Table 3: Exon-intron junctions of human (Hs) *XRRA1*, murine (Mm) *Xrra1* and rat (Rn) *Xrra1*

Exon No.		Exon Size (bp)	5' Splice Donor	Intron No.	Intron Size (bp)[a]	3' Splice Acceptor
1	Hs *XRRA1*	128	GGCTCAGGAG**gt**aaagccataa			ttgcactcac**ag**GGAGGCATCT
	Mm *Xrra1*	133	GGCTCAGGAG**gt**attgtcacag			ctgtgcccac**ag**GGAGGCATCC
	Rn *Xrra1*	133	GGCTCAGGAG**gt**attatcacag			ctgtgcccac**ag**GGAGGCATCT
2	Hs *XRRA1*	219	GACCGGACAG**gt**gagtgacacc	1	687	ccttaattga**ag**ACTGAAGAAG
	Mm *Xrra1*	206	GACCGCACAG**gt**gagctatggc		685	ccttaattga**ag**ACTGAAGAAG
	Rn *Xrra1*	206	GACCGCACAG**gt**gagcccgccc		-	tcttaattga**ag**ACTGAAGAAG
3	Hs *XRRA1*	43	AACCTCAGAG**gt**aggaactgac	2	43215	cttctgtttc**ag**AGGTTGTGTT
	Mm *Xrra1*	44	AACCTCAGAG**gt**aggaggacct		4118	tttctgtttc**ag**AGGTTGTGTT
	Rn *Xrra1*	44	AACCTCAGAG**gt**aggaggaacc		-	tttatgtttc**ag**AGGTTGTGTT
4	Hs *XRRA1*	127	CTACAACAAG**gt**gactttctgc	3	3696	cttctttcta**ag**ACAACCAAGA
	Mm *Xrra1*	126	CTACAACAAG**gt**aactgctttg		3706	tttcctctga**ag**ACAGCCAAGG
	Rn *Xrra1*	126	CTACAACAAG**gt**aaactgcttt		-	ttccctcctaagACAACCAAGG
5	Hs *XRRA1*	96	CATACACGAG**gt**atagtgcccg	4	7073	cattcattag**ag**ATCGCAAAAG
	Mm *Xrra1*	94	CACACACGAG**gt**atggtggcca		4393	cattcattag**ag**ATCGCAAAAG
	Rn Xrra1	94	CACACACGAG**gt**atggtggcca		-	tgcactttcc**ag**ATCGCAAAGG
6	Hs *XRRA1*	123	ATCATGGAAG**gt**agggcttcca	5	779	tcatggcacc**ag**GGGTCCCTCC
	Mm *Xrra1*	116	GTCACGGAAG**gt**aagccttcca		736	ccttcatacc**ag**GGATCCACAC
	Rn *Xrra1*	116	ATCGCGGAAG**gt**aagccttcca		-	ccctcacacc**ag**GGATCCCACC
7	Hs *XRRA1*	351	CCTGACCCAG**gt**acctgtatcc	6	2622	tctccatgac**ag**GTGAAAAGCG
	Mm *Xrra1*	324	CCTGACCCAG**gt**gacagccccc		2256	tctctcatcc**ag**GTGAAGACCT
	Rn *Xrra1*	324	CCTGACCCAG**gt**gacagccccc		-	tctctcatac**ag**GTGAAAACCT
8	Hs *XRRA1*	202	GAGCCAAAGG**gt**atgtgagggc	7	2839	tcctgggcac**ag**GTGAGTGAAC
	Mm *Xrra1*	196	GAGCCAAAGG**gt**atgggattaa		1421	ttcctggcac**ag**GTGAATGAGC
	Rn *Xrra1*	199	GAGCCAAAGG**gt**acgggattaa		-	ttcctggcac**ag**GTGAATGAGC
9	Hs *XRRA1*	117	AGAGAAACGG**gt**aaacatccag	8	788	ttcttgccct**ag**GGAATCCAGA
	Mm *Xrra1*	116	GGAGAAGAGG**gt**aagcacttca		594	tgcatcctct**ag**GCATCCAGAA
	Rn *Xrra1*	116	AGAGAAGAGG**gt**aagcacttca		-	tgcatcctcc**ag**GCATCCAGAA
10	Hs *XRRA1*	108	GCTCCACTAG**gt**acggctctgc	9	207	accttctgcc**ag**GCCCAGAGAA
	Mm *Xrra1*	106	GCTCCACTGG**gt**atggtgccag		169	tgcatcctct**ag**GCCGGGAGGA
	Rn *Xrra1*	106	GCTCCACTGG**gt**atggtgccag		-	cgtttgtgcc**ag**GTTGCGAGGA
11	Hs *XRRA1*	484		10	397	cttttcctgc**ag**GTGCTGTCCT
	Mm *Xrra1*	445			294	tcattcctgt**ag**GCACTGTGTT
	Rn *Xrra1*	445			-	tcatccctgt**ag**GCGCTGTGTG

[a]The intron sizes of Rn *Xrra1* gene could not be determined since the sequencing of rat genomic clone CH230-188L13 (GenBank AC127923) containing the gene was in progress. Nevertheless, their splicing sites can be identified, thus the exon sizes are also known.

Table 4: Percent identities of *XRRA1* cDNA and protein among several mammals

		Percent Identity (cDNA)*					
Percent Identity (Protein)*	ORGANISM	Hs *XRRA1*	Mf *XRRA1*	Mm *Xrra1*	Rn *Xrra1*	Sc *Xrra1*	Bt *Xrra1*
	HsXRRA1	-	***96.0***	***75.9***	***75.9***	***83.4***	***76.8***
	Mf XRRA1	95.0	-	***76.2***	***76.3***	***83.6***	***77.2***
	MmXrra1	71.8	74.1	-	***93.6***	***74.6***	***70.9***
	Rn Xrra1	69.6	71.7	90.2	-	***74.7***	***70.1***
	Sc Xrra1	77.7	79.4	67.7	62.6	-	***86.4***
	Bt Xrra1	68.3	69.5	62.0	61.6	70.1	-

*Hs, *Homo sapiens*; Mf, *Macaque fascicularis*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Sc, *Sus scrofa*; Bt, *Bos taurus*.

contrast, HCT116^{Clone10} cells down regulated *XRRA1* expression approximately ten minutes following XR. Interestingly, *XRRA1* expression in HCT116^{CloneK_XRS} cells was initially similar to HCT116^{Clone10} cells but regained the basal levels by 24 hours following XR.

XRRA1 was expressed predominantly in the testis in human and macaque

We carried out DNA microarray where PCR-amplified QtsA-20433 (Mf *XRRA1*) was included. Because we found that Mf *XRRA1* was more commonly present in testis than in brain cDNA libraries, we further investigated whether Mf *XRRA1* was predominantly expressed in the testis. The ratio of Mf *XRRA1* expression in testis to that in mixed tissues was 3.02. We also compared the expression of *XRRA1* between human and macaque testis. The ratio of *XRRA1* expression in human testis to macaque testis was 1.04. In addition, we examined Hs *XRRA1* expression from various normal human tissues/organs. We found that Hs *XRRA1* was expressed predominantly in testis followed by prostate and ovary (Fig. 4, Table 5). Other tissues/organs such as peripheral blood leukocyte, spleen, thymus, small intestine and colon demonstrated low expression of Hs *XRRA1*.

XRRA1 was expressed ubiquitously in various types of cancer cells

We evaluated the expression of Hs *XRRA1* in various cancer cell lines (neuroblastoma, glioma, breast, lung, leukemia, renal, ovarian, prostate, another colorectal). *XRRA1* was present in all of those cells, although the expression level was variable (Fig. 4, Table 5). We also detected *XRRA1* expression in immortalized normal fibroblast AG1522 and COS-7 cells, as well as in a dermal microvascular endothelial cell line (HDMEC). Surprisingly, besides mouse fibrosarcoma cells and tumor biopsy from a mouse model of fibrosarcoma, we were also able to detect Mm *Xrra1* expression in pluripotent cells such as murine embryonic stem cells R1 (Fig. 4, Table 5).

Over-expression of GFP-XRRA1 fusion protein was only achieved transiently and showed nucleo-cytoplasmic protein distribution

To understand further the possible function of the *XRRA1* gene, we fused the *XRRA1* cDNA immediately downstream of a green fluorescent protein (*GFP*) cDNA. We over-expressed the GFP-XRRA1 in HCT116 clones, but after 48 hours post-transfection the fluorescence clearly decreased both in the number of cells and intensity of fluorescence. We obtained similar results when we over-expressed GFP-XRRA1 in COS-7 cells. After three weeks selection in 500 μg/ml G418, very few COS-7 cells remained fluorescent (Fig. 5). Only transient expression of the GFP-XRRA1 was observed in both types of cells. The fluorescence of the GFP-XRRA1 did not show an exclusive nuclear localization (although the predicted XRRA1 protein contains NLS motif). Instead, the fluorescence appeared in most cells to be nucleo-cytoplasmic (Fig. 5).

Over-expressed COS-7 cells with GFP-XRRA1 might modulate Ku86 expression

Ku86 was exclusively localized in nuclei with or without XR (Fig 6). The over-expression of GFP-XRRA1 abolished nuclear immunostaining of Ku86 in COS-7 (Fig 6A). Immunostaining of α-Tubulin, as additional control for the experiment, did not change in the presence of GFP-XRRA1 (Fig 6B). The secondary antibody conjugated with Cyanine-3 (Cy3) was used for both Ku86 and α-Tubulin immunostaining.

Discussion

EBOA and *in silico* analysis helped us to identify a novel gene *XRRA1* from a previously unknown EST R40588 not only in human, but also in mouse, bovine, and rat. Recently, several clones that contained Hs *XRRA1* spliced variants have been deposited to GenBank database by others. Hs *XRRA1* clone from testis (GenBank accession no. BC037294) lacked exon 4; the one from spleen (GenBank accession no. AK074152) lacked exons 3, 4 and a possible partial deletion of exon 5, and the one from NT2 cells (GenBank accession no. AK056364) lacked exons 2, 3, 4, and 10 (Fig. 3A). The testis clone with GenBank no. BC037294 did not appear to have a proper open reading frame. EBOA and GrailExp methods failed to produce a start site further upstream of Hs *XRRA1* (GenBank no. BK000541).

Surprisingly, when we searched from approximately 63,400 macaque brain and testis cDNAs, we found only three testis clones that were homologous to Hs *XRRA1*, suggesting that the Mf *XRRA1* gene is expressed only at a very low level in brain. The Mf *XRRA1* is highly homologous with Hs *XRRA1*. Because the Hs *XRRA1* gene consists of 11 exons, it is likely that the Mf *XRRA1* gene also contains 11 exons (see Fig. 3A). In general, the *XRRA1* gene seems to be highly conserved among mammals (i.e. human, macaque, mouse, rat, pig, and bovine), suggesting that it has similar function(s). Moreover, we were unable to find any convincing evidence of *XRRA1* gene from fully sequenced genomic DNA of fish, fly, worm, yeast and microbes. This suggests that *XRRA1* might be specific for, or evolutionary-enriched in, mammals. A specific role(s) for *XRRA1* in mammals was further supported by the observation that the predicted protein sequences for motifs such as LRR, PEST, and NLS, as well as Tyr phosphorylation sites, are preserved among the above-mentioned mammals. Additionally, these motifs suggested that the *XRRA1* protein may interact dynamically with other proteins, or might be subjected to rapid degra-

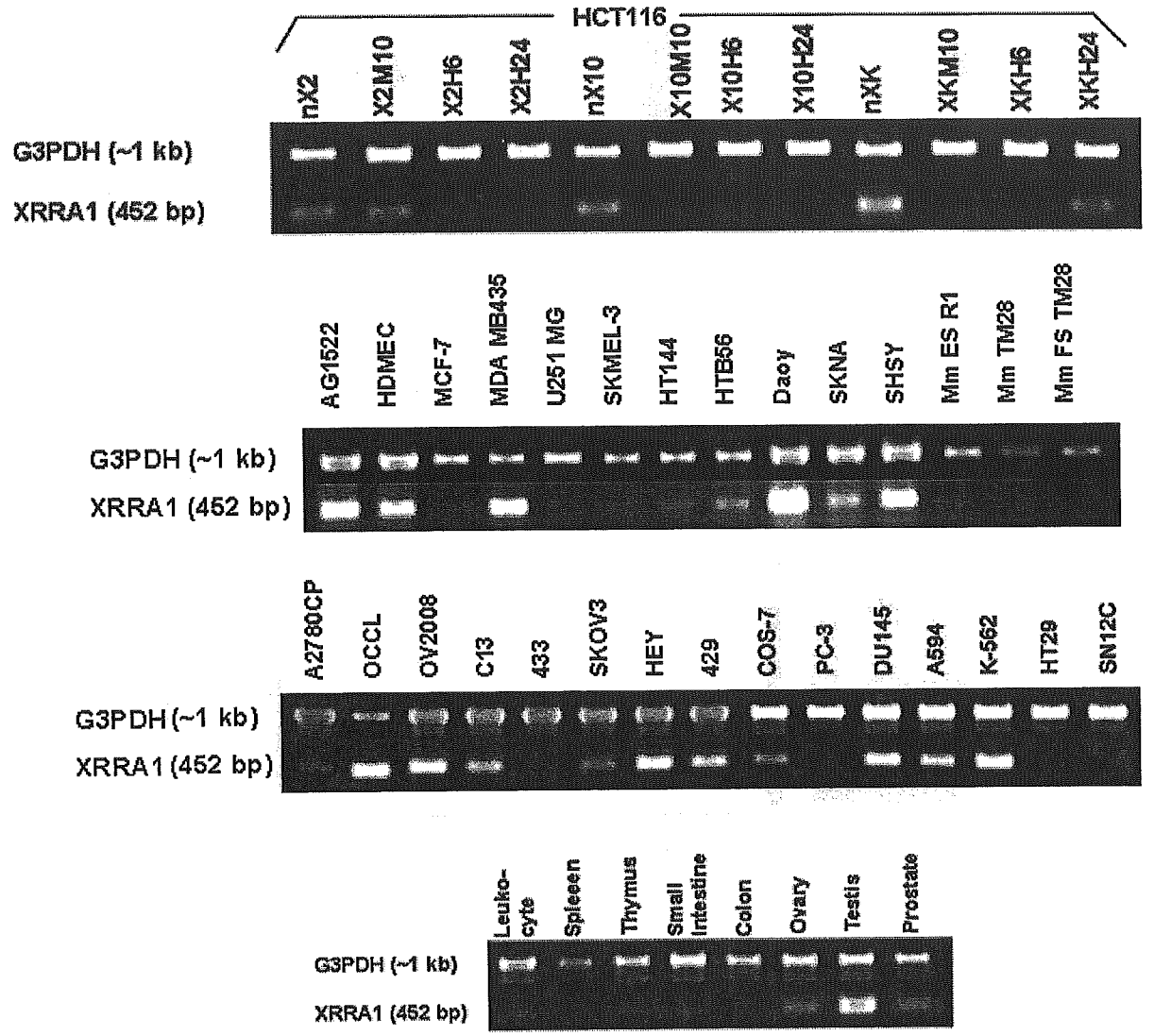

Figure 4
RT-PCR of *XRRA1* and a house-keeping gene *G3PDH* from HCT116 clones treated with X-radiation, various cancer and normal cells, and normal tissues/organs. The same set of primers was used to amplify corresponding mRNA from mouse embryonic stem cells and fibrosarcoma cells/tumor.

dation via the proteasomal pathway, or might suggest a possible role for *XRRA1* in cellular signaling.

Although the protein contains a NLS motif, it appears that XRRA1 is not exclusively located in the nucleus, as shown by the over-expression of the GFP-XRRA1 fusion protein. HCT116 clones and COS-7 cells containing the fusion protein (either with full length or truncated XRRA1 protein) had lost the auto-fluorescence. This suggested that the protein concentration above the basal expression levels of XRRA1 might be lethal to the cells.

It is noteworthy that the expression of *XRRA1* in HCT116 clones was variably affected by 4 Gy XR. The results sug-

Table 5: Differential expression of human *XRRA1* in XR-treated HCT116 clones, in normal tissues/organs and cells, as well as in various cancer cell types.

Cells/Tissues/Organs	Relative Expression Level (Normalized against *G3PDH*)*	Cell/Tissue/Organ type; Cancer Type
HCT116^{Clone2_XRR}	+++	Colorectal, more resistant to X-radiation than HCT116
HCT116^{Clone10}	++++	Colorectal, similar radiation response to HCT116
HCT116^{CloneK_XRS}	++++++	Colorectal, more sensitive to X-radiation than HCT116
HCT116^{Clone2_XRR}	+++	Without X-radiation
HCT116^{Clone2_XRR} M10	++++	10 Minutes after X-radiation
HCT116^{Clone2_XRR} H6	++	6 Hours after X-radiation
HCT116^{Clone2_XRR} H24	+	24 Hours after X-radiation
HCT116^{Clone10}	++++	Without X-radiation
HCT116^{Clone10} M10	+	10 Minutes after X-radiation
HCT116^{Clone10} H6	+	6 Hours after X-radiation
HCT116^{Clone10} H24	+	24 Hours after X-radiation
HCT116^{CloneK_XRS}	++++++	Without X-radiation
HCT116^{CloneK_XRS} M10	+	10 Minutes after X-radiation
HCT116^{CloneK_XRS} H6	+	6 Hours after X-radiation
HCT116^{CloneK_XRS} H24	++++	24 Hours after X-radiation
Testis	++++++++	Normal
Prostate	+++++	Normal
Ovary	+++++	Normal
Leukocyte	++	Normal
Spleen	+	Normal
Thymus	++	Normal
Small Intestine	++	Normal
Colon	+	Normal
AG1522	++++	Normal fibroblast (immortalized)
COS-7	+++	Normal fibroblast (immortalized)
HDMEC	++++	Normal microvascular endothelial (primary)
MCF7	+	Breast cancer
MDA MB435	+++++	Breast cancer
U251MG	+	Glioma
SKMEL-3	+	Melanoma
HT144	+	Melanoma
HTB56	+++	Lung cancer
A594	++++	Lung cancer
Daoy	++++	Neuroblastoma
SKNA	+++	Neuroblastoma
SHSY	++++	Neuroblastoma
A2780CP	+	Ovarian cancer, resistant to cisplatin
OCCL	++++	Ovarian cancer
OV2008	++++	Ovarian cancer
C13	+++	Ovarian cancer
433	+	Ovarian cancer
SKOV3	++	Ovarian cancer
HEY	++++	Ovarian cancer
429	+++	Ovarian cancer
PC-3	+	Prostate cancer
DU145	+++++	Prostate cancer
SN12C	+	Renal cancer
HT29	+	Colorectal cancer
K-562	+++++	Leukemia
Mm TM28	+++	(Murine) Fibrosarcoma
Mm FS-TM28	+	(Murine) Fibrosarcoma biopsy
Mm ES-R1	++	(Murine) Embryonic Stem Cell

*+, very low; ++, low; +++ moderate; >++++, high.

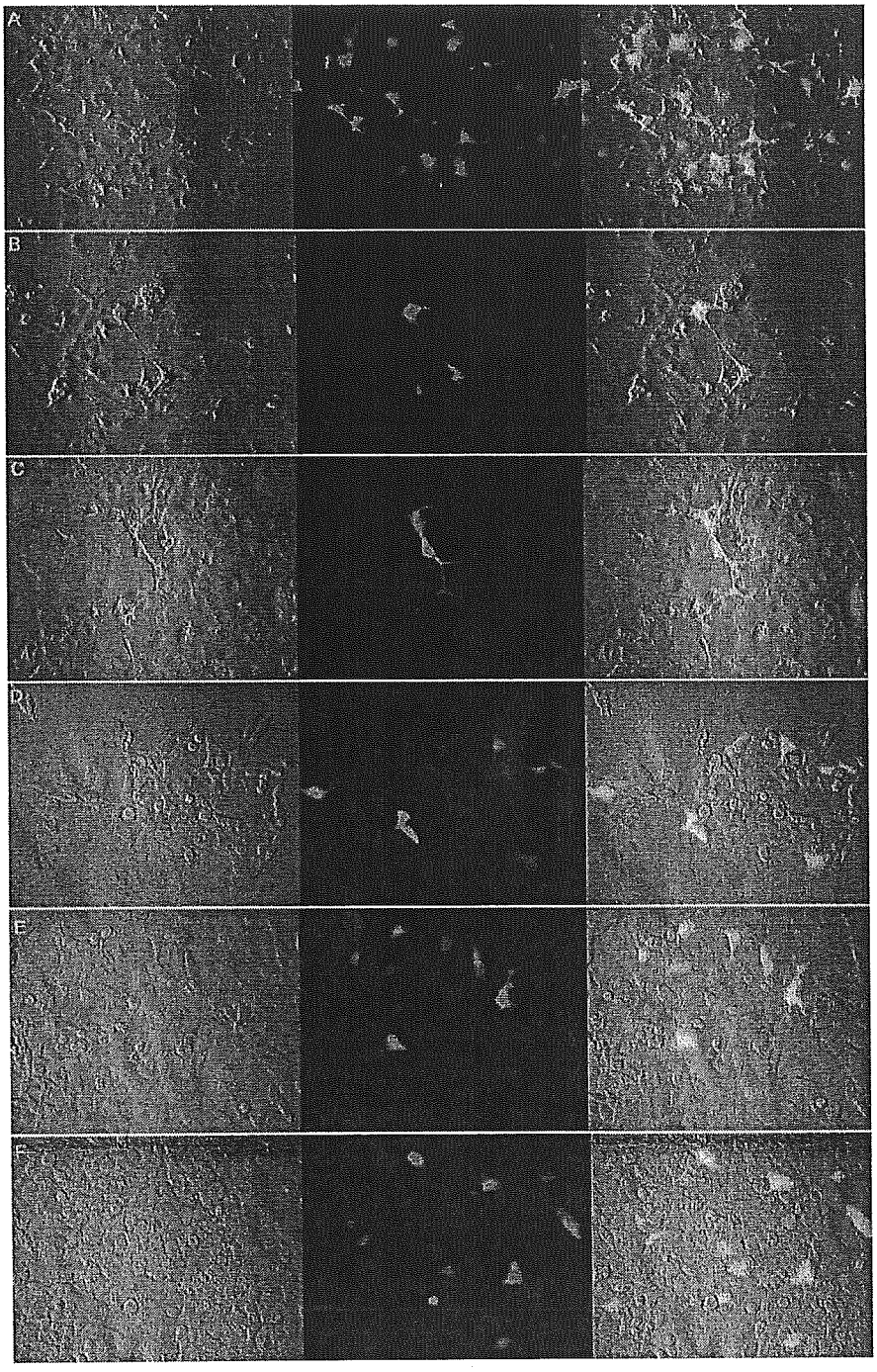

Figure 5
GFP-XRRA1 fusion protein in COS-7 and HCT116 clones. A, B, and C are transfected COS-7 cells. D, E, and F are transfected HCT116^{Clone2_XRR} cells. Control with GFP cassette alone is on panel A and D. Expression of GFP-truncated XRRA1 is shown in panels B and E. Expression of GFP-full length XRRA1 is shown in panels C and F. Transfected COS-7 cells were examined by fluorescence microscopy after 3 weeks of G418-selection. Expression of GFP-XRRA1 fusion protein was detected only transiently in HCT116 clones. The leftmost panel is a phase-contrast picture of the cells. The middle panel is green fluorescence shown with FITC filter. The rightmost panel is an overlay of the two previous panels. Magnification used was 400×.

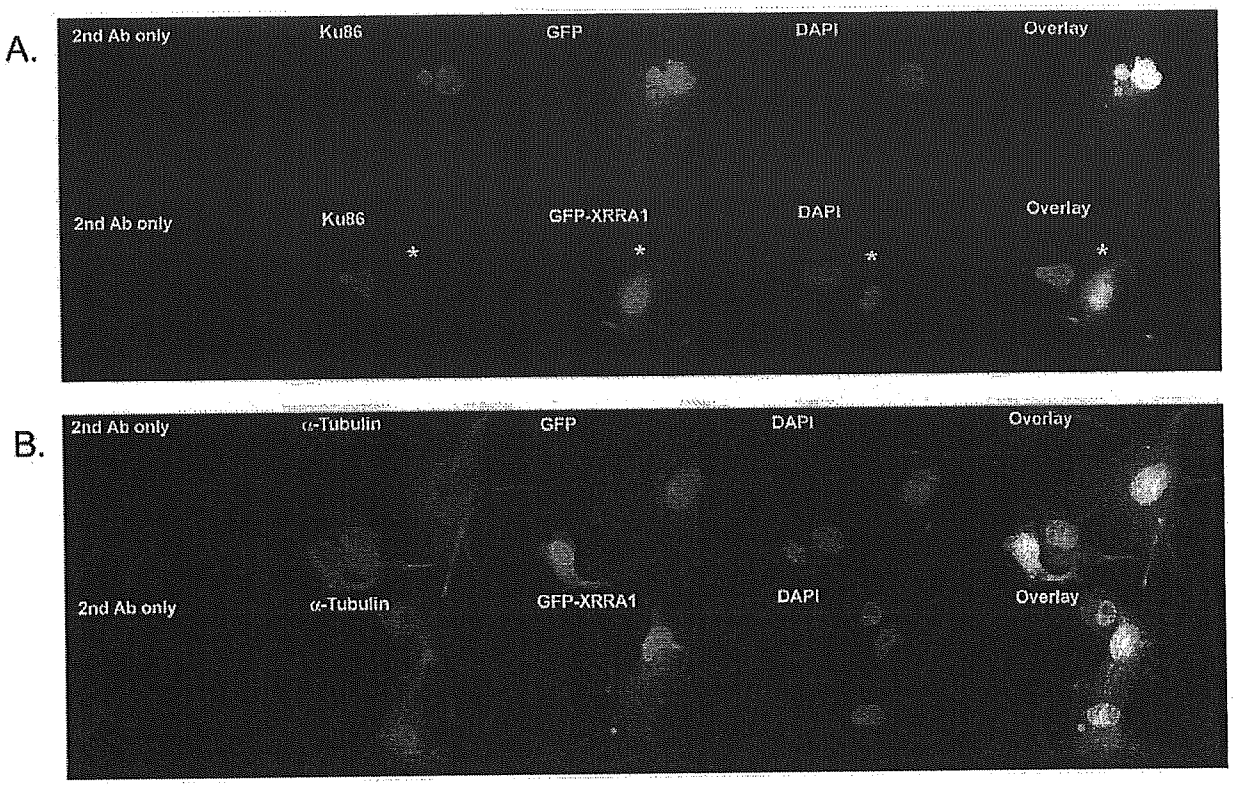

Figure 6
Immunocytochemistry of COS-7 cells over-expressing GFP-XRRA1 with Ku86 (A) and α-Tubulin (B). (A) Ku86 (red fluorescence) was localized in nuclei. Nuclear Ku86 immunostaining was eliminated (asterisk) in the presence of GFP-XRRA1 (green fluorescence), but not with the GFP alone. (B) Neither GFP nor GFP-XRRA1 changed the immunostaining of α-Tubulin (red fluorescence). DAPI-stained nuclei are shown in blue fluorescence. Magnification used was 1000×.

gested that the *XRRA1* gene might be involved in the immediate response following XR and might further be linked to XR resistance and/or sensitivity. Thus the return of *XRRA1* to almost pre-irradiated basal levels by 24 hours post-irradiation in radiosensitive HCT116^{CloneK_XRS} cells (but not in the other two clones) might be necessary for the subsequent manifestation of enhanced sensitivity to XR by this clone. One possibility is that the *XRRA1* gene may play a role either in the expression of cell death and/or signaling of DNA damage following XR stress. Therefore, genetic modulation of the expression of *XRRA1* in these clones may, specifically lead to either enhanced or decreased XR-induced lethality, and, generally, may be relevant for determining cellular response to XR in both tumor and normal cells.

In mammalian cells, DNA double strand breaks caused by XR are preferentially repaired by non-homologous end joining (NHEJ) mechanism. The major molecules involved in the NHEJ pathway are heterodimeric DNA end-binding complexes Ku86/Ku70 [14]. Therefore, we used Ku86 as the first candidate to test whether XRRA1 is involved in the modulation of the NHEJ pathway. The immunocytochemistry study on both HCT116^{Clone2_XRR} and COS-7 cells showed that there was no significant modulation of Ku86 with XR. However, COS-7 cells over-expressing GFP-XRRA1 were devoid of nuclear Ku86. This phenomenon suggested that over-expression of XRRA1 might have resulted in the down-regulation of Ku-86. Further studies are required to determine whether there is a negative correlation between XRRA1 and Ku86 expressions during cellular response to XR.

The varying expression of *XRRA1* following XR described above could be related to the role of transcription factors. In fact, we were able to detect a possible region between

nt -506 and -400 from the start codon for transcription factor binding sites that regulate the expression of *XRRA1*. These factors were C/EBPα, HSF-1, c-Jun, c-Fos, AP-1, and CREB. Interestingly, most of the previously mentioned transcription factors are members of the bZIP superfamily. The latter is associated with various functions, including mediating G1 arrests [15,16], regulating responses against environmental stresses including IR and UV [17], acting as an immediate-early response factor [18], controlling wide range plasticity processes [19], and serving as a promoter element that mediates transcriptional activation in response to increased levels of intracellular essential secondary messenger cAMP [20]. Thus, putative binding sites for these transcription factors further strengthen the suggestion that the expression of *XRRA1* was rapidly affected after XR treatment.

XRRA1 was expressed predominantly in the testis in both human and macaque. It may be relevant to note that the testis is generally regarded as one of the most sensitive organs to XR [21]. Thus upregulation of *XRRA1* in an XR-sensitive organ would appear to be generally consistent with our observation that it is downregulated in the XR-resistant HCT116^{Clone2_XRR} cells. Because there was comparable expression level of *XRRA1* in both human and macaque, it suggests that there is a functional necessity for *XRRA1* in mammalian testis (i.e. *XRRA1* might serve as a testis-specific molecule). By contrast, the low expression of *XRRA1* in healthy colon and relatively much stronger expression of *XRRA1* in HCT116 cells, may suggest an upregulation of this gene in cancer development. The ubiquitous expression of *XRRA1* in numerous cancer cells and in immortalized normal cells might be indicative of a role of this gene in tumor development.

Interestingly, from the SAGEmap, we identified three tags that specifically represent *XRRA1* anchored with restriction enzyme *Nla*III. The tags were TATTCAGGGG, ACCTGGTGCC, and GAATCAAGTG. SAGE libraries where *XRRA1* was less than 20 tags per million were excluded. We found that the *XRRA1* was present in mammary gland epithelium ductal *in situ* carcinoma, normal cerebellum, normal gastric body epithelium, well differentiated oligodendroglioma, pancreatic epithelium ductal adenocarcinoma, metastasis mammary gland carcinoma, ovarian clear cell poorly differentiated carcinoma, brain glioblastoma multiform cell line, ovarian serous adenocarcinoma, mammary gland ductal *in situ* high grade carcinoma, prostate carcinoma, brain juvenile ependymoma, and medulloblastoma cerebellum. The foregoing data further suggested that *XRRA1* might be correlated with carcinomas in breast, brain, prostate, pancreas and ovary. Finally, the expression of Mm *Xrra1* expression in mouse embryonic stem (ES) cells (R1) as well as in mouse fibrosarcoma cells and in fibrosarcoma biopsy suggest that *XRRA1* might function as an early expressed gene.

Conclusions

XRRA1 is a novel molecule that is expressed selectively in normal healthy tissues/organs. It is expressed at relatively higher levels in primary sex organs such as testis (of both human and macaque), prostate and ovary. The expression of *XRRA1* in normal proliferative cells, embryonic stem cells, and various tumor cells is also generally high and comparable to testicular expression of *XRRA1*. These findings suggest that *XRRA1* expression may be important for cell proliferation, development, and differentiation as well as carcinogenesis. In regard to XR, we found that *XRRA1* expression was rapidly modulated after treatment, suggesting the potential involvement of this gene/protein in the manifestation/expression of radioresistance/sensitivity by tumor cells such as human colorectal cancer cells. Over-expression of GFP-XRRA1 fusion protein was located in both nucleus and cytoplasm of HCT116 clones or COS-7 cells. A possible correlation with NHEJ pathway for repairing DNA double strand break after XR is shown by the lack of Ku86 immunostaining in COS-7 cells that over-expressed the GFP-XRRA1. Further experiments to determine the function of XRRA1 in DNA damage sensor, repair and apoptosis following radio- and or chemotherapy are ongoing. Also the function of XRRA1 in carcinogenesis as well as embryogenesis and reproduction biology needs to be elucidated.

Methods

Normal/Cancer cell cultures

Clones of colorectal cancer cells (HCT116^{Clone2_XRR}, HCT116^{Clone10}, and HCT116^{CloneK_XRS}), normal fibroblast cells (COS-7 and AG1522), breast cancer cells (MCF-7 and MDA MB435), glioma cells (U251 MG), melanoma cells (SKMEL-3 and HT144), lung cancer cells (HTB56 and A549), ovarian cancer cells (A2780CP and A2780S), prostate cancer cells (PC-3 and DU145), leukemia cells (K-562), colorectal cancer cells (HT29), renal cancer cells (SN12C), and mouse fibrosarcoma cells (TM28) were grown in DMEM/Ham' s F12 1:1 mix (Wisent Inc., St. Bruno, QC, Canada) supplemented with 10% fetal bovine serum (FBS) (Wisent Inc.) and 15 mM HEPES in a humidified atmosphere of 95% air, 5% CO_2 at 37°C. Various individuals from the ORCC, Ottawa, Canada, kindly provided us with the cells. They were: J.C. Bell (COS-7, PC-3, DU145, K-562, HT29, and SN12C cells), C.L. Addison (MCF-7 and MDA MB435 cells, neuroblastoma cells (Daoy, SKNA, and SHSY), normal human dermal microvascular cells (HDMEC)), HC Birnboim (TM28 cells), B. Vanderhyden (ovarian cancer cells OCCL, OV2008, C13, SKOV3, 433, HEY, and 429), and M. McBurney (mouse embryonic stem cells R1). H.C. Birnboim also donated a C57BL/6 mouse with a solid fibrosarcoma (TM28)

growing in the flank. A. Gatignon (Lady Davis Insitute, Montreal, Canada) kindly provided the U251 MG cells. Clones of HCT116, AG1522, SKMEL-3, HT144, A2780CP, and A2780S cells were from our own laboratory collection (C.E. Ng, ORCC, Ottawa, Canada).

Cloning and sequencing of Hs XRRA1 *splice variant*

A specific primer pair was created to amplify the 5' region of the *XRRA1* cDNA. They were 5'-GCGCTGGAGACACT-GATGCTGG ATGACAAC-3' and 5'-GGCCAGGCTAAGG-TATCTCAGCTCTGGG-3'. These primers generated a product of 452-bp that comprised the first 4 exons. Total RNA from HCT116^{Clone2_XRR} and HCT116^{Clone10} cells were used as template for RT-PCR. PCR fragments were gel purified using Gel Extraction Kit (Qiagen Inc., Mississauga, ON, Canada) and cloned into pTZ57R plasmid using InsT/Aclone PCR Product Cloning Kit (MBI Fermentas, Burlington, ON, Canada). Sequencing of the clones was performed by DNA Sequencer model 4000L according to the manufacturer' s protocol (LI-COR Biosciences, Lincoln, NE, USA) using M13 primer.

Construction of macaque cDNA libraries and DNA sequencing of their cDNA inserts

Oligo-capped cDNA libraries were made from macaque brain and testis. cDNAs were isolated according to the method described previously [22,23]. The 5'-end sequences of the clones were sequenced using ABI 3700 sequencer (Applied Biosystems, Tokyo, Japan) and categorized using DYNACLUST (Dynacom Co., Chiba, Japan) based on a BLAST search against the GenBank database. The entire sequences of clones were determined by the primer walking method with an ABI PRISM BigDye Terminator Sequencing kit (Applied Biosystems) according to the manufacturer' s instructions.

Construction of plasmid expressing Mf XRRA1 fused to GFP

Macaque cDNA clone of QtsA-20344 was found from the cDNA libraries to contain Mf*XRRA1* transcript. A 1709-bp *Xho*I fragment of Mf *XRRA1* was removed from pME18S-FL3 and sub-cloned into *Xho*I sites of pEGFP-C1 (BD Biosciences, Mississauga, ON, Canada), immediately downstream of the *GFP* sequence (pFM1709). To create a truncated Mf *GFP-XRRA1* (pFM584), 1.1-kb *Eco*RI fragment was removed from pFM1709 leaving a 584-bp of the 5' *XRRA1* cDNA fused with *GFP*.

Transfection of GFP-XRRA1 *into HCT116^{Clone2_XRR}, HCT116^{Clone10}, HCT116^{CloneK_XRS}, and COS-7 cells*

HCT116 clones and COS-7 cell were transfected with a mixture of pFM584, pFM1709 or pEGFP-C1 and lipid complexes FuGENE 6 (Roche Diagnostics, Laval, QC, Canada) transiently or stably. Briefly, HCT116 clones and COS-7 were grown on cover slips in a 6-well plate and reached 60–70% confluences before transfection. The ratio of FuGENE 6 and pure plasmid was 6:1 in serum free medium. G418 (Sigma-Aldrich, Oakville Ltd., ON, Canada) at a concentration of 400 μg/ml was used for selection. Cells were fixed with 3.7% formaldehyde in PBS for 10 min, and then washed three times with 1x PBS. Coverslips were mounted upside down onto slides with Dako Fluorescent Mounting Medium (DAKO, Carpinteria, CA, USA). Visualization was performed using a fluorescence microscope Axioskop 2 MOT (Carl Zeiss MicroImaging Inc., Thornwood, NY, USA). Images of cells were captured with a CCD camera fitted with a FITC filter.

Immunocytochemistry

Transfected COS-7 and HCT116^{Clone2_XRR} cells with pFM1709 or pEGFP-C1 were grown on cover slips until they reached 70–80% confluency before being treated with 4 or 10 Gy XR. Cells were fixed in 2% formaldehyde in 1x PBS for 15 min 24 h after XR treatment. Cells were then permeabilized with 0.2% Triton X-100 for 10 min. Ku86 antibody was obtained from Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA. The secondary antibody was conjugated with Cy3 (Amersham Biosciences Corp, Baie d' Urfe, QC, Canada). Controls for the immunostaining were the secondary antibody alone, and control for the experiment was α-Tubulin antibody that was obtained from Calbiochem, San Diego, CA, USA.

RT-PCR

Total RNAs of normal/cancer cells were extracted using RNAeasy kit (Qiagen Inc.) according to the manufacturer' s instructions. TRIZOL Reagent (Invitrogen Inc., Burlington, ON, Canada) was used to isolate RNA of mouse fibrosarcoma tumor biopsy. RNAs of human peripheral blood leukocyte, spleen, thymus, small intestine, colon, ovary, testis and prostate were purchased from BD Biosciences. First Strand cDNA Synthesis kit (MBI Fermentas) was used to create first strand cDNA from RNA samples. A specific primer pair that amplified ~1 kb house keeping gene glyceraldehyde-3-phosphate dehydrogenase (*G3PDH*) was used as the control. The primers were 5'-TGAAGGTCGGTGTGAACGGATTTGGC-3' and 5'-CATG-TAGGCCA TGAGGTCCACCAC-3'. Conditions of PCR were as follows: 94°C for 30 s, 61°C for 15 s, and 72°C for 40–60 s for 35 cycles. The PCR products were then separated on 1.5% agarose gels, stained with ethidium bromide, and visualized with the gel documentation imager Epi Chemi II (UVP Inc., Upland, CA, USA). Densitometry of bands were analyzed using Labworks™ software (UVP). Relative expression of *XRRA1* was calculated as percentages of the mean raw densities of grayscale level of *XRRA1* to *G3PDH*.

X-radiation of HCT116^{Clone2_XRR}, HCT116^{Clone10}, and HCT116^{CloneK_XRS} cells

Clones of HCT116 cells were cultured at 1 × 10^5 cells/ml. Cultures reached 70–80% confluence before being treated with XR. Each clone was treated with a single dose of 4 or 10 Gy using a 250 kVp X-ray unit (Pantak, CT, USA) at a dose rate of 150 cGy/min. Total RNAs were collected 5 minutes, 6 hours and 24 hours following the XR. Specific primers for Hs *XRRA1* as mentioned above were used. *G3PDH* was used as control. Relative expression of *XRRA1* to *G3PDH* was evaluated

cDNA microarray of Mf XRRA1 *from testis versus mixed tissues or human testis*

512 fully sequenced testis cDNA clones were amplified using 5'-CTTCTGCTCTAAAAGCTGCG-'3 as a forward primer and 5'-CGACCTGCAGCTCGAGCACA-'3 as a reverse primer. Successful amplification was confirmed by agarose gel electrophoresis. Approximately 300 μg /ml DNA in 2 × Solution-T reagent (Takara Bio Inc., Shiga, Japan) were printed on duplicate glass-slides with a GMS 417 arrayer (Genetic MicroSystems, Woburn, MA, USA). RNA was isolated with TRIZOL Reagent (Invitrogen K.K., Tokyo, Japan) and purified with Oligo-Tex (Takara Bio Inc.). Macaque testis RNA was labeled with Cy3-dUTP (Pharmacia K.K., Tokyo, Japan). A mixture of macaque RNA was also obtained from 10 tissues: brain, heart, skin, liver, spleen, renal, pancreas, stomach, small intestine, large intestine, and labeled with Cy5-dUTP (Pharmacia). We mixed equal amount of RNA of these ten tissues. In a separate experiment, RNA from human testis (BD Biosciences) was labeled with Cy5-dUTP. Both labeled RNAs were co-hybridized to DNA spots. After the hybridization and washing procedure, slides were scanned with ScanArray (PerkinElmer Life Sciences Co., Tokyo, Japan). Beta actin and/or total signal intensity was used as the control for hybridization. The methods and results of DNA microarray have been deposited in the Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo with the accession numbers GPL206, GSM2387, and GSM2388.

ESTs-based ORF assembling (EBOA) and other in silico *analysis*

In order to elucidate what gene R40588 might represent, we employed EBOA to obtain the full-length cDNA containing start-, stop-codon, and poly-A signal or a complete open reading frame (ORF). The sequence of R40588 was used as a template to do BLAST search for all possible overlapping ESTs from human dbEST http://www.ncbi.nlm.nih.gov/dbEST. These retrieved ESTs, either sense-antisense or 5'- and 3'-direction, were selected and were again used as template for searching other overlapping ESTs. To confirm the ORF candidates, (i) we carried out a BLAST search against full-length cDNAs that were available on public databases, and (ii) genomic sequences comprising the ORF candidates were analyzed with GrailEXP version 3.3 http://compbio.ornl.gov/grailexp[11]. Multiple alignments were done according to ClustalW method [24]. Transcription factors and their binding sites were predicted by employing the Alibaba2 version 2.1 program http://www.gene-regulation.com/pub/programs.html that used the TRANSFAC database http://transfac.gbf.de/TRANSFAC[25]. Possible tissues/organs expression of *XRRA1* was explored from the SAGEmap http://www.ncbi.nlm.nih.gov/SAGE[26]. Protein motifs were searched using software available on these web sites: http://www.expasy.ch/prosite[27], http://www.at.embnet.org/embnet/tools/bio/PESTfind[28], and http://www.cbs.dtu.dk/services/NetPhos[29].

Data deposition

The sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank databases under the accession numbers BK000541, AY163836, AB072776, BK000542, and BK000644, respectively. HUGO Gene Nomenclature Committee and MGD Nomenclature Committee have approved the names and symbols of the *XRRA1* and *Xrra1* genes.

List of abbreviations

Bt, *Bos taurus* (bovine); EBOA, ESTs-based ORF assembling; EST, expressed sequence tag; Hs, *Homo sapiens* (human), Gy, Gray; *Macaca fascicularis* (macaque); Mm, *Mus musculus* (mouse); ORF, open reading frame; Rn, *Rattus norvegicus* (rat); SAGE, serial analysis gene expression; Sc, *Sus scrofa* (pig); *XRRA1*, X-ray radiation resistance associated 1, XR, X-radiation.

Authors' contributions

FMM carried out the *in silico* analysis including the multiple sequence alignments, cloning of the splice variant, XR assay, cells cultures, RT-PCR of XR treated, various human/mouse cancer and normal cells/tissues/organs, construction and expression of the fusion protein, immunocytochemistry study, and drafted the manuscript. NO carried out the construction and sequencing of macaque testis and brain cDNA libraries, and cloning and cDNA microarray of Mf *XRRA1*. KH was funded to carry out the construction of cDNA libraries and cDNA microarray of macaque brain and testis, conceived of the Mf *XRRA1* study and the *in silico* analysis. QYL contributed technical expertise to the cDNA microarray analysis of HCT116 clones. CEN is the Principal Investigator and was funded to carry out the cDNA microarray of HCT116 clones, supervised this work, provided the technical radiobiological expertise and contributed to the writing of this manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Dr. A. Lagarde for assistance with the DNA sequencing. This work was supported in part by grants from the National Cancer Institute of Canada with funds from the Canadian Cancer Society (to CEN) and from the Ministry of Health, Labor, and Welfare of Japan (to KH).

References

1. Zhivotovsky B, Joseph B and Orrenius S: **Tumor radiosensitivity and apoptosis.** *Exp Cell Res* 1999, **248:**10-17.
2. Khanna KK and Jackson SP: **DNA double-strand breaks: signaling, repair and the cancer connection.** *Nat Genet* 2001, **27:**247-254.
3. Pandita TK: **ATM function and telomere stability.** *Oncogene* 2002, **21:**611-618.
4. Abraham RT: **Cell cycle checkpoint signaling through the ATM and ATR kinases.** *Genes Dev* 2001, **15:**2177-2196.
5. Jeggo PA: **DNA breakage and repair.** *Adv Genet* 1998, **38:**185-218.
6. Wood RD, Mitchell M, Sgouros J and Lindahl T: **Human DNA repair genes.** *Science* 2001, **291:**1284-1289.
7. Marcou Y, D'Andrea A, Jeggo PA and Plowman PN: **Normal cellular radiosensitivity in an adult Fanconi anaemia patient with marked clinical radiosensitivity.** *Radiother Oncol* 2001, **60:**75-79.
8. Mirzayans R, Aubin RA, Bosnich W, Blattner WA and Paterson MC: **Abnormal pattern of post-gamma-ray DNA replication in radioresistant fibroblast strains from affected members of a cancer-prone family with Li-Fraumeni syndrome.** *Br J Cancer* 1995, **71:**1221-1230.
9. Bennett CB, Lewis LK, Karthikeyan G, Lobachev KS, Jin YH, Sterling JF, Snipe JR and Resnick MA: **Genes required for ionizing radiation resistance in yeast.** *Nat Genet* 2001, **29:**426-434.
10. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
11. Xu Y and Uberbacher EC: **Automated gene identification in large-scale genomic sequences.** *J Comput Biol* 1997, **4:**325-338.
12. Kozak M: **An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.** *Nucleic Acids Res* 1987, **15:**8125-8148.
13. Florea L, Hartzell G, Zhang Z, Rubin GM and Miller W: **A computer program for aligning a cDNA sequence with a genomic DNA sequence.** *Genome Res* 1998, **8:**967-974.
14. Li G, Nelsen C and Hendrickson EA: **Ku86 is essential in human somatic cells.** *Proc Natl Acad Sci USA* 2002, **99:**832-837.
15. Friedman AD: **Runx1, c-Myb, and C/EBPalpha couple differentiation to proliferation or growth arrest during hematopoiesis.** *J Cell Biochem* 2002, **86:**624-629.
16. Wang H, Goode T, Iakova P, Albrecht JH and Timchenko NA: **C/EBPalpha triggers proteasome-dependent degradation of cdk4 during growth arrest.** *EMBO J* 2002, **21:**930-941.
17. Xie Y, Chen C, Stevenson MA, Auron PE and Calderwood SK: **Heat shock factor 1 represses transcription of the IL-1beta gene through physical interaction with the nuclear factor of interleukin 6.** *J Biol Chem* 2002, **277:**11802-11810.
18. Chinenov Y and Kerppola TK: **Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity.** *Oncogene* 2001, **20:**2438-2452.
19. Sanyal S, Sandstrom DJ, Hoeffer CA and Ramaswami M: **AP-1 functions upstream of CREB to control synaptic plasticity in Drosophila.** *Nature* 2002, **416:**870-874.
20. Servillo G, Della Fazia MA and Sassone-Corsi P: **Coupling cAMP signaling to transcription in the liver: pivotal role of CREB and CREM.** *Exp Cell Res* 2002, **275:**143-154.
21. de Rooij DG, van de Kant HJ, Dol R, Wagemaker G, van Buul PP, van Duijn-Goedhart A, de Jong FH and Broerse JJ: **Long-term effects of irradiation before adulthood on reproductive function in the male rhesus monkey.** *Biol Reprod* 2002, **66:**486-494.
22. Osada N, Hida M, Kusuda J, Tanuma R, Iseki K, Hirata M, Suto Y, Hirai M, Terao K, Suzuki Y, Sugano S, Hashimoto K and Kusuda J: **Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes.** *Gene* 2001, **275:**31-37.
23. Osada N, Hida M, Kusuda J, Tanuma R, Hirata M, Hirai M, Terao K, Suzuki Y, Sugano S and Hashimoto K: **Prediction of unidentified human genes on the basis of sequence similarity to novel cDNAs from cynomolgus monkey brain.** *Genome Biol* 2002, **3:**1-006.
24. Thompson JD, Higgins DG and Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
25. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, Pruss M, Schacherer F, Thiele S and Urbach S: **The TRANSFAC system on gene expression regulation.** *Nucleic Acids Res* 2001, **29:**281-283.
26. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ and Altschul SF: **SAGEmap: a public gene expression resource.** *Genome Res* 2000, **10:**1051-1060.
27. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K and Bairoch A: **The PROSITE database, its status in 2002.** *Nucleic Acids Res* 2002, **30:**235-238.
28. Rechsteiner M and Rogers SW: **PEST sequences and regulation by proteolysis.** *Trends Biochem Sci* 1996, **21:**267-271.
29. Blom N, Gammeltoft S and Brunak S: **Sequence and structure-based prediction of eukaryotic protein phosphorylation sites.** *J Mol Biol* 1999, **294:**1351-1362.

8

Construction of a Full-Length Enriched and a 5′-End Enriched cDNA Library Using the Oligo-Capping Method

Yutaka Suzuki and Sumio Sugano

1. Introduction

With the completion of the draft sequence of the human genome *(1,2)*, it is now essential to extract biological information from the large volumes of human genomic sequence data. A number of attempts, which can be comprehensively termed "functional genomics," are being carried out to decipher which parts of the human genome are transcribed, how the transcripts are spliced and translated, and what functions the eventual protein products conduct. For these purposes, a full-length complementary DNA (cDNA) that contains the entire sequence of the mRNA from the cap structure to the poly(A) tail is a unique resource because a variety of information about the gene functions is contained in a full-length cDNA sequence. The intensive analysis of a full-length cDNA would enable us to identify the following:

1. The exact position of the mRNA transcriptional start site, which is indispensable for the identification of the adjacent promoter.
2. The sequence of the complete 5′ untranslated region (5′ UTR), which is related to the translation efficiency and the cellular localization of mRNA.
3. The continuous protein coding region (CDS), which is required for producing a recombinant protein.
4. The sequence of the complete 3′ untranslated region (3′ UTR), which is related to the translation efficiency, the cellular localization, and the stability of mRNA.

Therefore, a collection of full-length cDNAs should serve as a valuable resource for the functional analysis of the human genome.

From: *Methods in Molecular Biology, vol. 221: Generation of cDNA Libraries: Methods and Protocols*
Edited by: S.-Y. Ying © Humana Press Inc., Totowa, NJ

However, a serious drawback exists in the cDNA libraries that are widely used for current cDNA analyses. In many cases, reverse transcriptase cannot make a full cDNA copy of an mRNA; instead, drops off somewhere in the middle, yielding an incomplete copy. Thus, cDNA libraries made by the conventional methods contain many incomplete cDNA clones that usually lack the 5′-end sequences of the template mRNA. This, at least, in part accounts for the current situation that most cDNA data cover mainly the 3′ ends of messenger RNA (mRNA) and the information about sequences near the 5′-ends remains limited.

In order to eliminate this drawback, we considered it essential to develop new technology to construct "a full-length cDNA library," which is a cDNA library consisting preferentially of full-length cDNAs. In order to make a full-length cDNA library, we need to devise a type of selection procedure to pick up full-length cDNAs from a cDNA pool predominantly occupied by truncated ones. "Selection of a full-length cDNA" is synonymous to "selection of a cDNA that contains both ends of the mRNA." Thus, in order to select a full-length cDNA, the features characteristic of the 3′ end and the 5′ end of an mRNA should be tagged for the selection. A full-length cDNA could be selected via selection steps for both the 3′-end and the 5′-end "tags."

The poly(A) stretch is a characteristic feature of the 3′ end of an mRNA. Conventional methods have generally utilized the poly(A) as a "sequence tag" to select the 3′-end of an mRNA. In the conventional methods, the first strand cDNA is usually synthesized starting from an oligo(dT) primer. Because dT primers mostly hybridize at the poly(A) tail, most of the cDNA is selectively synthesized from the 3′ end of the mRNA. Therfore, the conventional methods include the selection step for the 3′-end "tag" of the mRNA.

In contrast, conventional methods include no selection step for the 5′ end of the mRNA. As a result, the largest part of the cDNA library constructed by the conventional methods consists of cDNAs that lack the 5′ end of the mRNA. The main reason for this lies, in our view, in the fact that mRNA does not originally have a "sequence tag" at the 5′ end. The 5′ end of an mRNA does have a characteristic structure, called the cap structure, but, unfortunately, it is not a "sequence tag." Unlike the poly(A) at the 3′ end, the cap structure cannot be used for hybridization as it is. If the 5′-end "tag" of an mRNA were also a "sequence tag," it would be easy to utilize the tag to select the 5′ end of an mRNA.

In order to introduce a "sequence tag" to the 5′ end of an mRNA, we have developed a new method, which we named the "oligo-capping" method *(3)*. This method allowed us to enzymatically replace the cap structure of an mRNA with a synthetic oligonucleotide (*see* **Fig. 1**). Each mRNA product of the "oligo-capping" should contain the sequence tags at both its ends, namely poly(A) at the 3′ end and the cap-replacing oligo at the 5′ end. With the "oligo-

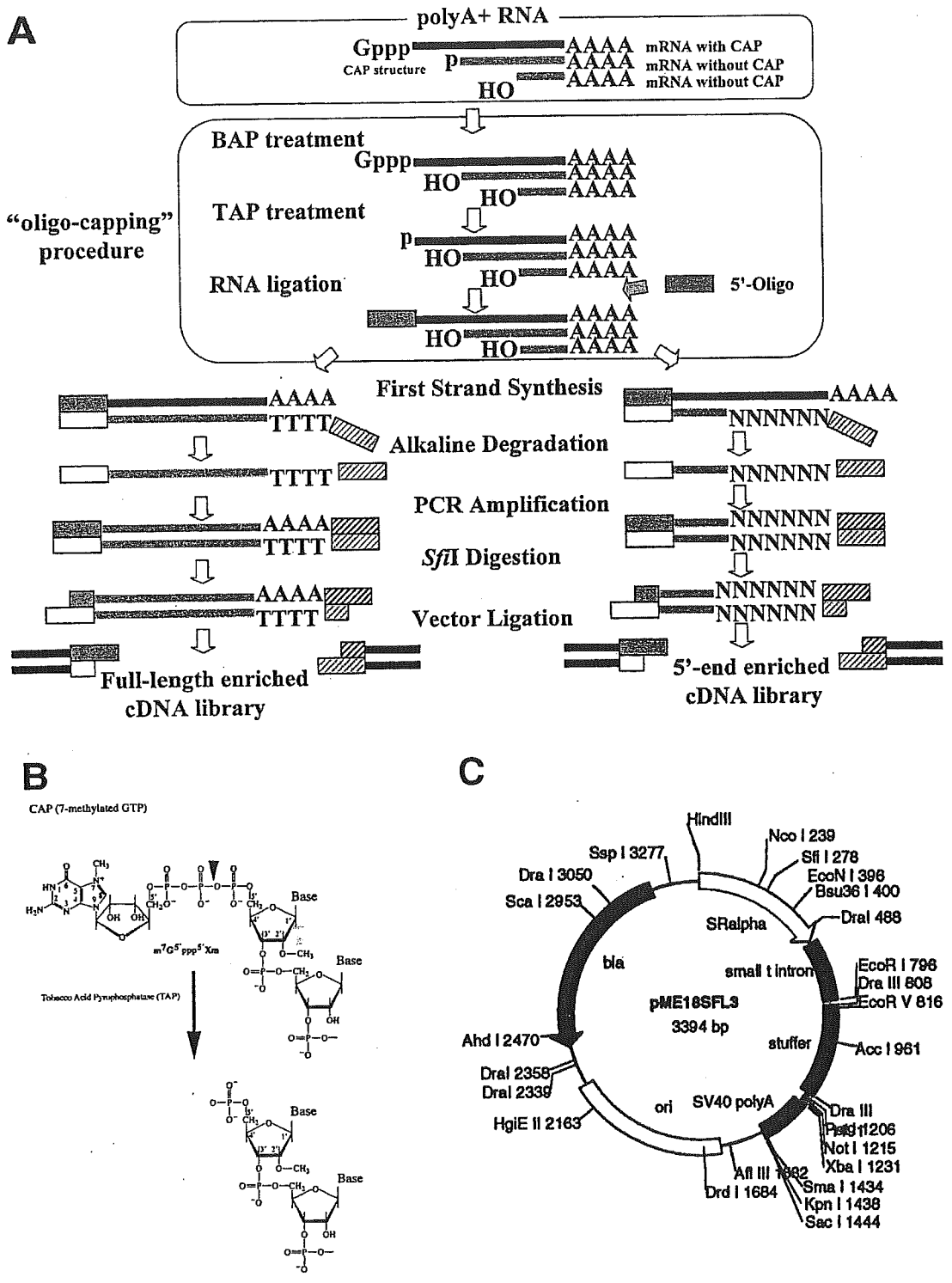

Fig. 1. (A) Schematic presentation of the construction of a full-length enriched and a 5′-end enriched cDNA library. According to the scheme shown, a full-length enriched and a 5′-end enriched cDNA library are constructed. Gppp: the cap structure; p: phosphate; OH: hydroxyl; AAAA: poly(A). (B) Enzymatic activity of TAP. TAP cleaves the triphosphate at the position shown by the arrowhead, leaving a phosphate at the 5′ end of the mRNA. (C) Plasmid map of the cloning vector pME18S-FL3. The cDNAs are inserted between the *Dra*III sites of the plasmid.

capped" mRNAs as starting material, we developed a new system to selectively clone the cDNAs that contain both of the sequence tags at the respective ends. Following the scheme shown in **Fig. 1**, we were able to construct a cDNA library in which the content of "full-length" cDNA is significantly enriched ("full-length enriched" cDNA library) *(4,5)*.

It is also possible that the full-length enriched cDNA library may not include cDNAs of long mRNAs, because the distance between the cap structure and the poly(A) may be beyond the limit that the reverse transcriptase can copy. In such cases, the cDNAs could not be cloned because of the lack of the 5′-end tags. This is especially likely when an mRNA longer than 5 kb is used as a template. In order to address this issue, we also developed a system to construct a "5′-end enriched cDNA library" to cover the 5′ ends of long mRNAs (*see* **Fig. 5**). In case the full-length cDNA could not be directly obtained from the full-length enriched library, the 5′-end enriched library and the conventional cDNA library, which is enriched for the 3′ end of the mRNA, could be used complimentarily with each other to isolate the full-length cDNA *(4)*.

1.1. Principles of the Construction of a Full-Length Enriched and a 5′-End Enriched cDNA Library

Figure 1A illustrates the scheme for the construction of a full-length enriched and a 5′-end enriched cDNA library. The "oligo-capping" procedure consists of three steps of enzymatic reactions. First, bacterial alkaline phosphatase (BAP) hydrolyzes the phosphate from the 5′ ends of truncated mRNAs that are noncapped. The cap structure on full-length mRNAs remains intact during this reaction. Second, tobacco acid pyrophosphatase (TAP) cleaves the cap structure itself at the position indicated by an arrow in **Fig. 1B**, leaving a phosphate at the 5′ ends of mRNAs. Finally, T4 RNA ligase selectively ligates the synthetic oligoribonucleotide to the phosphate at the 5′ end. As a result, the oligoribonucleotide is introduced only to the 5′ ends of mRNAs that originally had the cap structure.

With "oligo-capped" mRNA as the starting material, first-strand cDNA is synthesized using an oligo-(dT) adaptor primer for a full-length enriched cDNA library or using a random hexamer adaptor primer for a 5′-end enriched cDNA library. After first-strand cDNA synthesis, the template mRNA is subjected to alkaline degradation. Then, polymerase chain reaction (PCR) is performed with 3′- and 5′-end primers that contain part of the adaptor primer sequence and the cap-replaced oligonucleotide sequence, respectively. The amplified cDNA fragments are digested with restriction enzymes, size fractionated and cloned into a plasmid vector in an orientation-specific manner.

2. Materials (*see* Notes 1 and 2)

1. Thermal cycler: Zymoreactor II (Atto, Tokyo, Japan) for library construction and GeneAmp PCR System 9700 (ABI, Foster City, CA) for colony PCR and sequencing reaction.
2. Sequencer: 3700 DNA Analyzer (ABI).
3. Centrifuge: GS-6KR (Beckman) for centrifugation of 50-mL tubes and M150-IVD (Sakuma, Tokyo, Japan) for 1.5-mL microtubes. Any type of centrifuge with a refrigerator may be used.
4. Tubes: We have confirmed that the following tubes are RNase free when newly opened. They should be used without autoclaving or diethyl pyrocarbonate (DEPC) treatment: 50-mL tubes (cat. no. 227 261; Greiner, Frickenhausen, Gernamy); 1.5-mL tubes (cat. no. 0030 102.002; Eppendorf, Hamburg, Gernamy).
5. Mechanical homogenizer: Polytron (cat. no. PT10-35, Kinematica, Luzern, Switzerland).
6. Carrier for the ethanol precipitation (RNase-free): ethachinmate (cat. no. 312-01791; WAKO, Tokyo, Japan).
7. RNA extraction kit: RNeasy (cat. no. 75163; Qiagen, Chatsworth, CA); Trizol (cat. no. 15596-018; Invitrogen, Carlsbad, CA).
8. Poly(A) selection: Oligo-(dT) cellulose (cat. no. OT-125-B; Molecular Research Center, Cincinnati, OH).
9. Polypropylene column: Poly-Prep (cat. no. 731-1550; Bio-Rad, Hercules, CA).
10. RNasin (40 U/μL; cat. no. N2111; Promega, Madison, WI).
11. Bacterial alkaline phosphatase (0.25 U/μL; cat. no. 2110; TaKaRa, Kyoto, Japan).
12. Tobacco acid pyrophosphatase (20 U/μL; purified from tobacco cells, BY-2, following the procedure described in **ref. *6***). Alternatively, TAP is now commercially available (cat. no. 313-04021; Wako).
13. T4 RNA ligase (25 U/μL; cat. no. 2050; TaKaRa).
14. Polyethylene glycol (PEG) 8000: 50% (w/v) PEG 8000 (cat. no. P2139; Sigma, St. Louis, MO). Add dH_2O to PEG 8000 so that the concentration is 50% (w/v). Dissolve the PEG 8000 at 65°C. Sterilize the solution by filtration through a 0.22-μ*M* membrane: MILLEX-GV (cat. no. SLGV025LS, Millipore, Molsheim, France).
15. DNase I (RNase-free) (5.0 U/μL; cat. no. 2215; TaKaRa).
16. Spin column: S-400HR (cat. no. 27-5140; Amersham Pharmacia Biotech, Piscataway, NJ).
17. Superscript II (200 U/μL; cat. no. 18064-014; Invitrogen).
18. PCR kit: GeneAmp (cat. no. N808-0192; Perkin-Elmer, Norwalk, CT, USA) for library construction; ExTaq (cat. no. RR001A; TaKaRa) for colony PCR.
19. Restriction enzymes: *Sfi*I (20 U/μL; New England Biolabs Beverly, MA); *Dra*III (20 U/μL; New England Biolabs).
20. Agarose (cat. no. 312-01193; WAKO).

21. DNA ligation kit (cat. no. 6021; TaKaRa).
22. $MgCl_2$ (cat. no. 135-00165; WAKO), ATP (cat. no. A2, 620-9; Sigma), dATP, dCTP, dGTP, dTTP (cat. no. 4026-4029; TaKaRa), NaOH (cat. no. 197-02125; WAKO), mineral oil (cat. no. M5904, Sigma).
23. Competent cells: TOP10 (cat. no. C4040-50; Invitrogen).
24. 2X Loading buffer for poly(A) selection: 40 m*M* Tris-HCl (cat. no. T-1503; Sigma) (pH 7.0), 1 *M* NaCl (cat. no. 191-01665; WAKO), 2 m*M* EDTA (cat. no. 345-01865; WAKO) (pH 8.0), 0.2% (w/v) sodium dodecyl sulfate (SDS) (cat. no. 191-07145; WAKO).
25. 5X BAP buffer: 500 m*M* Tris-HCl (pH 7.0), 50 m*M* of 2-mercaptoethanol (cat. no. 137-06862; WAKO). Do not use the supplied buffer because the pH of the buffer is so high that it is hazardous to RNA.
26. 5X TAP buffer: 250 m*M* sodium acetate (pH 5.5), 50 m*M* of 2-mercaptoethanol, 5 m*M* EDTA (pH 8.0).
27. 10X Ligation buffer: 500 m*M* Tris-HCl (pH 7.0), 100 m*M* of 2-mercaptoethanol.
28. 10X STE: 100 m*M* Tris-HCl (pH 7.0), 1 *M* NaCl, 10 m*M* EDTA (pH 8.0).
29. QIAquick Gel Extraction kit (cat. no. 28704; QIAGEN, Hilden, Germany).
30. DNA ligation kit (cat. no. 6021; TaKaRa).
31. Cloning vector: pME18S-FL3 (Genbank acc. no. AB009864; **Fig. 1C**).
32. Buffers and columns marked with asterisks in the text are included in the corresponding kit.
33. 5′-Oligoribonucleotide A: 5′-AGCAUCGAGUCGGCCUUGUUGGCCUACU GG-3′ (100 ng/μL; custom order, TaKaRa).
34. Oligo-(dT) adapter primer B: 5′-GCGGCTGAAGACGGCCTATGTGGCC$(T)_{17}$-3′ (5 pmol/μL; custom order, Invitrogen).
35. Random hexamer adapter primer C: 5′-GCGGCTGAAGACGGCCTATGTG GCCNNNNNNC-3′ (10 pmol/μL; custom order, Invitrogen).
36. 5′ Primer D: 5′-AGCATCGAGTCGGCCTTGTTG-3′ (10 pmol/μL; custom order, Invitrogen).
37. 3′ Primer E: 5′-GCGGCTGAAGACGGCCTATGT-3′ (10 pmol/μL; custom order, Invitrogen).
38. 3′ Primer F for the EF1-α amplification: 5′-ACGTTCACGCTCAGCTTTCAG-3′ (10 pmol/μL; custom order, Invitrogen).
39. 3′ Primer G for the EF1-α amplification: 5′-AACACCAGCAGCAACAAT CAG-3′ (10 pmol/μL; custom order, Invitrogen).
40. Colony PCR primer H (forward): 5′-TCAGTGGATGTTGCCTTTAC-3′ (3.2 pmol/μL; custom order, Invitrogen).
41. Colony PCR primer I (reverse): 5′-TGTGGGAGGTTTTTTCTCTA-3′ (3.2 pmol/μL; custom order, Invitrogen).
42. Sequencing primer J (forward read): 5′-GGATGTTGCCTTTACTTCTA-3′ (3.2 pmol/μL; custom order, Invitrogen).
43. Sequencing primer K (reverse read): 5′-CGACCTGCAGCTCGAGCACA-3′ (3.2 pmol/μL; custom order, Invitrogen).

3. Method

3.1. RNA Isolation from Tissues or Cultured Cells (see Note 3)

3.1.1. Total RNA Isolation Using Trizol

1. Put 20 mL of Trizol* in a 50-mL tube.
2. Add up to 1 g of tissue or sample material to the tube and crash the sample using a Polytron (Kinematica). If a mechanical homogenizer is not available, grind the sample to a fine powder in a mortar filled with liquid nitrogen.
3. Centrifuge the tube at 3500*g* for 10 min at room temperature (RT) to remove the tissue that has remained undissolved.
4. Decant the supernatant to a new tube and discard the debris.
5. Add 4 mL of chloroform and rock the tube for 3 min at RT.
6. Centrifuge at 3500*g* for 20 min at RT.
7. Transfer the upper layer gently to a fresh tube (*see* **Note 4**).
8. Add 10 mL of 100% (v/v) isopropanol and let it stand for 10 min at RT.
9. Centrifuge at 3500*g* for 20 min.
10. Discard the supernatant and wash the pellet with 5 mL of 80% (v/v) ethanol.
11. Briefly spin down the pellet and remove the supernatant.
12. Dissolve the pellet in 1 mL of dH_2O.
13. To check the RNA, take 1 µL of the solution, add 4 µL of dH_2O and apply to a 2 % agarose gel in TAE for electrophoresis (*see* **Note 5**).

3.1.2. Purification of the Total RNA Using RNeasy

1. Add 15 mL of RLT* to the sample prepared in **Subheading 3.1.1.**
2. Add 15 mL of 70 % (v/v) ethanol to the tube and mix well.
3. Apply the solution to the RNeasy column* and let it pass through by brief centrifugation at 2300*g* at RT. Discard the flow through and repeat this step until the entire sample has been applied.
4. To wash the column, apply 10 mL of RW1* to the column and centrifuge briefly at 3000 rpm at RT. Discard the flowthrough.
5. To wash the column further more, apply 10 mL of RPE* to the column and centrifuge briefly at 2300*g* at RT. Discard the flow through and repeat this step once more.
6. Use a fresh tube to collect the eluate. To elute the sample, apply 1.2 mL of dH_2O to the column and let it stand for 1 min at RT. Centrifuge at 2300*g* for 1 min at RT. Collect the eluate.
7. To check the RNA, take 3 µL of the sample, add 2 µL of dH_2O and apply to a 2% agarose gel in TAE for electrophoresis (*see* **Note 5**).

3.1.3. Poly(A) Selection of the RNA (see **Notes 6** *and* **7***)*

1. Transfer the oligo-(dT) powder from two prepacked columns* to a polypropylene Poly-Prep column (Biorad). The bed volume of the powder should be approx 0.5 mL when the powder from two columns is used (*see* **Note 6**).

2. Denature the dT powder by washing with 3 mL of 0.1 *N* NaOH.
3. Wash out the alkaline solution with 5 mL of dH_2O.
4. Pre-equilibrate the column with 5 mL of 1X loading buffer.
5. Set a fresh tube to collect the flowthrough.
6. Add an equal volume (1.2 mL) of 2X loading buffer to the sample, mix well, and apply to the column.
7. Collect the flow through and apply to the column. Repeat this step two more times.
8. Wash the column with 5 mL of 1X loading buffer.
9. Set a fresh collection tube and elute the sample by applying 3 mL of dH_2O.
10. Add 8 mL of 100% ethanol and 330 µL of 3 *M* sodium acetate (pH 5.5) and centrifuge for ethanol precipitation.
11. Dissolve the sample in 100 µL of dH_2O and ethanol precipitate once more (as described in **Subheading 3.2.1., steps 4** and **5**).
12. Dissolve the sample with 72.3 µL of dH_2O.
13. Take 5 µL of the solution and apply to a 2% agarose gel in TAE for electrophoresis (*see* **Note 5**).

3.2. BAP Reaction

1. Set up a BAP reaction by combining: 67.3 µL of the sample [10–20 µg of poly(A)$^+$ RNA], 20.0 µL of 5X BAP buffer, 2.7 µL of RNasin, and 10.0 µL of BAP.
2. Incubate at 37°C for 60 min.
3. Add an equal volume of phenol:chloroform (1:1) to the sample and mix well. Centrifuge at 12,000*g* briefly at 4°C. Transfer the upper aqueous layer to a fresh tube.
4. Ethanol precipitate the RNA by adding 1 µL of ethachinmate, 11 µL of sodium acetate (pH 5.5), and 275 µL of 100% ethanol. Centrifuge at 12,000*g* at 4°C for 10 min.
5. Remove the supernatant and rinse the pellet with 150 µL of 80% (v/v) ethanol. Drying is not necessary (*see* **Note 8**).
6. Dissolve the sample in 75.3 µL of dH_2O.

3.3. TAP Reaction

1. Set up a TAP reaction by combining: 75.3 µL of the sample, 20.0 µL of 5X TAP buffer, 2.7 µL of RNasin, and 2.0 µL of TAP.
2. Incubate at 37°C for 60 min.
3. Extract with phenol:chloroform (1:1) and ethanol precipitate (as described in **Subheading 3.2., steps 3–5**).
4. Resuspend the sample in 11.0 µL of dH_2O.

3.4. RNA Ligation

1. Ligate the BAP/TAP-treated poly(A)$^+$ RNA to the 5′-oligoribonucleotide (sequence A) by combining 11.0 µL of the sample, 4.0 µL of the 5′-oligoribonucleotide, 10.0 µL of 10X ligation buffer, 10.0 µL of 50 m*M* $MgCl_2$, 2.1 µL of

24 m*M* ATP, 2.5 µL of RNasin, 10.0 µL of T4 RNA ligase, and 50.0 µL of 50% (w/v) PEG 8000.
2. Incubate at 20°C for 3 h.
3. Add 200 µL of dH_2O.
4. Extract with 300 µL of phenol : chloroform (1 : 1) (*see* **Note 9**). Ethanol precipitate by adding 1 µL of ethachinmate, 33 µL of sodium acetate (pH 5.5), and 825 µL of 100% ethanol (the rest of the procedure is as described in **Subheading 3.2., steps 3–5**).
6. Dissolve the sample in 70.3 µL of dH_2O.

3.5. DNase I Treatment

1. Remove the residual DNA with DNase I by combining 70.3 µL of the sample, 16.0 µL of 50 m*M* $MgCl_2$, 4.0 µL of 1 *M* Tris-HCl (pH 7.0), 5.0 µL of 0.1 *M* dithiothreitol (DTT) (use DTT supplied with SuperScript II), 2.7 µL of RNasin, and 2.0 µL of DNase I.
2. Incubate at 37°C for 10 min.
3. Extract with phenol : chloroform (1 : 1) and ethanol precipitate (as described in **Subheading 3.2., steps 3–5**).
5. Dissolve the sample in 45 µL of dH_2O.

3.6. Spin-Column Purification

Remove excess 5′-oligoribonucleotide and fragmented DNA by spin-column chromatography.

1. Resuspend the resin of the column S400HR* thoroughly by vortexing.
2. Centrifuge the column for 1 min at RT at 735*g* (*see* **Note 10**) to remove the pre-equilibrated water. Discard the flow through.
3. In order to completely remove the water, repeat the previous step once more.
4. Add 5 µL of 10X STE to the sample, mix briefly, and gently, and apply to the column.
5. Set a fresh tube for collecting the eluate and let the sample pass through the column by centrifugation at 735*g* for 2 min at RT.
6. Add 50 µL of dH_2O.
7. Ethanol precipitate (as described in **Subheading 3.2., steps 4** and **5**).
8. Dissolve the sample in 21 µL of dH_2O.

3.7. First-Strand cDNA Synthesis

1. Synthesize first-strand cDNA with reverse transcriptase, SuperScript II, by combining: 21 µL of the sample, 10.0 µL of 5X first-strand buffer,* 8.0 µL of the 4 dNTPs at 5 m*M* each, 6.0 µL of 0.1 *M* DTT,* 2.5 µL of the oligo-(dT) adapter primer (sequence B), 1.0 µL of RNasin, and 2.0 µL of SuperScript II for the full-length cDNA library. For the 5′-end enriched library, use 2.5 µL of the dR (random hexamer) adapter primer (sequence C) instead of the oligo-(dT) adapter primer.

2. Incubate at 42°C for more than 3 h for the full-length cDNA library. For the 5′-end library, incubate at 12°C for 1 h and 42°C for more than 3 h (*see* **Note 11**).
3. Add 50 µL of dH_2O and extract the solution with phenol:chloroform (1:1) (as described in **Subheading 3.2., step 3**).
4. Add 2 µL of 0.5 *M* EDTA (pH 8.0) to stop the reaction completely.

3.8. Alkaline Degradation of the Template mRNA

1. Degrade the template RNA by adding 15 µL of 0.1 *M* NaOH. Incubate at 65°C for 40 min.
2. Add 20 µL of 1 *M* Tris-HCl (pH 7.0) to neutralize.
3. In order to remove the fragmented RNA, ethanol precipitate the first-strand cDNA by adding 1 µL of ethachinmate, 70 µL of 7.5 *M* ammonium acetate (*see* **Note 12**), and 500 µL of ethanol (the rest of the procedure is as described in **Subheading 3.2., steps 3–5**).
4. Dissolve the sample in 50 µL of dH_2O.

3.9. Confirmation of the First-Strand cDNA

In order to confirm the integrity of the first-strand cDNA and the rate of success of the "oligo-capping," PCR amplify the 5′ end of the EF1-α mRNA.

1. Combine 1 µL (1/50) of the first-strand cDNA (for the full-length library and the 5′-end library) in 52.4 µL of dH_2O with 30.0 µL of 3.3X reaction buffer II,* 8.0 µL of the 4 dNTPs at 2.5 m*M* each,* 4.4 µL of 25 m*M* magnesium acetate,* 1.6 µL of the 5′ primer (sequence D), 1.6 µL of the 3′ primer (sequence F) or 1.6 µL of the 3′ primer (sequence G), and 2.0 µL of rTth DNA polymerase. All of the reaction components except for the primers are included in the GeneAmp kit. Overlay with 125 µL of mineral oil.
2. Thermocycle for 30 cycles at 94°C, 1 min; 52°C, 1 min; 72°C, 1 min.
3. Take 5 µL (1/10–1/20) of the PCR products and apply to a 2% (w/v) agarose gel in TAE for electrophoresis. Confirm the PCR products of 312 bp and 474 bp for primer pairs D + F and D + G, respectively (*see* **Fig. 2**).

3.10. PCR Amplification of the cDNA

1. Use 10–25 µL (1/5–1/2; depending on the quality and quantity of the starting RNA; usually 10 µL is sufficient) of the first strand cDNA in 52.4 µL of dH_2O with 30.0 µL of 3.3X reaction buffer II,* 8.0 µL of the 4 dNTPs at 2.5 m*M* each,* 4.4 µL of 25 m*M* magnesium acetate,* 1.6 µL of 5′ Primer (sequence D), 1.6 µL of 3′ Primer (sequence E), and 2.0 µL of rTth DNA polymerase. Overlay with 125 µL of mineral oil.
2. Thermocycle for 12–15 cycles (depending on the quality and quantity of the library; usually 12 cycles is adequate) at 94°C, 1 min; 58°C, 1 min; 72°C, 10 min.
3. Extract with phenol:chloroform (1:1) and ethanol precipitate (as described in **Subheading 3.2., steps 3–5**).
4. Dissolve the sample in 89 µL of dH_2O.