Available online at www.sciencedirect.com
ScienceDirect
Gene 405 (2007) 27–35
GENE
www.elsevier.com/locate/gene

Characterization of the porcine *ATM* gene: Towards the generation of a novel non-murine animal model for Ataxia-Telangiectasia ☆

Margarita B. Rogatcheva [a], Krista L. Fritz [a], Laurie A. Rund [a], Callie B. Pollock [a], Jonathan E. Beever [a], Christopher M. Counter [b], Lawrence B. Schook [a,c,*]

[a] *Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1201 W. Gregory Dr., Urbana, IL 61801, USA*
[b] *Department of Pharmacology and Oncology, Duke University Medical Center, Durham, NC 27710, USA*
[c] *Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 W. Gregory Dr., Urbana, IL 61801, USA*

Received 29 January 2007; received in revised form 20 August 2007; accepted 22 August 2007
Available online 30 August 2007
Received by M. D'Urso

Abstract

Ataxia-Telangiectasia (A-T) is a genetic disorder causing cerebellar degeneration, immune deficiency, cancer predisposition, chromosomal instability and radiation sensitivity. Among the mutations responsible for A-T, 85% represent truncating mutations that result in the production of shorter, highly unstable forms of ATM (AT-mutated) protein leading to a null ATM phenotype. Several ATM-deficient mice have been created however none reflects the extent of neurological degeneration observed in humans. In an attempt to identify an alternative animal model, we have characterized the porcine ortholog of *ATM* (*pATM*). When compared to the human *ATM* (*hATM*), the *pATM* showed a high level of homology in the coding region, particularly in the regions coding for functional domains, and had extensive alternative splicing of the 5′UTR, characteristic for the human *ATM* mRNA. Six different 5′UTRs resulting from alternative splicing of the first three exons were identified. The porcine 5′UTRs varied in size, had multiple ATG codons and different secondary structures, supporting the possibility of complex transcriptional regulation. Three of the six transcripts demonstrated alternative splicing of exon 3, the first putative coding exon, altering the translation start and giving rise to a putative protein lacking the N-terminus substrate binding domain (82–89 aa) involved in activation of human p53 and BRCA1 pathways. Real time-PCR analysis revealed variable expression levels of total *ATM* transcripts in individual tissues. Although each splice variant was ubiquitously expressed among the tissues, differences in the relative abundances of specific 5′UTRs were observed. The extensive alternative splicing of the *pATM* gene resembles the complex splicing observed in the *hATM* and could provide insights for differences observed between mice and humans with regards to the onset of A-T. Thus, the pig may provide a more relevant clinical model of A-T.

Keywords: Alternative splicing; Translation initiation; Pig; Transcriptional regulation

1. Introduction

Ataxia-Telangiectasia (A-T) is an autosomal recessive disorder characterized by progressive cerebellar degeneration, immunodeficiency, cancer predisposition, gonadal atrophy, growth retardation, and premature aging (Lavin and Shiloh, 1997; McKinnon, 2004). Cells from A-T patients generally display shorter telomeres and are highly sensitive to ionizing radiation (IR) (Pandita et al., 1999; Kishi and Lu, 2002; Pandita, 2002). A-T is caused by a deficiency in the ATM protein (AT-mutated) that mediates the cellular response to DNA damage through multiple transduction pathways (Pandita et al., 2000; Yang et al., 2004). In humans, ATM

Abbreviations: aa, amino acid(s); AS, alternative splicing; ATM, Ataxia-Telangiectasia mutated: BAC, bacterial artificial chromosome; bp, base pair; cDNA, complementary to RNA; IR, ionizing radiation; kb, kilobase pair or 1000 bp; nt, nucleotide; ORF, open reading frame; PCR, polymerase chain reaction; poly A, polyadenylation site: RACE, rapid amplification of cDNA ends; rRNA, ribosomal RNA; RT, reverse transcription; SV, splice variant(s); UTR, untranslated region.

☆ The nucleotide sequence data reported in this paper deposited in GenBank nucleotide sequence database under the accession numbers AY587061; DQ646385–DQ646390.

* Corresponding author. University of Illinois at Urbana-Champaign, Department of Animal Sciences, 1201 W. Gregory Dr., Edward R. Madigan Laboratory 382, Urbana, IL 61801, USA. Tel.: +1 217 265-5326; fax: +1 217 244 5617.
E-mail address: schook@uiuc.edu (L.B. Schook).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.08.014

is a large protein of 3056 aa with highly conserved catalytic domains for phosphatidylinositol-3-kinase (PIK), FAT, and FATC. In response to DNA damage, ATM displays kinase activity and initiates a signaling cascade by phosphorylating various substrates such as p53, Chk2, Mdm2, NBS1, BRCA1, 53BP1 Smc1, FANCD2, H2AX and Pin2/TRF1, that in turn control cell cycle-check points, DNA double-strand break repair pathways, apoptosis and telomere metabolism (Kastan and Lim, 2000; Shiloh, 2003; Ball and Xiao, 2005). The broad range of the ATM targeted substrates explains the multisystem phenotypic manifestation of ATM-deficiency.

The human *ATM* gene has 66 exons spanning over approximately 146 kb of genomic DNA (Savitsky et al., 1995; Platzer et al., 1997). The mutation sites reported in A-T patients span the entire open reading frame (ORF) without any apparent hot spots (Lavin et al., 2004). Approximately 85% of the mutations cause premature termination translation resulting in production of unstable and usually undetectable forms of the ATM protein (Gilad et al., 1996; Lakin et al., 1996; Ball and Xiao, 2005). In A-T patients, truncating mutations located downstream of the PI3-kinase domain close to the C-terminus, lack ATM activity (Bakkenist and Kastan, 2003). The remaining mutations are characterized either by decreased levels of functional protein or normal protein levels with markedly reduced kinase activity (Stewart et al., 2001).

ATM is constitutively expressed in numerous tissues with the highest level observed in proliferating tissues, reproductive organs and the nervous system (Chen and Lee, 1996). The same tissues that display higher protein activity are the most affected by loss of ATM protein activity in both A-T patients and ATM-knockout mice (Gueven et al., 2006). The major control of ATM activity is via activation of preexisting protein by autophosphorylation of the ATM dimer and release of the active monomer (Bakkenist and Kastan, 2003). This enables cells to rapidly manipulate ATM activity without new mRNA synthesis, processing or export. However, in addition to activation by autophosphorylation of existing protein, *ATM* gene activity can also be regulated at the level of transcription. The gene expression of *ATM* is driven by a TATA-less promoter that regulates bi-directional transcription of *ATM* and *NPAT* (nuclear protein mapped to the A-T locus) genes (Byrd et al., 1996). The promoter activity is tissue specific and can be increased in response to IR in a tissue specific manner, with the higher increase of activity observed in tissues with a lower basal level of ATM (Gueven et al., 2006). In response to IR, some tissues produce more protein without changes in the mRNA level (Fang et al., 2001) suggesting translational regulation of gene expression. Moreover, alternative splicing of the 5′UTR and alternative 3′ polyadenylation (polyA) sites in human *ATM* transcripts (Savitsky et al., 1997) provide an opportunity for additional posttranscriptional regulation of the *ATM* expression.

Alternative splicing is thought to play an important role in regulating the expression of many vertebrate genes and in the generation of proteomic diversity (Graveley, 2001; Modrek and Lee, 2003; Pan et al., 2005). This provides a potential source of complexity and species-specific differences. The currently available murine models of A-T may not reflect the complexity of human *ATM* gene as there have been no reports of either splice variants in the 5′UTR or alternative 3′ termination of the murine *ATM* mRNA (Barlow et al., 1996; Elson et al., 1996; Herzog et al., 1998). Phenotypically, *ATM*-deficient mice display a variety of growth defects, meiotic defects, immunological abnormalities, radiation hypersensitivity and cancer predisposition, similar to those seen in A-T patients, confirming the most common pleiotropic roles of ATM. In contrast, development of neurological defects in these mice is slower than clinically observed in humans. Rather than becoming ataxic, mice die of other consequences of ATM-deficiency (e.g., lymphoma). The neurological abnormalities seen in the *ATM*-mutant mice do not reflect the extent of human neurological abnormalities, thus suggesting that an alternative animal model of human A-T could be beneficial.

The domestic pig, *Sus scrofa domestica* has been recognized as a powerful and well established model for medical studies (Tumbleson and Schook, 1996; Schook et al., 2005). The pig has provided insight into various human diseases, such as cancer, diabetes and atherosclerosis, and could be a valuable model for studying A-T. Thus, in order to evaluate the appropriateness or validity of producing a porcine A-T model, we have characterized the porcine *ATM* gene both structurally and transcriptionally.

2. Materials and methods

2.1. Library screening and BAC sequencing

High-density membranes corresponding to the RPCI-44 porcine BAC library (Fahrenkrug et al., 2001) were screened by hybridization using an overgo approach (http://genome.wustl.edu/). Three probes corresponding to exons 8, 12 and 42 were designed within regions of high similarity between human and mouse sequences. Probes were labeled by a standard Klenow fill-in reaction and hybridization was performed at 65 °C for 18 h in Church buffer. Membranes were removed from hybridization solution, washed 3× in 0.1× SSC, 0.1% SDS at 60 °C and exposed to autoradiography film at −70 °C for 1–2 days. Positive clones were screened by PCR to confirm the presence of the porcine *ATM* gene using heterologous primers designed from exons 1 and 48 of the human *ATM* gene.

RPCI-44 clone 90A8 was selected and cultured for 20 h in 50 ml 2XLB supplemented with 20 μg/ml chloramphenicol. BAC DNA was purified using the NucleoBond BAC Maxi Kit (BD Biosciences, San Jose, CA). The quality and quantity of DNA were assessed by agarose gel electrophoresis and spectrophotometry. The EZ-TN™ (oriV/KAN-2) Insertion Kit (Epicentre Biotechnologies, Madison, WI) was used to produce a subclone library for transposon-mediated shotgun sequencing according to the manufacturer's recommendations. BAC DNA was sequenced bi-directionally using primers specific for the 5′ and 3′ ends of the inserted transposon. Samples were analyzed using an ABI 3730 automated capillary sequencer (Applied Biosystems, Foster City, CA). In addition, clone 90A8 was sent to Agencourt, Inc. for plasmid shotgun library construction and sequencing. Resulting sequence chromatograms were combined,

analyzed and assembled using Phred (Ewing and Green, 1998) and Phrap via Interphace (CodonCode Corporation, Dedham, MA).

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted from frozen porcine tissues of a Duroc–Berkshire cross using the TRIZOL™ according to the manufacturer's instructions (Invitrogen, Carlsbad, CA). Tissues included the brain, heart, thymus, lung, trachea, liver, spleen, kidney, adrenal gland, thyroid, stomach, intestine, colon, mesenteric and mandibular lymph nodes, uterus, ovary, testis, muscle, skin, bone marrow, spinal cord, and fat. Total RNA from the porcine primary fibroblast line PF-159 derived from a cloned fetus of female Duroc 2–14 was also obtained. In order to remove genomic DNA contamination, total mRNA samples (about 10 μg each) were treated with DNase I (Qiagen, Valencia, CA) and purified using a RNeasy Kit (Qiagen, Valencia, CA). Samples were reverse transcribed (+RT samples) with Oligo(dT)$_{12–18}$ and random primers (Stratagene, Grand Islands, NY) and Superscript II (Invitrogen, Carlsbad, CA) reverse transcriptase following the manufacturer's instructions. The controls (−RT) samples were processed identically to (+RT) samples, except the (−RT) reaction contained no reverse transcriptase.

2.3. cDNA sequencing

To determine the porcine *ATM* cDNA sequence, 0.5 kb to 1.8 kb overlapping cDNA fragments were amplified and sequenced. The porcine genomic sequence (AY587061) was used as a reference to design primers located in exons of the *ATM* gene (Table 1). The 5′ RACE analysis was performed on total RNA extracted from porcine brain and lymph nodes using the FirstChoice RLM-RACE kit (Ambion, Austin, TX) according to the manufacturer's protocol. Adaptor-ligated cDNAs were amplified by performing two rounds of PCR amplification (each for 30 cycles) using Phusion polymerase (MJ Research, Boston, MN) and then Q-Taq polymerase (Qiagen, Valencia, CA). First round PCRs were carried out with an adaptor-specific primer from the kit and a gene-specific antisense primer RACE ex8R (Table 1). Nested PCRs were performed with adaptor-specific primer and gene-specific antisense primer RACE ex7R (Table 1) derived from exon 7. Splicing of the 5′UTR was analyzed by amplifying fragments from exon 1 to exon 8 with the forward 5′-GAT TCC ACT TGG TTG TCC AGT C-3′ and reverse Race ex8R (Table 1) primers. PCR fragments were purified using a QIAquick PCR purification kit (Qiagen, Valencia, CA) and bi-directionally sequenced with PCR primers. Alternatively, if multiple amplicons were observed, the amplified fragments were cloned into the pCR®2.1-TOPO® vector system (Invitrogen, Carlsbad, CA) and then sequenced using M13 primers. DNA sequencing was performed with the ABI Prism® BigDye™ Terminator Cycle Sequencing Kit (version 3.1) (Applied Biosystems, Foster City, CA) and analyzed on an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA). Pair-wise comparisons were done using the Vector NTI 9 program. The putative promoter region was predicted with the Proscan software, version 1.7 (http://thr.cit.nih.gov/molbio/proscan/). In order to identify putative regulatory domains within the promoter, the 700-bp sequence upstream of the transcription initiation site of the pig, human and mouse *ATM* genes were aligned using DNAMAN analysis software, version 5.0 (http://www.lynnon.com). The 3′UTR was analyzed for polyA signal sites using polyAH (http://www.softberry.com). Prediction of secondary structures and calculation of free energy values of the 5′ UTRs was performed using the MFOLD program, version 3.2 developed by Zuker and Turner (Zuker, 2003; http://www.bioinfo.rpi.edu/applications/mfold).

Table 1
Primers for amplifying porcine ATM cDNA

Name	Position (mRNA)	Sequence 5′–3′	Amplicon size (bp)
ATMex1F	216	GATTCCACTTGGTTGTCCAGTC	
ATMex5R	734	CCTGGTGGCTTGTGTTGAGG	518
ATMex4F	555	AGGCACCTGATTCAAGATCC	
ATMex12R	2262	TGTGAAGAATTGGAGGCAGC	1707
ATMex11F	1819	CACCTTACGTGTTACGATGC	
ATMex19R	3205	GTATTGGACTGAGCGATGGT	1386
ATMex15F	2632	GTGTCCTTGGCTGCTATTGT	
ATMex27R	4251	ACCTCATCACCAGATGTGGA	1619
ATMex25F	3872	AGGAGGTGAGAACCGTGAAC	
ATMex36R	5671	CAGGTAACAGCTGCTGATCG	1799
ATMex35F	5600	CATGATGCTGACCTACCTGA	
ATMex44R	6781	TTCCGCCAGGCCACTTGATA	1181
ATMex43F	6631	CTTCATCAACTCGCCAGGCA	
ATMex55R	8384	AGTCCTCCACTGAGTGGCAT	1753
ATMex54F	8295	AACAGGAGACGTCAGATGGT	
ATMex64R	9499	CTCATCAAGACACGTTCAGC	1204
ATMex58F	8763	CTTGAATGGTGCACAGGAAC	
ATMex64R	9499	CTCATCAAGACACGTTCAGC	736
Race ex7R	1076	TCCATCAGTCTGTGAACAGCAT	
Race ex8R	1200	TAGCAGCCAAAGTCTTGAGGAA	

2.4. Quantitative RT-PCR of 5′UTR splice variants of ATM transcripts

For the identification and quantification of 5′UTR variants, quantitative RT-PCR analyses were performed on cDNA samples from twenty-five tissues using TaqMan 5′nuclease assays. Primers and probes were designed with Primer Express software (Applied Biosystems, Foster City, CA) to anneal with splice variant-specific sequences at exon–exon junctions. To provide more accurate splice variant discrimination, TaqMan® Probes with a non-fluorescent quencher and a minor groove binder at the 3′ end were used (Applied Biosystems, Foster City, CA). The RT-PCR was carried out with following primers and probes. The forward primer derived from the 3′-end of exon 1 (5′-CCA GGG ATT CCA CTT GGT TGT-3′) was used for amplification of all splice variants (SV). For the SV1 detection, the reverse primer annealed to the 5′-end of exon 2 (5′-TCA ACG AAT CAC TCG TTG ACT TAA A-3′) and the probe (6FAM-5′-AGT CCA AGC CGT CTT GTA-3′) hybridized across the exon 1–exon 2 junction. For the SV2, reverse primer hybridized to exon 3 (5′-GAC TCA TGG TTC AGA AGT TCA GAA AA-3′) and the probe (6FAM-5′-GTC CAG CCG TCT TGA CAA-3′) recognized the exon 1–exon 3 junction. To

detect SV3, a probe was designed to hybridize across the exon 1–exon 4 junction (TET-5′-TTC AAG ACG GCT GGG CT-3′) and a reverse primer was from exon 4 (5′-TCA GGT GCC TAA AGT TCT CAA CTG-3′). In order to estimate the total amount of *ATM* transcripts, a probe (TET-5′-CAA CAG TGG CGA GAG TT-3′) that recognized the exon 6–exon 7 junction, a sequence common to all splice variants, was used. The reverse primer was derived from exon 7 (5′-GAA TTA TTC TAG CCA CCA AAA GTC TAT TAA-3′) and the forward primer was from exon 6 (5′-TGC TGA AGG ACA TTC TTT CAG TGA-3′). The TaqMan® assay was performed using a 96-well plate format with a 20 μl reaction mix (TaqMan® Universal PCR Master Mix, No UNG, Applied Biosystems, Foster City, CA) and the ABI PRISM 7900 Sequence Detection System instrument and software (Applied Biosystems, Foster City, CA). The (+RT) samples were run in triplicate, whereas (−RT) control samples were run in duplicate. Each reaction was loaded with 25 ng of cDNA reverse transcribed from total mRNA, except for 18s rRNA where 12.5 ng per reaction gave optimal resolution.

The SV copy number per reaction was determined for each tissue and the absolute copy number was normalized for the relative amount of 18s rRNA and background subtracted using the (−RT) control reactions. To quantify the porcine 18s rRNA the TaqMan reagent Starter kit (Applied Biosystems, Foster City, CA) was used according to the manufacture's protocol, except that the probe–primer mix was diluted 10-fold. An absolute quantitation assay was used to estimate the copy number of each SV by interpolating their quantity from a standard curve. Standard curves were developed using serial dilutions (10^1 to 10^7 copies) of each SV standard. The pCR2.1 plasmids with cloned fragments of either SV 1, 2 or 3 were used as standards. These fragments were amplified as previously described from exon 1 to exon 7 of brain cDNA, cloned into the pCR®2.1-TOPO® vector system (Invitrogen, Carlsbad, CA) and were confirmed by sequencing. The molecular weight of each plasmid was used to calculate the number of plasmid molecules per dilution. A standard curve was constructed for each SV using threshold cycle (Ct) data plotted against the log copy numbers for the standard dilution sets.

3. Results and discussion

3.1. Genomic sequence

The genomic sequence of porcine *ATM* was determined by sequencing the BAC clone 90A8 from the RPCI-44 porcine BAC library (http://bacpac.chori.org/mporcine44.htm). This BAC clone contained the complete *ATM* gene. Comparison of the porcine genomic sequence with human genomic and mRNA sequences with 5′RACE analysis demonstrated the porcine *ATM* gene consisted of 65 exons (GenBank accession no. AY587061).

3.2. Characterization of ATM cDNA

In order to verify the sequence of the putative mRNA, *pATM* transcripts in several tissues were sequenced and characterized to determine the exon–intron boundaries and analyze untranslated regions. Sequence analysis of overlapping fragments, amplified from brain and mesenteric lymph node cDNAs provided a complete sequence for the coding region of the *pATM* mRNA. All exon–intron boundaries contained the consensus splice donor–acceptor sites (GT–AG). The cDNA sequences were in agreement with the putative mRNA sequence (AY587061) demonstrating the same exon–intron boundaries as predicted, with the exception of a 3 bp insertion at the junction between exons 24 and 25. The human–porcine genomic comparison predicted the intron/(exon 25) junction to be agaag/tc in pig, as it is in humans. Instead, the porcine splice acceptor site is shifted by 3 bp from exon 25 towards the intron to become ag/aagtc. Thus, the porcine protein gains an additional lysine amino acid at position 1135 in comparison with *hATM*. This insertion does not involve any known functional domain and likely does not affect any critical functions of the ATM protein.

Overall, the porcine ATM protein consists of 3058 aa and is highly homologous to the hATM protein (88% identity and 93% positive). Among all the non-primate mammalian species, the human protein has the highest homology with pATM, with identity greater than 92% in the FAT and PIK functional domains.

3.3. Analysis of 3′UTR

A human–porcine genomic sequence comparison predicted the 3′UTR to be 3060 bp long and the mRNA to be terminated after the polyadenylation (polyA) signal site at position 12,675. In addition to this site, four optimal AATAAA sites with equal or higher scores were predicted along the 3′UTR at positions 9987 nt, 11,658–11,682 nt, and 12,451 nt with the length of 3′UTRs expected to be 365 bp, 2060 bp and 2800 bp, respectively. In the 11,658 to 11,682 nt region two optimal termination sites reside only 24 bp apart, suggesting preferable termination, that would result in a transcript of 11.7 kb.

Analysis of human ESTs and cDNA libraries (Savitsky et al., 1997) revealed that even suboptimal termination sites of the human *ATM* 3′UTR were used for differential termination of *hATM* transcripts. Thus, four optimal porcine polyA signal sites are likely to lead to differential termination of porcine transcripts producing a combination of *ATM* mRNAs from 10 to 12.7 kb. Northern blot analysis of porcine fetal fibroblast total RNA demonstrated a single band of about 12.5 kb–13 kb corresponding to the longest predicted transcript (data not shown). Additional BLAST searches of available porcine resources found only one EST (BG894659) homologous to the longest 3′end. The prevalence of the longest transcript could reflect its abundance as well as the importance of the long 3′UTR for *pATM* mRNA stability. Although no direct evidence for alternate 3′UTRs in *pATM* was observed, other optimal polyA sites could still be used for differential termination, but may be expressed at levels below that of detection.

3.4. Analysis of the promoter region and 5′UTR

Efforts to define the ATM promoter region focused on a 700 bp intergenic fragment covering the first exon of the *ATM* gene and the complete 5′UTR of the *NPAT* gene. These two

genes share a bi-directional promoter in humans (Byrd et al., 1996). Database searches predicted a TATA-less porcine *ATM*/*NPAT* bi-directional promoter in this region and different consensus binding sites for transcription factors. The porcine core promoter contains two CCAAT boxes (NF1-binding sites) and three sites for GC-binding factor (GCF). Three GC boxes (SP1 binding sites), a cAMP-responsive element (CRE), which is binding site for the CRE-binding (CREB) protein, a xenobiotic responsive element (Xre), and a site for ETS transcription factors were revealed in the direction of the *pATM* gene transcription (Supplementary Fig. S1). Multiple alignments of the pig, human, and mouse promoter regions demonstrated that the common promoter components are highly conserved across species; although the human and porcine promoters shared more common regulatory elements, than when either was compared to the mouse promoter. In humans, two Sp1 binding sequences are located in the transcribed strand for the *NPAT* gene and two are located in the strand for the *ATM* gene. In pigs, there are three SP1 sites on both strands, and the pig promoter region has two additional GCF sites. Humans and pigs share the same Xre, Ets, and Cre binding sites, that together with the GCF and Sp1 sites have been shown to play important roles in basal and IR induced activity of the human promoter (Gueven et al., 2003).

The *pATM* 5′UTR was analyzed by amplification and sequencing the exon 1–exon 8 region of cDNAs obtained from 24 diverse tissues (Fig. 1). Six different fragments were identified resulting from alternative splicing of the first three exons (Fig. 1 and Supplementary Fig. S2). In addition to previously assigned exons (AY587061), a new exon 2a, adjacent to exon 2 was observed in two fragments. When the 5′UTRs of pig and human were compared, no similarities were detected between porcine leader exons 1, 2, 2a and human leader exons 1a, 1b, 2, 3. Only exons harboring the *ATM* translational start, human exon 4 and porcine exon 3, were evolutionary conserved and demonstrated a high level of homology. Thus, six different porcine 5′UTRs result from alternative splicing of pig-specific exons (GenBank accession nos: DQ646385–DQ646390).

A 5′RACE analysis of thymus and mesenteric lymph node transcripts detected four of six splice variants (SV1, SV2, SV3 and SV6). Both SV1 (DQ646385) and SV2 (DQ646386) had the same transcriptional start sites, whereas the transcriptional start in SV3 (DQ646387) was 13 bp downstream and in SV6 (DQ646390) was in the middle of exon 1 (Fig. 2). Variation in the transcription start sites may be the result of the absence of a TATA box in the *pATM* promoter, which has been shown to be important in determining the precise starting point for transcription (Lewin, 2004). The alternative splicing (AS) of *pATM* produced GC rich 5′UTRs of variable length (Fig. 1). The leader sequences were predicted to have various extensive secondary structures with free energies calculated ranging from −53.8 to −140.8; secondary structure stability of −50 kcal/mol can inhibit translation by 85–95% (Kozak, 1986). The porcine 5′UTRs also varied in the number of upstream AUG (uAUG) sites preceding the *ATM* initiator codon (Figs. 1 and 2). Of the porcine transcripts that included the putative exon 3 initiation site, the SV1 transcript had the longest 5′UTR of 464 bp and only one uAUG codon, that represents a potentially weak translational start site (Kozak, 2005) encoding a peptide (6 aa). The SV6 transcript differs from SV1 by the presence of exon 2a instead of exon 2. Among detected splice variants, the SV2 with a short and less structured 5′UTR and a single uAUG appears to be the most favorable for translation.

Splicing in *pATM* transcripts not only involves the 5′UTR but also the beginning of the coding region, however, the alternative splicing of the first coding exon of the hATM has not been documented. Three porcine transcripts without the first coding exon (exon 3) were detected, thus removing or altering the putative primary translational start site. Experimental data suggests that when the point mutation ablates the normal start codon or when it is eliminated via a change in splicing, the next downstream AUG, which had been silent, becomes the new initiating site (Kozak, 2005). Any porcine transcript carrying

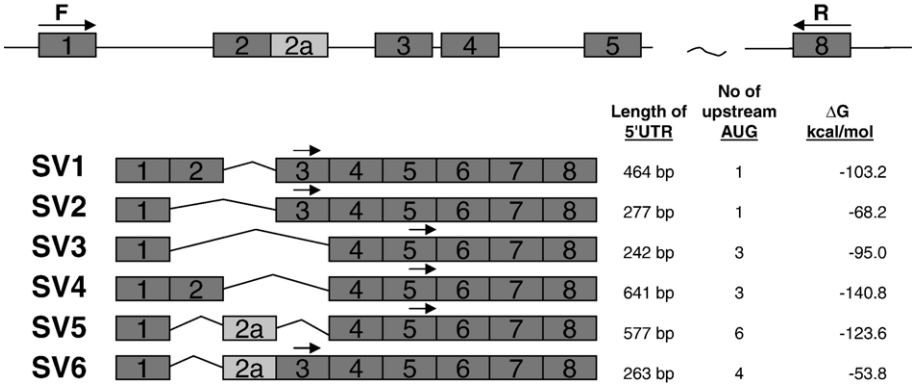

Fig. 1. The 5′UTRs of the porcine *ATM* gene. (a) A schematic representation of diverse 5′UTR of porcine ATM transcripts. The top scheme shows the genomic organization of the 5′UTR. Exons are boxed and numbered. Arrows (F, forward and R, reverse) indicate the primer positions for PCR amplification, presented in (b). Arrows above splice variants (SV) indicate either the *ATM* translation start site [in SV1 (DQ646385), SV2 (DQ646386) and SV6 (DQ646390)] or the first in frame Met codon [in SV3 (DQ646387), SV4 (DQ646388), and SV5 (DQ646389)]. The length of the 5′UTR, the free energy of the 5′UTR and the number of upstream AUGs are provided for each SV. (b) Visualization of SVs by PCR amplification of exon 1 to exon 8 fragments from cDNAs obtained from: 1, skin; 2, fat; 3, adrenal gland; 4, spinal cord; 5, muscles; 6, mandibular lymph node; 7, thyroid; 8, thymus; 9, trachea; 10, stomach; 11, brain; 12, uterus; 13, colon; 14, small intestine; 15, bone marrow; 16, heart; 17, kidney; 18, ovary; 19, bladder; 20, lung; 21, spleen; 22, testis; 23, mesenteric lymph node; 24, liver; M, 100 bp ladder.

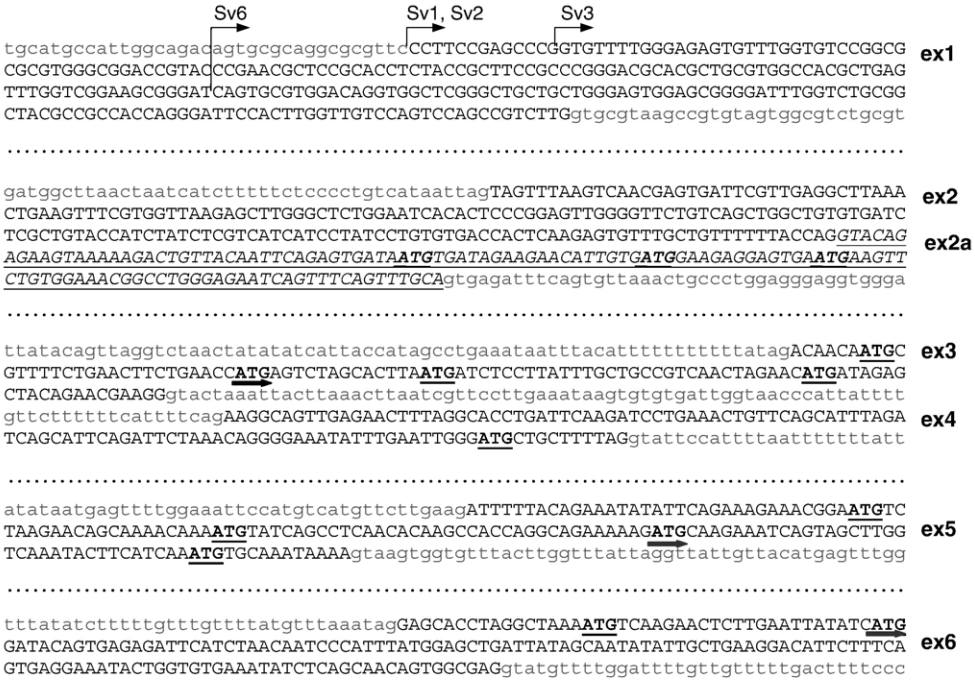

Fig. 2. Exonic nucleotide sequences spanning the 5′ end of the porcine ATM gene. Exonic sequences are in capital letters. Exon 2a is italicized and underlined. Only partial intronic sequences are shown. AUG codons are in bold and underlined. The AUG start codon of the ATM major ORF is underlined with a thick black arrow, whereas in frame AUG codons in exons 5 and 6 are underlined with thick gray arrows. Bent arrows denote transcriptional start sites determined by 5′ RACE.

the first coding exon (exon 3) would be translated into a protein with an N-terminus, similar to the hATM N-terminus and be considered as a 'normal' pATM protein. Thus, translation from the *ATM* transcripts lacking exon 3 could start from the next in frame methionine (Met) residue downstream of the *ATM* initiator codon giving rise to a shorter protein that would be missing 94 or more amino acids at the normal ATM N-terminus (Figs. 2 and 3). The hypothetical 'shorter' protein lacks the N-terminus substrate binding domain (82 to 89 aa) (Fig. 3), which is involved in activation of the p53 and BRCA1 pathways in human. It has been shown that in humans the deletion of only this domain diminished ATM involved cell responses to DNA damage caused by bleomicin (Turenne et al., 2001). Since the alteration of the N-terminal part of the hATM protein impacts or blocks ATM activity in humans, the functionality of transcripts lacking the N-terminus in the pig remains unclear.

In *ATM* transcripts missing exon 3, the 5′UTRs contained numerous uAUGs [three in SV3 and SV4 (DQ646388), six in SV5 (DQ646389)] preceding first *ATM*-in-frame Met codon (Figs. 1 and 2). Although each uAUGs does not have an optimum context (Kozak, 2005), their number could considerably slow down translation from *ATM* in-frame AUG codons or abolish it completely. Studies have shown that some cDNA sequences containing many uAUGs in their 5′UTR do not correspond to functional mRNAs (Bevitt et al., 2005; Kozak, 2005).

Altogether, the SVs missing the normal translation start are unlikely to be translated and they could serve as a ballast to slow or dampen total *ATM* translation. Only 5–10% of eukaryotic mRNAs have long (more than 100 nt) highly structured 5′UTRs that contain uATG codons. All of them are regulatory proteins and probably designed to be poorly translated (Kozak, 2005). The long 5′UTR with numerous uAUGs are common in proto-

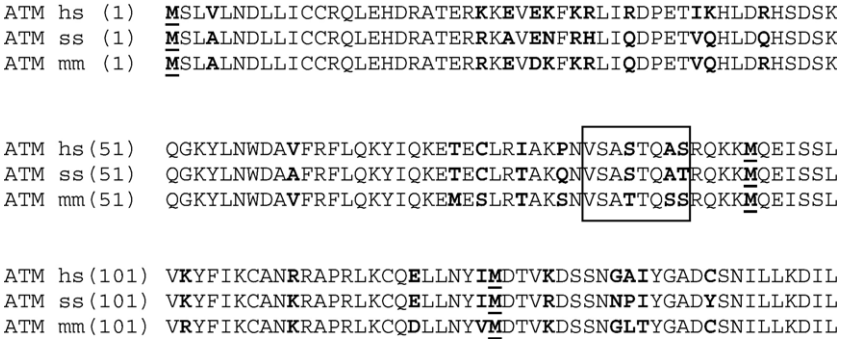

Fig. 3. Comparison of N-terminus of human (hs), pig (ss) and mouse (mm) ATM proteins. Amino acids that are not conserved in all three species are in bold. Methionine residues are in bold and underlined, and the N-terminus substrate binding domain is boxed.

oncogenes, genes for transcription factors, receptor proteins, signal transduction components, and proteins involved in immune response (Churbanov et al., 2005). Rare mRNAs with highly unfavorable initiation sites encode potent regulatory proteins, such as ATM, suggesting that a weak context might be an occasional ploy to modulate the yield of proteins that could be harmful if overproduced (Churbanov et al., 2005).

3.5. Tissue specific expression

Splicing provides a mechanism for regulating the level of translation and RNA processing in various tissues during development. It could potentially regulate the translation rate of *ATM* transcripts in different tissues by producing transcripts of differential translational potential. To determine the relative abundance for each SV, we conducted quantitative analyses of SV1 SV2, and SV3 in various porcine tissues. Our analysis was limited to these three transcripts since they represented the major SVs. Optimized SV1, SV2, SV3 assays were used to quantify cDNA from 24 different adult porcine tissues and primary fetal fibroblasts. Each SV was detected in all tissues analyzed, though in different proportions and abundances (Table 2). The predominant form of *ATM* transcript in these tissues was the SV1 transcript (average 74% and never less then 60% of the total *ATM* transcripts detected in each tissue). Both the SV2 and SV3 transcripts were detected at low levels in most of the tissues, the SV2 variant had a higher expression in fibroblasts, fat, spleen, adrenal gland, lung and trachea and in general was more abundant in the connective tissue containing organs (Table 2). The SV3 transcript had higher expression levels in fat, bladder, spinal cord, brain, colon, small intestine, and the uterus (Table 2).

Table 2
Relative abundance of the three major ATM transcripts amongst 25 tissues

	% of ATM transcripts		
	SV1	SV2	SV3
Skin	72.58	17.81	9.62
Fibroblast	67.75	26.04	6.20
Spinal cord	78.30	10.14	11.56
Brain	83.42	7.87	8.71
Adrenal gland	81.79	13.30	4.91
Lymph node, mandibular	85.02	8.04	6.94
Lymph node, mesenteric	67.09	13.22	19.69
Spleen	70.99	18.65	10.36
Ovary	79.74	8.70	11.56
Uterus	80.01	7.91	12.08
Testis	84.73	10.56	4.71
Fat	59.17	21.39	19.43
Bone marrow	74.38	13.58	12.03
Muscle, heart	74.19	9.98	15.83
Muscle, skeletal	84.75	10.14	5.11
Kidney	72.62	15.79	11.58
Thymus	82.06	11.53	6.41
Thyroid	81.61	12.99	5.39
Lung	68.80	20.78	10.41
Trachea	68.98	24.23	6.79
Small intestine	60.16	9.07	30.78
Colon	64.64	12.04	23.32
Stomach	86.31	6.38	7.32
Liver	65.71	21.28	13.01
Bladder	61.70	10.28	28.02
Average	74.26	13.67	12.07

To estimate the total amount of *ATM* transcripts in individual tissues, the copy number of transcripts with the exon 6–exon 7 junction was determined. Additionally we estimated the sum of each variant (SV1, SV2 and SV3). Those tissues with the highest total expression of *ATM* were lung, trachea, fibroblasts, bladder, fat, adrenal gland, spleen, thyroid, mesenteric lymph node, spinal cord, brain, ovary, uterus, and testis. We observed that the adrenal gland, thymus, mesenteric lymph node, spleen, colon, small intestine and thyroid contained more copies of ATM transcripts than the sum of SV1, SV2 and SV3 transcripts. This difference was attributed to additional transcripts (SV4, SV5 and SV6) that were undetectable by the specific SV1, SV2, and SV3 TaqMan assays.

Overall, higher total porcine *ATM* mRNA levels were observed in tissues with higher intrinsic proliferation, hormone-producing tissues, and in the tissues of neural origin. The onset of A-T disease includes the presence of a small embryonic thymus, hypoplasia of lymphoid tissues, immunodeficiency, premature aging of skin, cerebellar degeneration, spinal cord neuronal degeneration, endocrine abnormalities and incomplete sexual maturation (McKinnon, 2004). We observed the highest expression of total *ATM* mRNA in those pig tissues that are most affected by the loss of the ATM protein in A-T patients.

ATM is ubiquitously expressed in tissues consistent with the cardinal feature of a housekeeping gene. DNA damage activates the preexisting ATM protein stimulating the rapid cellular responses. Tissues that are more susceptible to DNA damage demonstrate a higher basal level of *ATM* mRNA and protein, thus providing a faster response. Further increases in protein activity are due to additional protein production and time dependent activation of transcription. In some tissues, the amount of protein raises without an apparent increase in the mRNA (Fang et al., 2001), indicating the regulation of the protein synthesis at the level of translation initiation or due to the shifting in transcription from less translationally available mRNAs to more active transcripts (without detectable changes in the total amount of mRNA). The ATM promoter is activated in response to IR in tissues with low basal ATM expression until it reaches saturation. Since the *ATM* mRNA pool consists of a variety of SV transcripts of approximately the same length, the total number of ATM transcripts could mask the amount of the translationally available transcripts. Another possible mechanism of posttranscriptional activation could involve autosplicing of 5′UTR exons. Shortening the 5′UTR and reducing secondary structure complexity as well as the number of uAUGs could provide transcripts more favorable for translation.

Finally, an extremely important role of alternative splicing is expanding protein diversity. Approximately 75% of human genes are alternatively spliced (Johnson et al., 2003) and the species-specific alternative splicing observed in human (Savitsky et al., 1997) and pig *ATM* represents the most common type of splicing in mammalian species (Caceres and Kornblihtt, 2002; Pan et al., 2005). The AS events that affect the

coding sequence change protein functions and thus provide proteomic diversity. The multiple splice variants of porcine *ATM* gene could possibly be translated into various proteins with yet unknown functions. Overall, the extensive alternative splicing of porcine *ATM* gene resembles complex AS in humans and could impact in differences observed between mice and humans with regards to the onset of A-T.

4. Conclusions

Progress toward useful interventions for ATM pathology progression in human patients would be advanced given a suitable animal model of the disease. To date, murine models have failed to recapitulate the human disorder and have provided only limited insights to therapies or pharmacological treatments. The biological and physiological similarities between humans and pigs have resulted in a number of porcine biomedical models that represent the human pathophysiology. Thus, we have chosen to investigate the potential use of a porcine model for ATM. Obligatory to the generation of such a model via gene manipulation are the development of molecular resources to construct a model and specifically a more complete understanding of the similarities between the human and porcine *ATM* genes. Genomic sequencing of the porcine *ATM* gene reveals that the similarity between the two species is greater than that between either and the mouse suggesting that from a molecular standpoint a porcine model may be promising. Furthermore, transcript analysis of the porcine gene has provided insights into alternative splicing events that may affect gene function in vivo. These data indicate that the development of gene constructs to recapitulate the human disease model need to be considered carefully to generate a model that would most likely elicit the affects of currently understood human genetic mutations.

Acknowledgements

This work was supported in part by USDA/NRI-CSREES grant AG2001-35205-11698 and USDA-ARS Cooperative Agreement AG58-5438-2-313 and the AT Children's Project.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2007.08.014.

References

Bakkenist, C.J., Kastan, M.B., 2003. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature 421, 499–506.

Ball, L.G., Xiao, W., 2005. Molecular basis of ataxia telangiectasia and related diseases. Acta. Pharmacol. Sin. 26, 897–907.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J.N., Ried, T., Tagle, D., Wynshaw-Boris, A., 1996. Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell 86, 159–171.

Bevitt, D.J., Li, Z., Lindrop, J.L., Barker, M.D., Clarke, M.P., McKie, N., 2005. Analysis of full length ADAMTS6 transcript reveals alternative splicing and a role for the 5′ untranslated region in translational control. Gene 359, 99–110.

Byrd, P.J., Cooper, P.R., Stankovic, T., Kullar, H.S., Watts, G.D., Robinson, P.J., Taylor, M.R., 1996. A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies. Hum. Mol. Genet. 5, 1785–1791.

Caceres, J.F., Kornblihtt, A.R., 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet. 18, 186–193.

Chen, G., Lee, E., 1996. The product of the ATM gene is a 370-kDa nuclear phosphoprotein. J. Biol. Chem. 271, 33693–33697.

Churbanov, A., Rogozin, I.B., Babenko, V.N., Ali, H., Koonin, E.V., 2005. Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. Nucleic Acids Res. 33, 5512–5520.

Elson, A., Wang, Y., Daugherty, C.J., Morton, C.C., Zhou, F., Campos-Torres, J., Leder, P., 1996. Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. Proc. Natl. Acad. Sci. U. S. A. 93, 13084–13089.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Fang, Z.M., Lee, C.S., Sarris, M., Kearsley, J.H., Murrell, D., Lavin, M.F., Keating, K., Clarke, R.A., 2001. Rapid radiation-induction of ATM protein levels in situ. Pathology 33, 30–36.

Fahrenkrug, S.C., Rohrer, G.A., Freking, B.A., Smith, T.P.L., Osoegawa, K., Shu, C.L., Catanese, J.J., de Jong, P.J., 2001. A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. Mamm. Genome 12, 472–474.

Gilad, S., Khosravi, R., Shkedy, D., Uziel, T., Ziv, Y., Savitsky, K., Rotman, G., Smith, S., Chessa, L., Jorgensen, T.J., Harnik, R., Frydman, M., Sanal, O., Portnoi, S., Goldwicz, Z., Jaspers, N.G., Gatti, R.A., Lenoir, G., Lavin, M.F., Tatsumi, K., Wegner, R.D., Shiloh, Y., Bar-Shira, A., 1996. Predominance of null mutations in ataxia-telangiectasia. Hum. Mol. Genet. 5, 433–439.

Graveley, B.R., 2001. Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17, 100–107.

Gueven, N., Keating, K., Fukao, T., Loeffler, H., Kondo, N., Rodemann, H.P., Lavin, M.F., 2003. Site-directed mutagenesis of the ATM promoter: consequences for response to proliferation and ionizing radiation. Genes, Chromosomes Cancer 38, 157–167.

Gueven, N., Fukao, T., Luff, J., Paterson, C., Kay, G., Kondo, N., Lavin, M.F., 2006. Regulation of the Atm promoter in vivo. Genes, Chromosomes Cancer 45, 61–71.

Herzog, K.H., Chong, M.J., Kapsetaki, M., Morgan, J.I., McKinnon, P.J., 1998. Requirement for Atm in ionizing radiation-induced cell death in the developing central nervous system. Science 280, 1089–1091.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., Shoemaker, D.D., 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302, 2141–2144.

Kastan, M.B., Lim, D.S., 2000. The many substrates and functions of ATM. Nat. Rev., Mol. Cell Biol. 1, 179–186.

Kishi, S., Lu, K.P., 2002. A critical role for Pin2/TRF1 in ATM-dependent regulation. Inhibition of Pin2/TRF1 function complements telomere shortening, radiosensitivity, and the G(2)/M checkpoint defect of ataxia-telangiectasia cells. J. Biol. Chem. 277, 7420–7429.

Kozak, M., 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. U. S. A. 83, 2850–2854.

Kozak, M., 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37.

Lakin, N.D., Weber, P., Stankovic, T., Rottinghaus, S.T., Taylor, A.M., Jackson, S.P., 1996. Analysis of the ATM protein in wild-type and ataxia telangiectasia cells. Oncogene 13, 2707–2716.

Lavin, M.F., Shiloh, Y., 1997. The genetic defect in ataxia-telangiectasia. Annu. Rev. Immunol. 15, 177–202.

Lavin, M.F., Scott, S., Gueven, N., Kozlov, S., Peng, C., Chen, P., 2004. Functional consequences of sequence alterations in the ATM gene. DNA Repair (Amst). 3, 1197–1205.

Lewin, B., 2004. GenesVIII. Pearson Education, Inc., New Jersey (USA), pp. 597–631.

McKinnon, P.J., 2004. ATM and ataxia telangiectasia. EMBO Rep. 5, 772–776.

Modrek, C.J., Lee, C.J., 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat. Genet. 34, 177–180.

Pan, Q., Bakowski, M.A., Morris, Q., Zhang, W., Frey, B.J., Hughes, T.R., Blencowe, B.J., 2005. Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet. 21, 73–77.

Pandita, T.K., 2002. ATM function and telomere stability. Oncogene 21, 611–618.

Pandita, T.K., Westphal, C.H., Anger, M., Sawant, S.G., Geard, C.R., Pandita, R.K., Scherthan, H., 1999. Atm inactivation results in aberrant telomere clustering during meiotic prophase. Mol. Cell Biol. 19, 5096–5105.

Pandita, T.K., Lieberman, H.B., Lim, D.S., Dhar, S., Zheng, W., Taya, Y., Kastan, M.B., 2000. Ionizing radiation activates the ATM kinase throughout the cell cycle. Oncogene 19, 1386–1391.

Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y., Rosenthal, A., 1997. Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. Genome Res. 7, 592–605.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., 1995. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268, 1749–1753.

Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Sartiel, A., Rosenthal, A., Elroy-Stein, O., Shiloh, Y., Rotman, G., 1997. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of ATM gene expression. Nucleic Acids Res. 25, 1678–1684.

Schook, L., Beattie, C., Beever, J., Donovan, S., Jamison, R., Niemi, S., Rothschild, M., Rutherford, M., Smith, D., Zuckerman, F., 2005. Swine in biomedical research: creating the building blocks of animal models. Anim. Biotechnol. 16, 183–190.

Shiloh, Y., 2003. ATM and related protein kinases: safeguarding genome integrity. Nat. Rev. Cancer 3, 155–168.

Stewart, G.S., Last, J.I., Stankovic, T., Haites, N., Kidd, A.M., Byrd, P.J., Taylor, A.M., 2001. Residual ataxia telangiectasia mutated protein function in cells from ataxia telangiectasia patients, with 5762ins137 and 7271T → G mutations, showing a less severe phenotype. J. Biol. Chem. 276, 30133–30141.

Tumbleson, M.E., Schook, L.B., 1996. Advances in Swine in Biomedical Research. Plenum Press, New York.

Turenne, G.A., Paul, P., Laflair, L., Price, B.D., 2001. Activation of p53 transcriptional activity requires ATM's kinase domain and multiple N-terminal serine residues of p53. Oncogene 20, 5100–5110.

Yang, J., Xu, Z.P., Huang, Y., Hamrick, H.E., Duerksen-Hughes, P.J., Yu, Y.N., 2004. ATM and ATR: sensing DNA damage. World J. Gastroenterol. 10, 155–160.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.