The Characterization of *atm-1* in *Caenorhabditis elegans*

by

Chin Huang

B.Sc. Honours, Simon Fraser University, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2011

Abstract

Loss of function of ATM (ataxia-telangiectasia, mutated) was discovered to be the genetic cause of the human disorder Ataxia-Telangiectasia (A-T). A-T is a rare, autosomal recessive human disorder that presents with multiple symptoms, including ataxia, telangiectasia, neurodegeneration, immunodeficiency, radiosensitivity, genomic instability, and a predisposition to developing cancer. The predicted ATM orthologue in *Caenorhabditis elegans* (*C. elegans*) is the gene *atm-1*. *C. elegans* is a model system that is easily amendable to molecular and genetic research. In this thesis, I have characterized the structure of *atm-1* and examined its mutant phenotype. The predicted gene model for *atm-1* was smaller than known orthologues. Using cDNA analysis and sequencing, I have shown that the three gene predictions *atm-1, K10E9.1,* and *F56C11.4* make up the complete coding region. I observed differences from the original prediction including unpredicted splice sites and exonic sequences. A single *atm-1* mutant allele, *gk186,* exists. I have shown that the *gk186* deletion produces two alternative transcripts, one of which retains the protein kinase domain. Thus, in contrast to previous expectations that this mutation generated a loss of function phenotype, it is likely that it produces a protein with kinase function. The phenotype of the mutant is radiation sensitivity, as expected, but is not as sensitive as *brd-1*, another loss of function mutant in the same double strand break repair pathway. This result is consistent with partial function retention of *atm-1*. In addition, mutant *atm-1* showed chromosomal instability in X-chromosome loss and subsequent sterility. Some of the mutational events were captured as lethals using the *eT1* balancer system documenting a mutator phenotype of *atm-1*. Characterization of *atm-1* gene function in *C. elegans* may provide additional information about its function in other organisms including humans.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A-T (Ataxia-Telangiectasia)

ATM (Ataxia-Telangiectasia, Mutated)

PI3K (Phosphatidylinositol 3-kinase)

PIKK (PI3K- Related protein kinase)

DSBs (Double strand breaks)

DDR (DNA damage response)

IR (Ionizing radiation)

NHEJ (Non-homologous end joining)

HR (Homologous recombination)

ESTs (Expressed sequence tags)

cDNA (Complementary DNA)

mRNA (messenger RNA)

NCBI (National Center for Biotechnology Information)

CDD (Conserved Domain Database)

Chr (Chromosome)

Dpy (Dumpy)

Unc (Uncoordinated)

pre-mRNA (primary mRNA)

NMD (Nonsense mediated decay)

PTC (Premature termination codon)

EJC (Exon-junction complexes)

TMP (4,5',8-trimethylpsoralen)

UV (Ultraviolet)

RNAi (RNA interference)

HU (Hydroxyurea)

EMS (Ethyl methanesulfonate)

SAGE (Serial analysis of gene expression)

RACE (Rapid amplification of cDNA ends)

Grays (Gy)

Acknowledgements

My deepest gratitude goes to Dr. Ann Rose. She has been an extraordinary supervisor, mentor, and most importantly, friend. Thank you for supporting me through the highs and the lows.

I would like to thank to my lab mates Dr. Nigel O'Neil, Dr. Martin Jones, Jessica McLellan, George Chung, Sanja Tarailo, Shir Hazir, Zoe Lohn, and Domena Tu for insightful contributions and moral support. As well, thanks to my collaborators Dr. David Baillie and Shu Yi Chua. Dr. Allan Mah for feedback and comments.

I'd also like to thank my committee members, Dr. Jan Friedman and Dr. Terry Snutch, for their curiosity, criticism, and input in the project.

Thank you Mathias, Candice, and Courtney. You've helped more than you could ever imagine.

And lastly, my family for not trying to understand science but appreciating the hard work that it entails.

Dedication

This thesis, all my hard work, and all my accomplishments are dedicated to my grandmother.

1. Introduction

1.1 Introduction to Ataxia-Telangiectasia

Ataxia-Telangiectasia (A-T) is a rare, autosomal recessive human disorder that presents with multiple symptoms [1-4]. Suspicion and diagnosis of A-T occurs at an early age when ataxia, lack of motor control, hinders the child's ability to walk normally. Ataxia progressively worsens and eventually immobilizes the child [4]. Telangiectasia, which is the constant dilation of the blood vessels of the eyes and face, ensues within the first decade of life [4]. In addition to the visible features, A-T patients have varying degrees of progressive neurodegeneration, immunodeficiency, abnormal or absent thymus, insulin-resistant diabetes, and are susceptible to developing cancer [4]. On a cellular level, their cells are radiosensitive, have cell-cycle checkpoint defects, and display chromosomal instability [1-5]. These chromosome defects are visible cytologically and are characterized by shortened telomeres, chromosomal rearrangements, and telomere fusions [5-9].

1.1.1 The molecular cloning of the Ataxia-Telangiectasia mutation - ATM

ATM (Ataxia-Telangiectasia, Mutated), was discovered by positional cloning, to be the genetic cause of A-T [10, 11]. ATM is a member of the phosphatidylinositol 3-kinase (PI3K) – related protein kinase (PIKK) family. The PIKK proteins functions as serine/threonine protein kinases [2, 3, 5, 12, 13]. In response to double-strand breaks (DSBs) in DNA, inactive ATM dimers localize to the break and disassociate into active monomers. Activation of ATM is facilitated by the MRE11-RAD50-NBS1 complex [2, 12-15]. The DNA-associated and active ATM then subsequently phosphorylates up- and down-stream targets that amplify the damage signal and initiate specific cell-cycle checkpoints, the DNA damage response, and the apoptosis pathways to appropriately address the DSB [2, 13]. The current understanding of activation and signal transduction pathways of ATM are often reviewed in the literature [2, 3, 12, 13].

1.1.2 The importance of studying ATM

The study of ATM is important because of the severe clinical manifestations of A-T patients. Moreover, ATM appears to be involved in a wide range of biological processes. Studying ATM has improved our fundamental understanding of intracellular signalling transduction and control. It has shed light on how molecules are able to "sense" specific cellular situations and facilitate post translational modifications in a spatial and temporal manner to elicit specific responses. ATM has also provided insight into how misregulation of signalling pathways, such as those involved in the DNA damage response (DDR) and cell-cycle checkpoints, can lead to tumorigenesis [16, 17]. Importantly, ATM facilitates DSB repair, a process essential for genomic stability and immunoglobulin generation, therefore this work will contribute to our understanding of cancer and immunity [2, 13].

1.2 Double-strand break repair

DSBs are complete breaks in DNA that can occur naturally, like those required for proper meiosis, or from external sources, such as ionizing radiation (IR) [18, 19]. As a result of improper DNA repair, DSBs can lead to genomic lesions including deletions, translocations, and fusions. These large scale genomic events can also lead to genomic instability which is a hallmark of cancer [16, 17]. DSBs are typically repaired in one of two ways: non-homologous end joining (NHEJ) or homologous recombination (HR) [18, 19]. During NHEJ, the nucleotides at the ends of the DSB are processed before ligation. This process can result in gains or losses of nucleotides at the break point resulting in mutations. HR on the other hand is a more accurate form of repair because it takes advantage of homologous sequences as template, typically from a sister chromatid, to direct repair.

1.3 Introduction to splicing and alternative splicing

Splicing is the process of removing introns from primary mRNA (pre-mRNA) transcripts to concatenate coding exons to ensure proper translation of the protein. The splicing machinery (spliceosome) is a large nucleoprotein complex comprised of five small nuclear RNAs and about 150 proteins. The spliceosome recognizes splice site sequences at the exon-intron boundaries and catalyzes the removal of introns while joining the adjacent

exons. Accessory splicing proteins aid in splicing by recognizing enhancer and suppressor sites associated with their particular splice site [20].

Alternative splicing is the processing of a single pre-mRNAs to generate different coding forms (isoforms) of mRNA [20, 21]. This allows a single gene to produce variants of a protein with different functions. There are several mechanisms by which alternative splicing can occur (Figure 1.1). The most common form of AS is exon skipping which is the exclusion of an exon [21]. The two flanking exons of the excluded exon are then spliced together. Another type of alternative splicing is the usage of alternative 5' and 3' splice sites that are not typically recognized for generating the native transcript. Other possibilities of alternative splicing include intron retention, mutually exclusive exons, alternative 5' starts, alternative 3' poly adenylation sites, and trans-splicing from two different mRNAs (not shown in Figure 1.1) [21, 22]. Because of alternative splicing, the number of proteins types produced in an organism is much larger than the number of genes in the genome [21].

1.4 Nonsense mediated decay is a quality control mechanism for mRNA

Nonsense medicated decay (NMD) is a surveillance process that recognizes premature termination codons (PTCs) in mRNA and initiates degradation of the aberrant transcript. NMD is important because translation of aberrant transcripts results in truncated proteins that could disrupt normal protein function; dysfunctional proteins with a dominant-negative function or a deleterious gain-of-function. Current studies suggest a model that requires the aberrant transcript be initially translated by the ribosomal complex. The first pass of translation by the ribosomal complex displaces exon-junction complexes (EJC) laid down during pre-mRNA splicing. If a PTC is present, then EJC downstream of the PTC will not be displaced from the pre-mRNA and the presence of EJC signals other factors to initiate degradation of the aberrant transcript [23].

1.5 National Center for Biotechnology Information

The National Center for Biotechnology Information (NCBI) is a central database that houses raw data and provides tools to help scientists access and utilize the information in a user friendly way. Information is constantly being added and NCBI is constantly updating its

website and providing new ways for accessing and interpreting new datasets [24]. One tool that is provided is HomoloGene. This tool automatically identifies and compares homologous and paralogous genes identified in all the sequenced eukaryotes in their database. HomoloGene reports also contain other information that is gathered through their other tools and resources. The Conserved Domain Database (CDD) is another tool that takes all the protein domains from NCBI databases and consolidates them into one searchable database. The Conserved Domain Search allows users to input protein sequences which are then compared to the CDD to identify known protein domains within the query sequence.

1.6 *Caenorhabditis elegans*

The microscopic nematode, *Caenorhabditis elegans* (*C. elegans*), was first proposed as a model organism by Sidney Brenner to understand mammalian nervous systems in a biologically relevant but simplified organism [25]. Since then, work with *C. elegans* has produced a complete high quality genomic sequence, a complete cell lineage map, and a large community dedicated to the use of *C. elegans* as a genetically tractable model organism. Some of the advantages of *C. elegans* as a model system include: a short three and half day life cycle, a relatively simple and transparent morphology, its ability to be easily amended to genetic and high throughput technologies, its informative phenotypes, and its ease of cultivation in the laboratory [26, 27]. Another advantage of *C. elegans* is that its genes are highly conserved with those of higher eukaryotes, including humans. Thus genetic and molecular studies can have a direct impact on the understanding of human biology and disease [28].

C. elegans are typically hermaphroditic, capable of propagating the species in a clonal fashion, and capable of producing approximately 300 progeny through self-fertilization. Males occur naturally at an approximately 0.1% frequency and are hemizygous for the X chromosome as a result of X nondisjunction or loss [29]. Importantly, the presence of male *C. elegans* allows for sexual reproduction and genetic manipulation through genetic crosses. Developmentally, worms hatch from eggs and develop through four larval stages (L1 through L4) before becoming adults. Interestingly, the hermaphrodite gonad produces oocytes in a production-line fashion starting with mitotically dividing cells originating at the distal end of

the gonad (Figure 1.2). The cells then progress into the transition zone where these cells become meiotic. Mature oocytes are fertilized by sperm as they pass through the spermatheca and are then expelled through the vulva [25-27].

1.6.1 The genomic translocation, *eT1,* is a genetic balancer system

One of the advantages of working with *C. elegans* is the existence of a well characterized system that can be used to study mutations. *eT1* is a stable reciprocal translocation of the *C. elegans* chromosomes (Chr) III and V [30-32]. Physically, the left portion of Chr III is attached to the left portion of Chr V and the right portion of Chr III is attached to the right portion of Chr V; named *eT1(III) and eT1(V)*, respectively (Figure 1.3). *eT1(III)* pairs with the normal Chr III and *eT1(V)* pairs with the normal Chr V during synapsis. The regions lacking homology during synapsis are recombinationally suppressed. These include the translocated regions of *eT1(III), eT1(V)*, the right portion of the normal Chr III, and the left portion of normal Chr V. The *eT1* balancer system is maintained as a heterozygote and contains the mutations *dpy-18* and *unc-46* in the recombinationally suppressed regions of the normal chromosomes (Figure 1.3). These visible markers allow the different chromosomes to be followed by observing the phenotypes of the progeny produced by a self-fertilizing *eT1* heterozygote. Heterozygous *eT1* worms produce a ratio of 4:1:1 wildtype, to Dumpy Uncoordinated (DpyUnc), to Unc-36 worms. Unc-36 worms are *eT1* homozygotes because the breakpoint in *eT1* disrupts *unc-36* on Chr III. A large portion of the *eT1* progeny arrest as embryos due to aneuploidy (10/16) (Table 2.1).

The *eT1* system is useful because it has been used to capture lethal mutation events that occur within the recombination suppressed regions of normal Chr III and Chr V. These mutagenesis screens have been used to obtain and characterize mutational events in addition to determining forward mutation frequencies [30, 31, 33, 34]. Mutations in essential genes located on Chr III or Chr V are easily identified because the mutations will cause lethality to progeny homozygous for the normal chromosomes, which are marked with *dpy-18* and *unc-46*. Therefore, failure of worms to produce progeny with DpyUnc worms indicates that a lethal mutation has occurred and been captured. The *eT1* balancer also facilitates the

maintenance of mutations because they are recombinationally suppressed in their respective regions of Chr III or V.

1.6.2 Curated *C. elegans* data are available at WormBase.org, an online database

The nearly complete sequence of *C. elegans* was published in 1998 by The *C. elegans* Sequencing Consortium, and the final gaps of the approximate 100 megabase genome filled in 2002 [35, 36]. The sequencing data are available at WormBase.org, a public database that aims to provide the scientific community with a curated resource based on published literature. In addition to the complete genomic sequence, WormBase.org also contains gene models, gene expression patterns, numerous other resources and tools, and the complete genomic data corresponding to four other *Caenorhabditis* species for comparative genomics [37].

The accuracy and integrity of the information on WormBase is achieved by integrating published data with bioinformatically generated information. An example of this is the curation of predicted gene models with the use of Expressed Sequence Tags (ESTs). First, programs like GeneFinder use known characteristics of gene structure to scan the genomic sequence and predict open reading frames (protein coding genes). Data such as ESTs can then be used to validate the predicted models. ESTs are short sequence reads obtained from the ends of vectors cloned with complementary DNA (cDNA). cDNAs are reverse transcribed from messenger RNA (mRNA), and thus sequencing obtained from the cDNA inserts represents exon sequences from genes and can support predicted protein coding gene models. Because only the 5' and 3' ends are sequenced, the ESTs represent the starts and stops of protein coding genes, while the remainder of the model is inferred. Sequencing the full cDNA would identify the complete gene structure [38]. A recent large scale example of curating gene models can be seen in the modENCODE project for *C. elegans* [39]. Because WormBase is periodically updated, every 5th data release is frozen to allow researchers to cite and reproduce research without having to worry about slight changes in annotation or sequence [37].

1.6.3 *atm-1* in *C. elegans*

atm-1 is a predicted homolog of ATM in *C. elegans* [40]. Based on the WormBase version WS205, *atm-1* produces a 5,431 base pair (bp) primary RNA, which is spliced into a 1,950bp mRNA. The *atm-1* mRNA is predicted to encode for a 649 amino acid (AA) protein. On the *C. elegans* physical map, *atm-1* is located on Chr I from 203,037bp to 208,467bp (Figure 1.4). One mutant strain exists for *atm-1*, named VC381. VC381 was created by the *C. elegans* Gene Knockout Consortium using the mutagen 4,5',8-trimethylpsoralen (TMP) activated by ultraviolet light (UV) [41]. The result is a 548bp deletion in *atm-1* that removes the 3' splice site of the 1st intron named *gk186* (Figure 1.4). Because the 3' intron splice site is lost in *atm-1(gk186)*, exon 2 lacks the required splice recognition sequences that are required for correct splicing. The lost splicing signal would most likely result in exon skipping of exon 2, which means splicing occurs from exon 1 to exon 3. If this is the case, the resulting transcript would result in a frameshift mutation producing multiple PTC. The protein would be truncated and missing the functional kinase domain. This would likely lead to a non-functional ATM-1 protein [42].

There have been numerous studies involving *atm-1,* either with the *gk186* allele or the use of RNA interference (RNAi) [42-46]. Garcia-Muse and Boulton found that *atm-1*(RNAi) did not elicit cell cycle arrest or recruit ATL-1 when replication fork stalling was induced VIA hydroxyurea (HU), indicating that *atm-1* does not participate in replication fork stalling resolution [44]. The authors also found that after a dose of 75 Grays (Gy) of IR, *atm-1* did not prevent ATL-1 or RPA-1 formation, indicating that *atm-1* is not responsible for recruiting these proteins to DSBs. After 75Gy IR, *atm-1*(RNAi) mutants were unable to induce cell cycle arrest or apoptosis, which contrasts wildtype mitotic and pachytene germ cells respectively, indicating that *atm-1* is required for proper DNA damage, cell cycle checkpoint, and apoptosis response to DSBs. Weidhaas *et al* (2006) showed that *atm-1(gk186)* mutants had an increased abnormal vulva phenotype when subjected to 400Gy IR treatment, suggesting that *atm-1* is involved in DSB response in the authors' model of preventing reproductive cell death [46]. Stergiou *et al* (2007) demonstrated that *atm-1(gk186)* mutants had defective apoptosis induction at lower doses of IR treatment (15Gy and 30Gy), which is consistent with Garcia-Muse and Boulton [42]. However, at a high dose (120Gy), *atm-*

1(gk186) resembled wildtype worms in apoptosis induction, suggesting that other repair pathways may compensate or aid in DSB repair at high IR dosages. Stergiou *et al* also report that *atm-1(gk186)* mutants are sensitive to UV-C radiation. A study by Lee *et al* (2010), consistent with Garica-Muse and Boulton and Stergiou *et al*, found that *atm-1(gk186)* is defective in arresting mitotic germ cells in mutant worms post 75Gy IR treatment [45]. Also consistent with Garcia-Muse and Boulton, mutant *atm-1* worms were not sensitive to HU treatment. A conflicting result from Bailly *et al* showed that L4 *atm-1(gk186)* mutants were not sensitive to IR. However, L1 worms were sensitive to IR treatment, suggesting differences in function for ATM-1 in meiotic verses mitotic germ cells, respectively [43].

1.7 Thesis Objectives: characterizing the *C. elegans* ATM orthologue, *atm-1*

These studies involving *atm-1* have shed light on its function in *C. elegans.* However, these studies do not focus specifically on the mechanism of ATM-1 but instead on other proteins. Thus, the purpose of my thesis is to provide further characterization of *atm-1*. My study will provide a foundation for current and future studies to build upon. In my thesis, I aim to provide characterization of *atm-1(gk186)* by studying it molecularly and by examining its phenotype(s). First, I address the validity of the current wildtype *atm-1* gene model. Then, I provide molecular characterization of the mutant *atm-1(gk186)* allele and the consequences of the *gk186* deletion on *atm-1* transcription and translation. Finally, I characterize *atm-1(gk186)* in the context of IR sensitivity in meiotic germ cells, genomic instability, and its effects on mutation rates.

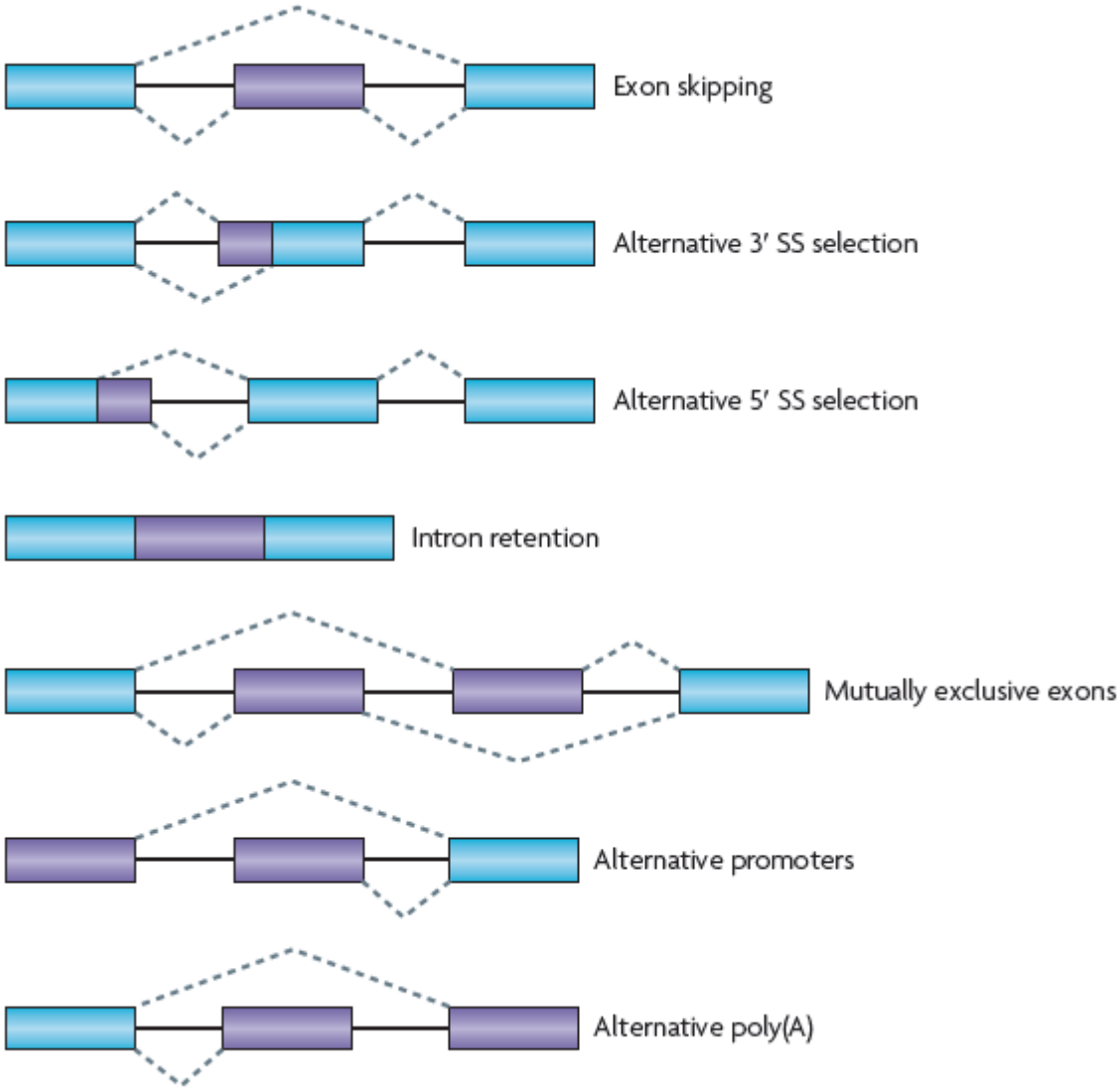

Figure 1.1 Different mechanisms of alternative splicing. Constitutive exons are shown in blue. Alternatively spliced regions are in purple. Solid lines indicates introns. Splicing is indicated by dotted lines. Adapted from [21]. *NOTE* Purple segment in Intron retention indicates the inclusion of an intron, now considered an exon.

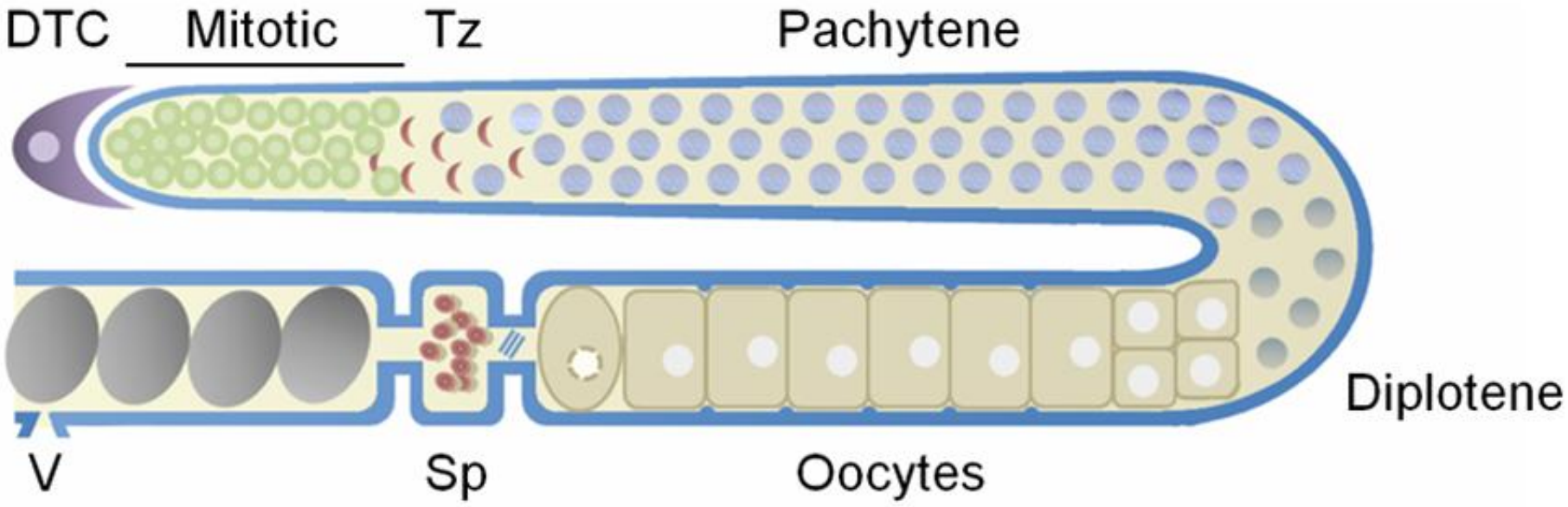

Figure 1.2 One arm of the C. elegans hermaphrodite germline. DTC, distal tip cell; Mitotic, mitotically proliferating region, Tz, transition zone; Sp, spermathecum containing sperm; V, vulva at the uterus containing embryos. Figure adapted from [44].

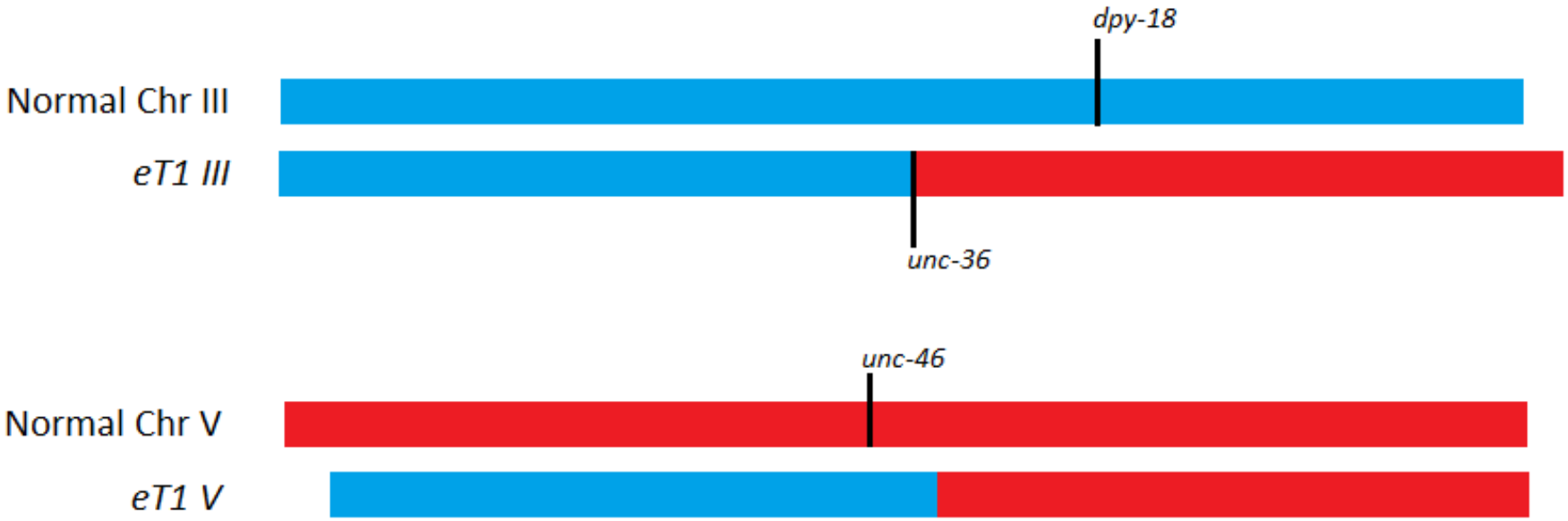

Figure 1.3 *eT1* balancer system. A phenotypically wildtype *eT1* worm with a heterozygous genotype is shown. Blue represents Chr III sequences. Red represents Chr V sequences. During synapsis, sequences of Chr III right and Chr V left lack homologous sequences for homologous recombination to occur. Normal chromosomes are marked with visible markers, *dpy-18* for Chr III and *unc-46* for Chr V. *unc-36* on Chr III is disrupted by the *eT1* translocation resulting in an Unc phenotype for *eT1* homozygotes. Homozygous normal Chr III and V worms appear Dpy Unc. Based on data from [31].

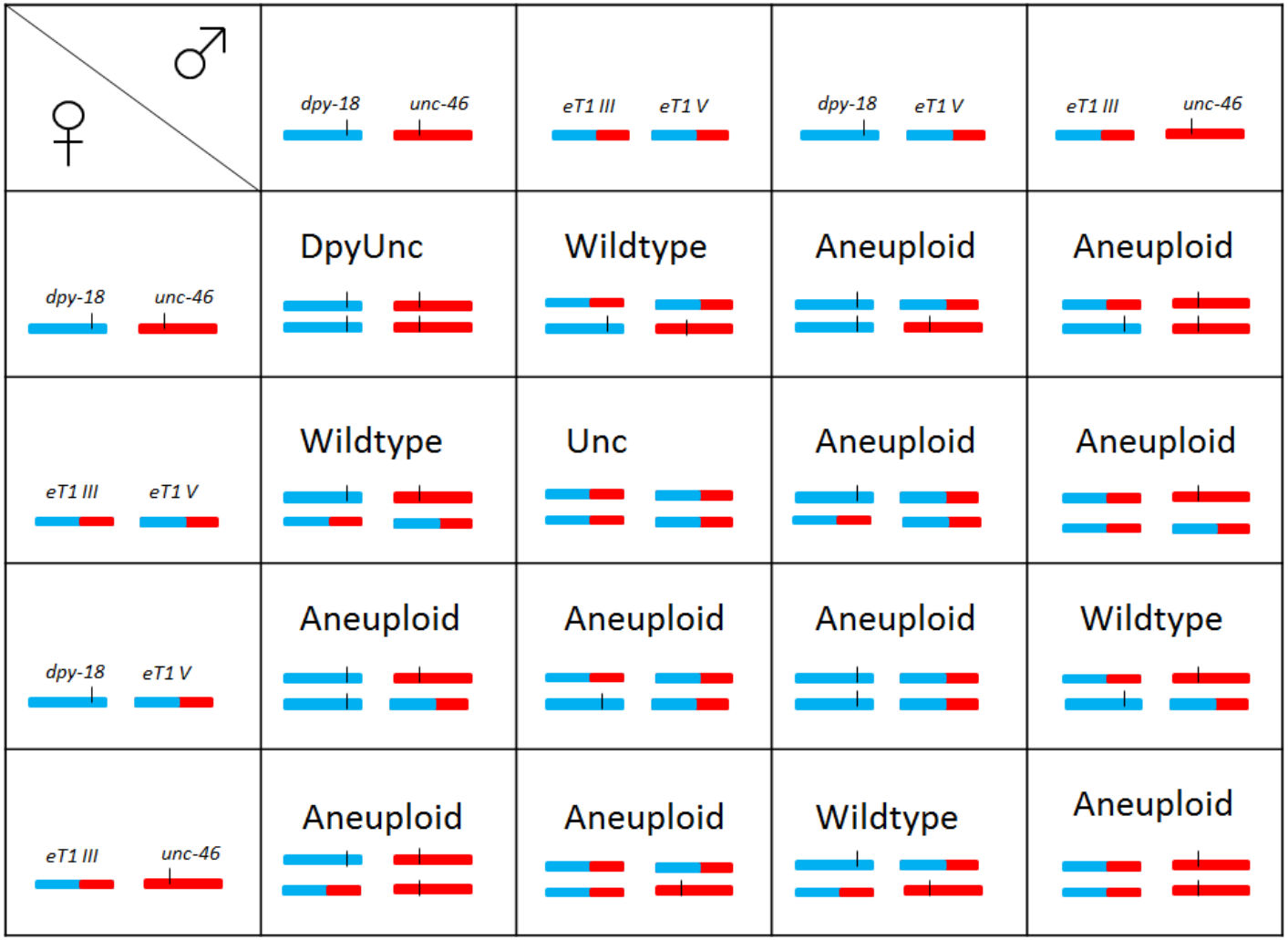

Table 1.1 All possible progeny genotypes combinations for a self-fertilizing *eT1* heterozygous hermaphrodite. Blue represents Chr III sequences. Red represents Chr V sequences. Each genotype is labelled with its corresponding phenotype. The surviving progeny has a 4:1:1 wildtype to DpyUnc to Unc (*unc-36*) ratio. 10/16 embryos arrest because of aneuploidy. Based on data from [31].

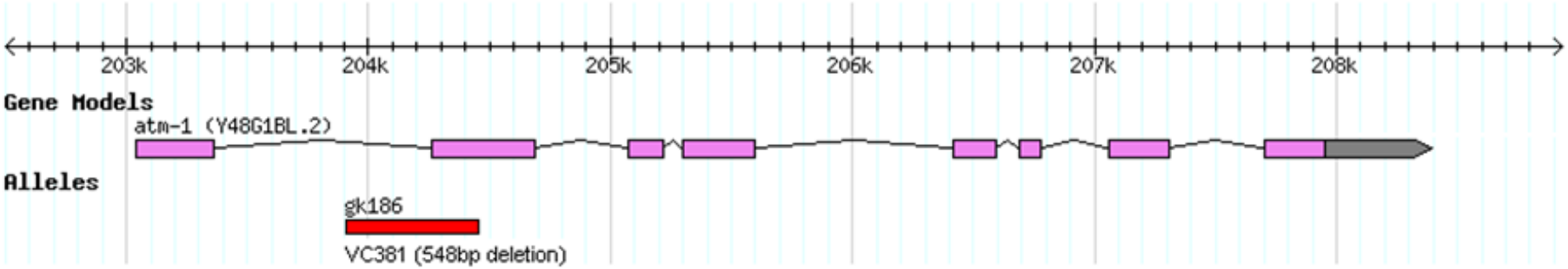

Figure 1.4 *atm-1* gene model as depicted in WormBase (WS205). The deletion in the *gk186* allele is indicated by a red bar below the gene model. Modified from www.wormbase.org

2. Materials and Methods

2.1 Nematode maintenance, culturing, and mating

C. elegans strains were maintained and cultured at 20°C on Petri plates containing standard nematode growth media (NGM) streaked with *Escherichia coli* OP50 as outlined by Brenner [25]. Worms were observed on a WiLD Heerbrugg M3 light microscope. Mating between strains was conducted at a 3:1 hermaphrodite to male ratio overnight on mating plates, and mated hermaphrodites were individually re-plated the following day.

2.2 Generation of complementary DNA (cDNA) library

10 saturated 35mm plates containing mixed stage worms for each strain were collected by washing with M9 buffer for total RNA isolation. The worms were pelleted by centrifugation. 0.5mL of packed worms were resuspended in 2mL of Trizol (Invitrogen, Catalog Number 15596-018). The worm/Trizol mix was vortexed, subjected to freeze-thaw, and incubated in a 37°C water bath. Nucleic acids were isolated by adding 1mL chloroform followed by centrifugation and isolation of the supernatant. 1mL isopropanol was added to the supernatant, mixed, and re-centrifugated. The nucleic acid pellet was washed with 0.5mL 75% ethanol and allowed to dry. The pellet was resuspended with DEPC-treated water. RNA quality was estimated by running through a 1% agarose gel. The nucleic acids were treated with DNase using the Fermentus DNase I, RNase-free kit and protocol (#EN0521). DNase was removed with phenol-chloroform followed by chloroform to isolate total RNA. RNA was precipitated overnight at three volumes of 100% ethanol to sample ratio with the addition of a 0.3M final concentration sodium acetate. Following centrifugation and removal of supernatant, the RNA pellet was washed with 0.5mL 75% ethanol and resuspended in DEPC-treated water. cDNA was generated using the RevertAid H Minus M-MuLV Reverse Transcriptase kit from Fermentas using poly-dT primers (#EP0451). The quality and concentration of the total RNA and cDNA were measured using a NanoDrop spectrophotometer.

2.3 Sequencing and sequence alignment

PCR Products that needed to be sequenced were first gel purified using QIAGEN's QIAquick Gel Extraction Kit and then eluted in double distilled water. Sequencing was performed by NAPS (located in the NCE building, UBC). Sequences were aligned manually to sequences obtained from WormBase (WS205) using BioEdit. Full contigs of my sequencing reactions are located in Appendix A and B.

2.4 Single worm DNA extraction

Single adult worms were individually placed in 5µL of lysis buffer (10mM Tris-HCl, 50mM KCl , 2.5mM MgCb, 0.45% NP40, 0.45% Tween20, 0.01% gelatin, 100µg/ml Proteinase K). Tubes were freeze thawed twice in liquid nitrogen and incubated at 56°C for 1 hour, then 95°C for 15 minutes to inactivate the proteinase K. Temperature cycling was done in a 96 well Eppendorf Mastercycler Gradient thermocycler.

2.5 Primer design and polymerase chain reaction (PCR)

Primers were designed using the Primer3 website interface http://frodo.wi.mit.edu/primer3/ and are listed in Table 2.1. PCR was conducted on a Dyad DNA engine machine using the appropriate temperatures for the type of polymerase and for the annealing temperatures of the primers. Gel electrophoresis resolution of DNA products was done in a 1% agarose gel with ethidium bromide in 1x TAE buffer.

cDNA PCRs were conducted using New England BioLabs Inc Phusion High-Fidelity DNA Polymerase to ensure high fidelity products for sequencing. Suggested Phusion protocol for PCR cycling was used: 98°C denaturing, 72°C elongation at 45 seconds/kb, primer specific annealing temperature (between 56-60°C) and cycled for 34 cycles. Phusion HF 5x buffer was used. 10mM stock dNTP solution was mixed from 100mM stock solutions of each dNTP bought from Applied Biological Materials (abm) Inc. A 20µL reaction consisted of 4µL 5x HF buffer, 0.5µL 10mM dNTPs stock solution, 0.4µL of each primer, 0.2µL of 1/10 dilution of cDNA library, 0.1µL Phusion polymerase, and 14.4µL of double distilled water.

2.6 *atm-1(gk186)* outcross and *gk186* deletion detection by PCR

VC381, *atm-1(gk186),* was outcrossed to wild type *C. elegans* males (strain – VC2010) to remove as many second site mutations induced by the TMP/UV treatment that was used to generate *gk186*. Two F_1's from the above cross were individually plated and 10 F_2's (5 from each F_1) were subsequently individually plated. F_2's were allowed to lay eggs before being single worm lysed. After confirmation of an F_2 homozygous for *gk186*, the outcross was repeated for a total of 10 outcrosses. The final 10^{th} outcrossed worm was used for all subsequent phenotypic characterizations. (Table 2.2)

The *gk186* deletion was detected using PCR with the poison primer technique [47]. The forward primer (AAATTCGATTTTTCGCGA), reverse primer (CAATTGACGCAATTTGCA), and poison primer (TCACTCCGACGTCCGTTT), were used to detect the deletion. The poison primer is designed from sequences that are located in the *gk186* deletion region. Therefore in wildtype worms, the poison primer is able to anneal and produce the smaller PCR product. *gk186* removes the sequences required for the poison primer to anneal to thus only the forward and reverse primers are used to produce PCR product in the mutant strain. The wildtype genotype produces a 956bp PCR product whereas the mutant *gk186* deletion strain produces a larger 1233bp product (Figure 2.1).

This PCR was done using Taq Plus DNA Polymerase (abm). The abm Taq suggested protocol was followed: 94°C denaturation, 72°C elongation at 1 minute/kb (2 minutes total), 58°C annealing, and cycled for 34 cycles. 2.5µL of 10x abm PCR buffer, 1µL 50mM $MgCl_2$, 0.5µL dNTP stock solution (abm), 0.4µL of each primer (forward, reverse, and poison primer), 0.1µL Taq polymerase, and 14.7µL of double distilled water was added to the 5µL lysis containing F_2 worms for the PCR reaction.

2.7 Ionizing radiation sensitivity assay

For each strain, 30 adult P_o's were plated on 3 plates (10 worms per plate) and allowed to lay eggs for 2 hours giving approximately 50 eggs per plate. These worms were synchronized to 1 day old adults and subjected to different IR dosages. IR was applied by the TORREX 150D X-Ray Inspection System. In the machine, the uncovered plates were

placed on tray 7. The settings were 5mA at 145kV which administers 15Gy in 135 seconds. For 0Gy, worms were still placed in the machine uncovered but not subjected to IR. 30 and 45Gy were subjected to 270s and 405s of IR, respectively. After IR, worms were allowed to rest in 20°C for 20 hours. 10 plates with 3 worms each were plated for each dose and a 5 hour brood was collected. The number of unhatched eggs and adults were scored 24 and 72 hours post brood collection. % survival was calculated by dividing the total number of adults by the total progeny (adults + unhatched eggs). Standard error was used for calculating statistical error.

2.8 DAPI staining

20-30 adult worms for each strain were picked onto a watch glass containing 5µL of M9 buffer. 250µL of 150mM DAPI in 96% ethanol was added and allowed to evaporate in the dark at room temperature for 1-2 hours. 500µL of M9 was used to destain the worms for 2 hours in room temperature. Worms were mounted on a microscope slide atop a 3% agarose pad. Diakinetic chromosomes were observed in oocytes located before the spermatheca using a Zeiss Axioscope fluorescent microscope with 40x objective using OpenLab software.

2.9 Quantifying genomic instability

20 *atm-1(gk186)* L4 worms were individually plated and the subsequent progeny scored for total progeny and the presence of males. A single L4 worm from each line was plated to propagate the line. This was repeated for 20 generations.

2.10 Creating *atm-1(gk186);eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)*

atm-1(gk186) males were crossed to homozygotes *eT1* hermaphrodites (derived from BC2200). F_1 males (genotype *atm-1/+;eT1(III)/+;eT1(V)/+*) were then crossed to *dpy-18(e364);unc-46(e177)* hermaphrodites (also derived from BC2200). Progeny from this cross were individually plated and allowed to self and produce progeny. The progeny was screened for the presence of Unc-36 worms indicating that the *eT1* chromosomes have been retained (*eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)*). Every parental giving progeny with some Unc-36 had their DNA extracted and genotyped by PCR for the *gk186* deletion.

Heterozygous *atm-1(gk186)* with *eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)* were obtained and allowed to self again. The progeny of these were genotyped again to identify and obtain a strain homozygous for *atm-1(gk186)* (genotype: *atm-1(gk186);eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)*)(Table 2.2).

Table 2.1 List of primers created and used. (#) indicates primers used for sequencing. (*) indicates primers designed by S.Y. Chua (Baillie Laboratory, Simon Fraser University).

Name	**Forward or Reverse**	**Sequence**	**Annealing T (°C)**
a *	Forward	ATCTGAAACACGCGATAGCTC	59.49
b #*	Forward	ACAATCAGGTGCCATGGATAG	59.83
c #*	Reverse	CAAGTGCCACGTCTCTTCTTC	60.04
d #*	Forward	GAAGAAGAGACGTGGCACTTG	59.99
e #	Forward	ATACGATGCGGCAATAAACC	59.82
f #	Reverse	ACCCGACGATACAGCTTTCC	61.40
g #	Forward	TCTGTATTGGCCAACTGTCG	59.72
h #*	Reverse	TACTCGCCGAAACTTCTTCTG	59.64
I *	Reverse	TCGGAAATTTGTCATCGTTTC	59.93
j #*	Reverse	CGCGTTCAATTTATGTTTTCC	59.49
seq1 #	Reverse	TCCAAATCCATTTCCATAAGC	58.89
seq2 #	Reverse	CAACTCCGAACTCTCTCATCG	60.00
seq3 #	Reverse	CGAGCTGATTGTCAAGATTCG	60.92
seq4 #	Forward	CCAATCATATTTGCGAAAACC	59.31
seq5 #	Forward	AGGCGAGAAACATCTACATCG	59.35
seq6 #	Forward	ATGGGTGATGAAGATGAGACG	59.94
seq7 #	Forward	ACCGAATTGACAGATCACAGG	59.98
seq8 #	Forward	GTCTGTATTGGCCAACTGTCG	60.57
seq9 #	Forward	GTCGAAGCTGTGGATAAGTGC	59.90
seq10 #	Reverse	CATCGTTTCCACGTGATTAGG	60.37
seq11 #	Forward	CCCGTTAGACACTGAATGTGG	60.42

Table 2.2 A list of strains used and created.

Name	**Genotype**	**Created by:**
KR4831	*atm-1(gk185)*5* (outcrossed 5 times)	Jim Huang
KR4941	*atm-1(gk186)*10* (outcrossed 10 times)	Jim Huang
KR5059	*atm-1(gk186);eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)* (Isolate 1)	Jim Huang
KR5060	*atm-1(gk186);eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)* (Isolate 2)	Jim Huang
KR4948	*atm-1(gk186)(h2680)*	Jim Huang
KR4947	*atm-1(gk186)(h2681)*	Jim Huang
KR4952	*atm-1(gk186)(h2682)*	Jim Huang
KR5104	*atm-1(gk186)(h2683)*	Jim Huang
KR5105	*atm-1(gk186)(h2684)*	Jim Huang
KR5106	*atm-1(gk186)(h2685)*	Jim Huang
VC381	*atm-1(gk186)*	Genome Knockout Center
VC2010	N2 Wildtype	Genome Knockout Center
VC655	*brd-1(gk297)*	Genome Knockout Center
BC2200	*eT1(III)/dpy-18(e364);eT1(V)/unc-46(e177)*	Dr. Robert Johnsen

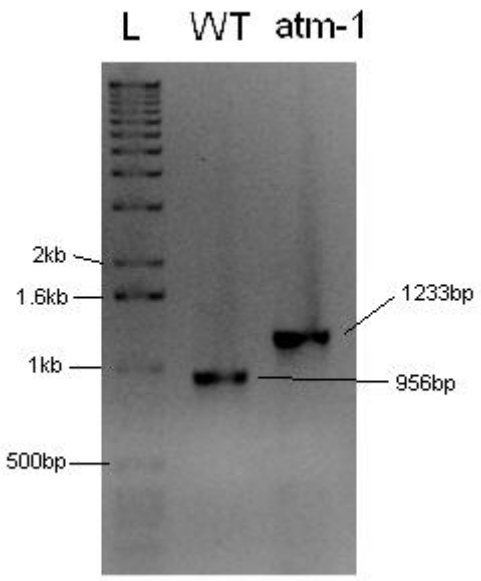

Figure 2.1 Agarose gel of wildtype (WT) and *atm-1(gk186)* PCR products using the poison primer technique.

3. Results

3.1 Molecular characterization of the wildtype *atm-1* gene

3.1.1 The predicted *atm-1* gene is much smaller than other ATM homologs

C. elegans atm-1 is predicted to produce a 649AA polypeptide (WormBase, WS205), whereas other orthologues have at least 2,700AA (Figure 3.1)[40]. Evident from this comparison, between ATM-1 and other known orthologues, is that ATM-1 is shorter and resembles the C-terminal ends of other ATMs (Figure 3.1). The summary of nematode ATM-1 on NCBI's CDD reveals that it contains the PI3Kc_like and FATC domains like other ATM homologues (Figure 3.2) [2, 13, 48]. However, ATM-1 is lacking domains known to be in other ATMs, including the FAT and TAN domains that are typically N-terminal of the kinase domain [48, 49]. Also evident from this comparison is that much of the N-terminal portion of ATMs appears to be species specific and lacks any known domains. Because of evolutionary divergence between ATM homologs, the domains may be less conserved and thus not recognizable through bioinformatic domain analysis and sequence comparisons. For example, the TAN domain is not present in the *Drosophila melanogaster* prediction but present in the *Saccharomyces cerevisiae* prediction, and the FAT domain is only predicted for the human prediction (Figure 3.1). These structural differences between the *C .elegans* ATM-1 and other known ATMs bring into question the accuracy of the *atm-1* prediction.

3.1.2 *K10E9.1* and *F56C11.4* are uncharacterized gene predictions upstream of *atm-1*

A mis-annotated N-terminal extension might be the reason why *atm-1* does not resemble other known ATMs. To address whether the N-terminal portion of *atm-1* might be mis-annotated, I analyzed the genomic region upstream of *atm-1* for evidence of an extended *atm-1* gene model. Directly upstream of *atm-1* are two predicted genes that are uncharacterized: *K10E9.1* and *F56C11.4* (Figure 3.3). *K10E9.1* is directly upstream of *atm-1* and is predicted to produce a large 12,836bp primary transcript. The coding sequence spans 16 exons, which corresponds to 4,356 nucleotides (nt) and results in an 1,451AA protein. Upstream of *K10E9.1* is *F56C11.4. F56C11.4* is predicted to be an 847bp transcript

encoding 3 exons totaling 690nt which translates into a 229AA protein. The close proximity and location of these two predictions suggested that *K10E9.1* and *F56C11.4* could be the mis-annotated N-terminal extension of *atm-1*. In fact, three ESTs also support the idea of a larger *atm-1* gene model (Figure 3.4). Each EST has a 5' and 3' sequence read from that specific clone and is represented by .5 and .3 at the end of the read names. EST clone yk1279h03 suggests that *F56C11.4* and *atm-1* are expressed as one mRNA. Likewise, clone yk175d02 suggests *K10E9.1* and *atm-1* are expressed as a single mRNA. A closer inspection through a BLAST search of yk220a2.5 reveals that it too suggests *K10E9.1* and *atm-1* are expressed as a single mRNA because a small portion of the .5 EST read aligns to the end of the *K10E9.1* gene prediction (Figure 3.5). Overall, these three ESTs that overlap with the *atm-1* gene prediction suggest that *atm-1, K10E9.1,* and *F56C11.4* may form a single transcript.

3.1.3 *F56C11.4, K10E9.1,* and *atm-1* gene predictions produce one single mRNA transcript

To provide support for a single, large transcript comprised of the merging of the three gene predictions, I created a wildtype (N2) cDNA library. To probe the coding sequences, I designed PCR primers that span my predicted junctions between adjacent gene models. Two sets of primers were designed: one pair between *F56C11.4* and *K10E9.1* (F-K Junction), and the other between *K10E9.1* and *atm-1* (K-A Junction)(Figure 3.6, Table 2.1). If the three gene models do in fact produce a single transcript, then splicing of the exons between the gene models will produce one continuous template. This single transcript will allow the PCR reaction to proceed and produce the predicted products. Conversely if the three gene models are indeed separate genes, then the genes will not be spliced together. In this case, the primers will not be able to amplify any PCR product. Panel "A" in Figure 3.7 shows the result of the PCR reaction for the F-K Junction. The arrow indicates a product close to the expected size of 674bp. Panel "B" shows the result of the PCR reaction for the K-A Junction. Again, the arrow indicates a product roughly of expected size of 682bp (Figure 3.7). My results indicate that both primer sets produce PCR products. These experiments support my prediction of a larger gene model, a merger of *atm-1, K10E9.1* and *F56C11.4* gene predictions, for the ATM homologue in *C. elegans*.

3.1.4 cDNA PCR products support a larger *atm-1* gene model

To understand the molecular structure of the merged gene model of *atm-1*, my collaborator Shu Yi Chua (Simon Fraser University, British Columbia) and I created primers to amplify products from the cDNA library spanning the entire new gene model (Figure 3.8A, Table 2.1). These primers successfully amplified products which overlapped each other and are roughly of predicted size indicating that the individual gene models represent the larger gene model well (Figure 3.8B, Table 3.1). Lanes 1 and 2 from Figure 3.8B also support the previous F-K Junction result and lanes 3 and 4 support the previous K-A Junction result. I observed no evidence for alternative transcription based on the single bands seen in each lane.

3.1.5 Sequencing reveals the nucleotide composition of the larger *atm-1* gene model with novel unpredicted changes

The above *atm-1* cDNA PCR products obtained were sequenced. The sequencing resulted in between 1- to 6-fold coverage of the entire *atm-1* gene model with an average of 2-fold coverage. The exact exonic sequence of the entire *atm-1* gene model and the intron-exon splice sites can be described by comparing the transcript sequence to the genomic sequence (Appendix B). The majority of the sequence aligns exactly with the predicted models of the three genes: *F56C11.4*, *K10E9.1*, and *atm-1*. Comparing both the contiguous cDNA sequence and the genomic sequence from WormBase (WS205), I was able to determine the differences between the predicted gene models and my merged *atm-1* gene model (Table 3.2). First, a novel 5' splice site occurs at the F-K boundary. This unpredicted splice site occurs at the predicted TAG stop of *F56C11.4* and disrupts the stop codon. Interestingly, this terminal exon of *F56C11.4* does not splice to the predicted first exon of the *K10E9.1* prediction. This last exon of *F56C11.4* splices to a novel exon I discovered, located 433nt upstream of the predicted first exon of *K10E9.1*. It contains 163nt and utilizes novel splice sites also previously unpredicted. This exon also explains the size discrepancy seen in Figure 3.7A. The F-K Junction PCR product's size appears to be closer to 900nt which is close to the sum of the prediction (674bp) and the novel exon (163bp). This novel exon splices to the predicted first exon of *K10E9.1*. However, another interesting discovery I have

made is that this first exon has unpredicted exonic sequences added to both the 5' and 3' ends of the predicted exon. Nine exonic nucleotides are attached to the 5' end and 42nt are added to the 3' end. These extensions on the predicted first exon of *K10E9.1* are also accompanied by novel splice sites. Finally, splicing between the *K10E9.1* and *atm-1* predictions is described. The novel 5' splice site removes 42nt from the end of the last predicted *K10E9.1* exon and the novel 3' splice site removes 30nt from the beginning of the first predicted *atm-1* exon. Overall, I have refined my extended *atm-1* gene model by discovering and annotating novel splice sites, a new exon, and additional coding sequences. The entire 7,137nt coding segment for my extended *atm-1* cDNA is in frame. The cDNA encodes for 2,378AA which is much larger than the 649AA initially predicted. My newly revised gene model of *atm-1* depicting the changes is illustrated in Figure 3.9B.

3.1.6 The extended gene model of *atm-1* does not reveal any additional protein domains

The revised 2,378AA ATM-1 sequence was reanalyzed through the CDD. The results from this analysis only revealed the PI3K_c and FATC domains (Figure 3.10). The TAN and FAT domains remain unpredicted. The extended *atm-1* gene model is most similar to the *Drosophila melanogaster* HomoloGene prediction (Figure 3.1). Again, the lack of predicted domains in the extended *atm-1* gene model could be a result of the chosen bioinformatic analyses, the current knowledge of these domains, and/or species divergence.

3.1.7 Section summary

I have discovered that the ATM homolog in *C. elegans* has been mis-annotated and is in fact represented by three gene predictions: *F56C11.4*, *K10E9.1*, and *atm-1*. Sequencing of PCR products spanning the three gene models reveals the exact nucleotide composition of the gene model and novel changes that were unpredicted. The revised 2,378AA ATM-1 protein is now more similar to other known ATMs than the previously predicted 649AA protein.

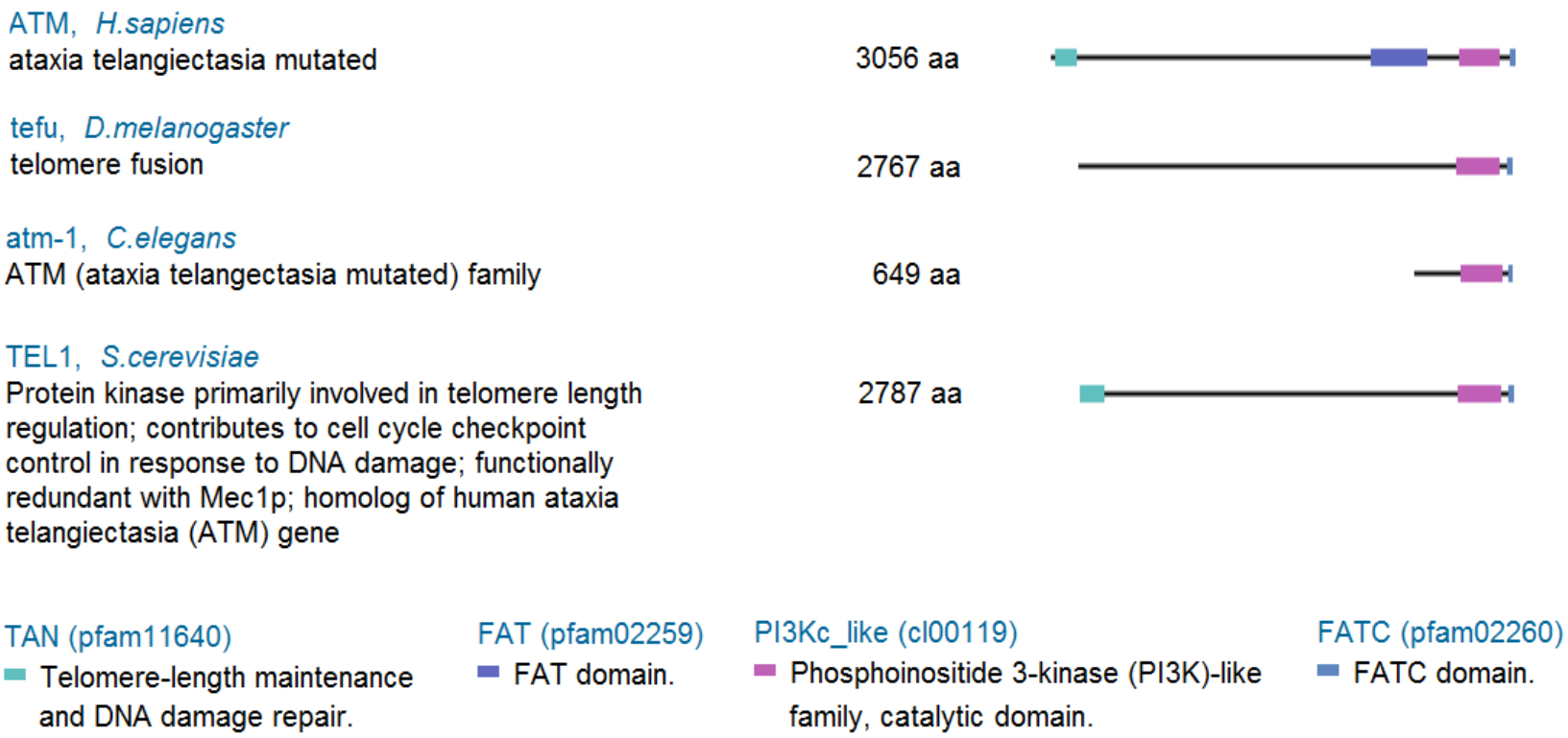

Figure 3.1 HomoloGene comparison between known ATM homologues. The *C. elegans* prediction resembles the C-termini of other ATMs in other organisms. Modified from NBCI HomoloGene. (http://www.ncbi.nlm.nih.gov/)

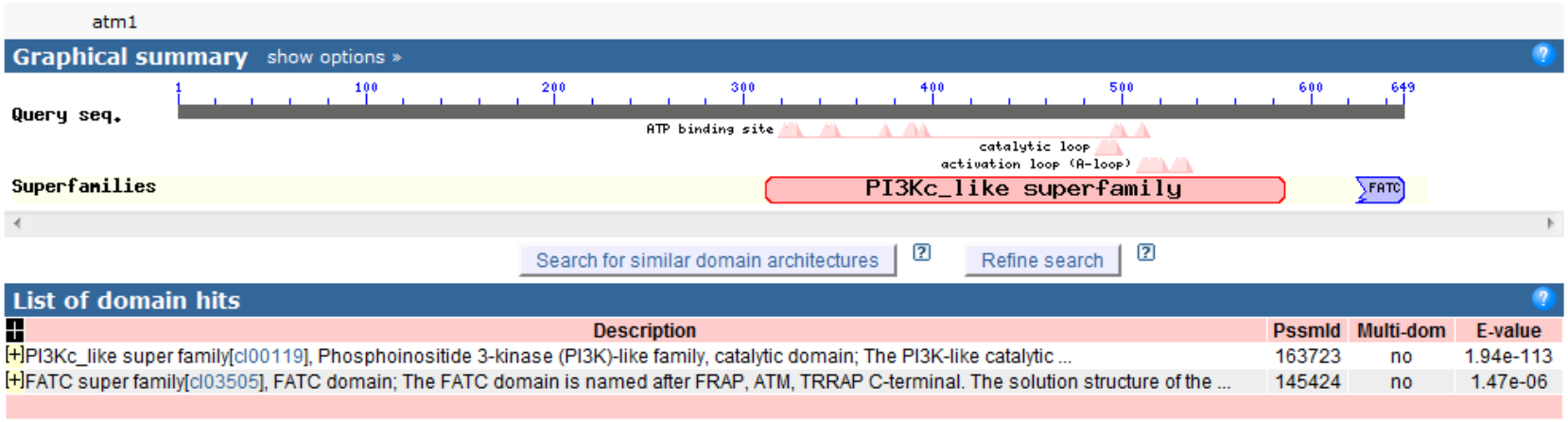

Figure 3.2 CCD result for the *atm-1* 649AA query. PI3Kc_like and FATC domains are predicted. The small size of ATM-1 is the most likely reason why the FAT and TAN domains are not predicted (see Figure 3.1). (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml)

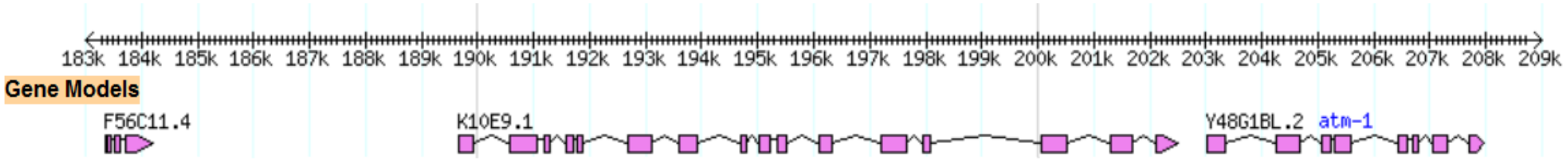

Figure 3.3 Genomic region 183,000bp to 209,000bp of chromosome I on wormbase.org (WS205). *F56C11.4* and *K10E9.1* are directly upstream of *atm-1* (Y48G1BL.2). Modified from wormbase.org (WS205).

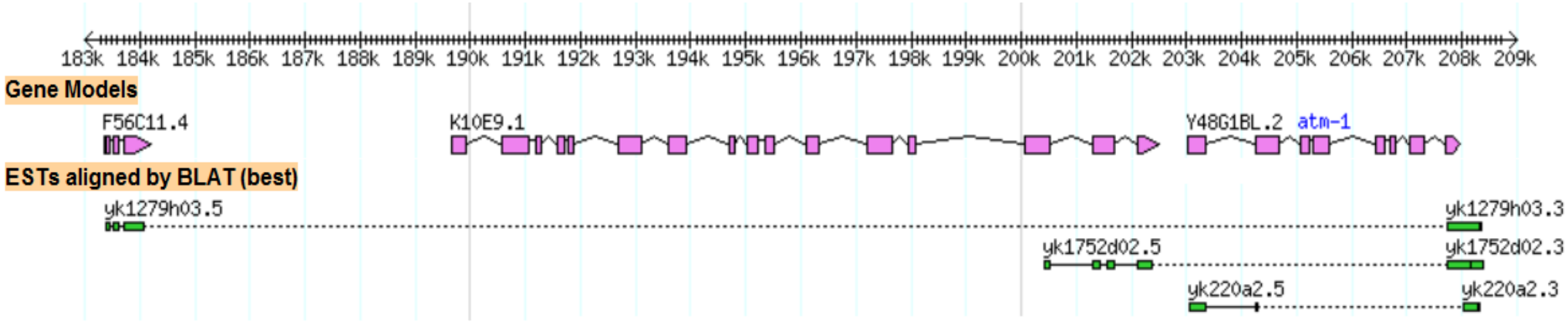

Figure 3.4 Genomic region 183,000bp to 209,000bp of chromosome I with the three ESTs that support linkage of the three gene models. The dotted lines infer that there are sequences between the two end reads that should be present within the full cDNA. Modified from wormbase.org (WS205).

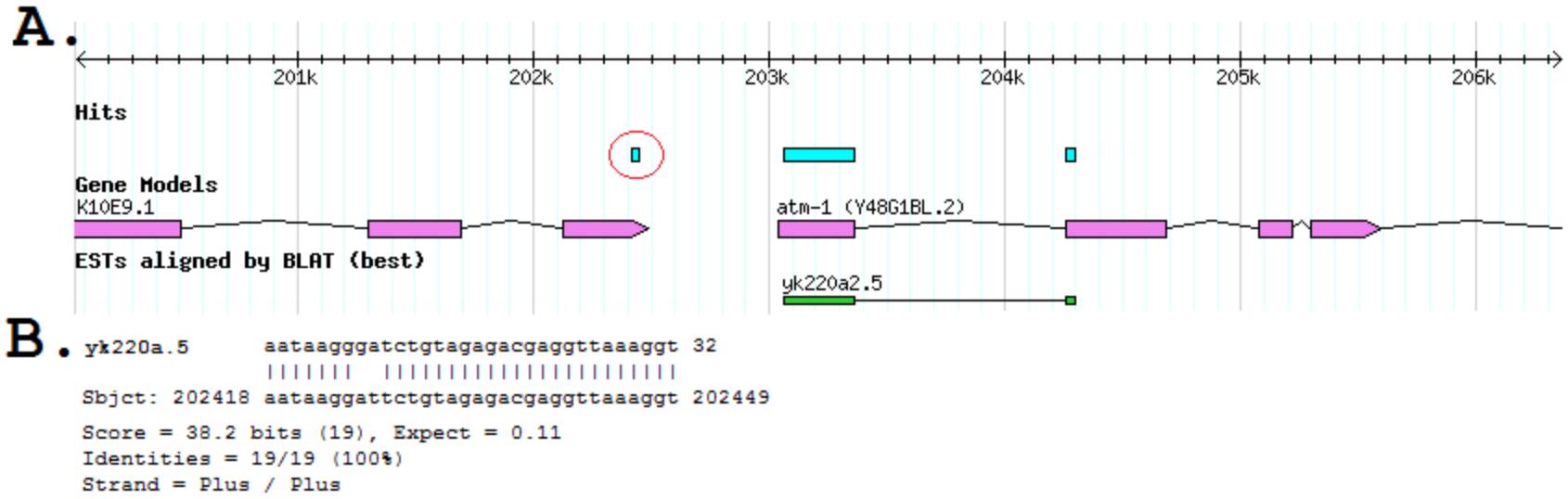

Figure 3.5 yk220a.5 also suggests linkage between *K10E9.1* and *atm-1*. A. BLAST search of yk220a2.5. Teal "Hits" represent homology between the EST read and the genome. The red circle indicates a portion of the sequence read located within the *K10E9.1* prediction. This indicates where the clone read starts. The start is continuous with the remainder of the read which aligns with *atm-1* suggesting a transcript spanning across these two gene models. B. The BLAST search sequence homology between the EST and the genome. Modified from wormbase.org (WS205).

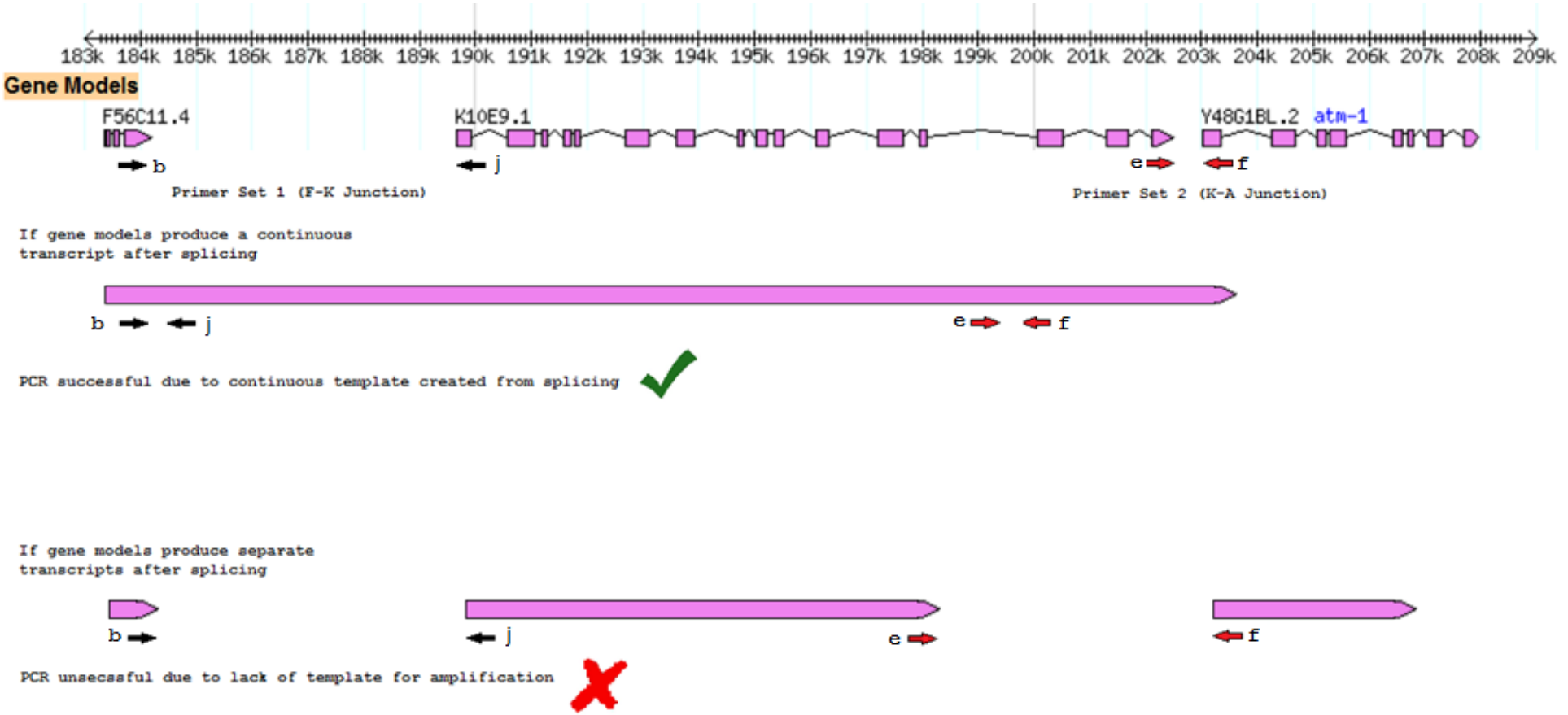

Figure 3.6 F-K and K-A Junction primers design. Black arrows indicate the rough locations of the F-K primers (b + j). Red arrows indicate the approximate locations of the K-A primers (e + f). Successful PCR products are amplified only if the three gene models produce a continuous transcript. If the upstream genes are not extensions of *atm-1*, the PCR reaction will fail to produce product.

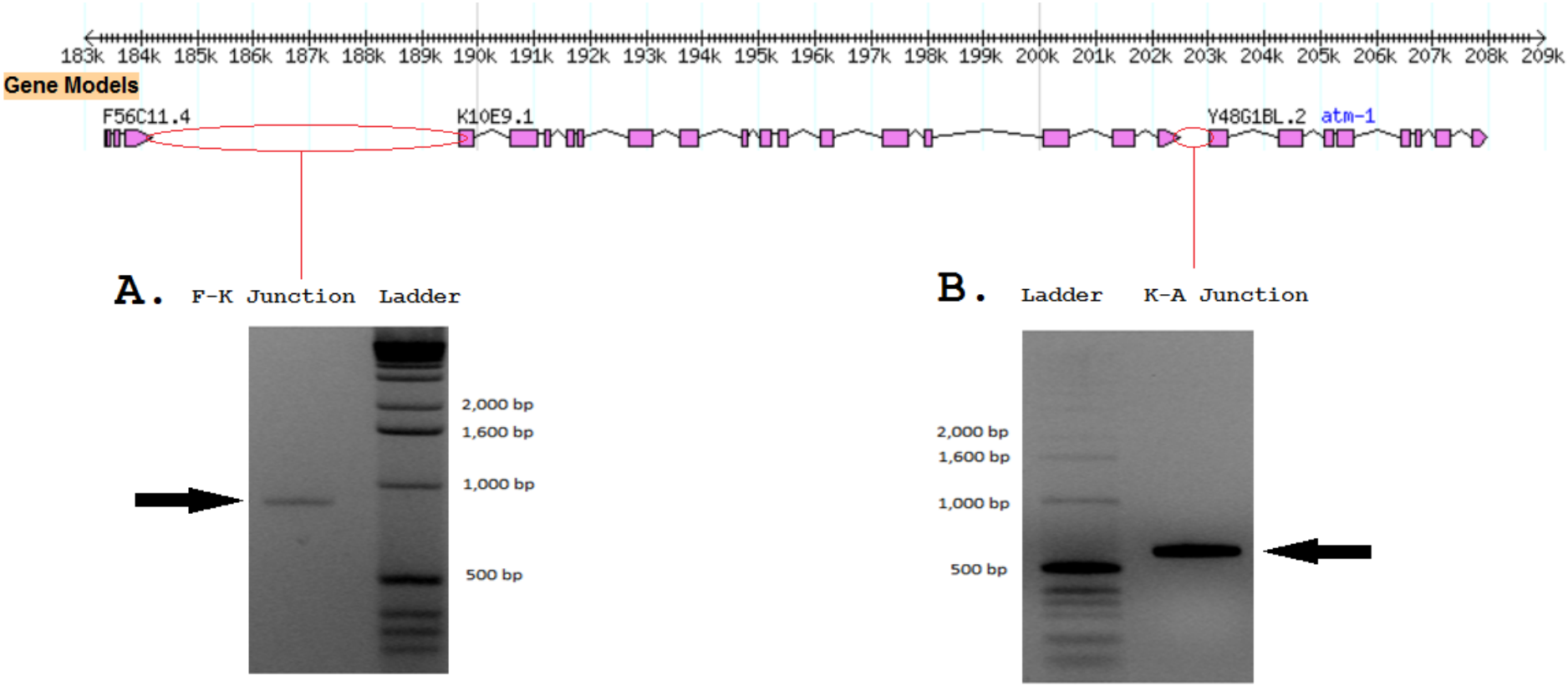

Figure 3.7 cDNA PCR results using the F-K and K-A Junction primers (b+j for F-K Junction and e+f for K-A Junction; Table 2.1). Black arrows indicate product of roughly expected size for each reaction. A. The F-K Junction product (predicted size: 674bp). B. The K-A Junction product (predicted size: 682bp).

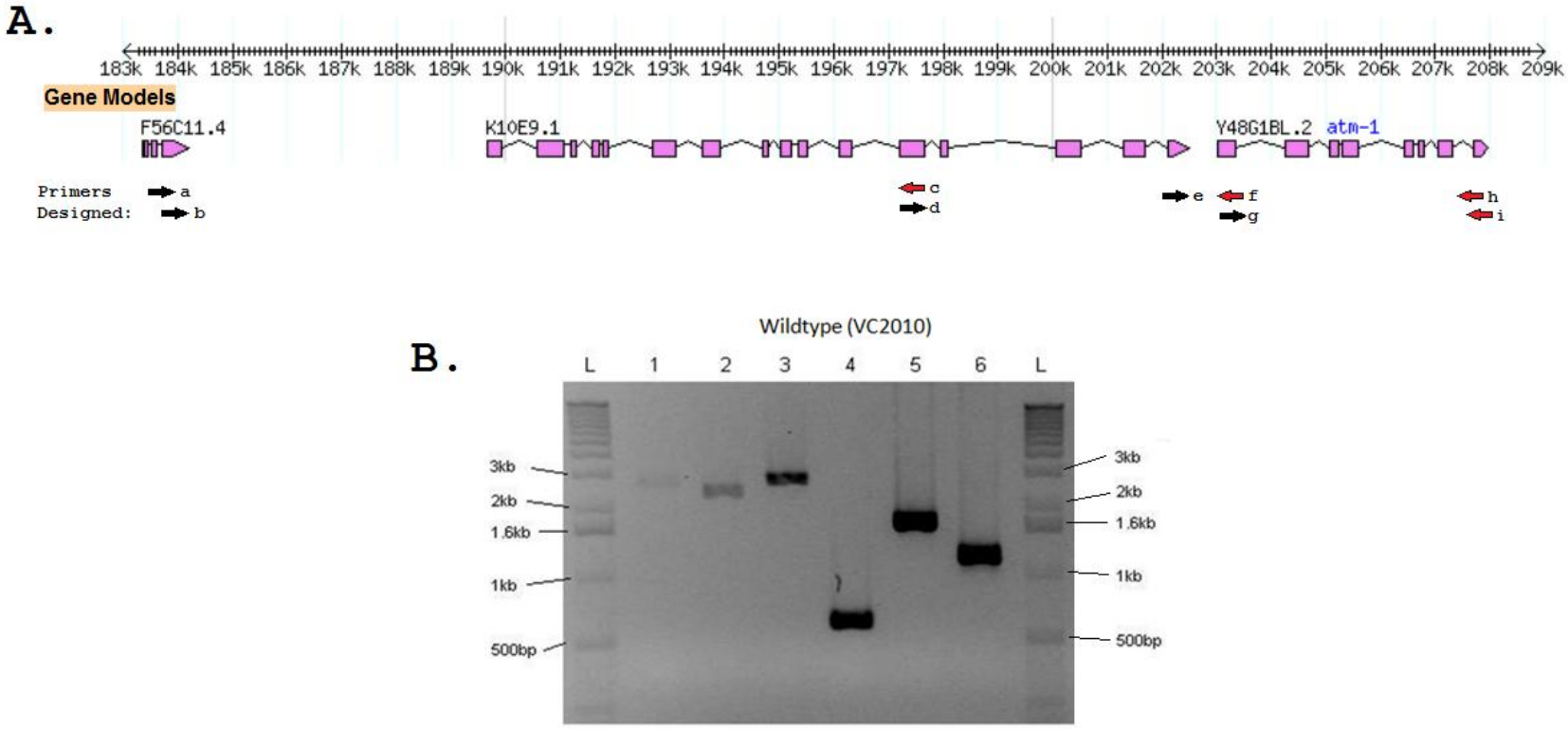

Figure 3.8 Wildtype cDNA primer design and products. A. Primers designed for amplifying overlapping PCR products from the wildtype cDNA library for the extended gene model. Primer sequences and exact location indicated in Table 2.1. Black arrows indicate forward primers and red arrows indicate reverse primers. B. Agarose gel with each lane representing products from a primer combination. Lane: 1. a+c, 2. b+c, 3. d+f, 4. e+f, 5. g+h 6. g+i. Expected product sizes given in Table 3.1.

Table 3.1 cDNA PCR primer combinations and expected product sizes. Primer sequences listed in Table 2.1.

Lane in Figure 3.8B	Primers used	Expected size based on predictions
1	a + c	2,495bp
2	b + c	2,176bp
3	d + f	2,859bp
4	e + f	682bp
5	g + h	1,703bp
6	g + i	1,212bp

Table 3.2 Summary of novel changes for the *atm-1* gene model. Underlined sequences indicate splice sites.

Location of Observed Change	**Type Of Observation**	**WormBase Prediction**	**Observed Sequence from cDNA Sequencing**
F56C11.4 and *K10E9.1* junction	Novel Splicing	*F56C11.4*: AAAGAATAG(STOP)TGCTGGA	*F56C11.4*: AAAGAATA(SPLICE) 5' intron splice sequence: TA//GTGCTGGA
433nt upstream of *K10E9.1*	Novel Exon and Novel Splicing	Genomic Sequence (not predicted to be exonic): ATTACAGTTGGTGTCATTACCGAAAATTCTC GGTCGTCCACATTCTCTCATATATCGTACATC GTTGGGGTCTCGAAGCGAGAGATTTCATTCT TCGTGAGATTTTCGAGCTCACTGGTAGTTTGT CCAACTTGATATCAGTGGCTCATAAGGATGG AGAACAGTCCAAGGTTTGAGT	Novel 163nt exon
			3' intron splice sequence: ATTACAG//TT
			5' intron splice sequence: AG//GTTTGAGT
First exon of *K10E9.1*	Additional Exonic Sequences and Novel Splicing	First Exon: TTTTCAGAAATGCATAATG(START) CGTTTGATT… …TTGAATTGACCG(SPLICE)GTGATAGCAGTA GTAAGAAGAAGGACGATGCCACGTTTGATAG TGCGTAA	5' 9nt addition to exon: AAATGCATA(EXON) 3' intron splice sequence: TTTTCAG//AA
			3' 42nt addition to exon: (EXON)GTGATAGCAGTAGT AAGAAGAAGGACGATGCCACGTTTGATA 5' intron splice sequence: TA//GTGCGTAA
K10E9.1 and *atm-1* junction	Novel Splicing	*K10E9.1*: AGGTACATGCGGTCGGGTCTTG CAGCGAAATAATGCATTTTTAA(STOP)	5' intron splice site: AG//GTACATGC (Removes 42nt including the stop codon)
		atm-1: ATG(START)AAACATTACAAAAAAA TTATTCATCAGAG	3' intron splice sequence: TCATCAG//AG (Removes 30nt including the start codon)

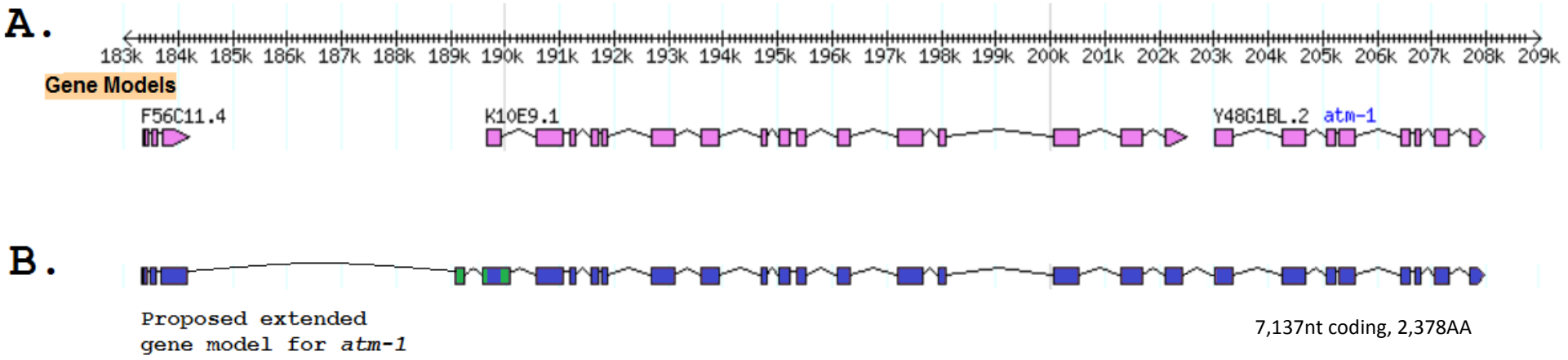

Figure 3.9 Comparison between the predicted gene models and the extended gene model. A. Wormbase.org predicted gene models of *F56C11.4*, *K10E9.1*, and *atm-1*. B. The revised extended gene model for *atm-1* merging *F56C11.4*, *K10E9.1*, and *atm-1*. The green highlights indicate novel *atm-1* coding sequences (Table 3.2).

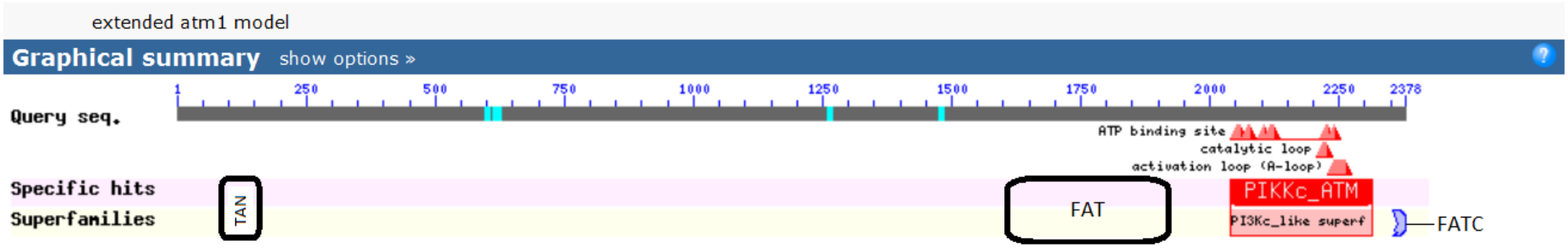

Figure 3.10 CDD prediction of the extended *atm-1* protein sequence. Only the PI3K_c and FATC domains are identified through bioinformatic prediction. Estimated locations of where the TAN and FAT domains should be predicted are indicated. Modified from Conserved Domains Database. http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

3.2 Molecular characterization of the *atm-1* mutant strain, VC381

3.2.1 *atm-1(gk186)* is a deletion that does not prevent transcription

The *atm-1* mutation, *gk186*, is a deletion that removes the 3' intron splice site for the 22nd exon of the revised *atm-1* gene model (Figure 3.9). This mutant strain, named VC381, is thought to have null ATM-1 function because exon 22 lacks the required sequence information for correct splicing. This will most likely result in exon skipping of exon 22 resulting in the splicing of exon 21 to exon 23. This transcript is predicted to produce a frameshift mutation resulting in a truncated protein lacking the kinase domain and null ATM-1 function [21]. Another argument for VC381 to have null ATM-1 function is through degradation of the *atm-1* mRNA due to NMD resulting from the multiple PTC that would be present in the mRNA if exon 22 was skipped [23]. To gain insight to which mechanism is utilized (exon skipping or NMD), I generated a cDNA library from the mutant strain VC381. From this cDNA library, I probed for *atm-1* transcripts. I used the same primers that were used to characterize wildtype *atm-1* (Table 3.1). The *atm-1* transcripts produced in VC381 are stable cDNAs, and can be used for PCR analysis (Figure 3.11). Consistent with the predictions, the PCR products encompassing the deletion region are shorter than their PCR products from wildtype worms (Figure 3.11, Lanes 5 and 6). These transcripts suggest that the *atm-1* mRNAs are not immediately degraded via NMD.

3.2.2 *atm-1(gk186)* produces two alternative transcripts near the deletion region

An unexpected result from the *atm-1(gk186)* cDNA PCRs was the production of two PCR products from primers that encompassed the deletion (Figure 3.11, Lanes 5 and 6). I verified these bands by running the PCR products at a lower concentration to reduce visual saturation (Figure 3.12). In comparison to wildtype cDNA PCR products amplified with the same primers, it is obvious that the *gk186* allele is a deletion that generates two alternative transcripts (Figure 3.12).

3.2.3 The *atm-1(gk186)* alternative transcripts do not skip exon 22 and utilize novel unpredicted splice sites

To further understand the molecular nature of these transcripts, the two *gk186* transcripts were purified and sequenced. To ensure sequence fidelity, 2-fold coverage was obtained for each transcript, which I have designated as *atm-1.a* and *atm-1.b* for the top and bottom bands (larger and smaller transcript), respectively (Appendix C). Surprisingly, both transcripts did not skip the full exon affected by the deletion. Sequencing of *atm-1.a* revealed the use of a novel 3' intron splice site within the affected exon (exon 22)(Figure 3.13B). The nucleotide sequence and changes are summarized in Table 3.3. *atm-1.b* incorporates the previous novel 3' intron splice seen in *atm-1.a* but also utilizes a novel 5' splice site within the previous exon (exon 21) (Figure 3.13C, Table 3.3).

3.2.4 The *gk186* transcripts are predicted to produce distinct products

To study the consequences of these new transcripts, the full *atm-1.a* and *atm-1.b* transcripts were translated (Appendix D). Multiple stop codons appear when *atm-1.a* is translated with the first stop occurring quickly after the new splice site. This results in a truncation of the protein at residue 1843. The effect of this truncation is the removal of the catalytic PI3K domain (Figure 3.14A). Presumably, this protein would lead to a kinase-null mutant. Interestingly, translation of *atm-1.b* reveals that the effect of the two novel spice sites results in an in-frame translated protein. The full *atm-1.b* transcript is predicted to be a 2,266AA polypeptide, which leads to a protein that is 112AA residues shorter than the wildtype polypeptide and would maintain the catalytic PI3K domain at the C-terminal end (Figure 3.14B). The implications of this transcript with the preserved kinase domain will be addressed in the Discussion.

3.2.5 Section summary

The *atm-1 gk186* allele is a deletion that results in unexpected consequences for the transcription and translation of *atm-1*. The deletion affects splicing but not by exon-skipping as expected. Stable *atm-1* transcripts are detected in the transcriptome corresponding to two alternative transcripts. These transcripts arise as a result of novel splice sites in the vicinity

of the deletion. The larger *atm-1.a* transcript uses a novel splice site within exon 22. This splice variant results in multiple PTC, the first stop truncating the protein immediately after the new splice site. The smaller *atm-1.b* transcript uses two novel splice sites; one splice site within exon 21 and the other within exon 22. This transcript translates in-frame with 112AA lost relative to its wildtype *atm-1* counterpart. The *atm-1.b* transcript is predicted to have an intact kinase domain.

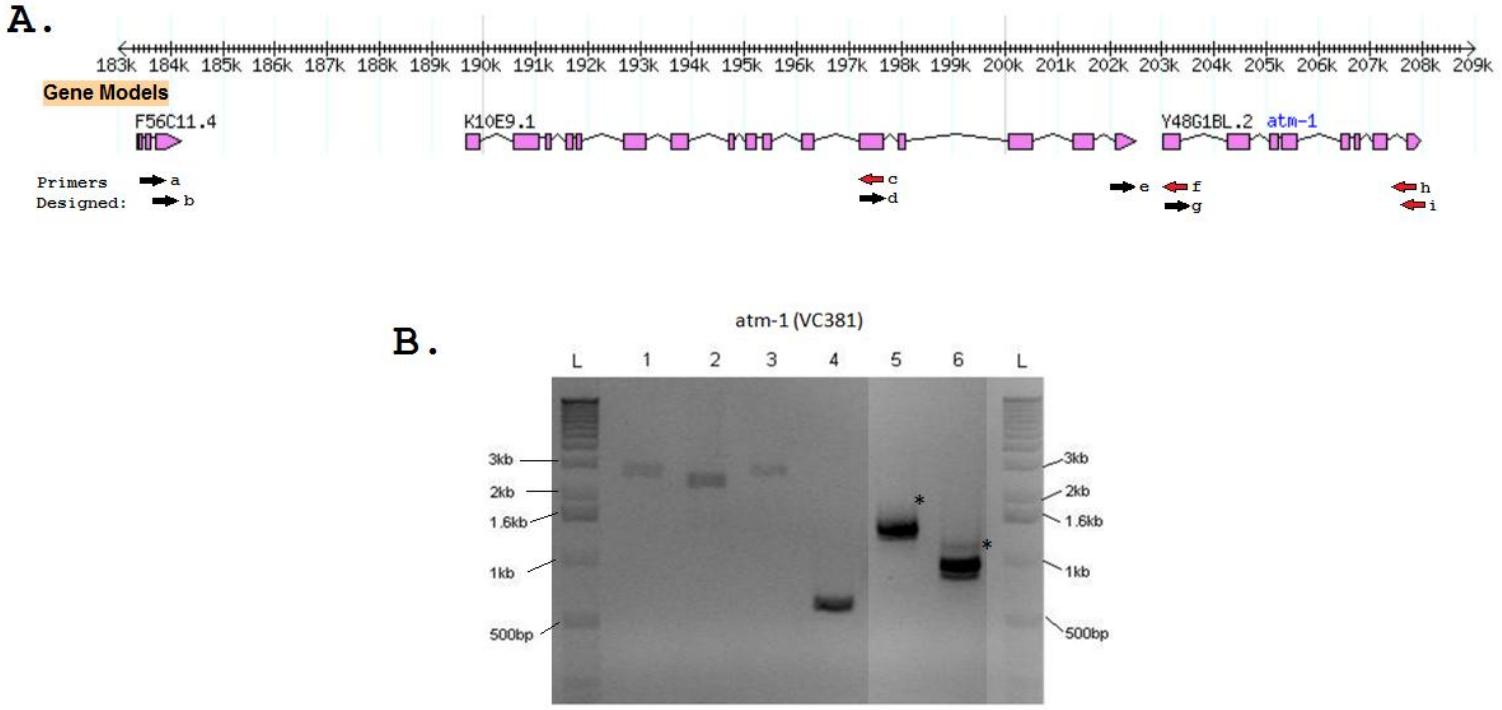

Figure 3.11 *atm-1* cDNA primer design and products. A. Primers designed for amplifying overlapping PCR products from the wildtype cDNA library for the extended gene model. Primer sequences are listed in Table 2.1. Black arrows indicate forward primers and red arrows indicate reverse primers. B. Agarose gel with products amplified from VC381 cDNA library for overlapping products of the extended gene model shown in Figure 3.9B. Each lane represents products from a primer combination. Lane: 1. a+c, 2. b+c, 3. d+f, 4. e+f, 5. g+h 6. g+i. Expected product sizes are given in Table 3.1. (*) Expected sizes of products in lanes 5 and 6 are 1,279bp and 788bp respectively if exon-skipping of the 22nd exon occurs. (*) Lanes 5 and 6 on this gel are flipped horizontally to be visually consistent with other figures

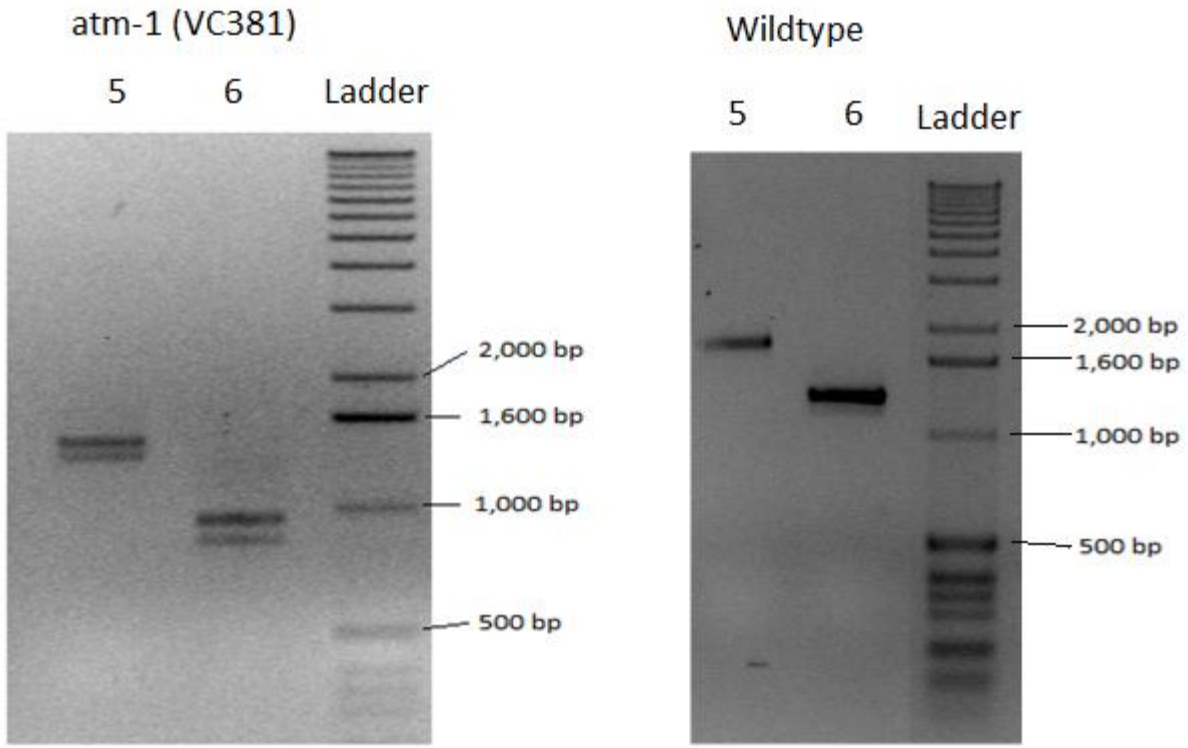

Figure 3.12 Comparison between *atm-1* and wildtype cDNA products using two primer sets; g + h (Lanes 5), and g+i (Lanes 6). Double bands are produced for both primer sets when product is amplified from the *atm-1* cDNA library. Single bands are produced from the wildtype cDNA library when using the same primer sets.

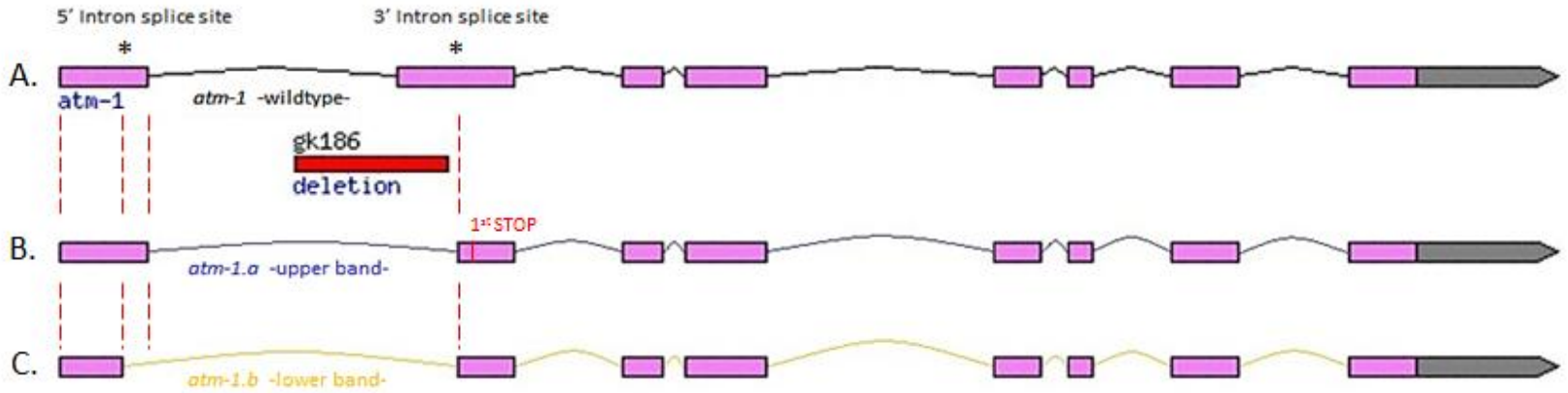

Figure 3.13 Abbreviated gene models for *atm-1*. (A) wildtype, (B) *atm-1.a* and (C) *atm-1.b* transcripts. The first stop encountered during translation for *atm-1.a* is indicated with a "1st STOP". (*) Indicates location of the novel splice sites used in splicing for the two transcripts.

Table 3.3 Summary of changes in the *atm-1(gk186)* transcripts. Exon designation based on the extended gene model (Figure 3.9B)

Transcript	5' Intron Splice Sequence : Location of change	3' Intron Splice Sequence : Location of change	Transcript Consequence
atm-1.a	AG/TATGCATC (Wildtype) :end of 21st exon	ATCACAG/AT :within the 22nd exon	Frameshift, multiple stop codons. Truncated protein lacking PI3K domain
atm-1.b	GG/GTGAGAAA :within the 21st exon	ATCACAG/AT :within the 22nd exon	In-frame splicing. 112AA removed but maintains the PI3K domain

Figure 3.14 Conserved Domains Database depiction of the two *atm-1* translations. A. *atm-1.a* translation reveals multiple stop codons, the first located at residue 1843 (Indicated by a red line and *) effectively truncating the protein product without the kinase domain (Red X). B. *atm-1.b* translation reveals a deletion of 112 AA starting at residue 1802 (Indicated by a red line and **). This protein maintains its PI3K_c (labelled PIKKc_ATM) and FATC domains but is 2266AA long (Indicated by a red circle). Estimated locations of where the TAN and FAT domains should be predicted are indicated. *atm-1.b*'s FAT domain may be affected by the deletion. Modified from Conserved Domains Database. http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

3.3 Phenotypic characterization of VC381

3.3.1 Mutant *atm-1(gk186)* does not have any obvious phenotypes

Before any phenotypic characterizations were conducted, the mutant strain VC381 was outcrossed to our laboratory wildtype strain, VC2010, to remove any background mutations left by the TMP/UV treatment. The outcrossing was done ten times and the homozygous *gk186* deletion was reisolated and confirmed using a poison primer PCR technique (Materials and Methods [47]). The 10^{th} outcross (Strain: KR4941) was subsequently used for all phenotypic characterizations (Table 2.2).

Morphologically, *atm-1(gk186)* appears wildtype and does not immediately exhibit any obvious phenotypes. However, brood analysis revealed that it has an average brood size of 206.6 (compared to 300 for wildtype) at 20°C and produces males at a 0.15% frequency (compared to 0.1% for wildtype [29]) (Table 3.4).

3.3.2 *atm-1(gk186)* is sensitive to double strand breaks caused by ionizing radiation

The ATM protein in other species is known to respond to DSB damage (Introduction). The sensitivity to DSBs of the *C. elegans* gene *atm-1(gk186)* was assessed using a protocol Youds *et al* established in the Rose lab (Youds, PhD Thesis 2007, University of British Columbia)[50]. This assay measures the percent survival of progeny of a parental hermaphrodite post ionizing radiation (IR). The rest period post IR allowed me to collect a specific brood and measure the impact of DSBs on the meiotic pachytene cells during IR treatment [43, 50]. Along with *atm-1,* I assayed VC2010 wildtype as a negative control and *brd-1* as the positive control. *brd-1,* which is one half of a heterodimer with *brc-1,* is required for homologous recombination (HR) repair of DSBs therefore worms deficient in *brd-1* are unable to repair DSBs via HR and the eggs arrest as embryos. The pooled results of three biologically replicate experiments are shown in Figure 3.15. At the highest dose (45Gy) tested, wildtype worms had a survival rate around 90%. The same dose on *brd-1* is nearly lethal. *atm-1* mutants are also sensitive to IR, although not as much as *brd-1*. Overall, the results indicate that *atm-1* is sensitive to IR.

3.3.3 *atm-1(gk186)* spontaneously produces a dominant high incidence of male (Him) phenotype and eventually becomes sterile

Maintaining *atm-1(gk186)* worms through generations has yielded an unexpected result. *atm-1* normally produces a low frequency (0.3%) of males (Table 2.2). However, over several generations, the frequency of males increases and the overall fecundity of the worms was negatively affected (data not shown). The increased male frequency, referred to as the Him phenotype is dominant and persists until the strain becomes sterile (data not shown). Multiple *atm-1(gk186)* Him strains were isolated and frozen (Table 2.2).

3.3.4 Him and sterile *atm-1(gk186)* strains have a reduced number of diakinetic chromosomes

Dominant Him strains have been characterized before [51]. Many of these Him strains yield translocations between the X chromosome and an autosome. More recently, Ahmed and Hodgkin have identified another mutant (*mrt-2*) that results in X:autosome fusions [52]. To understand how the Him phenotype arises in my strains, I visualized the chromosomes using DAPI staining. Normal wildtype and freshly outcrossed *atm-1* worms have the normal complement of 6 pairs seen as 6 distinct spots when visualized using DAPI (Figure 3.16A). In the dominant Him *atm-1(gk186)* worms, the DAPI staining revealed a reduced number of diakinetic chromosomes. The different Him strains collected in the lab show examples of worms with 5, 4, and 3 pairs of diakinetic chromosomes (Figure 3.16B, C, and D respectively, and summarized in Table 3.5).

3.3.5 Quantifying independent *atm-1(gk186)* lines indicate inherited genomic instability

Chromosomal fusions, translocations, Hims (X nondisjunction), and sterility are all indicators of genomic instability [16, 51, 52]. To gain insight and to quantify the genomic instability defect in *atm-1* mutants, I individually plated 20 L4 *atm-1(gk186)* worms. The progeny of each worm was scored and a subsequent generation for each line was propagated through a single L4 offspring. This was repeated for 20 generations and the results are summarized in Figure 3.17 and Table 3.6. The results indicate that *atm-1* worms do display genomic instability. Through 20 generations, 16 of 20 lines have become sterile. Various

outcomes are observed as the independent lines can become only sterile, Him, or first become Him and then sterile. These results also point to a model in which heritable mutations that confer Him and sterility are generated in an *atm-1(gk186)* background. The progeny numbers for each generation of different lines are also unpredictable with some gradually decreasing in progeny numbers where other lines have fluctuating progeny numbers (Table 3.6).

3.3.6 *atm-1(gk186)* produces lethal mutations that can be captured by the *eT1* balancer system

In order to capture mutational events induced by *atm-1(gk186)* for further analysis, I created *atm-1; dpy-18/eT1(III);unc-46/eT1(V)* which uses the *eT1* balancer system in an *atm-1(gk186)* background to try and capture mutational events that lead to lethality (Materials and Methods)(Table 2.2). With the help of Dr. Ann Rose, 1139 F_2's were plated and their F_3 progeny screened for the lack of DpyUncs. A total of 11 independent lines lacking DpyUncs were obtained after rescreening potential positives (Table 3.7). Thus, the mutational events created in the *atm-1(gk186)* background are amenable to capture by the *eT1* balancer.

3.3.7 The forward mutation frequency of *atm-1(gk186)* is similar to the general mutator *dog-1*

Our mutation screen revealed a forward mutation frequency of 0.97% (from the observation of 11 captured mutations in 1139 lines; Table 3.7). This is a very similar mutation rate to that of *dog-1* (0.9%) which is responsible for resolving G track secondary structures during replication [34]. My 0.97% is similar to 500 Roentgens of gamma radiation (0.96%), less than 0.004M ethyl methanesulfonate (EMS) (1.5%), less than 0.11% formaldehyde treatment (1.58%), less than 120J/m^2 UV exposure (3.1%), and greater than the spontaneous rate of 0.06% of the *eT1* balancer [30, 31, 33].

3.3.8 Section summary

The *atm-1* deletion strain, VC381, appears morphologically wildtype. However, the *atm-1(gk186)* allele demonstrates genomic instability and is sensitive to DSBs caused by IR. The observed genomic instability events appear to be rearrangements that reduce the chromosome number but maintain a full complement of genomic information for normal development.

Moreover, the *atm-1(gk186); eT1* balancer system was able to capture lethal mutations. Using this system we observed a forward mutation rate of 0.96%, which is similar to *dog-1* which acts as a general mutator due to its role in resolving G tract secondary structure in DNA and produces a variety of mutational events [34].

Table 3.4 VC381 characteristics at 20°C (*Data from Plates 6-10 done by Jessica McLellan).

Plate #	**Total Brood**	**Total Males**
1	193	0
2	154	0
3	203	0
4	186	1
5	196	1
6*	227	0
7*	237	0
8*	214	0
9*	238	0
10*	218	1

Totals	2066	3	Percent Males
Average	206.6	0.3	0.15%
Standard Error	8.17	0.15	0.07%

NOTE: 1 sterile was omitted from the progeny count scored by Jessica McLellan (Rose Lab, University of British Columbia).

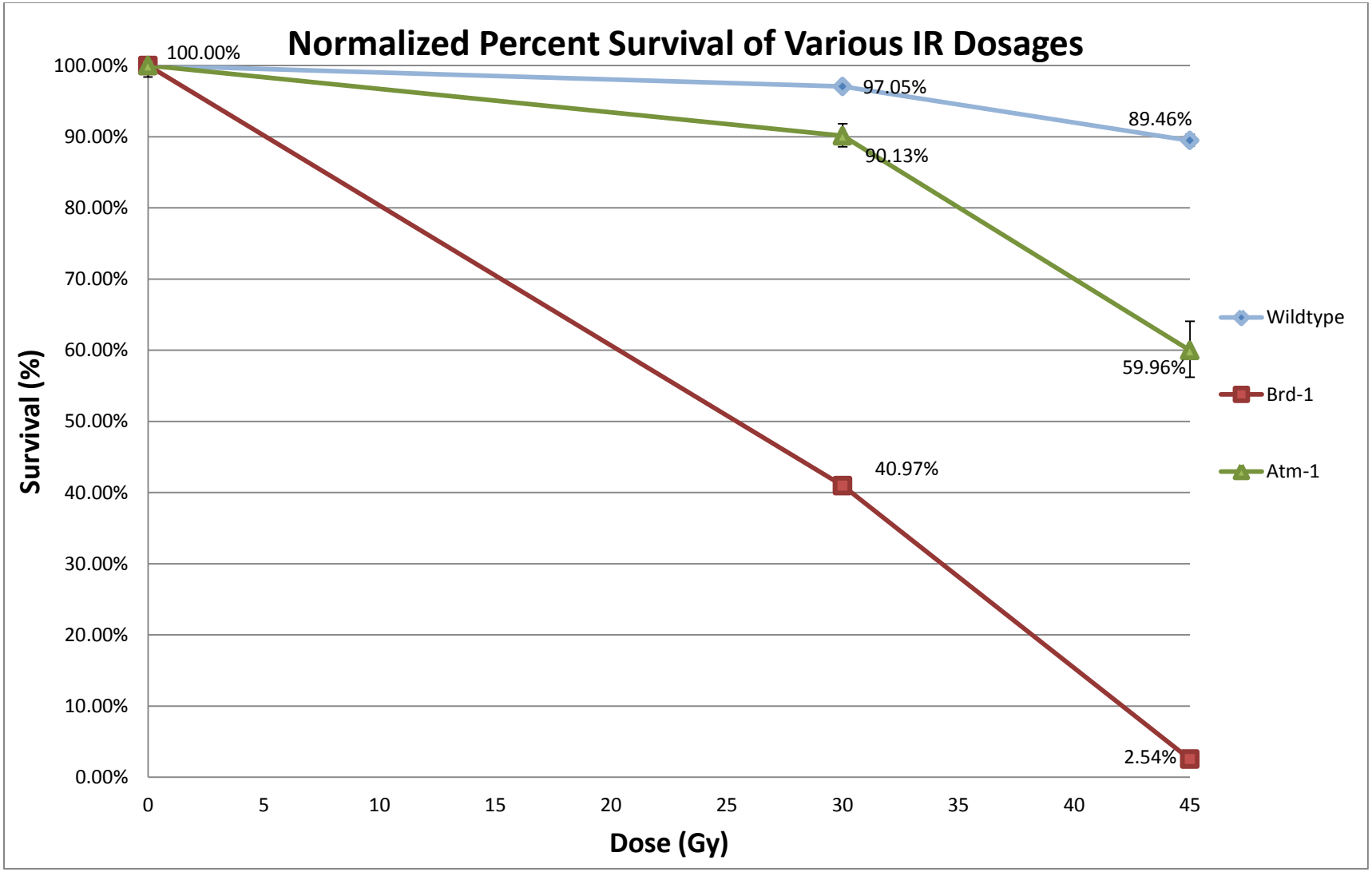

Figure 3.15 The *atm-1* mutant is sensitive to IR. The graph shows normalized percent survival for wildtype, *brd-1*, and *atm-1*, at three IR dosages: 0, 30, and 45 Gy. Standard error bars are indicated. These results were pooled from three biologically replicated experiments.

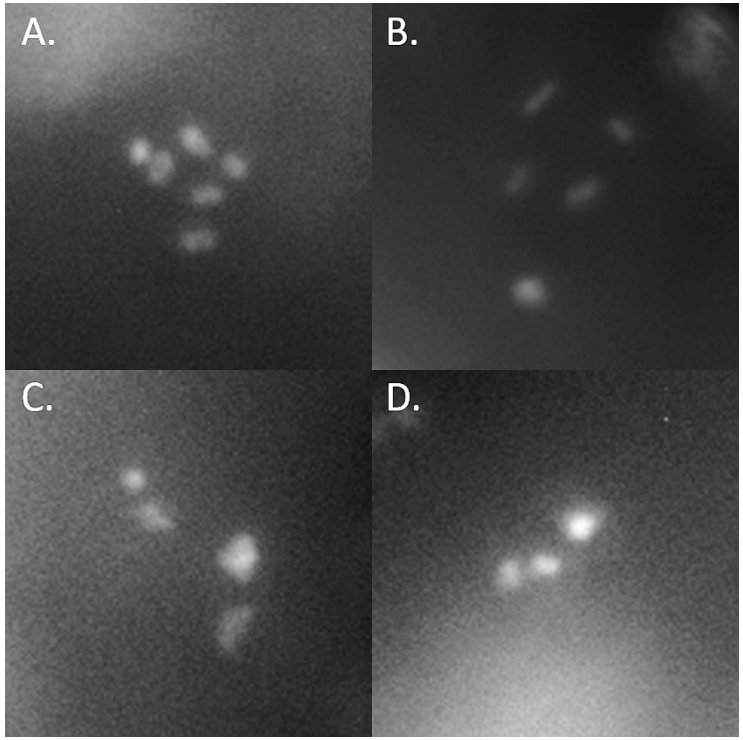

Figure 3.16 *atm-1* mutants have reduced number of diakinetic bodies. DAPI Staining of oocytes in the hermaphrodite gonad arm. Examples of various chromosome bodies observed: A. 6 bodies seen in progenitor strain *atm-1(gk186)*, B. 5 bodies seen in *atm-1(gk186;h2681)*, and C. 4 bodies seen in *atm-1(gk186;h2680)*, D. 3 bodies seen in *atm-1(gk186;h2685)*Table 3.5.

.

Table 3.5 List of obtained *atm-1(gk186)* Him strains with their associated diakinetic chromosome numbers VIA DAPI.

atm-1(gk186) strain allele	# of Diakinetic Chromosomes Observed
h2680	4
h2681	5
h2682	5
h2683	5
h2684	5
h2685	3

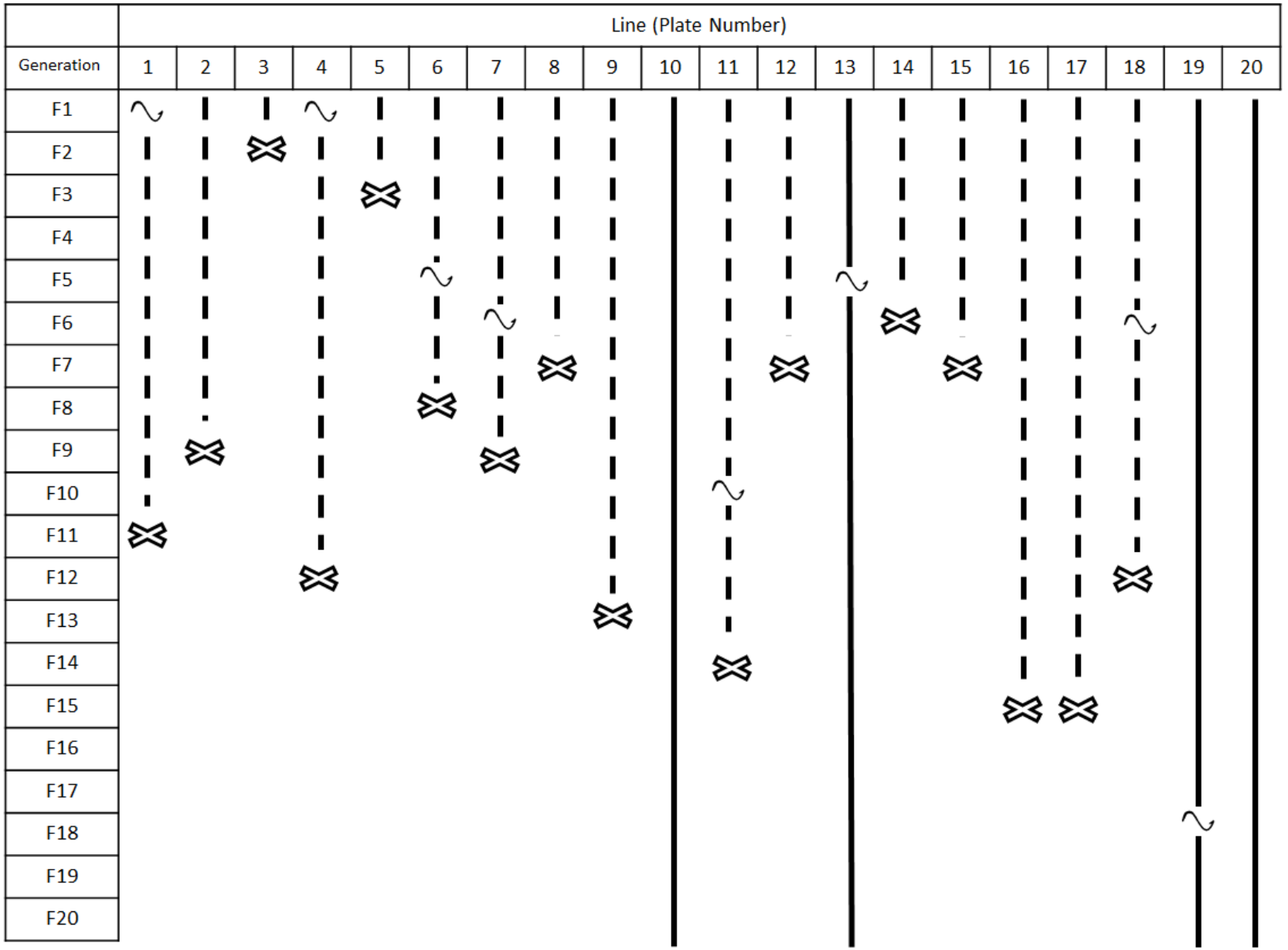

Figure 3.17 *atm-1(gk186)* mutants have genomic stability defects. 20 independent lines propagated over 20 generations. The generation at which the line became Him and/or sterile are indicated with a squiggly hook (∿) and X, respectively.

Table 3.6 The progeny of *atm-1(gk186)* fluctuates or decreases as a function of generations. Total progeny counts for the 20 independent *atm-1(gk186)* lines for 20 generations as seen in Figure 3.17. (-) indicates sterility.

Line	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}	F_{19}	F_{20}
1	108	116	142	144	167	4	37	11	12	5	-									
2	282	144	147	160	218	113	69	87	-											
3	245	-																		
4	110	49	41	24	78	91	48	135	160	49	97	-								
5	63	1	-																	
6	126	37	60	80	37	30	6	-												
7	79	138	144	202	126	107	61	79	-											
8	246	133	89	78	172	10	-													
9	136	77	125	130	230	61	114	37	103	183	88	113	-							
10	187	6	40	59	98	38	37	31	45	81	107	43	113	148	166	41	107	181	53	77
11	262	116	187	158	206	206	154	187	59	138	18	20	23	-						
12	118	58	79	74	84	20	-													
13	225	83	146	116	131	107	96	77	35	38	31	93	123	168	298	158	165	111	58	1
14	141	16	12	25	38	-														
15	16	30	32	29	24	14	-													
16	84	83	128	130	159	203	51	43	94	135	83	23	61	67	-					
17	240	125	95	44	125	57	119	115	37	68	51	60	20	3	-					
18	212	111	178	152	67	60	83	5	8	58	7	-								
19	108	151	189	121	95	232	187	117	144	61	37	104	4	90	155	47	46	82	24	29
20	91	61	107	31	44	86	85	60	40	81	74	40	50	33	34	14	9	100	33	24

Table 3.7 Summary of *atm-1; dpy-18/eT1(III);unc-46/eT1(V)* screen.

Number of F_2's Set Up	Number of Plates Lacking Dpy Uncs	Number of Steriles	Numbers of *eT1s*	Total of Lethals After Re-screening	Forward Mutation Rate = Number of mutations / Total screen (non- steriles and non-*eT1*s) * 100
1180	44	28	13	11	11 / (1180-28-13) * 100 = 0.97%

4. Discussion

4.1 The extended gene model requires further refinement

Revealing the structure and function of ATM is an important step in understanding the human disease A-T and requires a model organism that is capable of providing both aspects. Here, I have established that *C. elegans* can be used as a model for ATM by characterizing the nematode orthologue, *atm-1*. I have characterized the gene structure and documented the phenotype of the deletion mutant. Identifying the correct gene structure for any given gene is significant because it allows future molecular and genetic experiment to be designed and carried out correctly. I show that the predicted *atm-1* gene model lacked N-terminal sequences that are poorly conserved. The reinterpretation merges the *atm-1* gene prediction with the *K10E9.1* and *F56C11.4* predictions, producing a predicted protein that is similar to the *Drosophila* ATM (Figure 3.1). The size of the reinterpreted gene model is now more comparable to other known ATMs. In addition to the size difference, previous unpredicted exonic sequences were found and may be important for protein function.

I sequenced the cDNA for the reinterpreted *atm-1* gene. Unlike other RNA-seq techniques (EST, cDNA-seq, RNA-seq, serial analysis of gene expression (SAGE)) I designed gene specific primers to achieve this. Because of the primer design, my products do not amplify the end of genes. These sequences require 5' and 3' RACE (rapid amplification of cDNA ends) to identify the *atm-1* start and stop sequences. Thus, in principle, my extended gene structure is incomplete at the 5' and 3' ends and requires end sequences. However, there are new data sets that could be data-mined through WormBase to potentially identify the end sequences of *atm-1* if there is enough coverage [28, 37]. These un-biased approaches could potentially identify any alternative starts or poly-adenylation signals.

Although the gene structure for *atm-1* appears complete, the still unrecognized TAN and FAT domains in the *atm-1* CDD prediction are a bit of a concern. Because roughly two thirds of the protein is species specific, prediction programs like the one used in CDD may not recognize these domains in the *atm-1* sequence. These domains may have to be extrapolated first using other *Caenorhabditis* species to see which regions are most

conserved and then aligned to higher eukaryotes. These domains would be important to identify as they could be regulatory domains that could define ATM-1's spatial and temporal control and would be of interest for understanding its role in the DNA damage response.

4.2 The alternative transcripts of *atm-1(gk186)* may have residual ATM-1 function

Stergiou *et al* have reported a transcript that has a "cryptic" splice site within exon 2 (of the original prediction)[42]. However, the authors fail to provide the sequence for this transcript but based on its translation, assume a functionally null protein from the multiple PTC seen. Here, I first show that NMD does not immediately degrade the *atm-1(gk186)* transcripts. Secondly, I provide the sequence for this "cryptic" splice site in exon 2, which is exon 22 in the extended gene model. Comparing my transcript sequence and Stergiou *et al*'s transcript sequence would be the only way to verify that they are the same. The transcript containing this "cryptic" splice site is named *atm-1.a* in my report, and it does indeed produce multiple PTC when translated. However, I also provide sequence for a second transcript, *atm-1.b*, which is also a result of the *gk186* deletion and has not been reported in the literature. *atm-1.b* utilizes two alternative splice sites that translate in-frame. 112AA is deleted from the resulting transcript. This protein retains its PI3K domain and thus cannot be assumed to be functionally null. *atm-1.b* places doubt on whether *atm-1(gk186)* is a true knockout or a mere knockdown. The possibility of *atm-1.b* producing a neomorph would have to be experimentally tested as well. This new information would require further experimentation to identify the protein consequence of *gk186* and may require re-evaluation of the current knowledge regarding *atm-1*. Regardless, the phenotypes observed for *atm-1(gk186)* are genuine but may not confer true *atm-1* knockout function.

4.3 Function conservation of *atm-1(gk186)* phenotypes

atm-1(gk186) is sensitive to IR but not as much as *brd-1*. This may be due to functional redundancy of the related gene *atl-1,* the human ATR orthologue [17, 44]. Another explanation could be the presence of functional ATM-1 produced from the *atm-1.b* transcript described previously. The IR sensitivity experiment conducted is most similar to the L4 IR sensitivity done by Bailly *et al [43]*. The results Bailley *et al* obtained conflict with mine as they do not observe IR sensitivity for *atm-1(gk186)*. Experimental factors such

as the brood size collected and the rest period may explain the discrepancy (i.e., 5h vs 12h brood, 20h and 24h rest period post IR). It seems unlikely that there would be no effect for IR on meiotic cells as Stergiou *et al* and Garcia-Muse and Boulton both show a defect in apoptosis induction, which would result in oocytes with unrepaired DSBs [42, 44]. These unrepaired breaks would presumably lead to some form of damage, causing arrested embryos.

The indicators of genomic instability in *atm-1* mutant worms are consistent with what is reported for human A-T cell lines. In human cell lines, reduced telomere length, increased rate of telomere shortening, and telomere fusions are observed [7-9]. In both fission and budding yeast, a role for preventing telomere fusions and chromosome translocations has been documented [53, 54]. The *Drosophila* orthologue, *tefu*, also prevents telomere fusions [55]. Here, I have observed that some *atm-1* mutants eventually become Him. Him strains have been documented before and have been shown to be X:autosome rearrangements [51]. X:autosome telomere fusions, which also have a Him phenotype, have also been documented in *mrt-2* mutants [52]. These worms show a reduced number of diakinetic chromosomes when DAPI stained, which is consistent with my mutant *atm-1* strains. Thus, it is likely that the *atm-1* mutants I have collected could be X:autosome telomere fusions. This would have to be confirmed through experiments such as oligo array comparative genomic hybridization (oaCGH), but would be consistent with ATM function and *C. elegans* Him strains. A telomere function for *atm-1* would be novel for *C. elegans*.

ATM-1's function in preventing telomere fusions does not appear to be specific for the X chromosome. Some of the independent lines in the genomic instability assay become sterile without becoming Him (Figure 3.17). This suggests that autosome:autosome events might be occurring leading to sterility. These strains along could also be further characterized using oaCGH to understand the nature of the sterility and may provide insight to the function of *atm-1* in promoting genomic stability, possibly through telomere maintenance.

The use of *eT1* to try and capture lethal mutations caused in an *atm-1* background was successful. The forward mutation rate for *atm-1(gk186)* was calculated to be 0.96%, which

is comparable to another DDR gene, *dog-1*, which acts to resolve G tract secondary structure during replication [50]. This low rate is similar to low dosages of other mutagens (EMS, gamma radiation, and formaldehyde)[30, 31, 33]. These mutations would have had to occur on Chr. III right and Chr. V left in order to be captured by the *eT1* balancer system, and thus these rates realistically represent only $1/6^{th}$ of the genome. It would also be interesting to understand the nature of these captured mutations to see if mutations such as insertions, deletions, and inversions also occur in an *atm-1* background.

If we consider each of the 16 independent lines that became Him/sterile a mutational event, then 16 events occurred through 230 generations, which is the sum of all the generations lived by all 20 lines (Figure 3.17). This translates into a 7% forward mutation rate, which is comparable to the *atm-1(gk186); eT1* balancer rate of 5.8% (6x 0.96%). This suggests that *atm-1* functions ubiquitously throughout the genome to promote genomic stability. Thus, although not an essential gene, *atm-1* clearly has a significant role in maintaining genomic stability by responding to DSBs and possible telomere maintenance function in *C. elegans*.

5. Conclusions

Here I have shown that the *C. elegans* ATM orthologue, *atm-1*, has been mis-annotated. The reinterpreted gene model for *atm-1* is a merger of the *F56C11.4, K10E9.1,* and *atm-1* gene predictions. The *gk186* deletion for *atm-1* produces two alternative transcripts, *atm-1.a* and *atm-1.b.* The former translates into multiple PTC, while the ladder translates into a product that has 112AA removed but retains the PI3K domain. This product may have kinase function. *atm-1* is sensitive to IR and displays indicators of genomic instability. Lethal mutational events were captured through the *eT1* balancer system. Future work on *atm-1* in *C. elegans* may provide insight into the DDR and intracellular signalling and control. Perhaps the most significant contribution to understanding *atm-1,* and ATM in general, would be discovering novel therapeutic targets for ATM driven cancers. ATM interactions for synthetic lethality could lead to targeted therapy for appropriate cancers. Research in these areas would be greatly facilitated through *C. elegans* because of the worm's ability to be amendable to high throughput, cost effective methods.

Bibliography

1. Kurz, E.U. and S.P. Lees-Miller, *DNA damage-induced activation of ATM and ATM-dependent signaling pathways.* DNA Repair (Amst), 2004. **3**(8-9): p. 889-900.
2. Lavin, M.F., *Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer.* Nat Rev Mol Cell Biol, 2008. **9**(10): p. 759-69.
3. Lavin, M.F., D. Delia, and L. Chessa, *ATM and the DNA damage response. Workshop on ataxia-telangiectasia and related syndromes.* EMBO Rep, 2006. **7**(2): p. 154-60.
4. Lavin, M.F. and Y. Shiloh, *The genetic defect in ataxia-telangiectasia.* Annu Rev Immunol, 1997. **15**: p. 177-202.
5. Rotman, G. and Y. Shiloh, *ATM: from gene to function.* Hum Mol Genet, 1998. **7**(10): p. 1555-63.
6. Kojis, T.L., et al., *Tissue specificity of chromosomal rearrangements in ataxia-telangiectasia.* Hum Genet, 1989. **83**(4): p. 347-52.
7. Pandita, T.K., S. Pathak, and C.R. Geard, *Chromosome end associations, telomeres and telomerase activity in ataxia telangiectasia cells.* Cytogenet Cell Genet, 1995. **71**(1): p. 86-93.
8. Metcalfe, J.A., et al., *Accelerated telomere shortening in ataxia telangiectasia.* Nat Genet, 1996. **13**(3): p. 350-3.
9. Xia, S.J., M.A. Shammas, and R.J. Shmookler Reis, *Reduced telomere length in ataxia-telangiectasia fibroblasts.* Mutat Res, 1996. **364**(1): p. 1-11.
10. Savitsky, K., et al., *A single ataxia telangiectasia gene with a product similar to PI-3 kinase.* Science, 1995. **268**(5218): p. 1749-53.
11. Savitsky, K., et al., *The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species.* Hum Mol Genet, 1995. **4**(11): p. 2025-32.
12. Lavin, M.F. and S. Kozlov, *ATM activation and DNA damage response.* Cell Cycle, 2007. **6**(8): p. 931-42.
13. Shiloh, Y., *The ATM-mediated DNA-damage response: taking shape.* Trends Biochem Sci, 2006. **31**(7): p. 402-10.
14. Lee, J.H. and T.T. Paull, *ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex.* Science, 2005. **308**(5721): p. 551-4.
15. Uziel, T., et al., *Requirement of the MRN complex for ATM activation by DNA damage.* EMBO J, 2003. **22**(20): p. 5612-21.
16. Negrini, S., V.G. Gorgoulis, and T.D. Halazonetis, *Genomic instability--an evolving hallmark of cancer.* Nat Rev Mol Cell Biol, 2010. **11**(3): p. 220-8.
17. Liang, Y., et al., *DNA damage response pathways in tumor suppression and cancer treatment.* World J Surg, 2009. **33**(4): p. 661-6.
18. Houtgraaf, J.H., J. Versmissen, and W.J. van der Giessen, *A concise review of DNA damage checkpoints and repair in mammalian cells.* Cardiovasc Revasc Med, 2006. **7**(3): p. 165-72.
19. Lemmens, B.B. and M. Tijsterman, *DNA double-strand break repair in Caenorhabditis elegans.* Chromosoma, 2011. **120**(1): p. 1-21.
20. Hastings, M.L. and A.R. Krainer, *Pre-mRNA splicing in the new millennium.* Curr Opin Cell Biol, 2001. **13**(3): p. 302-9.

21. Keren, H., G. Lev-Maor, and G. Ast, *Alternative splicing and evolution: diversification, exon definition and function.* Nature Reviews Genetics, 2010. **11**(5): p. 345-355.
22. Pettitt, J., et al., *The evolution of spliced leader trans-splicing in nematodes.* Biochem Soc Trans, 2010. **38**(4): p. 1125-30.
23. Chang, Y.F., J.S. Imam, and M.F. Wilkinson, *The nonsense-mediated decay RNA surveillance pathway.* Annu Rev Biochem, 2007. **76**: p. 51-74.
24. Sayers, E.W., et al., *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Research, 2010. **39**(Database): p. D38-D51.
25. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
26. Riddle, D.L., et al., eds. *C. elegans II*. Cold Spring Harbor Mongraph Series, ed. C.S.H.L. Press. Vol. 33. 1997, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York.
27. Hall, D.H. and Z.F. Altun, *C. elegans Atlas*, ed. C.S.H.L. Press. 2008, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
28. Kaletta, T. and M.O. Hengartner, *Finding function in novel targets: C. elegans as a model organism.* Nat Rev Drug Discov, 2006. **5**(5): p. 387-98.
29. Rose, A.M. and D.L. Baillie, *The Effect of Temperature and Parental Age on Recombination and Nondisjunction in CAENORHABDITIS ELEGANS.* Genetics, 1979. **92**(2): p. 409-18.
30. Johnsen, R.C. and D.L. Baillie, *Formaldehyde mutagenesis of the eT1 balanced region in Caenorhabditis elegans: dose-response curve and the analysis of mutational events.* Mutat Res, 1988. **201**(1): p. 137-47.
31. Rosenbluth, R.E., C. Cuddeford, and D.L. Baillie, *Mutagenesis in Caenorhabditis elegans. II. A spectrum of mutational events induced with 1500 r of gamma-radiation.* Genetics, 1985. **109**(3): p. 493-511.
32. Rosenbluth, R.E. and D.L. Baillie, *The genetic analysis of a reciprocal translocation, eT1(III; V), in Caenorhabditis elegans.* Genetics, 1981. **99**(3-4): p. 415-28.
33. Stewart, H.I., R.E. Rosenbluth, and D.L. Baillie, *Most ultraviolet irradiation induced mutations in the nematode Caenorhabditis elegans are chromosomal rearrangements.* Mutat Res, 1991. **249**(1): p. 37-54.
34. Zhao, Y., et al., *Spectrum of mutational events in the absence of DOG-1/FANCJ in Caenorhabditis elegans.* DNA Repair (Amst), 2008. **7**(11): p. 1846-54.
35. Consortium*, *C.e.*S., *Genome sequence of the nematode C. elegans: a platform for investigating biology.* Science, 1998. **282**(5396): p. 2012-8.
36. Harris, T.W., et al., *WormBase: a multi-species resource for nematode biology and genomics.* Nucleic Acids Res, 2004. **32**(Database issue): p. D411-7.
37. Harris, T.W., et al., *WormBase: a comprehensive resource for nematode research.* Nucleic Acids Res, 2010. **38**(Database issue): p. D463-7.
38. Chen, N., et al., *WormBase as an integrated platform for the C. elegans ORFeome.* Genome Res, 2004. **14**(10B): p. 2155-61.
39. Gerstein, M.B., et al., *Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project.* Science, 2010. **330**(6012): p. 1775-87.
40. WormBase.Org. *WormBase web site, http://www.wormbase.org, release WS205, date 18 September 2009*.

41. Yandell, M.D., L.G. Edgar, and W.B. Wood, *Trimethylpsoralen induces small deletion mutations in Caenorhabditis elegans.* Proc Natl Acad Sci U S A, 1994. **91**(4): p. 1381-5.
42. Stergiou, L., et al., *The nucleotide excision repair pathway is required for UV-C-induced apoptosis in Caenorhabditis elegans.* Cell Death Differ, 2007. **14**(6): p. 1129-38.
43. Bailly, A.P., et al., *The Caenorhabditis elegans homolog of Gen1/Yen1 resolvases links DNA damage signaling to DNA double-strand break repair.* PLoS Genet, 2010. **6**(7): p. e1001025.
44. Garcia-Muse, T. and S.J. Boulton, *Distinct modes of ATR activation after replication stress and DNA double-strand breaks in Caenorhabditis elegans.* EMBO J, 2005. **24**(24): p. 4345-55.
45. Lee, S.J., et al., *The Caenorhabditis elegans Werner syndrome protein functions upstream of ATR and ATM in response to DNA replication inhibition and double-strand DNA breaks.* PLoS Genet, 2010. **6**(1): p. e1000801.
46. Weidhaas, J.B., et al., *A Caenorhabditis elegans tissue model of radiation-induced reproductive cell death.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9946-51.
47. Edgley, M., et al., *Improved detection of small deletions in complex pools of DNA.* Nucleic Acids Res, 2002. **30**(12): p. e52.
48. Bosotti, R., A. Isacchi, and E.L. Sonnhammer, *FAT: a novel domain in PIK-related kinases.* Trends Biochem Sci, 2000. **25**(5): p. 225-7.
49. Seidel, J.J., C.M. Anderson, and E.H. Blackburn, *A novel Tel1/ATM N-terminal motif, TAN, is essential for telomere length maintenance and a DNA damage response.* Mol Cell Biol, 2008. **28**(18): p. 5736-46.
50. Youds, J.L., et al., *DOG-1 is the Caenorhabditis elegans BRIP1/FANCJ homologue and functions in interstrand cross-link repair.* Mol Cell Biol, 2008. **28**(5): p. 1470-9.
51. Hartman, P.S. and R.K. Herman, *Radiation-sensitive mutants of Caenorhabditis elegans.* Genetics, 1982. **102**(2): p. 159-78.
52. Ahmed, S. and J. Hodgkin, *MRT-2 checkpoint protein is required for germline immortality and telomere replication in C. elegans.* Nature, 2000. **403**(6766): p. 159-64.
53. Lee, K., Y. Zhang, and S.E. Lee, *Saccharomyces cerevisiae ATM orthologue suppresses break-induced chromosome translocations.* Nature, 2008. **454**(7203): p. 543-6.
54. Naito, T., A. Matsuura, and F. Ishikawa, *Circular chromosome formation in a fission yeast mutant defective in two ATM homologues.* Nat Genet, 1998. **20**(2): p. 203-6.
55. Oikemus, S.R., et al., *Drosophila atm/telomere fusion is required for telomeric localization of HP1 and telomere position effect.* Genes Dev, 2004. **18**(15): p. 1850-61.

Appendices

Appendix A : WS205 exonic sequences

A.1 *F56C11.4*

ATGAGCGTTACTCAGCTAAAAAATCTGAAACACGCGATAGCTCAGCTTCTGGAGTGGGATGGAACGAAAACGGC
GAGAAAGAAAATCGTCGACGAGGTGGTTCTTCTCTATCACGCTCTGGGAGCGGAGGCATTAAGTGAGGATAACC
AGGAGATATATGATTTGTATGATTTATCTGCACGTATATTCAACTTAGCTAAGAAAGAAATCGAAGAAGCCAAC
CAACAATTCGAAAAAGAACGAAAAAAGGGCACAAGACGAAGCGAAAAACCAGTCCCAACTCCACTTTTCGAGCT
ATCAATACAACATTTAAAACGTTGCTGTCAGCAGGGAATCGATCACAATCAGGTGCCATGGATAGCATATTGTC
TGAAACTACTGGAATTTCCAATCACAATCACCGAAAAATCGATCGAAAACGAGATTTCCAACGTGCTCCTATTG
AGCTCCAACGCCTCACAGCTCCATTGGGCCGAGCATGCTCATTTGAGCAGCTTATGGAAATGGATTTGGAGCCG
TGTCGAGACCGCCGATATTGGGGCACTCGCCATGAGAAATTATATGGAATTGGCGGCGAACTTGCTGGAAAACG
TGGATTACGTGGTTTTCGAGAAGTCGCCTATTGACCTGATGGCGAAAGTGATGGGAACACTGAAGAAAAGTGTG
GAAATGGGAAATCCCAAAGAATAG

A.2 *K10E9.1*

ATGCGTTTGATTGATGATCTCGTGAAGCTCGCCATGATCGAAACCGTACACGGCCACCGTACCATGAACGAAGT
GACACGTGGAAATATTCAAAAACTCGTGAAAACCGGAATCCAAGAGTCTCTGAAATCGGCGCACCGAAATTTCT
CAAGGAGTTCGACATTTTCGATTTCCGAAGAATGCGTGAGATATCTGACGAGATGGTTGTTGGCCGAACGAAGA
CTTGAACAGCCGTCTGCGGCTATGAATGAATCGTTTGAATTGACCGGCCTTATCGATCTCTCCTTTGGCTCGAC
GATTTCCGGAAAACATAAATTGAACGCGTGGAATGGTGTAATGCAAATCCTGAATGAGCTCCTAAAAAGTCGAC
GACTCGAACTTCAAGTCACTGAAAAAATCGTGACAATCCTCTGGGAAAAGCGAAAATCCTACACAACGGAGCCA
CTCCGTACTGTGTTCTGCTCCATTCTCTCCACAGTCGTCTGCCAGGCCGATGTTCGATTCGGTCATCGGAAAGT
GCCGACAATCGACTCGATTCTCAAATATTCGCTGTCTCTAATGCCAAATGTCGCATCTCTTCCCAGTGCCGCTG
CGTTGACCGAAACGATTGTCAGATTCAGGACAGTATCACGAGAGGGTCTCCGTAACACGTGGGATACCGTATCC
CGAACTAGCTCCGGCTCATTTGAAGTTGTTCGGCTGATTTCGGCGTTGATCTCCGTTACGGAATTTGATGAGAA
TTCGAGATTCGCCAACGATGAGAGAGTTCGGAGTTGGAGTTTTCGAAAAGACATAATCGAATGGGTATTGCTGG
ATCCGAATGCACATAGTCACAAATTACTCTATCAGTTGTGCCAGTATCATCCAACGTATTGTTATGAATCAGAA
GCTTCTTCTAGTGACGACTCCCTGCTTCAAACTCTGAAACTATGTAAATTAGCTTGCTCTCCAGCTCCCCCATC
GGCTCCAAAAGCCCTCCGACCACTCGAAGCTTCAATTGAAGAGATTGTGAGATATGTGCATGATAAGCTCAAGA
GCATTCTCGCGACTGAAATCACTCTGCCTGCATTTGTGCTCTGCCACGAATTTGCTCTGAAGTATCCGGATAGA
TCTTATGAGTTTAATAAAATGTACAAAAAGCTCTACCAAATCATGGAAGATCAAGAAGAAGACGAGTTTCTCCA
ATCAGCTCGCCATTTCTCAAAATGGCCTCAAAATCTGACACTACCAATACAAAAACAGACAATAAATTGCATGG
CTGTCTTTTTCGAAGCGAATCTTGACAATCAGCTCGTCGATCTCTGTCAGTGGAGTGACCGACGAAAAGTGCTT
GTCGAGATGCTCGCCGAGCTGGCCGCCACAAGATCTGAAATTCGAGATAAACTTCAAAAATCGATGCCGTTCAA
CAAATTCGTCAAGGAGTGTATAATGGAGAATCGCGGTGATTTGTATGAAATGACAAAGAGATTTGAGAAATATT

CGTTTTTGCTCTCGATTCGGAATTTAATTGTTACTAGGATGATTATAACAAACGAAGCCGCCCGACTCCTAGGA
GATGGTGAAACAATCAGTGAAACCGATATCTTCATAATCGAAAAGCGTACTCTTTCCACGTGTATTCGTAATGT
GTCCGAAGGAAAAGAGTTGAGCGGCTACACACTGGACCCCTATACGGTAGCTGCCAACGTGCACAATGTGCATT
TCGATCACATAAACGTCGAAATCTATCTGGAATTGCTGAAAAAGTCGCCATTTTTCGCTCAGAACATTGTACGC
CACTTGTTACGGCAGAATGGAAAAGAAGCAGAAGAAGAGACGTGGCACTTGCATGCCACTGTGCTGAAAATTGT
GATGAAAGATGAAAAGTTGCTGGCGGTATGCGTGGCCACAATTCCAAATATGGTTCGATATCTCAAAGTCTATC
AGATTCATTTCAGTCCGAAATCCAACGCGGCGAAGTTCTTACACCTCGACATGGAATCGATTTCCCACTGCCAA
TCATATTTGCGAAAACCAACAAAATCATCCAATCTGATCACGGCCGCCAACTTTTTGACACTTTTCGGATGTGA
AAAGCGCACGTGGAAGCGCCCGATTCTCAGATTTTGGAGCATTTTCAAGCAGCAACCGGCTATGTGTTGCGAGA
AATTGCTCATTTTTGCTGAAGAATGTGTCGAACTTGGCCTGAACCACCGAATCGCTTGTCTTTTACGCGCACTG
ACAACCAGTGAATTCTGCCGAAAAGCTCTATGTGATGAATATCTGAAAATCGCGTTTCAGCTGACTTATCGATC
GATTTTTCTGATTTTAAGCAAAAATGAGTGCAGACCAGAAATTGTGGAGCTCTGCGATGACATGAATCTTCGGT
ACGATCTTCTCCAGCACCAGATCAAACATGTCGCGGCTCACCATTTGGAGCACTTTGAGCGTTTCGAAACGAAA
ATCGCATTTTCTGTTGAGAAATTTCTGAAATCTGGAATTGATGGAATCGATTTTGAAGATTTAGGATTAGTCGA
GTTTTATAAGCAGTTGAACGAAAATTTGACAGAAGACGCGATCAGGAGCAATGAGGCGAGAAACATCTACATCG
TCGACATTCTCTCAACAATTTGGCTCCAGCTCCCCTCAATTCGTCCTCAAATTCTACCGATTCTCGCCCGCTTC
AAGCACATTTCCCCAGCGTGGACTAATTTCCCACAGCCGCCTCATATTTCGACAAACGAGAAATCATTTCTTCA
ACATCTACGATTTCATCTTTATCTAAAAATGATGAATATCTCGAAATCCATGACGCAAGGCGAGTATGCCACGT
GTATTATGATGCTCCTGACGAGCTACGACTCGAGCCATTTCGTTGCGGATTTGATAGAGAAAAAGCAGCTGGGA
AAGCTGAAATTGCAGCAGAGAAGGAATGTTTTGTGCATTTTGAGCCGACTTTTGAAAGATCAAGCTGTGATGGG
TGATGAAGATGAGACGATTATCGATCCGATTCTTTTTAAGGCAATCACCAAAGCTTCCGCCGTTTTTGAAGACA
CCGCCGCGTGCATCGTACCATTTTTGTTCAAAATTTGTGTGGATTTCAAGGGAAAATACGATAAATGTGTGATA
AATTTGCTGGGATGTCTTAAGGGAGTTAATGCAGAAGACGAAATAGTAGTCCGCTGCCTTGCCGAATGCGTCGA
CTCCATTGGACTCAACGTGATTGCTCGTTACGAACGCCTGAACATTGAGACTCATTCGGAATTCGGTGTGAAAT
GGTTCTTCAAATTATCCCGTCTCTTTCTGAAACATGGATTTACTACGCATTCCTTCGCGATTGCGAATATATTA
TTTGACCGACTTTCGGCGAGAAAACGAAATACAATGATGATAGATCGTACGAGTTTAGACCGAATTGACAGATC
ACAGGAACTTATCAATCTTTTGGTGGAGATCTATGTGGCCGAGGGTAACTCGGTAGCGTTGTCCTCTCTGCCAC
CTGCGGTACAAAATAGACCGGATGTTCGGCAGGTTATGAATAAGAGTTCAAAGGAATGGTTGAAATTGCTGTCT
TCGAATCAAATGGACTCGTGGGAGTTGACTATTGTGCAGTGGATGTGTGGCATCCAATTCAATGCAATCACCGG
CGATAAATACCTGAATTCAATTCTTCGTTGCAATTTCAACGAGTACACCAAAAAAATCGATTCCCCATTAAAAT
TCGTCTATTTTCAACTATTCCATCTTTCGACGAGCACTCTGGAAATCGAAGAAGCCATTTCTAGCATGCCGTTG
GCTCCAACAATCGATCAAATGCGGCTTATGATTATAGCCAACGCGACGGCAAGCTTCGAGCCACAATCCGTTGA
AGAGCACGTTGTTCGAGCTGTTCGAGAGCTCCGAGAGACGTCAAATCGGCGAAAATCCGGTGGAAATGTGAAGG
GAATTAATGAAAAAACAACGAGAATGGTGAAACTGGCCGAGATGCTCACCGAGAATAAAGCATACGATGCGGCA
ATAAACCTGCTCGACACGTGGGAGCACGAGTGTCTCCAATGGACATCTGTCGCTGCCGAATCAATCGATATCGA
TTTAATTCGAATCTGCAAACAACACGTGACGTGTCGGTCGGGAGATCCAAGAATGGCGGACATAAATCTACGAA
CAATGCATCCACGTGTCCCGGTGATGAGTGACCTGGCGATTGCCGAATGGTCACTCGCGTTGAGCAAGATTACA

ATTGAATATCGGAATGATATGGAAGAGGGTATTCGGATTTTGGAATTTGGATGCAAACATTTACAGAATAAGGA
TTCTGTAGAGACGAGGTTAAAGGTACATGCGGTCGGGTCTTGCAGCGAAATAATGCATTTTTAA

A.3 *atm-1*

ATGAAACATTACAAAAAAATTATTCATCAGGTTCTCCTAAAACTCCATTCCGTCTGTATTGGCCAACTGTCGAA
ACTCGAAGAATATCGCGAAACGCGTACCTACCGTATGAAGCAGCAGGCGGTCACCGCATTCGAACAACAAATTC
AAAATTCGTGTCGAACCAGTCTGGCACGTGGCAATTCGGGTGACGAATGGACGAAAAAAACGGTGCAACGGGTG
AGAAAAGAGCATCAGTTTGAGAAGAATGATTTGGAAAAAGTGGATAATTCGTTGAATTCGGCGGCCCGGAAAGC
TGTATCGTCGGGTTTTGATGCACTTTTATGCATCAGCCAACTGGAAGACGACGACGAAGCGATCCGCGCTTCAT
CTCTCATAATATTTCCATTAATCGATGTGATCTACAAATACGAAACGGACGTCGGAGTGATCGCCTTGCTCAAA
GAGCACACCAAATCGAAGCTCCCGTCGAAGCTGTGGATAAGTGCCACCTCACACATTGCCTCTAAGTGCTTCTC
CATCGAAAAATCGCAAATCACGAGACACTTATCACAGATTCTGTGTCATCTCATCTATGACTACCCGTATCACG
TTTTGCACACAATTTTGATGTATGATGACGAGAAGAACGCTTCCAAGGTGAAAGGCTTCTTGAAAACGATATTT
GACGCGCGAGCTGACCAACGGGATTCGTCGAAGCTTAAGGAGATTGTGATTACTATTCGTGAGGCTCACCAGGC
TTATAGAGAAATTGCAATGCTTGACGTGAGAGGAAACGTTCGGATACAACGTGTTGAAATAAACGGAAAAACGA
TGTACCGATGGCCACATGATTTGAAGATTTTCAAGTGCAAATTGCGTCAATTGCCGATTCCCACCATTTCGCAA
AAGATTGGTTGCCCGGGTGACTACTCGACGACTGACCTAATCACGTGGAAACGATGGAAAGATGTATTCACAAT
TGCCGACGGAATTTCCACTCCAAAAATCTGGGAAATTGAGGGTAGCGATGGAAAATGGTATAAAACTGTGTGGA
AGAAGGACGACGTACGACAGGATGTGCTTGTTGAGCAAATGTTTGATGTTACGAATAATATGTTGGAGAAGGCG
ATGCTTCGGACGTATAATGTTGTCCCGTTAGACACTGAATGTGGAGTTATAGAGTTTTGTGGTGGAACTGTTAG
CTTAAAAGAAGTTATGTGTGGTGTGACACGAGAAGGCGGTCTCCACCGGGAGTTCAATTCAGAAGAAGTTTCGG
CGAGTAAAGTGTCGTCAATGATGAGACAAGTACAGACAGAGTCCACAGAGACACGACGACAAGTTTTTGTGGAG
ATTTGTCAGCAGTATTCTCCAGTTTTTAGACACTTCTTCTACACCAACTTTTCGACGGCGCAGATTTGGCGGCA
AAAAATCATAAATTACAGGCAGAGCTTGGCCACGTGGAGTATTGTGTGCTACATCGTTGGCCTCGGCGATCGAC
ACGCGTCGAATATTCTATTCGATCAGAAATTGTGCACATTTGTGCATATCGATTTGGGAATGATTTTGGAGTAT
AGTAAACGAACGTTGCCAGTTCCCGAACAAGTTCCATTCCGTATAACTCGAGATGTGCTCGACCCGATTCTGAT
TGAAGGAATCGAGAATGGACAGTTGGCTGAGGAATGTACGCAGATTATGGAGAAATTGAAGGAAAATGGAAAGG
TAATCCTCGGTGTCGCCTCTGCTCTACTCCGCGAAACGATGACAAATTTCCGAGAAGCCGAACAGGCTGCCGGC
CGCCCGTCCTACATTTCTGAAATGGCCATCGGACGGCTTCGAGAAAAGCTACGGGGCACCGATGACGGTGTGAC
GGCCCAATCGTCGAATCTTCAAATTCGGCGACTTTTACGAGAAGCTACGAGTGCTGACAATTTATCGCGAATGT
TCTGCGGATGGATGCCGTTTTTGTAG

Appendix B : Full contig of the extended gene model

B.1 *atm-1* full contig

ATGAGCGTTACTCAGCTAAAAAATCTGAAACACGCGATAGCTCAGCTTCTGGAGTGGGATGGAACGAAAACGGC
GAGAAAGAAAATCGTCGACGAGGTGGTTCTTCTCTATCACGCTCTGGGAGCGGAGGCATTAAGTGAGGATAACC
AGGAGATATATGATTTGTATGATTTATCTGCACGTATATTCAACTTAGCTAAGAAAGAAATCGAAGAAGCCAAC
CAACAATTCGAAAAAGAACGAAAAAAGGGCACAAGACGAAGCGAAAAACCAGTCCCAACTCCACTTTTCGAGCT
ATCAATACAACATTTAAAACGTTGCTGTCAGCAGGGAATCGATCACAATCAGGTGCCATGGATAGCATATTGTC
TGAAACTACTGGAATTTCCAATCACAATCACCGAAAAATCGATCGAAAACGAGATTTCCAACGTGCTCCTATTG
AGCTCCAACGCCTCACAGCTCCATTGGGCCGAGCATGCTCATTTGAGCAGCTTATGGAAATGGATTTGGAGCCG
TGTCGAGACCGCCGATATTGGGGCACTCGCCATGAGAAATTATATGGAATTGGCGGCGAACTTGCTGGAAAACG
TGGATTACGTGGTTTTCGAGAAGTCGCCTATTGACCTGATGGCGAAAGTGATGGGAACACTGAAGAAAAGTGTG
GAAATGGGAAATCCCAAAGAATATTGGTGTCATTACCGAAAATTCTCGGTCGTCCACATTCTCTCATATATCGT
ACATCGTTGGGGTCTCGAAGCGAGAGATTTCATTCTTCGTGAGATTTTCGAGCTCACTGGTAGTTTGTCCAACT
TGATATCAGTGGCTCATAAGGATGGAGAACAGTCCAAGAAATGCATAATGCGTTTGATTGATGATCTCGTGAAG
CTCGCCATGATCGAAACCGTACACGGCCACCGTACCATGAACGAAGTGACACGTGGAAATATTCAAAAACTCGT
GAAAACCGGAATCCAAGAGTCTCTGAAATCGGCGCACCGAAATTTCTCAAGGAGTTCGACATTTTCGATTTCCG
AAGAATGCGTGAGATATCTGACGAGATGGTTGTTGGCCGAACGAAGACTTGAACAGCCGTCTGCGGCTATGAAT
GAATCGTTTGAATTGACCGGTGATAGCAGTAGTAAGAAGAAGGACGATGCCACGTTTGATAGCCTTATCGATCT
CTCCTTTGGCTCGACGATTTCCGGAAAACATAAATTGAACGCGTGGAATGGTGTAATGCAAATCCTGAATGAGC
TCCTAAAAAGTCGACGACTCGAACTTCAAGTCACTGAAAAAATCGTGACAATCCTCTGGGAAAAGCGAAAATCC
TACACAACGGAGCCACTCCGTACTGTGTTCTGCTCCATTCTCTCCACAGTCGTCTGCCAGGCCGATGTTCGATT
CGGTCATCGGAAAGTGCCGACAATCGACTCGATTCTCAAATATTCGCTGTCTCTAATGCCAAATGTCGCATCTC
TTCCCAGTGCCGCTGCGTTGACCGAAACGATTGTCAGATTCAGGACAGTATCACGAGAGGGTCTCCGTAACACG
TGGGATACCGTATCCCGAACTAGCTCCGGCTCATTTGAAGTTGTTCGGCTGATTTCGGCGTTGATCTCCGTTAC
GGAATTTGATGAGAATTCGAGATTCGCCAACGATGAGAGAGTTCGGAGTTGGAGTTTTCGAAAAGACATAATCG
AATGGGTATTGCTGGATCCGAATGCACATAGTCACAAATTACTCTATCAGTTGTGCCAGTATCATCCAACGTAT
TGTTATGAATCAGAAGCTTCTTCTAGTGACGACTCCCTGCTTCAAACTCTGAAACTATGTAAATTAGCTTGCTC
TCCAGCTCCCCCATCGGCTCCAAAAGCCCTCCGACCACTCGAAGCTTCAATTGAAGAGATTGTGAGATATGTGC
ATGATAAGCTCAAGAGCATTCTCGCGACTGAAATCACTCTGCCTGCATTTGTGCTCTGCCACGAATTTGCTCTG
AAGTATCCGGATAGATCTTATGAGTTTAATAAAATGTACAAAAAGCTCTACCAAATCATGGAAGATCAAGAAGA
AGACGAGTTTCTCCAATCAGCTCGCCATTTCTCAAAATGGCCTCAAAATCTGACACTACCAATACAAAAACAGA
CAATAAATTGCATGGCTGTCTTTTTCGAAGCGAATCTTGACAATCAGCTCGTCGATCTCTGTCAGTGGAGTGAC
CGACGAAAAGTGCTTGTCGAGATGCTCGCCGAGCTGGCCGCCACAAGATCTGAAATTCGAGATAAACTTCAAAA
ATCGATGCCGTTCAACAAATTCGTCAAGGAGTGTATAATGGAGAATCGCGGTGATTTGTATGAAATGACAAAGA
GATTTGAGAAATATTCGTTTTTGCTCTCGATTCGGAATTTAATTGTTACTAGGATGATTATAACAAACGAAGCC
GCCCGACTCCTAGGAGATGGTGAAACAATCAGTGAAACCGATATCTTCATAATCGAAAAGCGTACTCTTTCCAC
GTGTATTCGTAATGTGTCCGAAGGAAAAGAGTTGAGCGGCTACACACTGGACCCCTATACGGTAGCTGCCAACG

TGCACAATGTGCATTTCGATCACATAAACGTCGAAATCTATCTGGAATTGCTGAAAAAGTCGCCATTTTTCGCT
CAGAACATTGTACGCCACTTGTTACGGCAGAATGGAAAAGAAGCAGAAGAAGAGACGTGGCACTTGCATGCCAC
TGTGCTGAAAATTGTGATGAAAGATGAAAAGTTGCTGGCGGTATGCGTGGCCACAATTCCAAATATGGTTCGAT
ATCTCAAAGTCTATCAGATTCATTTCAGTCCGAAATCCAACGCGGCGAAGTTCTTACACCTCGACATGGAATCG
ATTTCCCACTGCCAATCATATTTGCGAAAACCAACAAAATCATCCAATCTGATCACGGCCGCCAACTTTTTGAC
ACTTTTCGGATGTGAAAAGCGCACGTGGAAGCGCCCGATTCTCAGATTTTGGAGCATTTTCAAGCAGCAACCGG
CTATGTGTTGCGAGAAATTGCTCATTTTTGCTGAAGAATGTGTCGAACTTGGCCTGAACCACCGAATCGCTTGT
CTTTTACGCGCACTGACAACCAGTGAATTCTGCCGAAAAGCTCTATGTGATGAATATCTGAAAATCGCGTTTCA
GCTGACTTATCGATCGATTTTTCTGATTTTAAGCAAAAATGAGTGCAGACCAGAAATTGTGGAGCTCTGCGATG
ACATGAATCTTCGGTACGATCTTCTCCAGCACCAGATCAAACATGTCGCGGCTCACCATTTGGAGCACTTTGAG
CGTTTCGAAACGAAAATCGCATTTTCTGTTGAGAAATTTCTGAAATCTGGAATTGATGGAATCGATTTTGAAGA
TTTAGGATTAGTCGAGTTTTATAAGCAGTTGAACGAAAATTTGACAGAAGACGCGATCAGGAGCAATGAGGCGA
GAAACATCTACATCGTCGACATTCTCTCAACAATTTGGCTCCAGCTCCCCTCAATTCGTCCTCAAATTCTACCG
ATTCTCGCCCGCTTCAAGCACATTTCCCCAGCGTGGACTAATTTCCCACAGCCGCCTCATATTTCGACAAACGA
GAAATCATTTCTTCAACATCTACGATTTCATCTTTATCTAAAAATGATGAATATCTCGAAATCCATGACGCAAG
GCGAGTATGCCACGTGTATTATGATGCTCCTGACGAGCTACGACTCGAGCCATTTCGTTGCGGATTTGATAGAG
AAAAAGCAGCTGGGAAAGCTGAAATTGCAGCAGAGAAGGAATGTTTTGTGCATTTTGAGCCGACTTTTGAAAGA
TCAAGCTGTGATGGGTGATGAAGATGAGACGATTATCGATCCGATTCTTTTTAAGGCAATCACCAAAGCTTCCG
CCGTTTTTGAAGACACCGCCGCGTGCATCGTACCATTTTTGTTCAAAATTTGTGTGGATTTCAAGGGAAAATAC
GATAAATGTGTGATAAATTTGCTGGGATGTCTTAAGGGAGTTAATGCAGAAGACGAAATAGTAGTCCGCTGCCT
TGCCGAATGCGTCGACTCCATTGGACTCAACGTGATTGCTCGTTACGAACGCCTGAACATTGAGACTCATTCGG
AATTCGGTGTGAAATGGTTCTTCAAATTATCCCGTCTCTTTCTGAAACATGGATTTACTACGCATTCCTTCGCG
ATTGCGAATATATTATTTGACCGACTTTCGGCGAGAAAACGAAATACAATGATGATAGATCGTACGAGTTTAGA
CCGAATTGACAGATCACAGGAACTTATCAATCTTTTGGTGGAGATCTATGTGGCCGAGGGTAACTCGGTAGCGT
TGTCCTCTCTGCCACCTGCGGTACAAAATAGACCGGATGTTCGGCAGGTTATGAATAAGAGTTCAAAGGAATGG
TTGAAATTGCTGTCTTCGAATCAAATGGACTCGTGGGAGTTGACTATTGTGCAGTGGATGTGTGGCATCCAATT
CAATGCAATCACCGGCGATAAATACCTGAATTCAATTCTTCGTTGCAATTTCAACGAGTACACCAAAAAAATCG
ATTCCCCATTAAAATTCGTCTATTTTCAACTATTCCATCTTTCGACGAGCACTCTGGAAATCGAAGAAGCCATT
TCTAGCATGCCGTTGGCTCCAACAATCGATCAAATGCGGCTTATGATTATAGCCAACGCGACGGCAAGCTTCGA
GCCACAATCCGTTGAAGAGCACGTTGTTCGAGCTGTTCGAGAGCTCCGAGAGACGTCAAATCGGCGAAAATCCG
GTGGAAATGTGAAGGGAATTAATGAAAAAACAACGAGAATGGTGAAACTGGCCGAGATGCTCACCGAGAATAAA
GCATACGATGCGGCAATAAACCTGCTCGACACGTGGGAGCACGAGTGTCTCCAATGGACATCTGTCGCTGCCGA
ATCAATCGATATCGATTTAATTCGAATCTGCAAACAACACGTGACGTGTCGGTCGGGAGATCCAAGAATGGCGG
ACATAAATCTACGAACAATGCATCCACGTGTCCCGGTGATGAGTGACCTGGCGATTGCCGAATGGTCACTCGCG
TTGAGCAAGATTACAATTGAATATCGGAATGATATGGAAGAGGGTATTCGGATTTTGGAATTTGGATGCAAACA
TTTACAGAATAAGGATTCTGTAGAGACGAGGTTAAAGGTTCTCCTAAAACTCCATTCCGTCTGTATTGGCCAAC
TGTCGAAACTCGAAGAATATCGCGAAACGCGTACCTACCGTATGAAGCAGCAGGCGGTCACCGCATTCGAACAA
CAAATTCAAAATTCGTGTCGAACCAGTCTGGCACGTGGCAATTCGGGTGACGAATGGACGAAAAAAACGGTGCA

ACGGGTGAGAAAAGAGCATCAGTTTGAGAAGAATGATTTGGAAAAAGTGGATAATTCGTTGAATTCGGCGGCCC
GGAAAGCTGTATCGTCGGGTTTTGATGCACTTTTATGCATCAGCCAACTGGAAGACGACGACGAAGCGATCCGC
GCTTCATCTCTCATAATATTTCCATTAATCGATGTGATCTACAAATACGAAACGGACGTCGGAGTGATCGCCTT
GCTCAAAGAGCACACCAAATCGAAGCTCCCGTCGAAGCTGTGGATAAGTGCCACCTCACACATTGCCTCTAAGT
GCTTCTCCATCGAAAAATCGCAAATCACGAGACACTTATCACAGATTCTGTGTCATCTCATCTATGACTACCCG
TATCACGTTTTGCACACAATTTTGATGTATGATGACGAGAAGAACGCTTCCAAGGTGAAAGGCTTCTTGAAAAC
GATATTTGACGCGCGAGCTGACCAACGGGATTCGTCGAAGCTTAAGGAGATTGTGATTACTATTCGTGAGGCTC
ACCAGGCTTATAGAGAAATTGCAATGCTTGACGTGAGAGGAAACGTTCGGATACAACGTGTTGAAATAAACGGA
AAAACGATGTACCGATGGCCACATGATTTGAAGATTTTCAAGTGCAAATTGCGTCAATTGCCGATTCCCACCAT
TTCGCAAAAGATTGGTTGCCCGGGTGACTACTCGACGACTGACCTAATCACGTGGAAACGATGGAAAGATGTAT
TCACAATTGCCGACGGAATTTCCACTCCAAAAATCTGGGAAATTGAGGGTAGCGATGGAAAATGGTATAAAACT
GTGTGGAAGAAGGACGACGTACGACAGGATGTGCTTGTTGAGCAAATGTTTGATGTTACGAATAATATGTTGGA
GAAGGCGATGCTTCGGACGTATAATGTTGTCCCGTTAGACACTGAATGTGGAGTTATAGAGTTTTGTGGTGGAA
CTGTTAGCTTAAAAGAAGTTATGTGTGGTGTGACACGAGAAGGCGGTCTCCACCGGGAGTTCAATTCAGAAGAA
GTTTCGGCGAGTAAAGTGTCGTCAATGATGAGACAAGTACAGACAGAGTCCACAGAGACACGACGACAAGTTTT
TGTGGAGATTTGTCAGCAGTATTCTCCAGTTTTTAGACACTTCTTCTACACCAACTTTTCGACGGCGCAGATTT
GGCGGCAAAAAATCATAAATTACAGGCAGAGCTTGGCCACGTGGAGTATTGTGTGCTACATCGTTGGCCTCGGC
GATCGACACGCGTCGAATATTCTATTCGATCAGAAATTGTGCACATTTGTGCATATCGATTTGGGAATGATTTT
GGAGTATAGTAAACGAACGTTGCCAGTTCCCGAACAAGTTCCATTCCGTATAACTCGAGATGTGCTCGACCCGA
TTCTGATTGAAGGAATCGAGAATGGACAGTTGGCTGAGGAATGTACGCAGATTATGGAGAAATTGAAGGAAAAT
GGAAAGGTAATCCTCGGTGTCGCCTCTGCTCTACTCCGCGAAACGATGACAAATTTCCGAGAAGCCGAACAGGC
TGCCGGCCGCCCGTCCTACATTTCTGAAATGGCCATCGGACGGCTTCGAGAAAAGCTACGGGGCACCGATGACG
GTGTGACGGCCCAATCGTCGAATCTTCAAATTCGGCGACTTTTACGAGAAGCTACGAGTGCTGACAATTTATCG
CGAATGTTCTGCGGATGGATGCCGTTTTTGTAG

Appendix C : *atm-1(gk186)* transcripts

C.1 *atm-1.a*

ATGAGCGTTACTCAGCTAAAAAATCTGAAACACGCGATAGCTCAGCTTCTGGAGTGGGATGGAACGAAAACGGC
GAGAAAGAAAATCGTCGACGAGGTGGTTCTTCTCTATCACGCTCTGGGAGCGGAGGCATTAAGTGAGGATAACC
AGGAGATATATGATTTGTATGATTTATCTGCACGTATATTCAACTTAGCTAAGAAAGAAATCGAAGAAGCCAAC
CAACAATTCGAAAAAGAACGAAAAAAGGGCACAAGACGAAGCGAAAAACCAGTCCCAACTCCACTTTTCGAGCT
ATCAATACAACATTTAAAACGTTGCTGTCAGCAGGGAATCGATCACAATCAGGTGCCATGGATAGCATATTGTC
TGAAACTACTGGAATTTCCAATCACAATCACCGAAAAATCGATCGAAAACGAGATTTCCAACGTGCTCCTATTG
AGCTCCAACGCCTCACAGCTCCATTGGGCCGAGCATGCTCATTTGAGCAGCTTATGGAAATGGATTTGGAGCCG
TGTCGAGACCGCCGATATTGGGGCACTCGCCATGAGAAATTATATGGAATTGGCGGCGAACTTGCTGGAAAACG
TGGATTACGTGGTTTTCGAGAAGTCGCCTATTGACCTGATGGCGAAAGTGATGGGAACACTGAAGAAAAGTGTG
GAAATGGGAAATCCCAAAGAATATTGGTGTCATTACCGAAAATTCTCGGTCGTCCACATTCTCTCATATATCGT
ACATCGTTGGGGTCTCGAAGCGAGAGATTTCATTCTTCGTGAGATTTTCGAGCTCACTGGTAGTTTGTCCAACT
TGATATCAGTGGCTCATAAGGATGGAGAACAGTCCAAGAAATGCATAATGCGTTTGATTGATGATCTCGTGAAG
CTCGCCATGATCGAAACCGTACACGGCCACCGTACCATGAACGAAGTGACACGTGGAAATATTCAAAAACTCGT
GAAAACCGGAATCCAAGAGTCTCTGAAATCGGCGCACCGAAATTTCTCAAGGAGTTCGACATTTTCGATTTCCG
AAGAATGCGTGAGATATCTGACGAGATGGTTGTTGGCCGAACGAAGACTTGAACAGCCGTCTGCGGCTATGAAT
GAATCGTTTGAATTGACCGGTGATAGCAGTAGTAAGAAGAAGGACGATGCCACGTTTGATAGCCTTATCGATCT
CTCCTTTGGCTCGACGATTTCCGGAAAACATAAATTGAACGCGTGGAATGGTGTAATGCAAATCCTGAATGAGC
TCCTAAAAAGTCGACGACTCGAACTTCAAGTCACTGAAAAAATCGTGACAATCCTCTGGGAAAAGCGAAAATCC
TACACAACGGAGCCACTCCGTACTGTGTTCTGCTCCATTCTCTCCACAGTCGTCTGCCAGGCCGATGTTCGATT
CGGTCATCGGAAAGTGCCGACAATCGACTCGATTCTCAAATATTCGCTGTCTCTAATGCCAAATGTCGCATCTC
TTCCCAGTGCCGCTGCGTTGACCGAAACGATTGTCAGATTCAGGACAGTATCACGAGAGGGTCTCCGTAACACG
TGGGATACCGTATCCCGAACTAGCTCCGGCTCATTTGAAGTTGTTCGGCTGATTTCGGCGTTGATCTCCGTTAC
GGAATTTGATGAGAATTCGAGATTCGCCAACGATGAGAGAGTTCGGAGTTGGAGTTTTCGAAAAGACATAATCG
AATGGGTATTGCTGGATCCGAATGCACATAGTCACAAATTACTCTATCAGTTGTGCCAGTATCATCCAACGTAT
TGTTATGAATCAGAAGCTTCTTCTAGTGACGACTCCCTGCTTCAAACTCTGAAACTATGTAAATTAGCTTGCTC
TCCAGCTCCCCCATCGGCTCCAAAAGCCCTCCGACCACTCGAAGCTTCAATTGAAGAGATTGTGAGATATGTGC
ATGATAAGCTCAAGAGCATTCTCGCGACTGAAATCACTCTGCCTGCATTTGTGCTCTGCCACGAATTTGCTCTG
AAGTATCCGGATAGATCTTATGAGTTTAATAAAATGTACAAAAAGCTCTACCAAATCATGGAAGATCAAGAAGA
AGACGAGTTTCTCCAATCAGCTCGCCATTTCTCAAAATGGCCTCAAAATCTGACACTACCAATACAAAAACAGA
CAATAAATTGCATGGCTGTCTTTTTCGAAGCGAATCTTGACAATCAGCTCGTCGATCTCTGTCAGTGGAGTGAC
CGACGAAAAGTGCTTGTCGAGATGCTCGCCGAGCTGGCCGCCACAAGATCTGAAATTCGAGATAAACTTCAAAA
ATCGATGCCGTTCAACAAATTCGTCAAGGAGTGTATAATGGAGAATCGCGGTGATTTGTATGAAATGACAAAGA
GATTTGAGAAATATTCGTTTTTGCTCTCGATTCGGAATTTAATTGTTACTAGGATGATTATAACAAACGAAGCC
GCCCGACTCCTAGGAGATGGTGAAACAATCAGTGAAACCGATATCTTCATAATCGAAAAGCGTACTCTTTCCAC
GTGTATTCGTAATGTGTCCGAAGGAAAAGAGTTGAGCGGCTACACACTGGACCCCTATACGGTAGCTGCCAACG

TGCACAATGTGCATTTCGATCACATAAACGTCGAAATCTATCTGGAATTGCTGAAAAAGTCGCCATTTTTCGCT
CAGAACATTGTACGCCACTTGTTACGGCAGAATGGAAAAGAAGCAGAAGAAGAGACGTGGCACTTGCATGCCAC
TGTGCTGAAAATTGTGATGAAAGATGAAAAGTTGCTGGCGGTATGCGTGGCCACAATTCCAAATATGGTTCGAT
ATCTCAAAGTCTATCAGATTCATTTCAGTCCGAAATCCAACGCGGCGAAGTTCTTACACCTCGACATGGAATCG
ATTTCCCACTGCCAATCATATTTGCGAAAACCAACAAAATCATCCAATCTGATCACGGCCGCCAACTTTTTGAC
ACTTTTCGGATGTGAAAAGCGCACGTGGAAGCGCCCGATTCTCAGATTTTGGAGCATTTTCAAGCAGCAACCGG
CTATGTGTTGCGAGAAATTGCTCATTTTTGCTGAAGAATGTGTCGAACTTGGCCTGAACCACCGAATCGCTTGT
CTTTTACGCGCACTGACAACCAGTGAATTCTGCCGAAAAGCTCTATGTGATGAATATCTGAAAATCGCGTTTCA
GCTGACTTATCGATCGATTTTTCTGATTTTAAGCAAAAATGAGTGCAGACCAGAAATTGTGGAGCTCTGCGATG
ACATGAATCTTCGGTACGATCTTCTCCAGCACCAGATCAAACATGTCGCGGCTCACCATTTGGAGCACTTTGAG
CGTTTCGAAACGAAAATCGCATTTTCTGTTGAGAAATTTCTGAAATCTGGAATTGATGGAATCGATTTTGAAGA
TTTAGGATTAGTCGAGTTTTATAAGCAGTTGAACGAAAATTTGACAGAAGACGCGATCAGGAGCAATGAGGCGA
GAAACATCTACATCGTCGACATTCTCTCAACAATTTGGCTCCAGCTCCCCTCAATTCGTCCTCAAATTCTACCG
ATTCTCGCCCGCTTCAAGCACATTTCCCCAGCGTGGACTAATTTCCCACAGCCGCCTCATATTTCGACAAACGA
GAAATCATTTCTTCAACATCTACGATTTCATCTTTATCTAAAAATGATGAATATCTCGAAATCCATGACGCAAG
GCGAGTATGCCACGTGTATTATGATGCTCCTGACGAGCTACGACTCGAGCCATTTCGTTGCGGATTTGATAGAG
AAAAAGCAGCTGGGAAAGCTGAAATTGCAGCAGAGAAGGAATGTTTTGTGCATTTTGAGCCGACTTTTGAAAGA
TCAAGCTGTGATGGGTGATGAAGATGAGACGATTATCGATCCGATTCTTTTTAAGGCAATCACCAAAGCTTCCG
CCGTTTTTGAAGACACCGCCGCGTGCATCGTACCATTTTTGTTCAAAATTTGTGTGGATTTCAAGGGAAAATAC
GATAAATGTGTGATAAATTTGCTGGGATGTCTTAAGGGAGTTAATGCAGAAGACGAAATAGTAGTCCGCTGCCT
TGCCGAATGCGTCGACTCCATTGGACTCAACGTGATTGCTCGTTACGAACGCCTGAACATTGAGACTCATTCGG
AATTCGGTGTGAAATGGTTCTTCAAATTATCCCGTCTCTTTCTGAAACATGGATTTACTACGCATTCCTTCGCG
ATTGCGAATATATTATTTGACCGACTTTCGGCGAGAAAACGAAATACAATGATGATAGATCGTACGAGTTTAGA
CCGAATTGACAGATCACAGGAACTTATCAATCTTTTGGTGGAGATCTATGTGGCCGAGGGTAACTCGGTAGCGT
TGTCCTCTCTGCCACCTGCGGTACAAAATAGACCGGATGTTCGGCAGGTTATGAATAAGAGTTCAAAGGAATGG
TTGAAATTGCTGTCTTCGAATCAAATGGACTCGTGGGAGTTGACTATTGTGCAGTGGATGTGTGGCATCCAATT
CAATGCAATCACCGGCGATAAATACCTGAATTCAATTCTTCGTTGCAATTTCAACGAGTACACCAAAAAAATCG
ATTCCCCATTAAAATTCGTCTATTTTCAACTATTCCATCTTTCGACGAGCACTCTGGAAATCGAAGAAGCCATT
TCTAGCATGCCGTTGGCTCCAACAATCGATCAAATGCGGCTTATGATTATAGCCAACGCGACGGCAAGCTTCGA
GCCACAATCCGTTGAAGAGCACGTTGTTCGAGCTGTTCGAGAGCTCCGAGAGACGTCAAATCGGCGAAAATCCG
GTGGAAATGTGAAGGGAATTAATGAAAAAACAACGAGAATGGTGAAACTGGCCGAGATGCTCACCGAGAATAAA
GCATACGATGCGGCAATAAACCTGCTCGACACGTGGGAGCACGAGTGTCTCCAATGGACATCTGTCGCTGCCGA
ATCAATCGATATCGATTTAATTCGAATCTGCAAACAACACGTGACGTGTCGGTCGGGAGATCCAAGAATGGCGG
ACATAAATCTACGAACAATGCATCCACGTGTCCCGGTGATGAGTGACCTGGCGATTGCCGAATGGTCACTCGCG
TTGAGCAAGATTACAATTGAATATCGGAATGATATGGAAGAGGGTATTCGGATTTTGGAATTTGGATGCAAACA
TTTACAGAATAAGGATTCTGTAGAGACGAGGTTAAAGGTTCTCCTAAAACTCCATTCCGTCTGTATTGGCCAAC
TGTCGAAACTCGAAGAATATCGCGAAACGCGTACCTACCGTATGAAGCAGCAGGCGGTCACCGCATTCGAACAA
CAAATTCAAAATTCGTGTCGAACCAGTCTGGCACGTGGCAATTCGGGTGACGAATGGACGAAAAAAACGGTGCA

ACGGGTGAGAAAAGAGCATCAGTTTGAGAAGAATGATTTGGAAAAAGTGGATAATTCGTTGAATTCGGCGGCCC
GGAAAGCTGTATCGTCGGGTTTTGATGCACTTTATTCTGTGTCATCTCATCTATGACTACCCGTATCACGTTTT
GCACACAATTTTGATGTATGATGACGAGAAGAACGCTTCCAAGGTGAAAGGCTTCTTGAAAACGATATTTGACG
CGCGAGCTGACCAACGGGATTCGTCGAAGCTTAAGGAGATTGTGATTACTATTCGTGAGGCTCACCAGGCTTAT
AGAGAAATTGCAATGCTTGACGTGAGAGGAAACGTTCGGATACAACGTGTTGAAATAAACGGAAAAACGATGTA
CCGATGGCCACATGATTTGAAGATTTTCAAGTGCAAATTGCGTCAATTGCCGATTCCCACCATTTCGCAAAAGA
TTGGTTGCCCGGGTGACTACTCGACGACTGACCTAATCACGTGGAAACGATGGAAAGATGTATTCACAATTGCC
GACGGAATTTCCACTCCAAAAATCTGGGAAATTGAGGGTAGCGATGGAAAATGGTATAAAACTGTGTGGAAGAA
GGACGACGTACGACAGGATGTGCTTGTTGAGCAAATGTTTGATGTTACGAATAATATGTTGGAGAAGGCGATGC
TTCGGACGTATAATGTTGTCCCGTTAGACACTGAATGTGGAGTTATAGAGTTTTGTGGTGGAACTGTTAGCTTA
AAAGAAGTTATGTGTGGTGTGACACGAGAAGGCGGTCTCCACCGGGAGTTCAATTCAGAAGAAGTTTCGGCGAG
TAAAGTGTCGTCAATGATGAGACAAGTACAGACAGAGTCCACAGAGACACGACGACAAGTTTTTGTGGAGATTT
GTCAGCAGTATTCTCCAGTTTTTAGACACTTCTTCTACACCAACTTTTCGACGGCGCAGATTTGGCGGCAAAAA
ATCATAAATTACAGGCAGAGCTTGGCCACGTGGAGTATTGTGTGCTACATCGTTGGCCTCGGCGATCGACACGC
GTCGAATATTCTATTCGATCAGAAATTGTGCACATTTGTGCATATCGATTTGGGAATGATTTTGGAGTATAGTA
AACGAACGTTGCCAGTTCCCGAACAAGTTCCATTCCGTATAACTCGAGATGTGCTCGACCCGATTCTGATTGAA
GGAATCGAGAATGGACAGTTGGCTGAGGAATGTACGCAGATTATGGAGAAATTGAAGGAAAATGGAAAGGTAAT
CCTCGGTGTCGCCTCTGCTCTACTCCGCGAAACGATGACAAATTTCCGAGAAGCCGAACAGGCTGCCGGCCGCC
CGTCCTACATTTCTGAAATGGCCATCGGACGGCTTCGAGAAAAGCTACGGGGCACCGATGACGGTGTGACGGCC
CAATCGTCGAATCTTCAAATTCGGCGACTTTTACGAGAAGCTACGAGTGCTGACAATTTATCGCGAATGTTCTG
CGGATGGATGCCGTTTTTGTAG

C.2 *atm-1.b*

ATGAGCGTTACTCAGCTAAAAAATCTGAAACACGCGATAGCTCAGCTTCTGGAGTGGGATGGAACGAAAACGGC
GAGAAAGAAAATCGTCGACGAGGTGGTTCTTCTCTATCACGCTCTGGGAGCGGAGGCATTAAGTGAGGATAACC
AGGAGATATATGATTTGTATGATTTATCTGCACGTATATTCAACTTAGCTAAGAAAGAAATCGAAGAAGCCAAC
CAACAATTCGAAAAAGAACGAAAAAAGGGCACAAGACGAAGCGAAAAACCAGTCCCAACTCCACTTTTCGAGCT
ATCAATACAACATTTAAAACGTTGCTGTCAGCAGGGAATCGATCACAATCAGGTGCCATGGATAGCATATTGTC
TGAAACTACTGGAATTTCCAATCACAATCACCGAAAAATCGATCGAAAACGAGATTTCCAACGTGCTCCTATTG
AGCTCCAACGCCTCACAGCTCCATTGGGCCGAGCATGCTCATTTGAGCAGCTTATGGAAATGGATTTGGAGCCG
TGTCGAGACCGCCGATATTGGGGCACTCGCCATGAGAAATTATATGGAATTGGCGGCGAACTTGCTGGAAAACG
TGGATTACGTGGTTTTCGAGAAGTCGCCTATTGACCTGATGGCGAAAGTGATGGGAACACTGAAGAAAAGTGTG
GAAATGGGAAATCCCAAAGAATATTGGTGTCATTACCGAAAATTCTCGGTCGTCCACATTCTCTCATATATCGT
ACATCGTTGGGGTCTCGAAGCGAGAGATTTCATTCTTCGTGAGATTTTCGAGCTCACTGGTAGTTTGTCCAACT
TGATATCAGTGGCTCATAAGGATGGAGAACAGTCCAAGAAATGCATAATGCGTTTGATTGATGATCTCGTGAAG
CTCGCCATGATCGAAACCGTACACGGCCACCGTACCATGAACGAAGTGACACGTGGAAATATTCAAAAACTCGT
GAAAACCGGAATCCAAGAGTCTCTGAAATCGGCGCACCGAAATTTCTCAAGGAGTTCGACATTTTCGATTTCCG
AAGAATGCGTGAGATATCTGACGAGATGGTTGTTGGCCGAACGAAGACTTGAACAGCCGTCTGCGGCTATGAAT

GAATCGTTTGAATTGACCGGTGATAGCAGTAGTAAGAAGAAGGACGATGCCACGTTTGATAGCCTTATCGATCT
CTCCTTTGGCTCGACGATTTCCGGAAAACATAAATTGAACGCGTGGAATGGTGTAATGCAAATCCTGAATGAGC
TCCTAAAAAGTCGACGACTCGAACTTCAAGTCACTGAAAAAATCGTGACAATCCTCTGGGAAAAGCGAAAATCC
TACACAACGGAGCCACTCCGTACTGTGTTCTGCTCCATTCTCTCCACAGTCGTCTGCCAGGCCGATGTTCGATT
CGGTCATCGGAAAGTGCCGACAATCGACTCGATTCTCAAATATTCGCTGTCTCTAATGCCAAATGTCGCATCTC
TTCCCAGTGCCGCTGCGTTGACCGAAACGATTGTCAGATTCAGGACAGTATCACGAGAGGGTCTCCGTAACACG
TGGGATACCGTATCCCGAACTAGCTCCGGCTCATTTGAAGTTGTTCGGCTGATTTCGGCGTTGATCTCCGTTAC
GGAATTTGATGAGAATTCGAGATTCGCCAACGATGAGAGAGTTCGGAGTTGGAGTTTTCGAAAAGACATAATCG
AATGGGTATTGCTGGATCCGAATGCACATAGTCACAAATTACTCTATCAGTTGTGCCAGTATCATCCAACGTAT
TGTTATGAATCAGAAGCTTCTTCTAGTGACGACTCCCTGCTTCAAACTCTGAAACTATGTAAATTAGCTTGCTC
TCCAGCTCCCCCATCGGCTCCAAAAGCCCTCCGACCACTCGAAGCTTCAATTGAAGAGATTGTGAGATATGTGC
ATGATAAGCTCAAGAGCATTCTCGCGACTGAAATCACTCTGCCTGCATTTGTGCTCTGCCACGAATTTGCTCTG
AAGTATCCGGATAGATCTTATGAGTTTAATAAAATGTACAAAAAGCTCTACCAAATCATGGAAGATCAAGAAGA
AGACGAGTTTCTCCAATCAGCTCGCCATTTCTCAAAATGGCCTCAAAATCTGACACTACCAATACAAAAACAGA
CAATAAATTGCATGGCTGTCTTTTTCGAAGCGAATCTTGACAATCAGCTCGTCGATCTCTGTCAGTGGAGTGAC
CGACGAAAAGTGCTTGTCGAGATGCTCGCCGAGCTGGCCGCCACAAGATCTGAAATTCGAGATAAACTTCAAAA
ATCGATGCCGTTCAACAAATTCGTCAAGGAGTGTATAATGGAGAATCGCGGTGATTTGTATGAAATGACAAAGA
GATTTGAGAAATATTCGTTTTTGCTCTCGATTCGGAATTTAATTGTTACTAGGATGATTATAACAAACGAAGCC
GCCCGACTCCTAGGAGATGGTGAAACAATCAGTGAAACCGATATCTTCATAATCGAAAAGCGTACTCTTTCCAC
GTGTATTCGTAATGTGTCCGAAGGAAAAGAGTTGAGCGGCTACACACTGGACCCCTATACGGTAGCTGCCAACG
TGCACAATGTGCATTTCGATCACATAAACGTCGAAATCTATCTGGAATTGCTGAAAAAGTCGCCATTTTTCGCT
CAGAACATTGTACGCCACTTGTTACGGCAGAATGGAAAAGAAGCAGAAGAAGAGACGTGGCACTTGCATGCCAC
TGTGCTGAAAATTGTGATGAAAGATGAAAAGTTGCTGGCGGTATGCGTGGCCACAATTCCAAATATGGTTCGAT
ATCTCAAAGTCTATCAGATTCATTTCAGTCCGAAATCCAACGCGGCGAAGTTCTTACACCTCGACATGGAATCG
ATTTCCCACTGCCAATCATATTTGCGAAAACCAACAAAATCATCCAATCTGATCACGGCCGCCAACTTTTTGAC
ACTTTTCGGATGTGAAAAGCGCACGTGGAAGCGCCCGATTCTCAGATTTTGGAGCATTTTCAAGCAGCAACCGG
CTATGTGTTGCGAGAAATTGCTCATTTTTGCTGAAGAATGTGTCGAACTTGGCCTGAACCACCGAATCGCTTGT
CTTTTACGCGCACTGACAACCAGTGAATTCTGCCGAAAAGCTCTATGTGATGAATATCTGAAAATCGCGTTTCA
GCTGACTTATCGATCGATTTTTCTGATTTTAAGCAAAAATGAGTGCAGACCAGAAATTGTGGAGCTCTGCGATG
ACATGAATCTTCGGTACGATCTTCTCCAGCACCAGATCAAACATGTCGCGGCTCACCATTTGGAGCACTTTGAG
CGTTTCGAAACGAAAATCGCATTTTCTGTTGAGAAATTTCTGAAATCTGGAATTGATGGAATCGATTTTGAAGA
TTTAGGATTAGTCGAGTTTTATAAGCAGTTGAACGAAAATTTGACAGAAGACGCGATCAGGAGCAATGAGGCGA
GAAACATCTACATCGTCGACATTCTCTCAACAATTTGGCTCCAGCTCCCCTCAATTCGTCCTCAAATTCTACCG
ATTCTCGCCCGCTTCAAGCACATTTCCCCAGCGTGGACTAATTTCCCACAGCCGCCTCATATTTCGACAAACGA
GAAATCATTTCTTCAACATCTACGATTTCATCTTTATCTAAAAATGATGAATATCTCGAAATCCATGACGCAAG
GCGAGTATGCCACGTGTATTATGATGCTCCTGACGAGCTACGACTCGAGCCATTTCGTTGCGGATTTGATAGAG
AAAAAGCAGCTGGGAAAGCTGAAATTGCAGCAGAGAAGGAATGTTTTGTGCATTTTGAGCCGACTTTTGAAAGA
TCAAGCTGTGATGGGTGATGAAGATGAGACGATTATCGATCCGATTCTTTTTAAGGCAATCACCAAAGCTTCCG

CCGTTTTTGAAGACACCGCCGCGTGCATCGTACCATTTTTGTTCAAAATTTGTGTGGATTTCAAGGGAAAATAC
GATAAATGTGTGATAAATTTGCTGGGATGTCTTAAGGGAGTTAATGCAGAAGACGAAATAGTAGTCCGCTGCCT
TGCCGAATGCGTCGACTCCATTGGACTCAACGTGATTGCTCGTTACGAACGCCTGAACATTGAGACTCATTCGG
AATTCGGTGTGAAATGGTTCTTCAAATTATCCCGTCTCTTTCTGAAACATGGATTTACTACGCATTCCTTCGCG
ATTGCGAATATATTATTTGACCGACTTTCGGCGAGAAAACGAAATACAATGATGATAGATCGTACGAGTTTAGA
CCGAATTGACAGATCACAGGAACTTATCAATCTTTTGGTGGAGATCTATGTGGCCGAGGGTAACTCGGTAGCGT
TGTCCTCTCTGCCACCTGCGGTACAAAATAGACCGGATGTTCGGCAGGTTATGAATAAGAGTTCAAAGGAATGG
TTGAAATTGCTGTCTTCGAATCAAATGGACTCGTGGGAGTTGACTATTGTGCAGTGGATGTGTGGCATCCAATT
CAATGCAATCACCGGCGATAAATACCTGAATTCAATTCTTCGTTGCAATTTCAACGAGTACACCAAAAAAATCG
ATTCCCCATTAAAATTCGTCTATTTTCAACTATTCCATCTTTCGACGAGCACTCTGGAAATCGAAGAAGCCATT
TCTAGCATGCCGTTGGCTCCAACAATCGATCAAATGCGGCTTATGATTATAGCCAACGCGACGGCAAGCTTCGA
GCCACAATCCGTTGAAGAGCACGTTGTTCGAGCTGTTCGAGAGCTCCGAGAGACGTCAAATCGGCGAAAATCCG
GTGGAAATGTGAAGGGAATTAATGAAAAAACAACGAGAATGGTGAAACTGGCCGAGATGCTCACCGAGAATAAA
GCATACGATGCGGCAATAAACCTGCTCGACACGTGGGAGCACGAGTGTCTCCAATGGACATCTGTCGCTGCCGA
ATCAATCGATATCGATTTAATTCGAATCTGCAAACAACACGTGACGTGTCGGTCGGGAGATCCAAGAATGGCGG
ACATAAATCTACGAACAATGCATCCACGTGTCCCGGTGATGAGTGACCTGGCGATTGCCGAATGGTCACTCGCG
TTGAGCAAGATTACAATTGAATATCGGAATGATATGGAAGAGGGTATTCGGATTTTGGAATTTGGATGCAAACA
TTTACAGAATAAGGATTCTGTAGAGACGAGGTTAAAGGTTCTCCTAAAACTCCATTCCGTCTGTATTGGCCAAC
TGTCGAAACTCGAAGAATATCGCGAAACGCGTACCTACCGTATGAAGCAGCAGGCGGTCACCGCATTCGAACAA
CAAATTCAAAATTCGTGTCGAACCAGTCTGGCACGTGGCAATTCGGGTGACGAATGGACGAAAAAAACGGTGCA
ACGGATTCTGTGTCATCTCATCTATGACTACCCGTATCACGTTTTGCACACAATTTTGATGTATGATGACGAGA
AGAACGCTTCCAAGGTGAAAGGCTTCTTGAAAACGATATTTGACGCGCGAGCTGACCAACGGGATTCGTCGAAG
CTTAAGGAGATTGTGATTACTATTCGTGAGGCTCACCAGGCTTATAGAGAAATTGCAATGCTTGACGTGAGAGG
AAACGTTCGGATACAACGTGTTGAAATAAACGGAAAAACGATGTACCGATGGCCACATGATTTGAAGATTTTCA
AGTGCAAATTGCGTCAATTGCCGATTCCCACCATTTCGCAAAAGATTGGTTGCCCGGGTGACTACTCGACGACT
GACCTAATCACGTGGAAACGATGGAAAGATGTATTCACAATTGCCGACGGAATTTCCACTCCAAAAATCTGGGA
AATTGAGGGTAGCGATGGAAAATGGTATAAAACTGTGTGGAAGAAGGACGACGTACGACAGGATGTGCTTGTTG
AGCAAATGTTTGATGTTACGAATAATATGTTGGAGAAGGCGATGCTTCGGACGTATAATGTTGTCCCGTTAGAC
ACTGAATGTGGAGTTATAGAGTTTTGTGGTGGAACTGTTAGCTTAAAAGAAGTTATGTGTGGTGTGACACGAGA
AGGCGGTCTCCACCGGGAGTTCAATTCAGAAGAAGTTTCGGCGAGTAAAGTGTCGTCAATGATGAGACAAGTAC
AGACAGAGTCCACAGAGACACGACGACAAGTTTTTGTGGAGATTTGTCAGCAGTATTCTCCAGTTTTTAGACAC
TTCTTCTACACCAACTTTTCGACGGCGCAGATTTGGCGGCAAAAAATCATAAATTACAGGCAGAGCTTGGCCAC
GTGGAGTATTGTGTGCTACATCGTTGGCCTCGGCGATCGACACGCGTCGAATATTCTATTCGATCAGAAATTGT
GCACATTTGTGCATATCGATTTGGGAATGATTTTGGAGTATAGTAAACGAACGTTGCCAGTTCCCGAACAAGTT
CCATTCCGTATAACTCGAGATGTGCTCGACCCGATTCTGATTGAAGGAATCGAGAATGGACAGTTGGCTGAGGA
ATGTACGCAGATTATGGAGAAATTGAAGGAAAATGGAAAGGTAATCCTCGGTGTCGCCTCTGCTCTACTCCGCG
AAACGATGACAAATTTCCGAGAAGCCGAACAGGCTGCCGGCCGCCCGTCCTACATTTCTGAAATGGCCATCGGA

CGGCTTCGAGAAAAGCTACGGGGCACCGATGACGGTGTGACGGCCCAATCGTCGAATCTTCAAATTCGGCGACT
TTTACGAGAAGCTACGAGTGCTGACAATTTATCGCGAATGTTCTGCGGATGGATGCCGTTTTTGTAG

Appendix D : *atm-1* protein translations

D.1 *atm-1*

MSVTQLKNLKHAIAQLLEWDGTKTARKKIVDEVVLLYHALGAEALSEDNQEIYDLYDLSARIFNLAKKEIEEAN
QQFEKERKKGTRRSEKPVPTPLFELSIQHLKRCCQQGIDHNQVPWIAYCLKLLEFPITITEKSIENEISNVLLL
SSNASQLHWAEHAHLSSLWKWIWSRVETADIGALAMRNYMELAANLLENVDYVVFEKSPIDLMAKVMGTLKKSV
EMGNPKEYWCHYRKFSVVHILSYIVHRWGLEARDFILREIFELTGSLSNLISVAHKDGEQSKKCIMRLIDDLVK
LAMIETVHGHRTMNEVTRGNIQKLVKTGIQESLKSAHRNFSRSSTFSISEECVRYLTRWLLAERRLEQPSAAMN
ESFELTGDSSSKKKDDATFDSLIDLSFGSTISGKHKLNAWNGVMQILNELLKSRRLELQVTEKIVTILWEKRKS
YTTEPLRTVFCSILSTVVCQADVRFGHRKVPTIDSILKYSLSLMPNVASLPSAAALTETIVRFRTVSREGLRNT
WDTVSRTSSGSFEVVRLISALISVTEFDENSRFANDERVRSWSFRKDIIEWVLLDPNAHSHKLLYQLCQYHPTY
CYESEASSSDDSLLQTLKLCKLACSPAPPSAPKALRPLEASIEEIVRYVHDKLKSILATEITLPAFVLCHEFAL
KYPDRSYEFNKMYKKLYQIMEDQEEDEFLQSARHFSKWPQNLTLPIQKQTINCMAVFFEANLDNQLVDLCQWSD
RRKVLVEMLAELAATRSEIRDKLQKSMPFNKFVKECIMENRGDLYEMTKRFEKYSFLLSIRNLIVTRMIITNEA
ARLLGDGETISETDIFIIEKRTLSTCIRNVSEGKELSGYTLDPYTVAANVHNVHFDHINVEIYLELLKKSPFFA
QNIVRHLLRQNGKEAEEETWHLHATVLKIVMKDEKLLAVCVATIPNMVRYLKVYQIHFSPKSNAAKFLHLDMES
ISHCQSYLRKPTKSSNLITAANFLTLFGCEKRTWKRPILRFWSIFKQQPAMCCEKLLIFAEECVELGLNHRIAC
LLRALTTSEFCRKALCDEYLKIAFQLTYRSIFLILSKNECRPEIVELCDDMNLRYDLLQHQIKHVAAHHLEHFE
RFETKIAFSVEKFLKSGIDGIDFEDLGLVEFYKQLNENLTEDAIRSNEARNIYIVDILSTIWLQLPSIRPQILP
ILARFKHISPAWTNFPQPPHISTNEKSFLQHLRFHLYLKMMNISKSMTQGEYATCIMMLLTSYDSSHFVADLIE
KKQLGKLKLQQRRNVLCILSRLLKDQAVMGDEDETIIDPILFKAITKASAVFEDTAACIVPFLFKICVDFKGKY
DKCVINLLGCLKGVNAEDEIVVRCLAECVDSIGLNVIARYERLNIETHSEFGVKWFFKLSRLFLKHGFTTHSFA
IANILFDRLSARKRNTMMIDRTSLDRIDRSQELINLLVEIYVAEGNSVALSSLPPAVQNRPDVRQVMNKSSKEW
LKLLSSNQMDSWELTIVQWMCGIQFNAITGDKYLNSILRCNFNEYTKKIDSPLKFVYFQLFHLSTSTLEIEEAI
SSMPLAPTIDQMRLMIIANATASFEPQSVEEHVVRAVRELRETSNRRKSGGNVKGINEKTTRMVKLAEMLTENK
AYDAAINLLDTWEHECLQWTSVAAESIDIDLIRICKQHVTCRSGDPRMADINLRTMHPRVPVMSDLAIAEWSLA
LSKITIEYRNDMEEGIRILEFGCKHLQNKDSVETRLKVLLKLHSVCIGQLSKLEEYRETRTYRMKQQAVTAFEQ
QIQNSCRTSLARGNSGDEWTKKTVQRVRKEHQFEKNDLEKVDNSLNSAARKAVSSGFDALLCISQLEDDDEAIR
ASSLIIFPLIDVIYKYETDVGVIALLKEHTKSKLPSKLWISATSHIASKCFSIEKSQITRHLSQILCHLIYDYP
YHVLHTILMYDDEKNASKVKGFLKTIFDARADQRDSSKLKEIVITIREAHQAYREIAMLDVRGNVRIQRVEING
KTMYRWPHDLKIFKCKLRQLPIPTISQKIGCPGDYSTTDLITWKRWKDVFTIADGISTPKIWEIEGSDGKWYKT
VWKKDDVRQDVLVEQMFDVTNNMLEKAMLRTYNVVPLDTECGVIEFCGGTVSLKEVMCGVTREGGLHREFNSEE
VSASKVSSMMRQVQTESTETRRQVFVEICQQYSPVFRHFFYTNFSTAQIWRQKIINYRQSLATWSIVCYIVGLG
DRHASNILFDQKLCTFVHIDLGMILEYSKRTLPVPEQVPFRITRDVLDPILIEGIENGQLAEECTQIMEKLKEN
GKVILGVASALLRETMTNFREAEQAAGRPSYISEMAIGRLREKLRGTDDGVTAQSSNLQIRRLLREATSADNLS
RMFCGWMPFL

D.2 *atm-1.a*

*** Indicates a PTC**

MSVTQLKNLKHAIAQLLEWDGTKTARKKIVDEVVLLYHALGAEALSEDNQEIYDLYDLSARIFNLAKKEIEEAN
QQFEKERKKGTRRSEKPVPTPLFELSIQHLKRCCQQGIDHNQVPWIAYCLKLLEFPITITEKSIENEISNVLLL
SSNASQLHWAEHAHLSSLWKWIWSRVETADIGALAMRNYMELAANLLENVDYVVFEKSPIDLMAKVMGTLKKSV
EMGNPKEYWCHYRKFSVVHILSYIVHRWGLEARDFILREIFELTGSLSNLISVAHKDGEQSKKCIMRLIDDLVK
LAMIETVHGHRTMNEVTRGNIQKLVKTGIQESLKSAHRNFSRSSTFSISEECVRYLTRWLLAERRLEQPSAAMN
ESFELTGDSSSKKKDDATFDSLIDLSFGSTISGKHKLNAWNGVMQILNELLKSRRLELQVTEKIVTILWEKRKS
YTTEPLRTVFCSILSTVVCQADVRFGHRKVPTIDSILKYSLSLMPNVASLPSAAALTETIVRFRTVSREGLRNT
WDTVSRTSSGSFEVVRLISALISVTEFDENSRFANDERVRSWSFRKDIIEWVLLDPNAHSHKLLYQLCQYHPTY
CYESEASSSDDSLLQTLKLCKLACSPAPPSAPKALRPLEASIEEIVRYVHDKLKSILATEITLPAFVLCHEFAL
KYPDRSYEFNKMYKKLYQIMEDQEEDEFLQSARHFSKWPQNLTLPIQKQTINCMAVFFEANLDNQLVDLCQWSD
RRKVLVEMLAELAATRSEIRDKLQKSMPFNKFVKECIMENRGDLYEMTKRFEKYSFLLSIRNLIVTRMIITNEA
ARLLGDGETISETDIFIIEKRTLSTCIRNVSEGKELSGYTLDPYTVAANVHNVHFDHINVEIYLELLKKSPFFA
QNIVRHLLRQNGKEAEEETWHLHATVLKIVMKDEKLLAVCVATIPNMVRYLKVYQIHFSPKSNAAKFLHLDMES
ISHCQSYLRKPTKSSNLITAANFLTLFGCEKRTWKRPILRFWSIFKQQPAMCCEKLLIFAEECVELGLNHRIAC
LLRALTTSEFCRKALCDEYLKIAFQLTYRSIFLILSKNECRPEIVELCDDMNLRYDLLQHQIKHVAAHHLEHFE
RFETKIAFSVEKFLKSGIDGIDFEDLGLVEFYKQLNENLTEDAIRSNEARNIYIVDILSTIWLQLPSIRPQILP
ILARFKHISPAWTNFPQPPHISTNEKSFLQHLRFHLYLKMMNISKSMTQGEYATCIMMLLTSYDSSHFVADLIE
KKQLGKLKLQQRRNVLCILSRLLKDQAVMGDEDETIIDPILFKAITKASAVFEDTAACIVPFLFKICVDFKGKY
DKCVINLLGCLKGVNAEDEIVVRCLAECVDSIGLNVIARYERLNIETHSEFGVKWFFKLSRLFLKHGFTTHSFA
IANILFDRLSARKRNTMMIDRTSLDRIDRSQELINLLVEIYVAEGNSVALSSLPPAVQNRPDVRQVMNKSSKEW
LKLLSSNQMDSWELTIVQWMCGIQFNAITGDKYLNSILRCNFNEYTKKIDSPLKFVYFQLFHLSTSTLEIEEAI
SSMPLAPTIDQMRLMIIANATASFEPQSVEEHVVRAVRELRETSNRRKSGGNVKGINEKTTRMVKLAEMLTENK
AYDAAINLLDTWEHECLQWTSVAAESIDIDLIRICKQHVTCRSGDPRMADINLRTMHPRVPVMSDLAIAEWSLA
LSKITIEYRNDMEEGIRILEFGCKHLQNKDSVETRLKVLLKLHSVCIGQLSKLEEYRETRTYRMKQQAVTAFEQ
QIQNSCRTSLARGNSGDEWTKKTVQRVRKEHQFEKNDLEKVDNSLNSAARKAVSSGFDALYSVSSHL*LPVSRF
AHNFDV**REERFQGERLLENDI*RAS*PTGFVEA*GDCDYYS*GSPGL*RNCNA*RERKRSDTTC*NKRKNDV
PMAT*FEDFQVQIASIADSHHFAKDWLPG*LLDD*PNHVETMERCIHNCRRNFHSKNLGN*G*RWKMV*NCVEE
GRRTTGCAC*ANV*CYE*YVGEGDASDV*CCPVRH*MWSYRVLWWNC*LKRSYVWCDTRRRSPPGVQFRRSFGE
*SVVNDETSTDRVHRDTTTSFCGDLSAVFSSF*TLLLHQLFDGADLAAKNHKLQAELGHVEYCVLHRWPRRSTR
VEYSIRSEIVHICAYRFGNDFGV**TNVASSRTSSIPYNSRCARPDSD*RNREWTVG*GMYADYGEIEGKWKGN
PRCRLCSTPRNDDKFPRSRTGCRPPVLHF*NGHRTASRKATGHR*RCDGPIVESSNSATFTRSYEC*QFIANVL
RMDAVFV

D.3 *atm-1.b*

MSVTQLKNLKHAIAQLLEWDGTKTARKKIVDEVVLLYHALGAEALSEDNQEIYDLYDLSARIFNLAKKEIEEAN
QQFEKERKKGTRRSEKPVPTPLFELSIQHLKRCCQQGIDHNQVPWIAYCLKLLEFPITITEKSIENEISNVLLL
SSNASQLHWAEHAHLSSLWKWIWSRVETADIGALAMRNYMELAANLLENVDYVVFEKSPIDLMAKVMGTLKKSV
EMGNPKEYWCHYRKFSVVHILSYIVHRWGLEARDFILREIFELTGSLSNLISVAHKDGEQSKKCIMRLIDDLVK
LAMIETVHGHRTMNEVTRGNIQKLVKTGIQESLKSAHRNFSRSSTFSISEECVRYLTRWLLAERRLEQPSAAMN
ESFELTGDSSSKKKDDATFDSLIDLSFGSTISGKHKLNAWNGVMQILNELLKSRRLELQVTEKIVTILWEKRKS
YTTEPLRTVFCSILSTVVCQADVRFGHRKVPTIDSILKYSLSLMPNVASLPSAAALTETIVRFRTVSREGLRNT
WDTVSRTSSGSFEVVRLISALISVTEFDENSRFANDERVRSWSFRKDIIEWVLLDPNAHSHKLLYQLCQYHPTY
CYESEASSSDDSLLQTLKLCKLACSPAPPSAPKALRPLEASIEEIVRYVHDKLKSILATEITLPAFVLCHEFAL
KYPDRSYEFNKMYKKLYQIMEDQEEDEFLQSARHFSKWPQNLTLPIQKQTINCMAVFFEANLDNQLVDLCQWSD
RRKVLVEMLAELAATRSEIRDKLQKSMPFNKFVKECIMENRGDLYEMTKRFEKYSFLLSIRNLIVTRMIITNEA
ARLLGDGETISETDIFIIEKRTLSTCIRNVSEGKELSGYTLDPYTVAANVHNVHFDHINVEIYLELLKKSPFFA
QNIVRHLLRQNGKEAEEETWHLHATVLKIVMKDEKLLAVCVATIPNMVRYLKVYQIHFSPKSNAAKFLHLDMES
ISHCQSYLRKPTKSSNLITAANFLTLFGCEKRTWKRPILRFWSIFKQQPAMCCEKLLIFAEECVELGLNHRIAC
LLRALTTSEFCRKALCDEYLKIAFQLTYRSIFLILSKNECRPEIVELCDDMNLRYDLLQHQIKHVAAHHLEHFE
RFETKIAFSVEKFLKSGIDGIDFEDLGLVEFYKQLNENLTEDAIRSNEARNIYIVDILSTIWLQLPSIRPQILP
ILARFKHISPAWTNFPQPPHISTNEKSFLQHLRFHLYLKMMNISKSMTQGEYATCIMMLLTSYDSSHFVADLIE
KKQLGKLKLQQRRNVLCILSRLLKDQAVMGDEDETIIDPILFKAITKASAVFEDTAACIVPFLFKICVDFKGKY
DKCVINLLGCLKGVNAEDEIVVRCLAECVDSIGLNVIARYERLNIETHSEFGVKWFFKLSRLFLKHGFTTHSFA
IANILFDRLSARKRNTMMIDRTSLDRIDRSQELINLLVEIYVAEGNSVALSSLPPAVQNRPDVRQVMNKSSKEW
LKLLSSNQMDSWELTIVQWMCGIQFNAITGDKYLNSILRCNFNEYTKKIDSPLKFVYFQLFHLSTSTLEIEEAI
SSMPLAPTIDQMRLMIIANATASFEPQSVEEHVVRAVRELRETSNRRKSGGNVKGINEKTTRMVKLAEMLTENK
AYDAAINLLDTWEHECLQWTSVAAESIDIDLIRICKQHVTCRSGDPRMADINLRTMHPRVPVMSDLAIAEWSLA
LSKITIEYRNDMEEGIRILEFGCKHLQNKDSVETRLKVLLKLHSVCIGQLSKLEEYRETRTYRMKQQAVTAFEQ
QIQNSCRTSLARGNSGDEWTKKTVQRILCHLIYDYPYHVLHTILMYDDEKNASKVKGFLKTIFDARADQRDSSK
LKEIVITIREAHQAYREIAMLDVRGNVRIQRVEINGKTMYRWPHDLKIFKCKLRQLPIPTISQKIGCPGDYSTT
DLITWKRWKDVFTIADGISTPKIWEIEGSDGKWYKTVWKKDDVRQDVLVEQMFDVTNNMLEKAMLRTYNVVPLD
TECGVIEFCGGTVSLKEVMCGVTREGGLHREFNSEEVSASKVSSMMRQVQTESTETRRQVFVEICQQYSPVFRH
FFYTNFSTAQIWRQKIINYRQSLATWSIVCYIVGLGDRHASNILFDQKLCTFVHIDLGMILEYSKRTLPVPEQV
PFRITRDVLDPILIEGIENGQLAEECTQIMEKLKENGKVILGVASALLRETMTNFREAEQAAGRPSYISEMAIG
RLREKLRGTDDGVTAQSSNLQIRRLLREATSADNLSRMFCGWMPFL