005.
10. Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, Easton D, Evans C, Deacon J and Stratton M: Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst 91: 943-949, 1999.
11. Rebbeck T: Inherited genetic predisposition in breast cancer - a population-based perspective. Cancer 86 (Suppl S): 2493-2501, 1999.
12. Fu Y, Yu J, Cheng T, Lou M, Hsu G, Wu C, Chen S, Wu H, Wu P and Shen C: Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: A multigenic study on cancer susceptibility. Cancer Res 63: 2440-2446, 2003.
13. Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA and Easton DF: A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 8: 843-854, 1999.
14. Bau D, Fu Y, Chen S, Cheng T, Yu J, Wu P and Shen C: Breast cancer risk and the DNA double-strand break end-joining capacity of nonhomologous end-joining genes are affected by BRCA1. Cancer Res 64: 5013-5019, 2004.
15. Bau D, Mau Y, Ding S, Wu P and Shen C: DNA double-strand break repair capacity and risk of breast cancer. Carcinogenesis 28: 1726-1730, 2007.
16. Kuschel B, Auranen A, McBride S, Novik KL, Antoniou A, Lipscombe JM, Day NE, Easton DF, Ponder BAJ, Pharoah PDP and Dunning A: Variants in DNA double-strand break repair genes and breast cancer susceptibility. Hum Mol Genet 11: 1399-1407, 2002.
17. Willems P, Claes K, Baeyens A, Vandersickel V, Werbrouck J, De Ruyck K, Poppe B, van den Broecke R, Makar A, Marras E, Perletti G, Thierens H and Vral A: Polymorphisms in non-homologous end-joining genes associated with breast cancer risk and chromosomal radiosensitivity. Genes Chromosomes Cancer 47: 137-148, 2008.
18. Roy D, Cai Q, Felty Q and Narayan S: Estrogen-induced generation of reactive oxygen and nitrogen species, gene damage, and estrogen-dependent cancers. J Toxicol Environ Health B Crit Rev 10: 235-257, 2007.
19. Valerie K and Povirk L: Regulation and mechanisms of mammalian double-strand break repair Oncogene 22: 5792-5812, 2003.
20. Burma S, Chen BP and Chen DJ: Role of non-homologous end joining in maintaining genomic integrity. DNA Repair 5: 1042-1048, 2006.
21. Iliakis, G, Wang H, Perrault AR, Boecker W, Rosidi B, Windhofer F, Wu W, Guan J, Terdouzi G and Pantelias G: Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenet Genome Res 104: 14-20, 2004.
22. Tsai YY, Bau DT, Chiang CC, Cheng YW, Tseng SH and Tsai FJ: Pterygium and genetic polymorphism of DNA double strand break repair gene Ku70/XRCC6. Mol Vis 13: 1436-1440, 2007.
23. Shapiro MB and Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15: 7155-7174, 1987.
24. Rogan PK, Faux BM and Schneider TD: Information analysis of human splice site mutations. Hum Mutat 12: 153-171, 1998.
25. Khan SG, Levy HL, Legerski R, Quackenbush E, Reardon JT, Emmert S, Sancar A, Li L, Schneider TD, Cleaver JE and Kraemer KH: Xeroderma pigmentosum group C splice mutation associated with autism and hypoglycinemia. J Invest Dermatol 111: 791-796, 1998.
26. Steitz JA: Splicing takes a holliday. Science 257: 888-889, 1992.
27. Hosoi Y, Watanabe T, Nakagawa K, Matsumoto Y, Enomoto A, Morita A, Nagawa H and Suzuki N: Up-regulation of DNA-dependent protein kinase activity and Sp1 in colorectal cancer. Int J Oncol 25: 461-468, 2004.
28. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T and Sugano S: Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 16: 55-65, 2006.
29. Bau DT, Tseng HC, Wang CH, Chiu CF, Hua CH, Wu CN, Liang SY, Wang CL, Tsai CW and Tsai MH: Oral cancer and genetic polymorphism of DNA double strand break gene Ku70 in Taiwan. Oral Oncol, May 16, 2008 (Epub ahead of print).
30. Pearman AT, Chou WY, Bergman KD, Pulumati MR and Partridge NC. Parathyroid hormone induces c-fos promoter activity in osteoblastic cells through phosphorylated cAMP response element (CRE)-binding protein binding to the major CRE. J Biol Chem 271: 25715-25721, 1996.