HUMAN MUTATION 23:67–76 (2004)

RESEARCH ARTICLE

Nonclassical Splicing Mutations in the Coding and Noncoding Regions of the ATM Gene: Maximum Entropy Estimates of Splice Junction Strengths

Laura Eng,[1] Gabriela Coutinho,[1] Shareef Nahas,[1] Gene Yeo,[2] Robert Tanouye,[1] Mahnoush Babaei,[1] Thilo Dörk,[3] Christopher Burge,[2] and Richard A. Gatti[1]*

[1]*Department of Pathology and Laboratory Medicine, The David Geffen School of Medicine at UCLA, Los Angeles, California;* [2]*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts;* [3]*Department of Radiation Oncology and Clinics of Obstetrics and Gynecology, Medical School Hannover, Hannover, Germany*

Communicated by Mark H. Paalman

Ataxia-telangiectasia (A-T) is an autosomal recessive neurological disorder caused by mutations in the *ATM* gene. Classical splicing mutations (type I) delete entire exons during pre-mRNA splicing. In this report, we describe nine examples of nonclassical splicing mutations in 12 A-T patients and compare cDNA changes to estimates of splice junction strengths based on maximum entropy modeling. These mutations fall into three categories: pseudoexon insertions (type II), single nucleotide changes within the exon (type III), and intronic changes that disrupt the conserved 3′ splice sequence and lead to partial exon deletion (type IV). Four patients with a previously reported type II (pseudoexon) mutation all shared a common founder haplotype. Three patients with apparent missense or silent mutations actually had type III aberrant splicing and partial deletion of an exon. Five patients had type IV mutations that could have been misinterpreted as classical splicing mutations. Instead, their mutations disrupt a splice site and use another AG splice site located nearby within the exon; they lead to partial deletions at the beginning of exons. These nonclassical splicing mutations create frameshifts that result in premature termination codons. Without screening cDNA or using accurate models of splice site strength, the consequences of these genomic mutations cannot be reliably predicted. This may lead to further misinterpretation of genotype–phenotype correlations and may subsequently impact upon gene-based therapeutic approaches. Hum Mutat 23:67–76, 2004. © **2003 Wiley-Liss, Inc.**

KEY WORDS: ataxia-telangiectasia; ATM; mutation screening, splicing, entropy modeling

DATABASES:
ATM – OMIM: 208900; GenBank: U82828.1;
ARG2 – OMIM: 107830; GenBank: NM_001172;
CFTR – OMIM: 602421; GenBank: NM_000492;
www.benaroyaresearch.org/bri_investigators/atm.htm (ATM mutation database); http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html (MaxENT website); http://exon.cshl.edu/ESE/ (ESE finder website)

INTRODUCTION

Ataxia-telangiectasia (A-T; MIM# 208900) is an autosomal recessive neurological disorder that is caused by mutations in the ataxia-telangiectasia mutated (*ATM*) gene [Boder and Sedgwick, 1958; Gatti et al., 1991; Gatti et al., 2001]. *ATM* spans approximately 150 kb of genomic DNA; it contains 66 exons and a 9,168-bp coding region [Savitsky et al., 1995; Uziel et al., 1996; Platzer et al., 1997]. Most A-T patients inherit a different mutation from each parent, i.e., they are compound heterozygotes.

Because sequencing the large *ATM* gene is laborious and costly, we use a series of screening tests for mutation detection. Initially, cDNA is synthesized from lymphoblastoid cell lines (LCLs) and analyzed by the protein truncation test (PTT) [Telatar et al., 1996]. If mutations remain undetected, we continue to screen either gDNA or cDNA by the SSCP assay [Castellvi-Bel et al., 1999]. Other labs have also used dHPLC [Thorstenson et al.,

Received 9 June 2003; accepted revised manuscript 29 August 2003.

*Correspondence to: R.A. Gatti, The David Geffen School of Medicine at UCLA, Department of Pathology and Laboratory Medicine, Los Angeles, CA 90095-1732.
E-mail: rgatti@mednet.ucla.edu

Grant sponsor: A-T Medical Research Foundation; Grant sponsor: NIH; Grant numbers: NS35322; CA76513; Grant sponsor: DFG; Grant number: Do-371/2-1.

DOI 10.1002/humu.10295
Published online in Wiley InterScience (www.interscience.wiley.com).

2001; Bernstein et al., 2003] or DOVAM [Buzin et al., 2003] to screen for ATM mutations. Although these two methods have reached the highest efficiency rates for mutation detection, both use gDNA as a template and do not detect unconventional splicing defects.

To date over 400 mutations have been described for this gene, and they are observed over the entire coding region with no true "hot spots" (www.benaroyaresearch.org/bri_investigators/atm.htm) [Mitui et al., 2003]. In A-T patients, most of these changes are predicted to give rise to truncated proteins that are highly unstable, effectively producing a null phenotype (~85%). However, a significant number of missense mutations have been recorded (~10%) and recent data suggest that many of these have dominant interfering effects [Scott et al., 2002; Spring et al., 2002; Concannon, 2002]. Some of the mutations described below would not have been detected at the gDNA level and, more importantly, some would have been misconstrued to be missense mutations.

Unconventional splicing defects often occur at exons with weak homology to canonical splicing sequences. Scoring all splice sites in the ATM gene with a splice-site model that accounts for adjacent as well as nonadjacent pairwise dependencies (maximum entropy [MaxENT] website: http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) [Yeo and Burge, 2003] suggests that many "weak exons" exist, making the gene more susceptible to splicing mutations. In addition, these weak exons may render the *ATM* gene especially amenable to alternative splicing in various cell types. An understanding of such mechanisms may allow better interpretation of genotype/phenotype relationships and may impact upon subsequent gene-based therapies.

Herein we report a series of nonclassical splicing mutations in the *ATM* gene. We describe three types of nonclassical splicing mutations in 12 A-T patients, and their consequences. We compare the MaxENT scores of wild-type and mutant splice junctions to predict usage of the splice site. Classical (type I) splicing mutations in the *ATM* gene delete entire exons during pre-mRNA splicing. Nonclassical type II splicing mutations introduce cryptic or pseudoexons into the mRNA and have been described previously by McConville et al. [1996] and Pagani et al. [2002]. We have found three type III splice mutations that would be incorrectly interpreted as missense or silent mutations if cDNA changes were not analyzed or if the newly created splice site were not scored with the MaxENT model. Five new type IV mutations have been analyzed that involve intronic changes that delete only portions of exons. We provide here the first comprehensive analysis of such splicing mutations in the ATM gene.

MATERIALS AND METHODS

Subjects

We selected from our mutation inventory nine mutations in 12 A-T patients whose PTT truncation results did not appear to be caused by classical exon skipping or whose genomic mutations were suspected to involve aberrant splicing (Table 1). The patients were representative of various ethnic groups: German, Polish, Brazilian, Guatemalan, Hispanic American, American, Costa Rican, Turkish, and Iranian. LCLs were established for most patients so that both cDNA and gDNA could be analyzed.

Mutations

DNAs were prescreened by PTT using eight overlapping cDNA regions; this only identifies mutations that lead to premature termination codons (PTC) [Telatar et al., 1996, Den Dunnen and Van Ommen, 1999]. SSCP prescreening followed; this method efficiently identifies mutations in >70% of samples if fragments are <300 nucleotides in length [Castellvi-Bel et al., 1999]. Lastly, targeted regions of cDNA or gDNA were directly sequenced in forward and reverse directions by automated methods.

Mutation nomenclature follows that required by the journal. The cDNA numbering, where appropriate, is based on +1 being the A of the initiation codon for GenBank sequence version U82828.1.

Haplotypes. Standardized short tandem repeat (STR) haplotyping was used to identify founder effect haplotypes in patients with type II mutation, IVS20-579delAAGT [Mitui et al., 2003]. Four markers were used: A4 (S1819), NS22, S2179, and A2 (S1818) [Rotman et al., 1994; Uhrhammer et al., 1995; Vanagaite

TABLE 1. A-T Patients With Non-Classical Splicing Defects in the *ATM* gene*

Proband	Ethnicity	Splicing mutation	Type	Consequence	ATM protein levels (%)
GAT1	German	IVS20-579delAAGT[a]	II	Insertion of 65 nt between Ex 20 and 21	10
WAR28	Polish	IVS20-579delAAGT	II	Insertion of 65 nt between Ex 20 and 21	0
AT195LA	Hispanic American	IVS20-579delAAGT	II	Insertion of 65 nt between Ex 20 and 21	NT
AT10LA	Guatemalan	IVS20-579delAAGT	II	Insertion of 65 nt between Ex 20 and 21	NT
CRAT [C]	Costa Rican	7449>A (W2483X)[a,b]	III	Deletion of last 70 nt of Ex 52	0
TAT [C]	Turkish	7865C>T (A2622V)[a,b]	III	Deletion of last 64 nt of Ex 55	0
IRAT9	Iranian	513C>T (Y171Y)[a,b]	III	Deletion of first 22 nt of Ex 8	0
AT201LA	Hispanic American	IVS29-IG>A	IV	Deletion of first nt of Ex 30	0
BRAT [T]	Brazilian	IVS11-2A>G	IV	Deletion of first 7 nt of Ex 12	0
AT202LA	Caucasian American	IVS37-5delTCTA[a]	IV	Deletion of first 7 nt of Ex 38	0
TIAT28	Turkish	IVS21-2A>G[b]	IV	Deletion of first 32 nt of Ex 22	0
GAT3	German	IVS38-2A>C	IV	Deletion of first 61 nt of Ex 39	0

*Mutation nomenclature: nucleotide numbering is based on +1 being the A of the initiation codon for GenBank sequence version U82828.1.
[a]Mutation reported previously.
[b]Indicates homozygosity.
NT, not tested.

et al., 1995; Udar et al., 1999]. In addition, single nucleotide polymorphism (SNP) haplotyping was performed, as described by Campbell et al. [2003].

Maximum entropy scores. In order to estimate the strengths of 5′ and 3′ splice junctions in the *ATM* gene, splice site sequence motifs were scored using the splice site models introduced by Yeo and Burge [in press] and the software available at: http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html. Briefly, splice site models that take into account adjacent and nonadjacent dependencies are built under the MaxENT framework using large datasets of human splice sites [Yeo and Burge, in press]. These splice site models assign a log-odd ratio (MaxENT score) to a 9-mer (5′ splice site) or a 23-mer (3′ splice site) sequence. The higher the score, the higher the probability that the sequence is a true splice site. Also, it can be argued that given two sequences of differing scores, the higher scoring sequence has a higher likelihood of being used.

RESULTS

We analyzed nine examples of nonclassical splicing mutations (type II, III, and IV) in 12 A-T patients. Ethnicity, mutations, and consequences are listed in Table 1, as well as ATM protein expression in LCLs from these patients. We also tried to compare the strength of splice-site junctions for wild-type and mutated ATM sequences, as estimated by MaxENT scores, to actual changes observed in cDNA for each mutation. As can be surmised from the analyses below, MaxENT scores generally correlated well with actual cDNA changes for these splicing mutations, and appeared to provide a possible alternative to cDNA analysis in interpreting gDNA changes.

Weak Exons in the ATM Gene

MaxENT scores were calculated for each of the 5′ and 3′ splice sites in the *ATM* gene (GenBank U82828.1) and compared to those of *ARG2* (MIM# 107830; GenBank NM_001172) and *CFTR* (MIM# 602421; GenBank NM_000492), two unrelated genes selected at random (Table 2). This identified weak exons that might be prone to splicing mutations or contain auxiliary *cis*-acting splicing elements [Faustino and Cooper, 2003]. Note, for example, the weak splice sites at IVS1, IVS34, IVS37, and IVS51. The *ATM* gene appeared to contain a greater number of weak exons. The average MaxENT scores for all 5′ splice sites in *ATM* was 7.74, as compared with the MaxENT score averages of 9.63 for *ARG2* and 8.59 for *CFTR*. The 3′ splice sites of *ATM* are only slightly weaker in comparison: MaxENT average of 7.46 for *ATM* vs. 8.76 and 8.60 for *ARG2* and *CFTR*, respectively.

Type II (Pseudoexon) Splicing Defects

Mutation IVS20-579delAAGT, found in four patients of diverse ethnicity, defined a new intronic splicing processing element (ISPE) within intron 20 [Pagani et al., 2002]. This mutation deletes part of a U1 snRNA binding site, which normally suppresses the retention of a 65-nucleotide segment in intron 20 during pre-mRNA splicing. When the U1 snRNA cannot bind, a pseudoexon is retained in the cDNA. This ultimately leads to a frameshift and PTC downstream and truncates the ATM protein (Fig. 1).

Pagani et al. [2002] first described the IVS20-579delAAGT mutation in a German A-T patient (GAT1). We have detected the same mutation in three additional unrelated A-T families of Polish, Hispanic-American, and Guatemalan backgrounds. All four patients carry this mutation on the same SNP H2 haplotype. The Polish patient's STR haplotype is identical to the original patient; however, the two New World A-T patients have a slightly altered allele at S2179, probably reflecting polymerase slippage during a previous generation. The other alleles are identical, suggesting that IVS20-579delAAGT can be traced to a common ancestor and represents a single mutational event (Fig. 2) [Campbell et al., 2003; Mitui et al., 2003]. LCLs from two of the four patients were deficient in the expression of ATM protein and were radiosensitive by colony survival assay [Huo et al., 1994; Sun et al., 2002] (data not shown).

Type III (Exonic) Splicing Defects

We reanalyzed three mutations that had been misinterpreted as missense or silent changes. These single nucleotide changes, which occur within exonic regions, did not appear to involve exonic splicing enhancer (ESE) motifs for SR proteins SF2ASF, SC35, SRp40, SRp55 (http://exon.cshl.edu/ESE/) (data not shown). After cDNA and MaxENT analyses, these mutations are now interpretable as nonclassical type III (exonic) splicing mutations.

The mutation 7449G>A (W2483X) is found on the CRAT [C] haplotype and has been observed in approximately 13% of Costa Rican patients [Telatar et al., 1998b]. It results in a deletion of the last 70 nucleotides of exon 52. It converts the low-scoring hexamer GTGGGT sequence (MaxENT = −0.50) to the high scoring GTGAGT (5.56) and competes as a 5′ splice site with the normal, but weak GCAAGT (3.24) (Fig. 3a).

The mutation 7865C>T (A2622V) is found on the TAT [C] haplotype and has been identified in approximately 4.5% of Turkish patients [Teraoka et al., 1999]. It deletes the last 64 nucleotides of exon 55. It converts a GCACTT. (MaxENT = 0.07) to a GTACTT (MaxENT = 7.82) and competes as a 5′ splice site with the normal GTATGC (MaxENT = 6.97). This effect on aberrant splicing was confirmed in LCLs from different A-T patients who were homozygous for the 7865C>T (A2622V) mutation (Fig. 3b).

A third example of a type III (exonic) splicing mutation, 513C>T (Y171Y), was found in the homozygous patient, IRAT9. It converts the ACTTCAG heptamer (MaxENT = 8.54) to ATTTCAG (MaxENT = 10.42), which then competes more readily as a 3′ splice site with the normal TTTTAAG (MaxENT = 7.84). This gDNA change deletes the first 22 nucleotides of exon 8 in the cDNA (Fig. 3c).

TABLE 2. Strength of Splice Junctions (Maximum Entropy)*

A. *ATM* exons

IVS	5′ SS	3′ SS	IVS	5′ SS	3′ SS	IVS	5′ SS	3′ SS
1	4.20	−19.92	23	10.03	8.27	45	10.77	7.92
2	8.55	8.29	24	7.57	7.43	46	7.26	7.83
3	6.45	8.35	25	6.49	7.90	47	9.76	6.70
4	5.69	9.72	26	7.84	8.52	48	8.76	8.17
5	4.57	6.72	27	9.66	9.94	49	4.02	9.04
6	9.82	7.19	28	9.99	11.38	50	10.57	7.01
7	7.76	7.84	29	8.34	5.58	51	8.63	1.87
8	7.05	9.42	30	7.52	10.50	52	3.24	6.03
9	7.08	7.63	31	8.88	9.92	53	8.62	6.96
10	8.68	10.82	32	8.55	8.27	54	7.63	9.87
11	4.05	12.11	33	7.96	11.06	55	6.97	5.38
12	11.01	9.78	34	−2.26	5.52	56	8.83	7.11
13	8.55	8.85	35	5.88	2.46	57	9.60	4.97
14	8.35	2.57	36	9.27	8.39	58	9.01	6.65
15	7.66	5.66	37	0.90	8.60	59	10.13	6.53
16	9.45	6.62	38	6.42	10.05	60	3.39	10.09
17	10.57	8.06	39	6.49	4.90	61	9.66	8.74
18	7.79	8.81	40	8.73	8.80	62	11.00	8.29
19	6.62	9.90	41	8.99	3.03	63	8.27	11.33
20	10.13	8.07	42	8.76	10.44	64	9.60	10.46
21	6.29	8.55	43	6.62	8.44			

B. *ARG2* exons

IVS	5′ SS	3′ SS
1	9.22	4.90
2	10.24	7.95
3	9.88	7.90
4	8.68	11.34
5	9.49	7.35
6	10.08	12.59
7	9.80	9.29

C. *CFTR* exons

IVS	5′ SS	3′ SS	IVS	5′ SS	3′ SS	IVS	5′ SS	3′ SS
1	9.11	9.32	10	9.88	5.10	19	6.99	9.12
2	7.83	12.20	11	8.88	7.99	20	10.47	8.97
3	8.49	9.40	12	9.79	7.50	21	6.65	12.30
4	6.97	12.13	13	7.84	6.62	22	9.60	8.27
5	6.13	4.16	14	11.08	9.58	23	10.28	11.27
6	9.35	9.58	15	8.40	10.92	24	8.31	7.09
7	8.63	8.65	16	8.17	7.06	25	10.27	9.87
8	7.66	5.66	17	7.44	9.47			
9	8.56	7.84	18	8.04	4.89			

*Ideal splice site scores: 11.81 (5′SS) and 13.59 (3′SS).

Type IV (Intronic) Splicing Defects

Type IV (intronic) splicing mutations would be interpreted as classic (type I) splice defects if only gDNA were sequenced; however, further analyses demonstrated that these mutations do not involve exon skipping but instead result in the activation of new 5′ or 3′ splice sites and only partial exonic deletions.

The mutation IVS29–1G>A, identified in a Hispanic-American patient, disrupts the original 3′ splice site and activates a cryptic splice site only one nucleotide downstream of the mutation site. This change causes the deletion of the first nucleotide of exon 30 and leads to a frameshift (Fig. 4a).

The mutation IVS11–2A>G was observed in two ostensibly unrelated Brazilian A-T patients [Coutinho et al., in press]. It disrupts the original 3′ splice site and activates a new cryptic 3′ splice site seven nucleotides downstream of the mutation site, resulting in the deletion of the first seven nucleotides of exon 12 and a frameshift (Fig. 4b). Both patients shared the same haplotype.

The mutation IVS37-5delTCTA was previously reported by Gilad et al. [1996] in an American patient (AT2SF) as the cDNA change 5320del7. We observed the same change in another American A-T patient (AT202LA) (Table 1). In gDNA, the mutation deletes part of the intronic 3′ AG splice site. The cDNA showed that the mutation leads to splicing at another nearby downstream AG within the exon, resulting in the deletion of the first seven nucleotides of exon 38 (Fig. 4c).

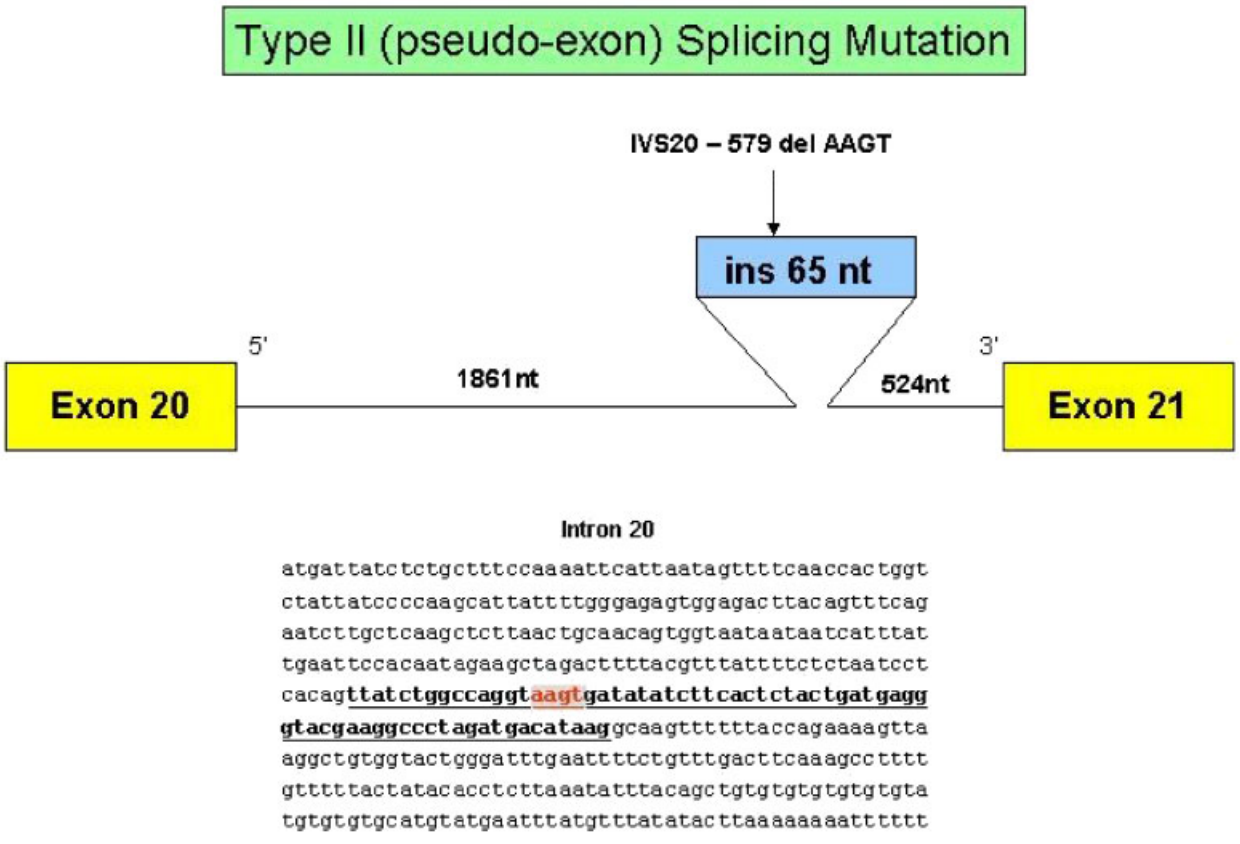


FIGURE 1. **Type II (pseudoexon) splicing mutation IVS20-579delAAGT results in the insertion of 65 nucleotides between exons 20 and 21 in the cDNA. Adjacent sequences are shown at bottom; the retained intronic sequences are underlined and the genomic deleted sequences are shaded. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

		GAT1	WAR28	AT195LA	AT10LA
STR haplotype	A4	131	131	131	131
	NS22	165	165	165	165
	S2179	139	139	141	141
	A2	160	160	160	160
SNP haplotype		*H2*	*H2*	*H2*	*H2*
Ethnicity		German	Polish	Hispanic	Guatamalan
Haplovariant		A	A	B	B

FIGURE 2. **Comparison of haplotypes for four patients with IVS20-579delAAGT. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

The mutation IVS21–2A>G was found in a homozygous Turkish patient. It disrupts the 3′ AG splice site, instead using the next downstream AG, which is in the exon. This results in the deletion of the first 32 nucleotides of exon 22 (Fig. 4d).

The mutation IVS38–2A>C was identified in a heterozygous German patient. Analysis of cDNA showed that it leads to splicing at a downstream AG within the exon, resulting in the deletion of the first 61 nucleotides of exon 39. It is interesting that in this case the next potential downstream AG site after the mutation was skipped over and not used; this would have resulted in an in-frame deletion of only 21 nucleotides (Fig. 4e).

DISCUSSION

Splice site consensus sequences are located at exon–intron junctions and are conserved phylogenetically [Shapiro and Senapathy, 1987; Cartegni et al., 2002]. Each nucleotide in the splice site motif has a differing rate of variance, with the 5′ GT and the 3′ AG motifs being the most highly conserved. Recognition of such

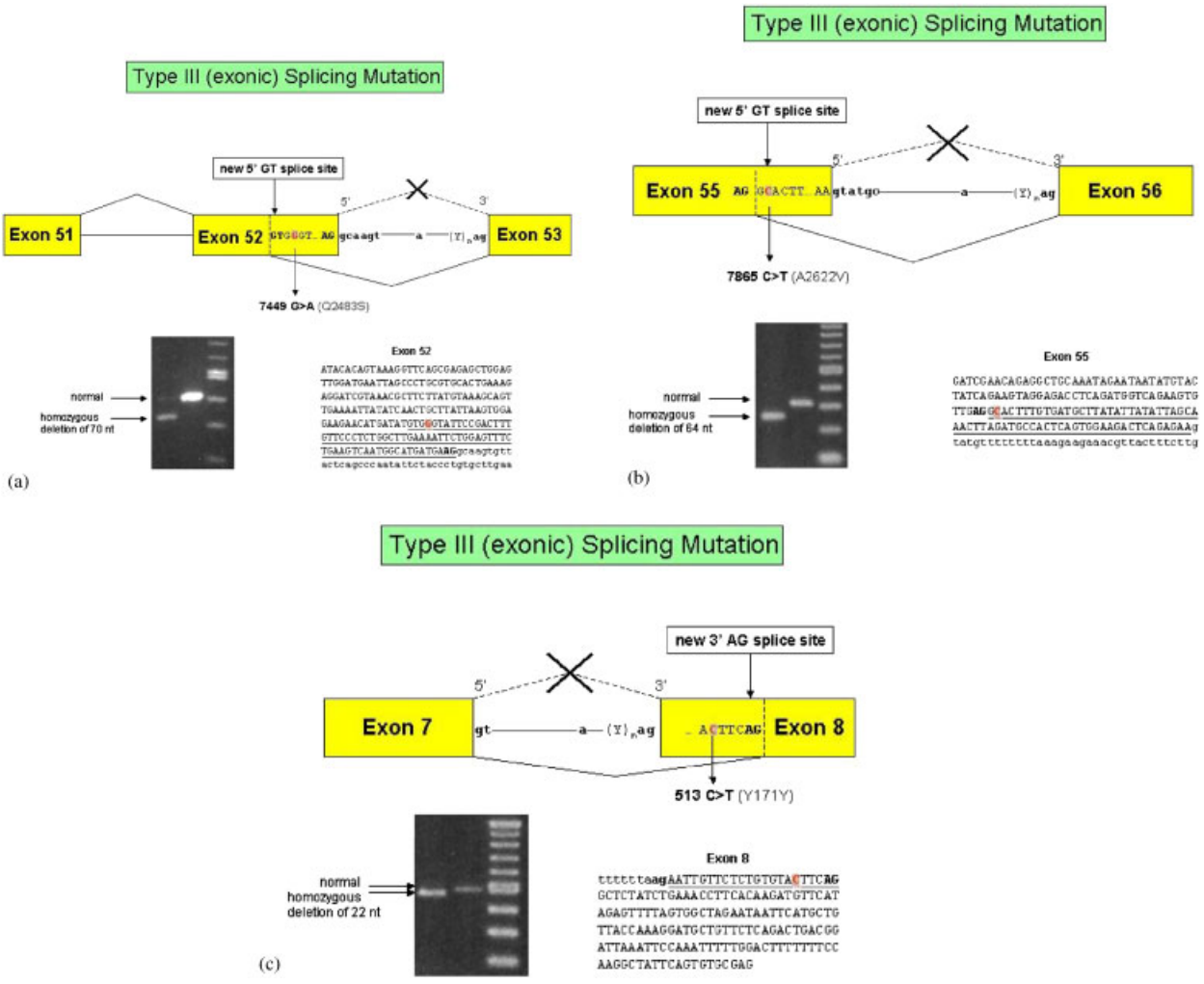


FIGURE 3. **a: Type III (exonic) splicing mutation 7449G>A (W2483X) results in the deletion of the last 70 nucleotides of exon 52, most likely due to the increased splicing efficiency of GTGAGT over GTGGGT or GCAAGT. Adjacent sequences are shown at bottom right, with the deleted nucleotides underlined and the mutated site shaded. An agarose gel showing the homozygous cDNA deletion of 70 nucleotides is shown at bottom left. b: Type III (exonic) splicing mutation 7865C>T (A2622V) results in the deletion of the last 64 nucleotides of exon 55, most likely due to the increased splicing efficiency of AG/GTACTT over AG/GCACTT or AA/GTATGC. Adjacent sequences are shown at bottom right with the deleted nucleotides underlined and the mutated site shaded. An agarose gel showing the homozygous cDNA deletion of 64 nucleotides is at bottom left. c: Type III (exonic) splicing mutation 513C>T (Y171Y) results in the deletion of the first 22 nucleotides of exon 8, most likely due to the increased splicing efficiency of ATTTCAG over ACTTCAG or TTTTAAG. Adjacent sequences are shown at bottom right, with the deleted nucleotides underlined and the mutated site shaded. An agarose gel showing the homozygous cDNA deletion of 22 nucleotides is at bottom left. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

motifs is important for understanding splicing mechanisms and helps to predict the consequences of changes identified in genomic DNA.

The *ATM* gene has many exons. We attempted to analyze the relative strength of the splice sites for each exon using newly-available software developed by Yeo and Burge [2003] (Table 2). The ideal MaxENT score is 11.81 for a 5′ splice site and 13.59 for a 3′ splice site. Theoretically, the larger the MaxENT value, the more efficient the splicing. The estimated MaxENT average for all the 5′ splice sites in *ATM* was 7.74. When compared with the MaxENT score averages of *ARG2* (9.63) and *CFTR* (8.59), this lower average suggests that weak 5′ splice sites exist in the *ATM* gene and implies a susceptibility to both splice mutations and to physiological alternative splicing. Furthermore, exons with weak splice sites have been predicted to contain more abundant auxiliary *cis*-elements, such as ESEs and ISEs [Faustino and Cooper, 2003], perhaps allowing for even more intricate regulation of alternative splicing.

We describe herein a group of nine nonclassical mutations in the *ATM* gene in 12 A-T patients (Table 1). The mutations have been grouped into three subtypes: type II, type III, and type IV. A maximum entropy model [Yeo and Burge, in press] was used to estimate the strengths of new splice sites created by these mutations as a partial explanation for the splices observed.

Type II (pseudoexon) splicing mutations occur in the deep intron and involve the insertion of a pseudoexon at the cDNA level [McConville et al., 1996; Pagani et al., 2002]. We observed three patients (AT195LA, WAR28, and AT10LA) with the same type II IVS20-579del-AAGT mutation that Pagani et al. [2002] described in a German patient (GAT1). All four patients had identical H2 SNP haplotypes. The STR haplotypes were similar but not identical (Fig. 2). Human migration patterns would suggest that the German and Polish haplotypes are ancestral to the haplovariants of the Guatemalan and Hispanic-American patients [Underhill et al., 2001; Campbell et al., 2003].

Type III (exonic) splicing mutations involve mutations that lie within the coding region. Some of these are easily misinterpreted as missense mutations. This distinction becomes especially important because ATM missense mutations have been associated with cancer as well as with milder A-T phenotypes. Interestingly, missense mutations constitute only ~10% of mutations in A-T patients but are far more frequent in breast cancer patients, some of them creating dominant interfering effects [Gilad et al., 1998; Stankovic et al., 1998; Gatti et al., 1999; Li and Swift 2000; Chenevix-Trench et al., 2002; Scott et al., 2002; Spring et al., 2002; Concannon, 2002].

The type III mutation, 7449G>A, deletes the last 70 nucleotides of exon 52. The normal 5′ splice site of intron 52 is predicted to be weak, with a MaxENT of 3.24 (Fig. 3a). The normal intronic sequence is GCAAGT, instead of the highly conserved GTAAGT. The G>A mutation at 7449 creates a new and stronger 5′ splice site motif upstream within the exon, becoming GTGAGT, with a MaxENT value of 5.56. Although the new site contains an AT in the exonic sequence preceding the splice junction (shown at the lower right in Fig. 3a), instead of the highly conserved AG, the new AT/GTGAGT splice site is still considerably stronger than the original (5.56 vs. 3.24). We postulate that if this mutation occurred in a strong exon (i.e., in an exon with a strong 5′ splice site motif), this change would be unlikely to result in aberrant splicing or disease because the original 5′ splice site would have had a higher 5′ splice site MaxENT score and, thus, would have been used preferentially. On the other hand, Shapiro and Senapathy [1987] suggested that 5′ splice sites, which do not contain GT, might involve special recognition mechanisms for gene regulation or cell differentiation.

Another type III (exonic) splicing mutation, 7865C>T, appears at first to be a missense mutation (A2622V). However, in reality it causes the deletion of 64 nucleotides at the end of exon 55 (Fig. 3b). The normal 5′ intronic donor splice site of exon 55 (AA/GTATGC) is weakened by the presence of a T at position four (underlined) instead of the conserved A. It is further weakened by the exonic AA, which differs from the highly conserved AG. The 7865C>T mutation creates a new 5′ splice site upstream within exon 55, with the motif, AG/GTACTT. Although this new site is not much stronger (MaxENT of 7.82, versus 6.97), it is apparently strong enough in this A-T patient to bypass the normal 5′ splice site and use the new splice site within the exon, thereby causing an out-of-frame deletion, truncating the protein, and causing disease.

The last type III mutation is 513C>T. This genomic change preserves the amino acid codon, thus suggesting that it is a silent change (Y171Y) [Sandoval et al., 1999]. However, cDNA analysis indicated that, more importantly, the mutation results in the aberrant splicing within exon 8. The change creates a polypyrimidine tract, which typically precedes a 3′ splice site, and induces the usage of a new intronic 3′ splice site using a nearby downstream AG. The normal 3′ splice site has a MaxENT score of 7.84, while the new 3′ splice site has a score of 8.54 when not mutated and a MaxENT score of 10.42 when mutated. In the absence of RNA, it may be possible to utilize MaxENT scores to analyze nearby splice sites for interpreting potential type III splicing mutations.

Type IV (intronic) splicing mutations are somewhat difficult to describe succinctly in the absence of a subtype classification. They are also easily misinterpreted as classic splicing mutations that would delete an entire exon unless mRNA/cDNA were analyzed. Instead, they lead to partial deletions of exons. The IVS29–1G>A mutation disrupts the 3′ conserved splice site sequence where the normal AG is mutated to an AA. However, the first nucleotide of exon 30 is a G, creating another 3′ splice site by using the mutated A from the intron and the G in the exon. The mutation then deletes the first nucleotide of exon 30, causing a frameshift.

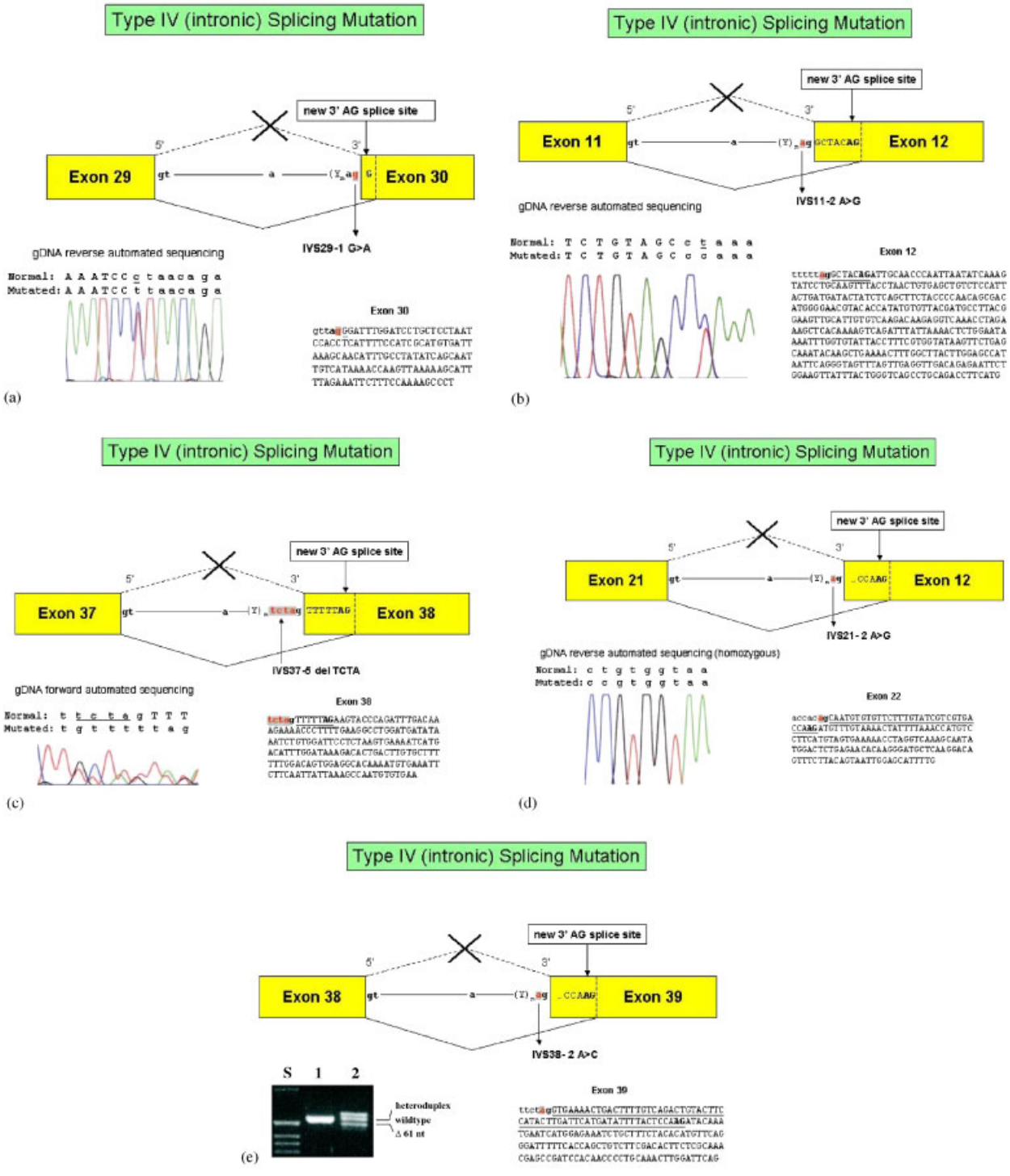
In the cases of type IV mutations IVS11–2A>G, IVS37-5delTCTA, IVS21–2A>G, and IVS38–2A>C, AG sites downstream of the normal 3′ splice sites were used. Typically, the next AG site is located 35–40 nucleotides away from the 3′ splice site, with less than 2% of species genomes containing an AG within 10 nucleotides of the 3′ splice site [Shapiro and Senapathy, 1987]. Despite this rule, in both IVS11–2A>G and IVS37-5delTCTA mutations, AG sites were present within six nucleotides. In the case of IVS38–2A>C, an AG site at position 20 with a MaxENT of 2.51 was bypassed; use of this site would have resulted in the in-frame cDNA deletion of 21 nucleotides. Instead, 61 nucleotides are deleted by the usage of the next AG site further downstream, with a MaxENT of 2.59. The mechanism for how one AG is selected over another within the exon is unclear. These limited data suggest that small in-frame deletions that result from type IV splicing mutations may produce a milder phenotype that is not recognizable as A-T. If true, a possible therapeutic approach in a patient with the IVS38–2A>C mutation might be to encourage the use of the bypassed AG splice site.

Splicing mutations are a prevalent and important class of mutations in A-T patients [Teraoka et al., 1999]. They are difficult to predict without analyzing mRNA/cDNA. The exact mechanism of a mutation may prove to be important in predicting the effects of

future therapeutic agents on specific A-T patients, such as in the possible use of complementary U1 snRNPs [Pagani et al., 2002] or in the use of agents that effect read-through of PTCs [Howard et al., 2000; Lai et al., 2003]. We conclude that the ATM transcript is susceptible to aberrant splicing, which can be induced by subtle sequence alterations. This requires that cDNA analyses be performed whenever missense or silent variants are encountered. Maximum entropy analysis also appears to be informative and may offer an alternative to cDNA studies in some circumstances.

ACKNOWLEDGMENTS

We thank Patrick Concannon and Jiuyong Xie for their critical comments in the preparation of this manuscript. We thank the NIH (NS 35322 and CA76513), the DFG (Do-371/2-1), and the A-T Medical Research Foundation for their grant support.



REFERENCES

Bernstein JL, Teraoka S, Haile RW, Borresen-Dale AL, Rosenstein BS, Gatti RA, Diep AT, Jansen L, Atencio DP, Olsen JH, Bernstein L, Teitelbaum SL, Thompson WD, Concannon P. 2003. Designing and implementing quality control for multicenter screening of mutations in the ATM gene among women with breast cancer. Hum Mutat 21:542–550.

Boder E, Sedgwick RP. 1958. Ataxia-telangiectasia: a familial syndrome of progressive cerebellar ataxia, oculocutaneous telangiectasia and frequent pulmonary infection. Pediatrics 21:526–554.

Buzin CH, Gatti RA, Nguyen VQ, Wen CY, Mitui M, Sanal O, Chen JS, Nozari G, Mengos A, Li X, Fujimura F, Sommer SS. 2003. Comprehensive scanning of the ATM gene with DOVAM-SSCP. Hum Mutat 21:123–131.

Campbell C, Mitui M, Coutinho G, Eng L, Thorstenson Y, Gatti RA. 2003. ATM mutations on distinct SNP and STR haplotypes in ataxia-telangiectasia patients of differing ethnicities reveal ancestral founder effects. Hum Mutat 21:80–85.

Cartegni L, Chew S, Krainer A. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298.

Castellvi-Bel S, Sheikhavandi S, Telatar M, Tai L-Q, Hwang M, Wang Z, Yang Z, Cheng R, Gatti RA. 1999. New mutations, polymorphisms, and rare variants in the ATM gene detected by a novel SSCP strategy. Hum Mutat 14:156–162.

Chenevix-Trench G, Spurdle AB, Gatei M, Kelly H, Marsh A, Chen X, Donn K, Cummings M, Nyholt D, Jenkins MA, Scott C, Pupo GM, Dork T, Bendix R, Kirk J, Tucker K, McCredie MR, Hopper JL, Sambrook J, Mann GJ, Khanna KK. 2002. Dominant negative ATM mutations in breast cancer families. J Natl Cancer Inst 94:205–215.

Concannon P. 2002. ATM heterozygosity and cancer risk. Nat Genet 32:89–90.

Coutinho G, Mitui M, Campbell C, Carvalho C, Nahas S, Sun X, Huo Y, Lai CH, Thorstenson Y, Tanouye R, Raskin S, Kim CA, Llerena J, Gatti RA. Five haplotypes account for fifty-five percent of ATM mutations in Brazilian patients with ataxia-telangiectasia: seven new mutations. Am J Med Genet (in press).

Den Dunnen J, Van Ommen G-J. 1999. The protein truncation test: a review. Hum Mutat 14:95–102.

Faustino NA, Cooper T. 2003. Pre-mRNA splicing and human disease. Genes Dev 17:419–437.

Gatti RA, Boder E, Vinters HV, Sparkes RS, Norman A, Lange K. 1991. Ataxia-telangiectasia: an interdisciplinary approach to pathogenesis. Medicine 70:99–117.

Gatti RA, Tward A, Concannon P. 1999. Cancer risk in ATM heterozygotes: a model of phenotypic and mechanistic differences between missense and truncating mutations. Mol Genet Metab 68:419–423.

Gatti RA, Becker-Catania S, Chun H, Sun X, Mitui M, Lai C-H, Khanlou N, Babaei M, Cheng R, Clark C, Huo Y, Udar N, Iyer R. 2001. The pathogenesis of ataxia-telangiectasia. Clin Rev Allergy Immunol 20:87–108.

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

Gilad S, Chessa L, Khosravi R, Russell P, Galanty Y, Piane M, Gatti RA, Jorgensen TJ, Shiloh Y, Bar-Shira A. 1998. Genotype-phenotype relationships in ataxia-telangiectasia and variants. Am J Hum Genet 62:551–561.

Howard MT, Shirts BH, Petros LM, Flanigan KM, Gesteland RF, Atkins JF. 2000. Sequence specificity of aminoglycoside-induced stop codon readthrough: potential implications for treatment of Duchene Muscular Dystrophy. Ann Neurol 48:164-169.

Huo YK, Wang Z, Hong JH, Chessa L, McBride WH, Perlman SL, Gatti RA. 1994. Radiosensitivity of ataxia-telangiectasia, X-linked agammaglobulinemia, and related syndromes using a modified colony survival assay. Cancer Res 54:2544–2577.

Lai C-H, Chun HH, Nahas S, Gatti RA. 2003. Aminoglycoside-induced ATM readthrough expression: a potential treatment for ataxia-telangiectasia. Amer J Hum Genet 73A:621.

Li A, Swift M. 2000. Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients. Am J Med Genet 92:170–177.

McConville CM, Stankovic T, Byrd PJ, McGuire GM, Yao Q-Y, Lennox GG, Taylor AMR. 1996. Mutations associated with variant phenotypes in ataxia-telangiectasia. Am J Hum Genet 59:320–330.

Mitui M, Campbell C, Coutinho G, Sun X, Lai CH, Thorstenson Y, Castellvi-Bel S, Fernandez L, Monros E, Tavares Costa Carvalho B, Porras O, Fontan G, Gatti RA. 2003. Independent mutational events are rare in the ATM gene: haplotype

FIGURE 4. **a: Type IV (intronic) splicing mutation IVS29–1G>A results in the deletion of the first nucleotide of exon 30, causing a frameshift. The mutant nucleotide (A) and the first nucleotide of exon 30 (G) are then used as the 3′ AG splice site. Adjacent sequences are shown at bottom right, with the deleted nucleotide underlined and the mutated site shaded. A chromatogram (reverse sequence) showing the genomic mutation (C>T) is at bottom left. b: Type IV (intronic) splicing mutation IVS11–2A>G results in the deletion of the first seven nucleotides of exon 12, causing a frameshift. The mutation disrupts the 3′ AG splice site; an adjacent AG seven nucleotides downstream, within the exon, is used as a new 3′ AG splice site. Adjacent sequences are shown at bottom right, with the deleted nucleotides underlined and the mutated site shaded. A chromatogram (reverse sequence) showing the genomic mutation is at bottom left. c: Type IV (intronic) splicing mutation IVS37-5delTCTA results in the deletion of the first seven nucleotides of exon 38. The 3′ AG sequence is disrupted by the TCTA deletion; another AG seven nucleotides downstream, within exon 38, is used as a new 3′ AG splice site. Adjacent sequences are shown at bottom right, with the deleted nucleotides underlined and the mutated site shaded. A chromatogram (forward sequence) showing the genomic mutation is at bottom left. d: Type IV (intronic) splicing mutation IVS21–2A>G results in the deletion of the first 32 nucleotides of exon 22. The mutation disrupts the 3′ AG splice site; an AG sequence 32 nucleotides downstream, within the exon, is used as a new 3′ AG splice site. Adjacent sequences are shown at bottom right, with the deleted nucleotides underlined and the mutated site shaded. A chromatogram (reverse sequence) showing the homozygous genomic mutation is at bottom left. e: Type IV (intronic) splicing mutation IVS38–2A>C results in the deletion of the first 61 nucleotides of exon 39. The mutation disrupts the 3′ AG splice site; the *second* AG within the exon, 61 nucleotides downstream, is used as a new 3′ AG splice site. Adjacent sequences are shown at bottom right, with the deleted sequences underlined and the mutated site shaded. An agarose gel showing the heterozygous cDNA deletion of 61 nucleotides is at bottom left. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

prescreening enhances mutation detection rate. Hum Mutat 22:43–50.

Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE. 2002. A new type of mutation causes a splicing defect in ATM. Nat Genet 30:426–429.

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

Rotman G, Vanagaite L, Collins RS, Shiloh Y. 1994. Three dinucleotide repeat polymorphisms at the ataxia-telangiectasia locus. Hum Mol Genet 3:2079.

Sandoval N, Platzer M, Rosenthal A, Dork T, Bendix R, Skawran B, Stuhrmann M, Wegner R-D, Sperling K, Banin S, Shiloh Y, Baumer A, Mernthaler U, Sennefelder H, Brohm M, Weber BHF, Schindler D. 1999. Characterization of ATM gene mutations in 66 ataxia telangiectasia families. Hum Mol Genet 8:69–79.

Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali SR, Simmons A, Clines GA, Sartiel A, Gatti RA, Chessa L, Sanal O, Lavin MF, Jaspers NGJ, Taylor MR, Arlett CF, Miki T, Weissman SM, Lovett M, Collins FS, Shiloh Y. 1995. A single ataxia-telangiectasia gene with a product similar to PI-3 kinase. Science 268:1749–1753.

Scott S, Bendix R, Chen P, Clark R, Dork T, Lavin MF. 2002. Missense mutations but not allelic variants alter the function of ATM by dominant interference in patients with breast cancer. Proc Natl Acad Sci USA 99:925–930.

Shapiro M, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes, sequence statistics and functional implications. Nucleic Acids Res 15:7155–7174.

Spring K, Ahangari F, Scott SP, Waring P, Purdie DM, Chen PC, Hourigan K, Ramsay J, McKinnon PJ, Swift M, Lavin MF. 2002. Mice heterozygous for mutation in ATM, the gene involved in ataxia-telangiectasia, have heightened susceptibility to cancer. Nat Genet 32:185–190.

Stankovic T, Kidd AMJ, Sutcliffe A, McGuire GM, Robinson P, Weber P, Bedenham T, Bradwell AR, Easton DF, Lennox GG, Haites N, Byrd PJ, Taylor AMR. 1998. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. Am J Hum Genet 62:334–345.

Sun X, Becker-Catania SG, Chun HH, Hwang MJ, Huo Y, Wang Z, Mitui M, Sanal O, Chessa L, Crandall B, Gatti RA. 2002. Early diagnosis of ataxia-telangiectasia using radiosensitivity testing. J Pediatr 140:724–731.

Telatar M, Wang Z, Udar N, Liang T, Bernatowska-Matuszkiewicz E, Lavin M, Shiloh Y, Concannon P, Good RA, Gatti RA. 1996. Ataxia-telangiectasia: mutations in ATM cDNA detected by protein-truncation screening. Am J Hum Genet 59:40–44.

Telatar M, Teraoka S, Wang Z, Chun HH, Liang T, Castellvi-Bel S, Udar N, Borresen-Dale A, Chessa L, Bernatowska-Matuszkiewicz E, Porras O, Watanabe M, Junker A, Concannon P, Gatti RA. 1998a. Ataxia-telangiectasia: identification and detection of founder-effect mutations in the ATM gene in ethnic populations. Am J Hum Genet 62:86–97.

Telatar M, Wang Z, Castellvi-Bel S, Tai L-Q, Sheikhavandi S, Regueiro JR, Porras O, Gatti RA. 1998b. A model for ATM heterozygote identification in a large population: four founder-effect ATM mutations identify most of Costa Rican patients with ataxia telangiectasia. Mol Genet Metab 64: 36–43.

Teraoka S, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chess L, Sanal O, Bernatowska E, Gatti RA, Concannon P. 1999. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet 64:1617–1631.

Thorstenson YR, Shen P, Tusher VG, Wayne TL, Davis RW, Chu G, Oefner PJ. 2001. Global analysis of ATM polymorphism reveals significant functional constraint. Am J Hum Genet 69:396–412.

Udar N, Farzad S, Tai L-Q, Bay J-O, Gatti RA. 1999. NS22: A highly polymorphic complex microsatellite marker within the ATM gene. Am J Med Genet 82:287–289.

Uhrhammer N, Lange E, Porras O, Naeim A, Chen X, Sheikhavandi S, Chiplunkar S, Yang L, Dandekar S, Liang T, Patel N, Teraoka S, Udar N, Calvo N, Concannon P, Lange K, Gatti RA. 1995. Sublocalization of an ataxia-telangiectasia gene distal to D11S384 by ancestral haplotyping in Costa Rican families. Am J Hum Genet 57:103–111.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogenography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62.

Uziel T, Savitsky K, Platzer M, Ziv Y, Helbitz T, Behls M, Boehm T, Rosenthal A, Shiloh Y, Rotman G. 1996. Genomic organization of the ATM gene. Genomics 33:317–320.

Vanagaite L, James MR, Rotman G, Savitsky K, Var-Shira A, Gilad S, Ziv Y, Uchenik V, Sartiel A, Collins FS, Sheffield VC, Richard CWIII, Weissenback J, Shiloh Y. 1995. A high-density microsatellite map of the ataxia telangiectasia locus. Hum Genet 95:451–454.

Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol (in press).