Journal of Pathology
J Pathol September 2022; **258:** 83–101
Published online 15 July 2022 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/path.5979

ORIGINAL ARTICLE

Minigene-based splicing analysis and ACMG/AMP-based tentative classification of 56 *ATM* variants

Elena Bueno-Martínez[1], Lara Sanoguera-Miralles[1], Alberto Valenzuela-Palomo[1], Ada Esteban-Sánchez[2], Víctor Lorca[2], Inés Llinares-Burguet[1], Jamie Allen[3], Alicia García-Álvarez[1], Pedro Pérez-Segura[2], Mercedes Durán[4], Douglas F Easton[3], Peter Devilee[5], Maaike PG Vreeswijk[5], Miguel de la Hoya[2†] and Eladio A Velasco-Sampedro[1†*]

[1] *Splicing and Genetic Susceptibility to Cancer, Unidad de Excelencia Instituto de Biología y Genética Molecular, Consejo Superior de Investigaciones Científicas (CSIC-UVa), Valladolid, Spain*
[2] *Molecular Oncology Laboratory CIBERONC, Hospital Clínico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain*
[3] *Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK*
[4] *Cancer Genetics, Instituto de Biología y Genética Molecular, Valladolid, Spain*
[5] *Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands*

**Correspondence to: EA Velasco-Sampedro, Grupo de Splicing y Cáncer, Instituto de Biología y Genética Molecular (IBGM), Consejo Superior de Investigaciones Científicas (CSIC)-UVa, Sanz y Forés 3, 47003 Valladolid, Spain. E-mail: eavelsam@ibgm.uva.es*

[†]These authors are joint senior authors.

Abstract

The ataxia telangiectasia-mutated (ATM) protein is a major coordinator of the DNA damage response pathway. *ATM* loss-of-function variants are associated with 2-fold increased breast cancer risk. We aimed at identifying and classifying spliceogenic *ATM* variants detected in subjects of the large-scale sequencing project BRIDGES. A total of 381 variants at the intron–exon boundaries were identified, 128 of which were predicted to be spliceogenic. After further filtering, we ended up selecting 56 variants for splicing analysis. Four functional minigenes (mgATM) spanning exons 4–9, 11–17, 25–29, and 49–52 were constructed in the splicing plasmid pSAD. Selected variants were genetically engineered into the four constructs and assayed in MCF-7/HeLa cells. Forty-eight variants (85.7%) impaired splicing, 32 of which did not show any trace of the full-length (FL) transcript. A total of 43 transcripts were identified where the most prevalent event was exon/multi-exon skipping. Twenty-seven transcripts were predicted to truncate the ATM protein. A tentative ACMG/AMP (American College of Medical Genetics and Genomics/Association for Molecular Pathology)-based classification scheme that integrates mgATM data allowed us to classify 29 *ATM* variants as pathogenic/likely pathogenic and seven variants as likely benign. Interestingly, the likely pathogenic variant c.1898+2T>G generated 13% of the minigene FL-transcript due to the use of a noncanonical GG-5'-splice-site (0.014% of human donor sites). Circumstantial evidence in three *ATM* variants (leakiness uncovered by our mgATM analysis together with clinical data) provides some support for a dosage-sensitive expression model in which variants producing ≥30% of FL-transcripts would be predicted benign, while variants producing ≤13% of FL-transcripts might be pathogenic.

Keywords: hereditary breast cancer; susceptibility genes; ATM; VUS; splicing; aberrant splicing; splicing assay; minigenes; variant classification

Received 24 March 2022; Revised 11 May 2022; Accepted 8 June 2022

No conflicts of interest were declared.

Introduction

The ataxia telangiectasia-mutated (*ATM*) gene (MIM#607585), located on chromosome 11q22–23, is composed of 62 coding exons and encodes a large serine/threonine kinase of 3,056 amino acids [1,2]. This protein plays an essential role in cellular homeostasis, being responsible for global orchestration of the cellular response to double-strand breaks. Biallelic germline mutations in *ATM* result in the autosomal recessive A-T (ataxia–telangiectasia) syndrome, characterized by neurodegeneration, progressive ataxia, immunodeficiency, ocular telangiectasia, regular respiratory infections, gonadal atrophy, and infertility [3], as well as increased cancer susceptibility, mostly lymphoid cancer [4,5].

Two recent large-scale studies of breast cancer (BC) patients have estimated that at least eight genes

are significantly associated with BC susceptibility, including *ATM* [6–8]. *ATM* is widely tested on commercial gene panels; heterozygous protein-truncating variants have been associated with a BC risk of around 2-fold [6,7], and a pancreatic cancer risk of 6.5-fold [9]. Possible increased risks of melanoma, stomach, and prostate cancers have also been reported [10].

Splicing is an essential and highly regulated RNA processing mechanism that is carried out by the spliceosome, an ensemble of ribonucleoproteins and other splicing factors that identify the cis-acting sequences needed for exon recognition, which include, among others, the basic donor or 5' splice-site (5'ss) and the acceptor or 3' splice-sites (3'ss) [11]. Historically, the role of splicing disruptions has been underestimated in genetic diseases, because often only variants in the "canonical" ± 1,2 positions have been considered potentially disease-causing. However, most potential spliceogenic variants are classified as variants of uncertain significance (VUS) because splicing outcomes cannot be accurately predicted. In this regard, RNA assays provide information that might become critical for accurate clinical classification. In principle, RNA analysis of variant carriers can be used to classify variants as likely spliceogenic but, unfortunately, these samples cannot always be collected, and analysis in patient RNA is hampered by the presence of the wildtype allele. Certainly, the latter can be somehow overcome by RNA-seq approaches that detect allele-specific expression and calculate percent splicing index [12]. Alternatively, minigene assays provide a valuable approach to perform functional analysis of variants [13,14].

The BRIDGES project (Breast Cancer Risk after Diagnostic Gene Sequencing; https://bridges-research.eu/) is an international initiative that has sequenced 34 known or suspected BC genes in more than 113,000 women. Previously, we performed comprehensive studies of BRIDGES' splice-site variants in *RAD51C*, *RAD51D*, and *PALB2* by the splicing reporter minigene technology [15–17]. Here we selected and functionally analyzed 56 potential spliceogenic variants in *ATM* identified in BRIDGES subjects, using four different *ATM* splicing reporter minigenes. Further, we integrated minigene data into an ACMG/AMP-based classification scheme that allows us to propose a tentative classification of all 56 tested variants.

Materials and methods

Ethical statement

Ethical approval for this study was obtained from the Ethics Committee of the Spanish National Research Council-CSIC (28/05/2018).

Annotation

All splicing events and predicted protein products were described according to the Human Genome Variation Society (HGVS) guidelines, using the Ensembl reference transcript ID ENST00000278616.8 (Genbank NM_000051.4). For clarity, we also used abbreviated notations using any of the following symbols [18,19]: ▼ (incorporation of intronic sequences not present in the reference transcript), Δ (deletion of exonic sequences present in the reference transcript), E (exon), p (alternative 3' splice-site, new acceptor site), q (alternative 5' splice-site, new donor site), and a number representing the exact number of nucleotides incorporated or skipped. For example, transcript ▼(E8q5) denotes the use of an alternative donor site 5 nucleotides downstream of exon 8, causing the incorporation of 5-nt into the mature mRNA.

Selection of candidate *ATM* variants

A total of 381 unique variants at the *ATM* exon/intron boundaries were identified in the BRIDGES consortium sequencing data [6]. *In silico* splicing predictions were performed in all 381 variants using MaxEntScan (MES) [20] (supplementary material, Table S1). We selected likely spliceogenic variants based on: (i) ≥20% decrease of MES scores [21,22], (ii) creation of putative *de novo* sites (MES cutoff ≥3.0), or (iii) changes at conserved positions (−3, −2, −1, exon 5'-3'-ends, +3,-+4, +5, and +6) of the consensus splice-site, regardless of MES predictions [17]. The latter included eight variants with scores above the −20% threshold (c.3994-3C>T, c.902G>A, c.3577G>C, c.3746+4A>C, c.4436+4A>G, c.3993+5G>T, c.4109+6T>G, and c.7788+6T>G).

Based on these criteria, we ended up selecting 136 likely spliceogenic variants spread all over the gene (supplementary material, Table S1). Since cloning all 63 *ATM* exons into minigenes was not feasible, we focused our attention on four exon clusters (4–9, 11–22, 25–29, and 49–52) in which a substantial proportion of candidate variants (61%) occur. After discarding candidate variants located in exons 17 to 22 (the dedicated minigene did not perform well, see Results), and filtering-out several candidate variants located at the same splice-site positions with similar MES impact (e.g. c.1898+3A>G and c.1898+3A>T), we ended up with a list of 56 variants to be tested in minigenes (Table 1 and supplementary material, Table S1).

Minigene construction and site-directed mutagenesis

Given that RNA from BRIDGES carriers had not been collected, we envisioned a minigene-based strategy similar to that we adopted in other BC susceptibility genes [15–17]. Minigenes mgATM_ex4–9, mgATM_ex11–17, mgATM_ex17–22, mgATM_ex25–29, and mgATM_ex49–52 were designed to include *ATM* exons 4 to 9, 11 to 17, 17 to 22, 25 to 29, and 49 to 52, respectively, and 200 nucleotides of flanking intronic sequences upstream and downstream from each exon (supplementary material, Figure S1). Subsequently, each insert was

Table 1. Bioinformatics analysis and splicing outcomes of *ATM* variants.

Variant (HGVS)[†]	MaxEntScan[‡]	Minigene FL-transcript	PTC-transcripts	In-frame transcripts	Unknown transcripts
mgATM_ex4-9					
Wild type		(65.7 ± 0.7%)	Δ(E7) (34.2 ± 0.7%)		
c.332-5A>G	[−] 3′ss (7.2→1.6)	(40.0 ± 8.4%)	Δ(E7) (36.7 ± 5.7%) ▼(E5p4)Δ(E7) (16.9 ± 2.2%) Δ[(E5)(E7)] (6.4 ± 0.3%)		
c.332-1G>A	[−] 3′ss (7.2→1.6)	–	Δ[(E5)(E7)] (36.6 ± 2.0%) Δ(E5p1) (19.8 ± 1.2%)	Δ(E5) (43.6 ± 0.8%)	
c.496G>A (p.Glu166Lys)	[↓] 5′ss (7.8→4.6)	(71.8 ± 1.0%)	Δ(E7) (28.2 ± 1.0%)		
c.496+5G>A	[↓] 5′ss (7.8→4.9)	(25.2 ± 0.3%)	Δ(E7) (5.7 ± 0.1%) Δ[(E5)(E7)] (33.0 ± 0.5%)	Δ(E5) (36.1 ± 0.4%)	
c.901G>T (p.Gly301Cys)	[−] 5′ss (7.1→−3.6)	–	Δ(E7) (91.0 ± 0.2%)		970-nt (9.0 ± 0.2%)
c.901+2T>C	[−] 5′ss (7.1→−.7)	–	Δ(E7) (100.0 ± 0%)		
c.901+3A>T	[−] 5′ss (7.08→1.88)	–	Δ(E7) (90.5 ± 1.1%)		970-nt (9.5 ± 1.1%)
c.902-1G>T	[−] 3′ss (7.6→−1.0)	–	Δ(E8) (37.4 ± 0.2%) Δ(E7_8) (62.6 ± 0.2%)		
c.902G>A (p.Gly301Asp)	[↓] 3′ss (7.6→6.9)	(64.8 ± 0.9%)	Δ(E7) (22.6 ± 0.5%) Δ(E7_8) (12.6 ± 1.0%)		
c.903T>G (p.Gly301=)	[↓] 3′ss (7.6→6)	(61.5 ± 0.3%)	Δ(E7) (21.3 ± 0.3%) Δ(E7_8) (17.2 ± 0.1%)		
c.1065+1G>T	[−] 5′ss (8.7→0.2)	–	▼(E8q5) (54.8 ± 0.6%) [Δ(E7)▼(E8q5)] (45.2 ± 0.6%)		
c.1065+3A>G	[−] 5′ss (8.7→5.0)	(25.3 ± 0.6%)	Δ(E7) (5.7 ± 0.1%) [Δ(E7)▼(E8q5)] (8.9 ± 0.5%) Δ(E7_8) (28.8 ± 0.8%) Δ(E8) (6.1 ± 0.4%) ▼(E8q5) (25.2 ± 1.3%)		
c.1066-6T>G	[↓] 3′ss (10.8→8.3)	(26.7 ± 1.2%)	Δ(E9) (44.8 ± 0.9%) Δ[(E7)(E9)] (14.9 ± 0.4%) Δ(E7) (3.3 ± 0.1%)	Δ(E7_9) (10.3 ± 0.2%)	
c.1235+4_1235+5del	[−] 5′ss (4.1→−1.3)	(45.1 ± 0.4%)	Δ(E9) (19.4 ± 0.2%) Δ[(E7)(E9)] (11.3 ± 0.1%) Δ(E7) (12.8 ± 0.2%)	Δ(E7_9) (11.4 ± 0.2%)	
mgATM _ex11-17					
Wild type		(84.1 ± 0.6%)		Δ(E11) (15.9 ± 0.6%)	
c.1898G>T (p.Cys633Phe)	[↓] 5′ss (8.4→4.3)	(62.6 ± 2.9%)		Δ(E12) (37.4 ± 2.9%)	
c.1898+2T>G	[−] 5′ss (8.4→0.7)	(13.0 ± 1.3%)*		Δ(E12) (87.0 ± 1.3%)	
c.1898+3A>T	[↓] 5′ss (8.4→4.9)	(100.0 ± 0.0%)			
c.1898+3_1898+4del	[−] 5′ss (8.4→−5.1)	–		Δ(E12) (100.0 ± 0.0%)	
c.2251-1G>C	[−] 3′ss (6.6→−1.4) [+] 5′ss (4.5)	–	Δ(E15p19) (57.4 ± 0.4%)	Δ(E15) (11.7 ± 0.7%) Δ[(E11)(E15)] (18.2 ± 1.1%) Δ(E15_16) (12.7 ± 0.8%)	
c.2376+1G>A	[−] 5′ss (10.6→2.4)	–		Δ(E15) (81.3 ± 0.6%)	

(Continues)

Table 1. Continued

Variant (HGVS)[†]	MaxEntScan[‡]	Minigene FL-transcript	PTC-transcripts	In-frame transcripts	Unknown transcripts
				Δ[(E11)(E15)] (18.7 ± 0.6%)	
c.2376+3A>T	[↓] 5′ss (10.6→6.1)	–		Δ(E15) (80.0 ± 0.0%)	
				Δ[(E11)(E15)] (20.0 ± 0.0%)	
c.2377-6T>A	[↓] 5′ss (8.1→6.1)	(83.3 ± 0.2%)		Δ(E11) (16.7 ± 0.2%)	
c.2377-2A>G	[–] 3′ss (8.1→0.1)	–		Δ(E16) (45.2 ± 0.4%)	
				Δ(E16p3) (38.7 ± 0.6%)	
				Δ(E11)Δ(E16) (16.1 ± 0.2%)	
c.2467-3A>G	[–] 3′ss (8.8→0.9)	(100.0 ± 0.0%)			
c.2638+3A>G	[↓] 5′ss (6.6→3.8)	(100.0 ± 0.0%)			
mgATM _ex25_29					
Wild type		(100.0 ± 0.0%)			
c.3577-1G>A	[–] 3′ss (8.5→–0.2) [+] 3′ss (3.6)	–	Δ(E25) (100 ± 0.0%)		
c.3577G>C (p.Val1931Leu)	[↓] 3′ss (8.5→7.4)	(37.3 ± 0.5%)	Δ(E25) (47.1 ± 0.8%)	Δ(E25p159) (7.9 ± 0.1%)	
				Δ(E25_26) (7.7 ± 1.0%)	
c.3746+1G>A	[–] 5′ss (9.7→1.5)	–	Δ(E25) (100 ± 0.0%)		
c.3746+4A>C	[↓] 5′ss (9.7→8.2)	(71.4 ± 1%)	Δ(E25) (28.6 ± 1%)		
c.3746+5G>A	[↓] 5′ss (9.7→6.6)	–	Δ(E25) (100 ± 0.0%)		
c.3993G>A (p.Gln1331=)	[–] 5′ss (10.0→5.0)	–	Δ(E26) (27.1 ± 0.5%)	Δ(E26q120) (72.9 ± 0.5%)	
c.3993+1G>A	[–] 5′ss (10.0→1.8)	–	Δ(E26) (32.7 ± 0.7%)	Δ(E26q120) (67.3 ± 0.7%)	
c.3993+5G>T	[↓] 5′ss (10.0→8.4)	(76.2 ± 0.8%)		Δ(E26q120) (23.8 ± 0.8%)	
c.3994-3C>T	[↓] 3′ss (11.4→9.9)	(100 ± 0.0%)			
c.3994-2A>G	[–] 3′ss (11.4→3.4)	–	Δ(E27) (100 ± 0.0%)		
c.4109+1G>T	[–] 5′ss (8.3→–0.2)	–	Δ(E27) (83.4 ± 0.4%)		
			Δ(E27q1) (16.6 ± 0.4%)		
c.4109+3A>G	[↓] 5′ss (8.3→4.5)	(45.7 ± 0.9%)	Δ(E27) (54.3 ± 0.9%)		
c.4109+5G>A	[–] 5′ss (8.3→3.9)	–	Δ(E27) (100.0 ± 0.0%)		
c.4109+6T>G	[↓] 5′ss (8.3→7.0)	(68.5 ± 0.8%)	Δ(E27) (31.5 ± 0.8%)		
c.4110-9C>G	[–] 3′ss (5.6→–0.1) [+] 3′ss (5.6)	–	Δ(E28) (15.2 ± 2.5%)		
			▼(E28p8) (84.8 ± 2.5%)		
c.4110-2A>C	[–] 3′ss (5.6→–2.5)	–	Δ(E28) (84.2 ± 0.1%)		
			Δ(E28p53) (15.8 ± 0.1%)		
c.4236+1G>A	[–] 5′ss (7.5→–0.7)	–	Δ(E28) (100 ± 0.0%)		
c.4236+5G>A	[–] 5′ss (7.5→–0.6)	–	Δ(E28) (100 ± 0.0%)		
c.4236+6T>C	[–] 5′ss (7.5→3.7)	–	Δ(E28) (100 ± 0.0%)		
c.4436+4A>G	[↓] 5′ss (8.9→7.2)	(72.8 ± 3.1%)	Δ(E29) (21.1 ± 0.3%)	Δ(E28_29) (6.1 ± 3.3%)	
mgATM _ex49-52					
Wild type		(75.8 ± 0.04%)		Δ(E52) (24.2 ±0.04%)	
c.7307+1G>A	[–] 5′ss (8.6→0.4)	–	Δ(E49q38) (62.5 ± 1.0%)		
			[Δ(E49q38) Δ(E52)] (17.7 ± 0.6%)		
			Δ(E49) (8.7 ± 0.1%)		
			[Δ(E49) Δ(E52)] (8.1 ± 0.2%)		
c.7307+4A>G	[↓] 5′ss (8.6→6.5)	–	Δ(E49q38) (86.5 ± 0.2%)		

(Continues)

Table 1. Continued

Variant (HGVS)†	MaxEntScan‡	Minigene FL-transcript	PTC-transcripts	In-frame transcripts	Unknown transcripts
			[Δ(E49q38) Δ(E52)] (13.5 ± 0.2%)		
c.7515G>A (p.Lys2505=)	[↓] 5'ss (3.2→2.0)	–	Δ(E50) (65.6 ± 1.1%)		
			[Δ(E50) Δ(E52)] (34.4 ± 1.1%)		
c.7515+6T>C	[↓] 5'ss (3.2→2.5)	(48.4 ± 2.1%)	Δ(E50) (34.3 ± 0.8%)		
			[Δ(E50) Δ(E52)] (17.3 ± 1.3%)		
c.7629+2T>G	[–] 5'ss (8.6→1.0)	–		Δ(E51) (100 ± 0%)	
c.7630-3C>T	[↓] 3'ss (7.0→5.5)	(22.1 ± 0.2%)		Δ(E52) (77.9 ± 0.2%)	
c.7630-2A>C	[–] 3'ss (7.0→−1.1)	–	Δ(E52p11) (33.6 ± 1.3%)	Δ(E52) (66.4 ± 1.3%)	
	[+] 3'ss (5.4)				
c.7787A>T (p.Glu2596Val)	[↓] 5'ss (7,6→3,1)	–		Δ(E52) (100 ± 0%)	
c.7788G>A (p.Glu2596=)	[–] 5'ss (7.6→0,9)	–		Δ(E52) (100 ± 0%)	
c.7788+1G>C	[–] 5'ss (7.6→0.9)	–		Δ(E52) (100 ± 0%)	
c.7788+6T>G	[–] 5'ss (7.6→−0.6)	(50.9 ± 4.2%)		Δ(E52) (49.1 ± 4.2%)	

*Use of a noncanonical GG 5'ss [29].
†Variants without any trace of the full-length transcripts are underlined.
‡[–] site disruption; [+] New site; [↓] Decrease of the splice-site score.

synthesized (Genewiz, South Plainfield, Waltham, MA, USA) and subcloned into the splicing plasmid pSAD (Patent P201231427-CSIC) (Figure 1; supplementary material, Figure S1, Supplementary materials and methods) [23,24]. The mgATM_ex4–9 minigene was obtained by inserting exon 9 into mgATM_ex4_8 minigene using HindIII/SalI restrictions enzymes. The final minigenes were confirmed by sequencing (Macrogen, Madrid, Spain) and functionally checked (i.e. expressing the expected transcripts) in MCF-7 cells. All DNA variants were introduced into the wildtype minigenes by site-directed mutagenesis using the QuikChange Lightning kit (Agilent, Santa Clara, CA, USA) (supplementary material, Table S2). All mutant constructs were confirmed by sequencing (Macrogen).

Transfection

Approximately 2 $\times 10^5$ MCF-7 cells (human breast adenocarcinoma cell line) were grown to 90% confluence in 4-well plates (Nunc, Roskilde, Denmark) in 0.5 ml of medium (MEME, 10% fetal bovine serum, 2 mM glutamine, 1% nonessential amino acids, and 1% penicillin/streptomycin stock solution). The reproducibility of the minigene outcomes was tested in MDA-MB-231 (triple-negative BC cell line) cells that were transfected with the wildtype and mutant minigenes with variants c.901+2T>C, c.2377–2A>G, c.3746+5G>A, and c.7629+2T>G. Cells were transiently transfected with 1 μg of each minigene and 2 μl of lipofectamine LTX (Life Technologies, Carlsbad, CA, USA). Nonsense mediated decay (NMD) was inhibited by incubating cells with cycloheximide 300 μg/ml (Sigma-Aldrich, St. Louis, MO, USA) for 4 h. RNA was purified using the Genematrix Universal RNA Purification Kit (EURx, Gdansk, Poland) with on-column DNAse I digestion following the manufacturer's instructions.

Reverse transcription polymerase chain reaction (RT-PCR) and cDNA amplification

The specific minigene-exon V2 primer RTPSPL3-RV (5'-TGAGGAGTGAATTGGTCGAA-3') was used to carry out a reverse transcription using 400 ng of RNA with the RevertAid First-Strand cDNA Synthesis Kit (Life Technologies). Two microliters of the resultant cDNA were used for amplification of the regions of interest using Platinum Taq DNA polymerase (Life Technologies). For all variants, the amplification was performed using the primers SD6-PSPL3_RTFW (5'-TCACCTGGACAACCTCAAAG-3') and RTpSAD-RV (Patent P201231427, CSIC). Samples were denatured at 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 59 °C for 30 s, and 72 °C (1 min/kb), and a final extension step at 72 °C for 5 min. RT-PCR products were sequenced by Macrogen. The expected minigene full-length (mgFL) transcripts were the following: mgATM_ex4–9 ($mgFL^{4-9}$: 1231 nt); mgATM_ex11–17 ($mgFL^{11-17}$: 1212 nt); mgATM_ex17–22 ($mgFL^{17-22}$: 999 nt); mgATM_ex25–29 ($mgFL^{25-29}$: 1041 nt); mgATM_ex49–52 ($mgFL^{49-52}$: 880 nt).

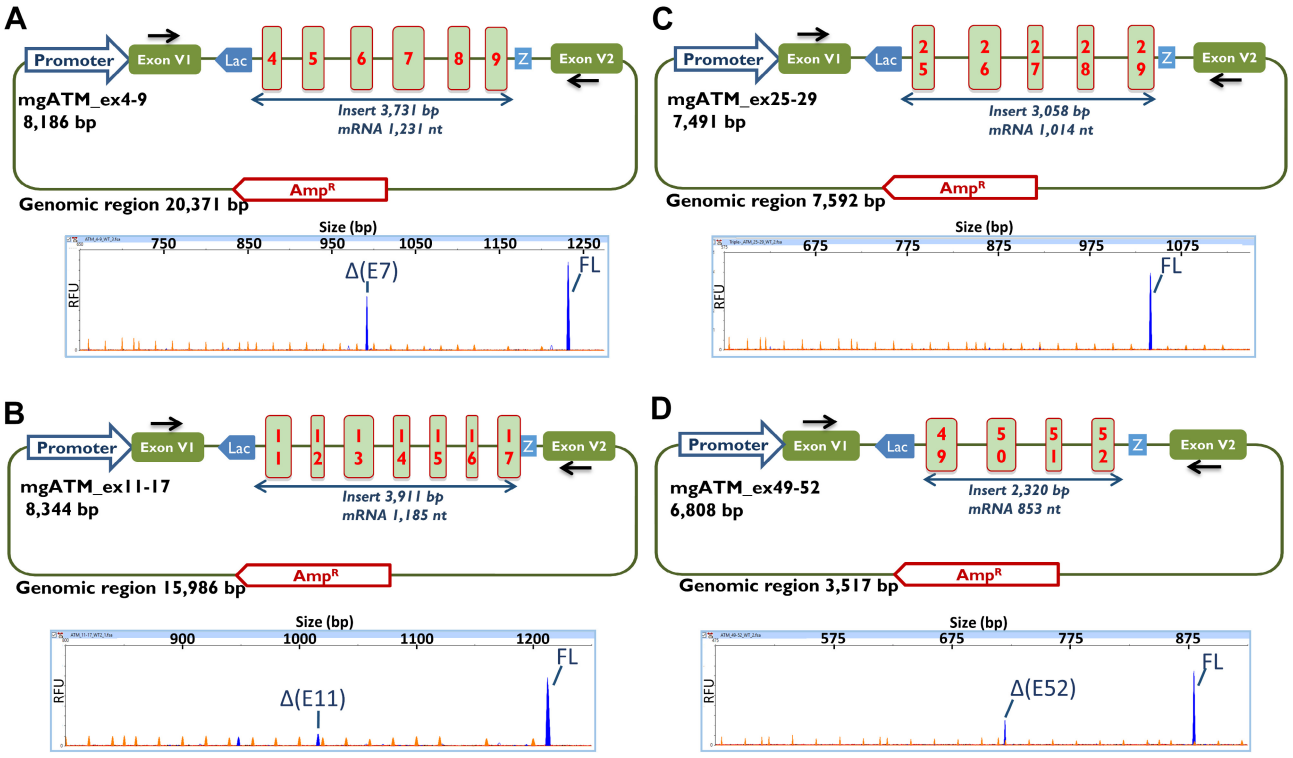

Figure 1. Structure and functional validation of the WT *ATM* minigenes used in this work. Schematic representation of the *ATM* minigenes with (A) exons 4 to 9 (mgATM_ex4–9), (B) 11 to 17 (mgATM_ex11–17), (C) 25 to 29 (mgATM_ex25–29), (D) 49 to 52 (mgATM_ex49–52). Exons are boxed; black arrows locate specific vector RT-PCR primers. Functional assays of the WT minigene are shown below. Fluorescent RT-PCR products were analyzed by capillary electrophoresis, where the full-length and alternative transcripts are shown as blue peaks and the Liz1200 size standard is shown as orange/faint peaks. The x-axis indicates size in bp (electropherograms on the top) and the y-axis represents relative fluorescence units (RFU).

In order to assess the relative contribution of each transcript to the overall mgATM expression, semiquantitative fluorescent RT-PCRs (26 amplification cycles) were performed in triplicate (in the case of c.1898+2T>G, experiments were replicated six times) using Platinum Taq DNA polymerase (Life Technologies) and the primers PSPL3_RTFW and RTpSAD-RV (both FAM-labeled) under standard conditions [24]. FAM-labeled products were run with a LIZ-1200 Size Standard at the Macrogen facility (Seoul, Korea) and analyzed using the Peak Scanner software V1.0 (Life Technologies). Only peak heights ≥50 RFU (relative fluorescence units) were considered. The protocol is summarized in the supplementary material, Figure S2.

ACMG/AMP-based tentative classification of *ATM* genetic variants

We classified 56 *ATM* genetic variants according to a recently proposed ACMG/AMP point system, a Bayesian framework that outperforms the original classification guidelines, and allows for increased flexibility and accuracy in combining different ACMG/AMP criteria and strengths of evidence [25,26]. In this framework, point-based variant classification categories are defined as follows: Pathogenic **(P)** ≥ +10; Likely Pathogenic **(LP)** + 6 to +9; Variant of Uncertain Significance **(VUS)** 0 to +5; Likely Benign **(LB)** −1 to −6; and Benign **(B)** ≤ −7.

To assign ACMG/AMP scores [27] to individual variants, we based our analysis primarily on recently released (19 January 2022) *ATM* specifications defined by the ClinGen Hereditary Breast, Ovarian and Pancreatic Cancer Variant Curation Expert Panel (clinicalgenome.org/affiliation/50039/). For some specific variants, we also used *ATM* specifications elaborated by the Spanish ATM Cancer Susceptibility Variant Interpretation Working Group [28]. Finally, we introduced some *ad-hoc* rules, in particular to incorporate mgATM complex readouts (≥2 transcripts) into the classification scheme as PVS1_O/BP7_O codes of variable strength depending on the actual outcome. As a result, we do not intend to provide an ACMG/AMP or ClinGen endorsed final classification of any *ATM* variant ready to be used in the clinical setting, but rather to highlight the complexity of incorporating complex minigene readouts into an ACMG/AMP-based classification scheme. A comprehensive description of the classification scheme is provided in Supplementary materials and methods, and supplementary material, Table S3.1–S3.3, and Figure S3A–C. For comparative purposes only, we performed an alternative classification incorporating predictive splicing codes PVS1/PP3/BP4 rather than experimental splicing codes PVS1_O/BP7_O (see supplementary material, Table S3.4).

Results

A total of 381 unique variants at the *ATM* exon/intron boundaries were identified in the BRIDGES cohort. After filtering, we selected for minigene analysis up to 56 likely pathogenic variants clustering in a subset of *ATM* exons (exon 4–9, 11–17, 25–29, and 49–52) (see Materials and methods for further details).

ATM minigenes

We constructed five *ATM* minigenes (mgATM_ex4_9, mgATM_ex11_17, mgATM_17-22, mgATM_ex25_29, and mgATM_ex49_52) that we tested in MCF-7 cells. Four minigenes mimic the reference transcript NM_000051.3, producing as main outcomes the expected FL-transcripts: V1-*ATM* exons 4 to 9-V2, 1,231-nt (65.7%); V1-*ATM* exons 11 to 17-V2, 1,212 nt (84.1%);

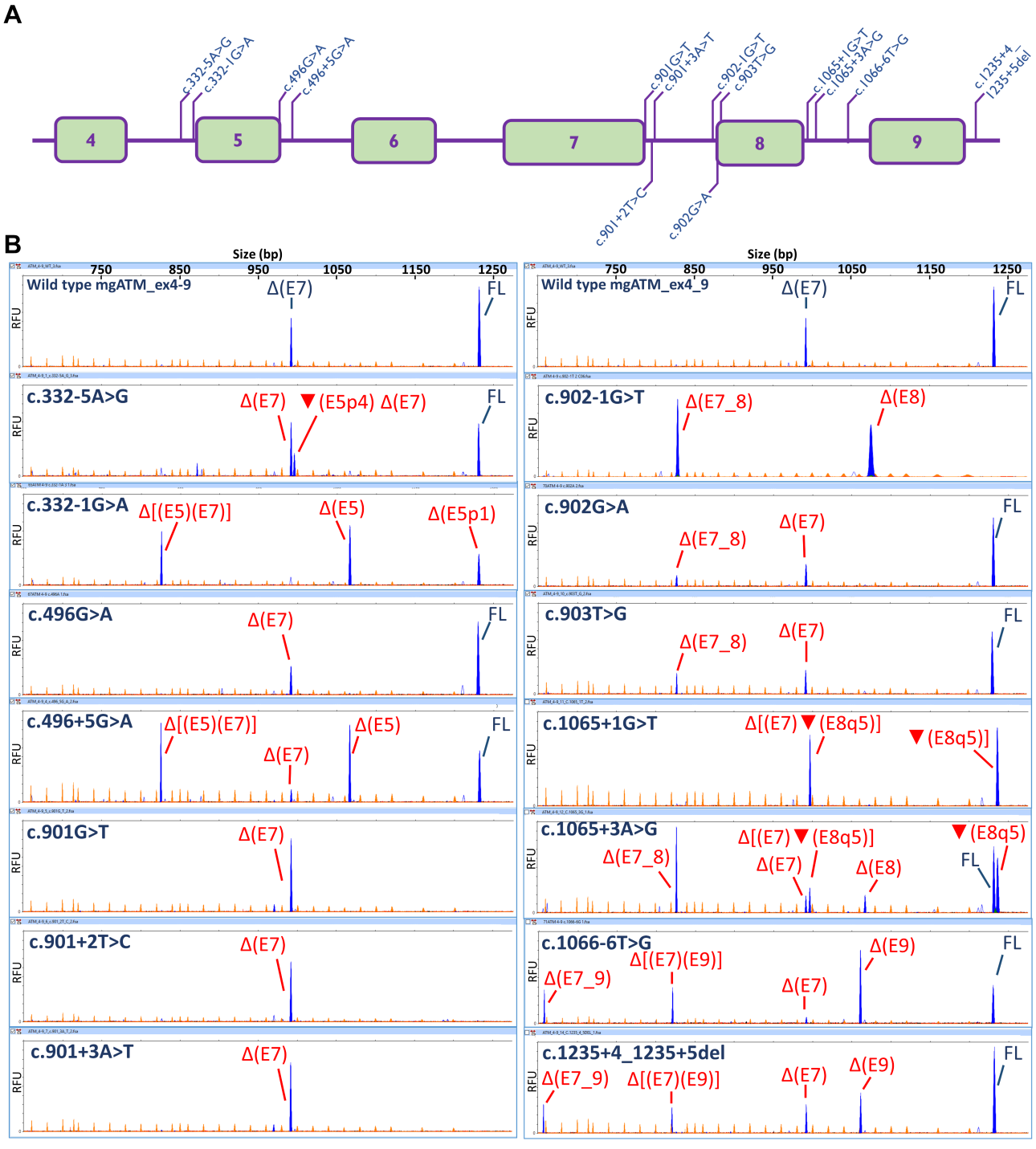

Figure 2. Splicing functional assays of selected *ATM* variants in mgATM_ex4–9 minigene. (A) Location of tested variants. (B) Fluorescent fragment analysis of transcripts generated by the wildtype and mutant minigenes. FAM-labeled products (blue peaks) were run with LIZ1200 (orange peaks) as size standard. FL, full-length transcript. The x-axis indicates size in bp (electropherograms on the top) and the y-axis represents relative fluorescence units (RFU).

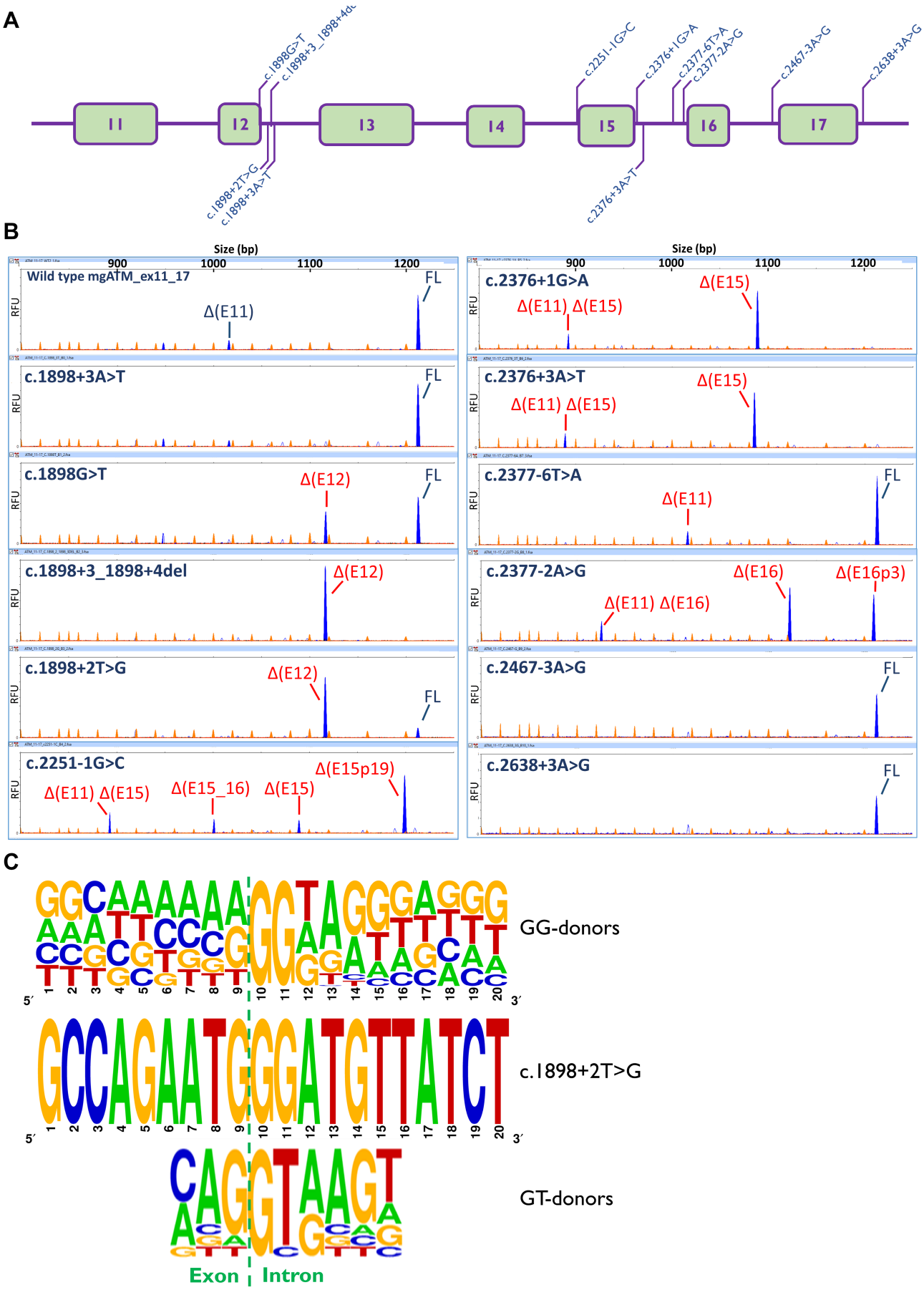

Figure 3 Legend on next page.

V1-*ATM* exons 25 to 29-V2, 1,041-nt (100%); V1-*ATM* exons 49 to 52-V2, 880-nt (75.8%). Likewise, the alternative isoforms Δ(E7) (34.2%; mgATM_ex4_9), Δ(E11) (15,9%; mgATM_ex11_17), Δ(E52) (24.2%; mgATM_ex49_52), and other uncharacterized transcripts were also detected (Figure 1 B,D,F,H, and Table 1). On the other hand, we discarded for variant testing minigenes mgATM_17–22 and mgATM_ex49–54 (insertion of exons 53–54 into mgATM_ex49–52), because they did not produce clean splicing profiles (see Supplementary materials and methods).

Splicing assays of *ATM* variants

Fifty-six variants were genetically engineered into the four minigenes: 14 in mgATM_ex4_9, 11 in mgATM_ex11–17, 20 in mgATM_ex25_29, and 11 in mgATM_ex49–52. For the purpose of the present analysis, splicing was considered "impaired" if the proportion of the corresponding mgFL-transcript was at least 10% lower than in the WT construct. After RNA isolation, a semiquantitative cDNA-amplification was performed to analyze the impact of each variant. Forty-eight out of 56 (86%) variants disrupted splicing (Table 1; Figures 2–5). –5). Twenty variants affected the ±1,2 positions and 28 targeted other splice-site positions: the polypyrimidine tract (three variants), −3 (one variant), +3 (four variants), +4 (four variants), +5 (five variants), +6 (four variants), as well as the first (one variant), and the two last exonic nt (six variants).

Up to 32 spliceogenic variants (underlined in Table 1) demonstrated strong impact on splicing (i.e. mgFL-transcripts not detected, or representing <5% of the overall signal), including one variant predicted missense [c.7787A>T (p.Glu2596Val)], and three variants predicted synonymous [c.3993G>A (p.Gln1331=), c.7515G>A (p.Lys2505=) and c.7788G>A (p.Glu2596=)] that did not produce any trace of the mgFL-transcript. The remaining 16 spliceogenic variants demonstrated weak to moderate splicing impacts, producing a nonnegligible proportion of mgFL-transcripts (13–71.4% of the overall signal). Curiously, four out of eight nonspliceogenic variants (c.2377-6T>A, c.2467-3A>G, c.2638+3A>G, and 3994-3C>T) improved the inclusion efficiency of the corresponding exons (i.e. the proportion of mgFL-transcripts were increased relative to their wildtype counterpart). Unexpectedly, variant c.1898+2T>G produced mgFL-transcripts (up to 13%; average of six replicas) that might be explained by the use of the atypical GG-5'ss (0.01% of human exons) [29] created by this variant (Figure 3C). Finally, to check splicing reproducibility, one variant of each minigene (c.901+2T>C, c.2377-2A>G, c.3746+5G>A, and c.7629+2T>G) was tested in MDA-MB-231 cells, showing identical outcomes (supplementary material, Figure S4).

Transcript analysis

Fluorescent-fragment analysis of minigene readouts allowed us to characterize the mgFL-transcripts produced by the four WT minigenes, and up to 43 other transcripts (Table 1 and supplementary material, Table S4). The latter includes three alternative splicing isoforms, Δ(E7), Δ(E11), and Δ(E52), produced by the corresponding WT minigenes. Twenty-seven transcripts, including Δ(E7), introduced a premature termination codon (PTC), while 15, including Δ(E11) and Δ(E52), kept the reading frame (Table 1, supplementary material, Table S4). One transcript of 970 nucleotides could not be characterized.

It is important to highlight the distinction between variant-induced transcripts (i.e. transcripts not produced by WT minigenes) and variant-induced splicing events (i.e. splicing events not detected in WT minigenes). For instance, the *ATM* variant c.332-1G>A (targeting exon five acceptor site) produces up to three variant-induced transcripts [Δ(E5) + [Δ(E5),Δ(E7)] + Δ(E5p1)], but only two variant-induced splicing events [Δ(E5) and Δ(E5p1)]. The Δ[(E5)(E7)] transcript combines variant-induced exon five skipping with exon seven skipping, a splicing event already observed in WT minigenes (see Figure 2B, and supplementary material, Tables S3 and S4).

A significant proportion of variants ($N = 32$) induced two or more splicing events, and/or demonstrated a partial effect on splicing, producing a nonnegligible amount of mgFL-transcripts (leaky variants). These complex readouts represented a challenge for variant interpretation (see below). Exon (or multi-exon) skipping, observed in 48 variants, was the most frequent variant-induced event. Alternative site-usage was observed in 15 variants. Five leaky variants produced FL-transcripts that harbor missense (r.496G>A, r.902G>A, r.1898G>U, r.3557G>C) or synonymous (r.903U>G) changes (Figures 2–4, supplementary material, Table S3.2).

ACMG/AMP-based tentative classification of 56 *ATM* variants

Once mgATM data were available, we decided to classify all 56 ATM variants according to ACMG/AMP

Figure 3. Splicing functional assays of selected *ATM* variants in mgATM_ex11–17 minigene. (A) Location of tested variants. (B) Fluorescent fragment analysis of transcripts generated by the wildtype and mutant minigenes. FAM-labeled products (blue peaks) were run with LIZ1200 (orange peaks) as size standard. FL, full-length transcript. The x-axis indicates size in bp (electropherograms on the top) and the y-axis represents relative fluorescence units (RFU). (C) Consensus sequence of exon–intron boundaries of 101 noncanonical human GG-splice junctions [29] (top panel) versus the sequence of the atypical GG-splicing donor used in 13% of transcripts induced by variant c.1898+2T>G (middle panel) and the consensus sequence of canonical GT-donors (bottom panel). The size of each letter represents the nucleotide frequency at each position. Pictograms were obtained using WebLogo (https://weblogo.berkeley.edu/logo.cgi).

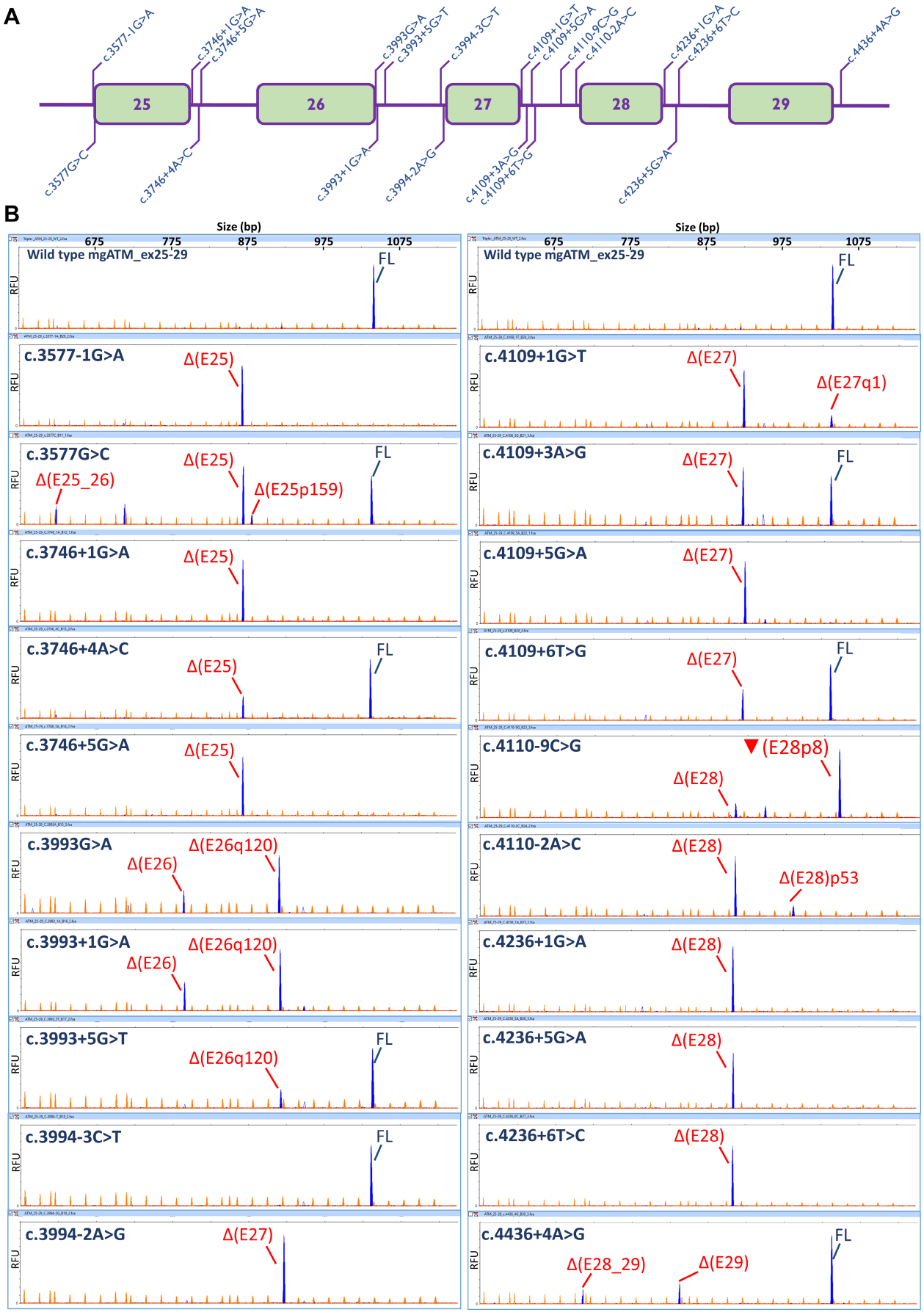

Figure 4 Legend on next page.

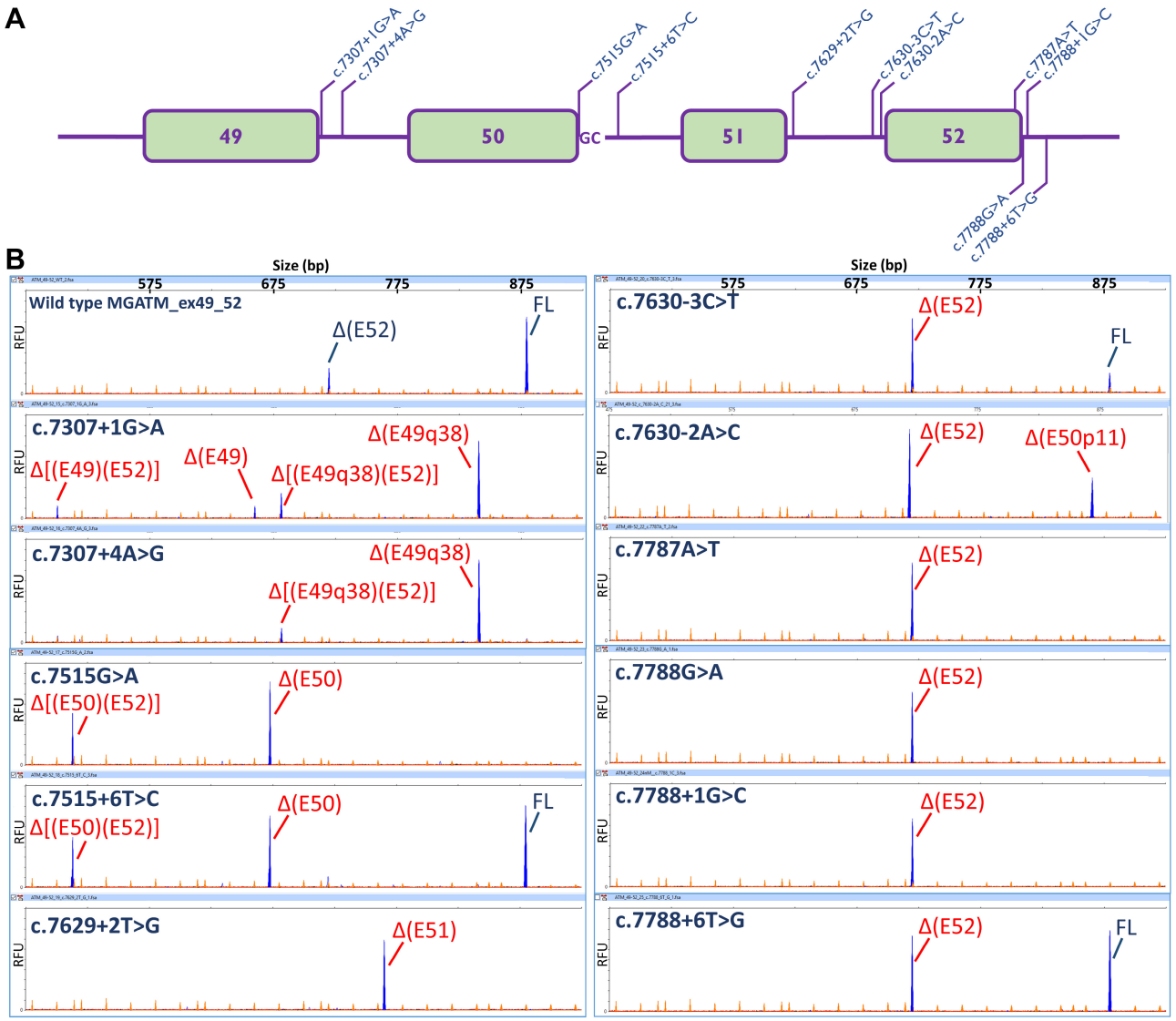

Figure 5. Splicing functional assays of selected *ATM* variants in mgATM_ex49–52 minigene. (A) Location of tested variants. (B) Fluorescent fragment analysis of transcripts generated by the wildtype and mutant minigenes. FAM-labeled products (blue peaks) were run with LIZ1200 (orange peaks) as size standard. FL, full-length transcript. The x-axis indicates size in bp (electropherograms on the top) and the y-axis represents relative fluorescence units (RFU).

variant classification guidelines [27], integrating mgATM data as PVS1_O/BP7_O evidence codes. We classified 29 variants as **P/LP** (six of them as pathogenic) and seven non-GT-AG intronic variants as **LB**. Up to 20 variants (36%) were classified as **VUS** (Table 2 and supplementary material, Table S3.1).

Overall, 37 of the 56 *ATM* variants here analyzed have been reported previously in ClinVar (last accessed 09/02/2022, see supplementary material, Table S3.4), but only 21 of them by multiple submitters with no conflicts (two-star review status). Focusing our analysis on the subgroup of 21 ClinVar no conflicting-variants, we conclude that our classification scheme (integrating mgATM data) does not reduce the number of **VUSs**, but rather reclassifies variants (seven variants, 33%) in both directions. Specifically, three variants reported in ClinVar as **VUSs** are upgraded to **P/LP**, while three variants reported as **P/LP** are downgraded to **VUSs**, and one variant reported as **LB/B** is upgraded to **VUS**. Supplementary material, Table S5 shows a comparative analysis in this subgroup of variants.

To evaluate the contribution of mgATM data to variant classification, we compared our final classification (Table 2, supplementary material, Table S3.1) with an

Figure 4. Splicing functional assays of selected *ATM* variants in mgATM_ex25–29 minigene. (A) Location of tested variants. (B) Fluorescent fragment analysis of transcripts generated by the wildtype and mutant minigenes. FAM-labeled products (blue peaks) were run with LIZ1200 (orange peaks) as size standard. FL, full-length transcript. The x-axis indicates size in bp (electropherograms on the top) and the y-axis represents relative fluorescence units (RFU).

Table 2. ACMG/AMP-based tentative classification of 56 *ATM* variants.

c.HGVS*	p.HGVS*	ClinVar†	Point-based classification‡		PVS1_O§	PM2¶	PM3¶¶	BP7_O§	BS1¶¶
c.332-5A>G		Not reported	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.332-1G>A		LP (2)	Pathogenic	+11 (+8 + 1 + 2)	PVS1_O(+8)	PM2_P (+1)	PM3(+2)	(−)	(−)
c.496G>A	p.(Glu166Lys)	VUS (3)	Uncertain	0 (+1-1)	(−)	PM2_P (+1)	(−)	BP7_O(−1)	(−)
c.496+5G>A		LP(7)/P(1)	Uncertain	+5 (+1 + 4)	(−)	PM2_P (+1)	PM3_S(+4)	(−)	(−)
c.901G>T	p.(Gly301Cys)	Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.901+2T>C		LP (1)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.901+3A>T		VUS (2)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.902-1G>T		P (5)	Pathogenic	+11 (+8 + 1 + 2)	PVS1_O(+8)	PM2_P (+1)	PM3(+2)	(−)	(−)
c.902G>A	p.(Gly301Asp)	VUS (17)	Uncertain	(+1)	PVS1_O_P(+1)	(−)	(−)	(−)	(−)
c.903T>G	p.(Gly301=)	LB(3)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.1065+1G>T		LP(7)/P(1)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.1065+3A>G		VUS (1)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.1066-6T>G		B(6)/LB(6)/VUS(5)	Likely Benign	(−4)	(−)	(−)	(−)	(−)	BS1(−4)
c.1235+4_1235+5del		Not reported	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.1898G>T	p.(Cys633Phe)	Not reported	Uncertain	+2 (+1 + 1)	PVS1_O_P(+1)	PM2_P (+1)	(−)	(−)	(−)
c.1898+2T>G		LP(2)/P(7)	Likely Pathogenic	+9 (+1 + 8)	(−)	PM2_P (+1)	PM3_VS(+8)	(−)	(−)
c.1898+3A>T		Not reported	Likely Benign	-3 (+1-4)	(−)	PM2_P (+1)	(−)	BP7_O_S(−4)	(−)
c.1898+3_1898+4del		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.2251-1G>C		LP(1)/P(1)	Uncertain	+5 (+4 + 1)	PVS1_O_S(+4)	PM2_P (+1)	(−)	(−)	(−)
c.2376+1G>A		LP(4)	Likely Pathogenic	+7 (+4 + 1 + 2)	PVS1_O_S(+4)	PM2_P (+1)	PM3(+2)	(−)	(−)
c.2376+3A>T		VUS(1)	Uncertain	+5 (+4 + 1)	PVS1_O_S(+4)	PM2_P (+1)	(−)	(−)	(−)
c.2377-6T>A		LB(1)/VUS(3)	Likely Benign	-3 (+1-4)	(−)	PM2_P (+1)	(−)	BP7_O_S(−4)	(−)
c.2377-2A>G		LP(4)	Uncertain	+2 (+1 + 1)	PVS1_O_P(+1)	PM2_P (+1)	(−)	(−)	(−)
c.2467-3A>G		Not reported	Likely Benign	-3 (+1-4)	(−)	PM2_P (+1)	(−)	BP7_O_S(−4)	(−)
c.2638+3A>G		LB(2)/VUS(2)	Likely Benign	-3 (+1-4)	(−)	PM2_P (+1)	(−)	BP7_O_S(−4)	(−)
c.3577-1G>A		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.3577G>C	p.(Val1193Leu)	VUS(8)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.3746+1G>A		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.3746+4A>C		LB(1)/VUS(2)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.3746+5G>A		VUS(3)	Likely Pathogenic	+9(+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.3993G>A	p.(Gln1331=)	VUS(1)	Uncertain	+3(+2 + 1)	PVS1_O_M(+2)	PM2_P (+1)	(−)	(−)	(−)
c.3993+1G>A		LP(1)/P(4)	Likely Pathogenic	+6(+2 + 4)	PVS1_O_M(+2)	(−)	PM3_S(+4)	(−)	(−)
c.3993+5G>T		B(9)/LB(3)	Likely Benign	(−4)	(−)	(−)	(−)	(−)	BS1(−4)
c.3994-3C>T		Not reported	Likely Benign	-3 (+1-4)	(−)	PM2_P (+1)	(−)	BP7_O_S(−4)	(−)
c.3994-2A>G		LP(3)/P(1)	Likely Pathogenic	(+8)	PVS1_O(+8)	(−)	(−)	(−)	(−)
c.4109+1G>T		LP(2)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.4109+3A>G		VUS(1)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.4109+5G>A		VUS(1)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.4109+6T>G		LB(4)/VUS(1)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.4110-9C>G		VUS(4)/P(1)	Pathogenic	+11(+8 + 1 + 2)	PVS1_O(+8)	PM2_P (+1)	PM3(+2)	(−)	(−)
c.4110-2A>C		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.4236+1G>A		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)

(Continues)

Table 2. Continued

c.HGVS*	p.HGVS*	ClinVar†	Point-based classification‡		PVS1_O§	PM2¶	PM3¶¶	BP7_O§	BS1¶¶
c.4236+5G>A		VUS(1)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.4236+6T>C		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.4436+4A>G		Not reported	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.7307+1G>A		LP (1)	Pathogenic	+11(+8 + 1 + 2)	PVS1_O(+8)	PM2_P (+1)	PM3(+2)	(−)	(−)
c.7307+4A>G		VUS(3)	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.7515G > A	p.(Lys2505=)	Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.7515+6T>C		VUS(1)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.7629+2T>G		Not reported	Likely Pathogenic	+9 (+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.7630-3C>T		LB(2)/VUS(2)	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)
c.7630-2A>C		LP(1)/P(15)	Pathogenic	+16 (+8 + 8)	PVS1_O(+8)	(−)	PM3_VS(+8)	(−)	(−)
c.7787A>T	p.(Glu2596Val)	Not reported	Likely Pathogenic	+9(+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.7788G>A	p.(Glu2596=)	LP(2)/P(5)	Pathogenic	+13(+8 + 1 + 4)	PVS1_O(+8)	PM2_P (+1)	PM3_S(+4)	(−)	(−)
c.7788+1G>C		Not reported	Likely Pathogenic	+9(+8 + 1)	PVS1_O(+8)	PM2_P (+1)	(−)	(−)	(−)
c.7788+6T>G		Not reported	Uncertain	(+1)	(−)	PM2_P (+1)	(−)	(−)	(−)

The table shows 56 ATM variants identified in the BRIDGES cohort, its current ClinVar clinical classification, and the ACMG/AMP-based tentative classification that we have performed by combining existing pathogenic and benign evidence with the mgATM data produced in the present study. For each individual *ATM* variant, we have evaluated all applicable evidence, but the table shows only evidence contributing to the final classification.

*NM_000051.3.

†ClinVar last accessed 2 February 2022. LB (Likely Benign), **VUS** (variant of uncertain significance), LP (Likely Pathogenic), P (Pathogenic). In brackets, N submitters supporting each classification. (−) not reported.

‡We used an ACMG/AMP point system Bayesian framework to combine all pathogenic and benign evidence.

§We integrated mgATM splicing functional data in the classification scheme as a pathogenic evidence code PVS1_O (or as a benign evidence code BP7_O) of variable strength, as per ClinGen *ATM* expert panel ACMG-AMP specifications. Of note, we used some ACMG-AMP evidence (not shown in the present table) to decide the PVS1_O or BP7_O code strength applicable to complex mgATM readouts).

¶Rarity evidence PM2 downgraded to supporting strength, as *per* ClinGen *ATM* expert panel ACMG-AMP specifications.

¶¶We assigned the recessive disorders evidence PM3 to variants identified in *trans* with a pathogenic variant in A-T patients (as reported in the scientific literature). Code strength as *per* ClinGen recommendations. BS1 evidence (allele frequency greater than expected for disease) applied as per ClinGen *ATM* expert panel ACMG-AMP specifications. See supplementary material, Table S3 and Supplementary materials and methods for further details.

alternative classification in which we simply replaced PVS1_O/BP7_O evidence with PVS1/PP3/BP4 predictive splicing evidence (supplementary material, Table S3.5). Supplementary material, Table S6 summarizes the comparative analysis. Experimental splicing data have an impact on the final classification of 16 *ATM* variants (29% of the tested variants). Overall, experimental splicing data have a positive effect on variant classification, reducing uncertainty (reducing the number of **VUS**s from 29 to 20).

Discussion

Next-Generation Sequencing (NGS) technology is an efficient screening approach to detect variants associated with cancer risk with high sensitivity, cost effectiveness, and speed. Genetic testing has become available for larger groups of patients, allowing more variant carriers to be identified, better management of risk and, in some cases, better treatment [30]. However, NGS also presents some challenges. One of them is the high rate of **VUS**s, for which the association with cancer risk in unclear, and genetic counselling of carriers is difficult [31]. A recent large-scale sequencing study reported that the prevalence of **VUS**s in 12 BC genes was 18.9% [7]. Classification of **VUS** may be improved through large-scale splicing or functional assays.

The *ATM* gene is one of the eight "core" genes that displayed a significant association with BC in the two large-scale studies already mentioned [6,7]. *ATM* pathogenic variants are associated with a moderate risk of BC (1.8–2.1) and an overall lifetime female BC risk above 20%. Protein truncating variants in *ATM* accounts for 0.63% of all BC cases [6,7]. We performed a comprehensive evaluation of potential spliceogenic variants to aid their interpretation. Here we analyzed *in silico* 381 *ATM* splice-site variants identified in BRIDGES patients and controls, through which 128 candidate variants were predicted to impair splicing.

NGS-based RNA-seq provides high-quality qualitative and quantitative data for the characterization of splicing variants in hereditary cancer genes [32]. However, lacking RNA from carriers, we designed five minigenes covering 27 out of the 62 *ATM* coding exons in which most of the preselected variants were located, although, unfortunately, minigenes mgATM_17-22 and mgATM_49-54 did not show the correct transcript profiles and variants at exons 18 to 22; 53 and 54 were excluded from the analysis. It is conceivable that smaller constructs, such as those with exons 19–20 or 21–22 (both with short introns), may be functional. Indeed, we introduced deletions of exons 17–18 and 17–20 into the WT mgATM_ex17–22 that generated the corresponding full-length transcripts without any other isoform (see Supplementary materials and methods).

Hybrid minigene technology has proven to be efficient for the description of the splicing outcomes of variants in the absence of patient RNA [33,34], as is the case in the present study. There are many examples verifying the reproducibility of this strategy, including previous minigene studies of our group [19,35]. In this regard, the present study included 10 *ATM* variants for which previous experimental RNA data in carriers have been published: c.496+5G>A [36]; c.901+3A>T [37]; c.902-1G>T [38]; c.1066-6T>G [39]; c.1898+2T>G [40]; c.1898+3_+4del [37]; c.3993+1G>A [37,41]; c.4110-9C>G [42], c.7630-2A>C [37,38,41,43], and c.7788G>A [44]. Supplementary material Table S7 shows a comparative analysis with mgATM data. Overall, the concordance was high and did not affect PVS1_O/BP7_O code strengths (supplementary material, Table S3.2). The only possible exception are three variants (c.496+5G>A, c.1066-6T>G, c.1898+2T>G) in which mgATM data have uncovered leaky effects not reported by previous RNA studies in carriers. In brief, the present study further supports the notion that hybrid minigenes are very good proxies for splicing assays in carriers.

Splice AI is a neural network that predicts splicing from a pre-mRNA sequence [45]. Recent evaluations have identified SpliceAI as the best predictor of variants that impact splicing, here termed spliceogenic variants [46–49]. To further evaluate our analysis, we compared mgATM data with SpliceAI predictions (note that SpliceAI was not used for the initial bioinformatics selection of *ATM* likely spliceogenic variants). The comparative analysis is shown supplementary material, Table S7. Taken together, the data further supports the robustness of the mgATM assay and, equally relevant, the accuracy of SpliceAI in predicting the actual outcome of spliceogenic variants. In relation to the latter, it is worth highlighting that:

i. Four *ATM* variants targeting consensus positions of the splice-sites (c.2377-6T>A, c.2467-3A>G, c.2638+3A>G, and c.3994-3C>T) do not disturb splicing (a remarkable finding correctly predicted by SpliceAI).
ii. Eight *ATM* variants (c.1065+1G>T, c.1065+3A>G, c.3577G>C, c.3993G>A, c.3993+1G>A, c.3993+5G>T, c.7307+1G>A, and c.7307+4A>G) contributed to the incorrect recognition of natural donor sites and to the use of cryptic 5'ss. SpliceAI predicted these variant impacts correctly, except for c.1065+3A>G.

The vast majority of human introns (~99%) are of the GT-AG type, while the most frequent atypical 5'ss is a GC-donor [29]. It is known that GC donor splice-sites are related to alternative splicing events [50,51]. However, the mechanisms underlying the GC-5'ss recognition are not completely understood yet, so variants disrupting GC donor sites are particularly interesting. We have focused our attention on the normal splicing of GC-exons of BC genes (e.g. *BRCA2* exon 17 or *PALB2* exon 12) as well as the anomalous GC usage induced by variants [17,19,34]. Here we analyzed two variants, c.7515G>A and c.7515+6T>C, affecting

ATM exon 50 GC-5'ss. Given the intrinsic weakness of these 5'ss, it is expected that any sequence change may disrupt splicing. Indeed, both variants impaired exon recognition and provoked exon skipping, especially c.7515G>A (last exon nt), where the mgFL$^{49\text{-}52}$-transcript could not be detected. Taking into account the predicted effect on protein translation (p.Lys2505=), this should be *a priori* reclassified as a spliceogenic variant that produces two likely deleterious transcripts (discussed below). On the other hand, variant c.1898+2T>G induced the use of an extremely rare GG 5'ss, which functions as a donor site in 0.01% of human introns [29], so that it partially restored the canonical splicing, generating 13% of the expected mgFL$^{11\text{-}17}$-transcript (V1-*ATM* exons 11 to 17-V2; Figure 3B,C).

Our classification schema integrates mgATM data (PVS1_O/BP7_O evidence) and provides an informative classification (P/LP or LB) for 36 variants, but 20 (36%) remained as **VUS**, including 14 intronic and 2 synonymous variants (the type of variants in which splicing data are expected to be a major contributor to classification). The relatively high proportion of **VUS**s in our study is (partly) explained by the high proportion of mgATM readouts for which inferring a pathogenic or benign evidence (i.e. deciding the appropriate PVS1_O or BP7_O code strength) is far from obvious, a complexity that we summarize as follows:

i. For several variants, mgATM readouts produced two or more altered transcripts with different coding potential and a different contribution to the overall expression.
ii. mgATM analysis identified 24 variants that produced altered transcripts, but also a significant proportion of full-length transcripts ("leaky variants").

To deal with these issues, we have been very conservative (as per ACMG/AMP recommendations), assuming that mgATM readouts are noninformative (i.e. not adding points to the classification scheme) if both transcripts supporting a pathogenic call and transcripts supporting a benign call represent >10% of the overall expression (see supplementary material, Figure S3A). Further, if a variant produces only transcripts supporting a pathogenic call (different strength), we selected the most conservative option for overall PVS1_O code strength, even if representing only 10% of the overall expression (see supplementary material, Figure S3C).

This conservative approach is reflected in the poor contribution of mgATM readouts to the final point-based classification: adding only ≥(−1) and ≤(+2) points to the final classification of 22 variants, including 16 variants for which no points were added (i.e. PVS1_O not applicable and BP7_O not applicable). In brief, many mgATM readouts are noninformative, reflecting the complexity of integrating mgATM readouts into an ACMG/AMP-based classification scheme.

In this regard, "leaky variants" are particularly challenging, as we do not know the precise relationship between *ATM* allele-specific expression levels of full-length transcripts and phenotype. It is conceivable to postulate a dosage-sensitive expression model in which some leaky variants producing full-length transcripts above a certain threshold are benign, leaky variants producing full-length transcripts below a certain threshold are pathogenic, and leaky variants in between associate with an intermediate phenotype. Yet, as far as we know, there are no clinical and/or functional data in the scientific literature supporting (or addressing) this issue.

That said, we noticed that evidence of leakiness uncovered by our mgATM analysis in three variants (c.1898+2T>G, c.496+5G>A, and c.1066-6T>G) together with clinical data available in the scientific literature [36,52] provides some support for an *ATM* dosage-sensitive expression model (see supplementary material, Figure S5 for further details). According to this tentative model, leaky variants producing ≥30% of full-length transcripts are predicted benign, leaky variants producing ≤13% are predicted pathogenic, and leaky variants in between might be associated with an intermediate phenotype. At present, this is just a tentative model based on circumstantial evidence. If confirmed by clinical evidence for a sufficient number of leaky variants and/or by functional studies, the dosage-sensitive expression model might be relevant to refine future iterations of the ACMG/AMP specifications for *ATM*.

In summary, here we carried out an exhaustive study of *ATM*, in which 56 preselected variants were tested using minigene assays (85.7% spliceogenic). Once again, minigenes have proven to be a robust and useful tool to assess potential spliceogenic variants. These splicing assays provide key data for the interpretation of variants, so, despite the complexity of the *ATM* gene (63 exons), efforts should be made to test additional variants identified in the clinical setting (minigene approach, or NGS based RNA-seq analysis of patient RNA whenever possible). According to our ACMG/AMP-based tentative classification scheme, 29 variants end up as pathogenic/likely pathogenic and seven variants as likely benign. Finally, we provide circumstantial evidence supporting a dosage-sensitive model that might be relevant to classify leaky variants.

Acknowledgements

PD, MPGV, DFE, MdlH and EAV received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 634935. The EAV lab is supported by grants from the Spanish Ministry of Science and Innovation, Plan Nacional de I+D+I 2013-2016, ISCIII (PI17/00227 and PI20/00225) co-funded by the FEDER from Regional Development European Funds (European Union) and from the Consejería de Educación, Junta de Castilla y León, ref. CSI242P18 (actuación cofinanciada P.O. FEDER 2014-2020 de Castilla y León). The MdlH lab is supported by a grant from the Spanish Ministry of Science and Innovation, Plan Nacional de I+D+I

2013-2016, ISCIII (PI20/00110) co-funded by FEDER from Regional Development European Funds (European Union). Programa Estratégico Instituto de Biología y Genética Molecular (IBGM), Escalera de Excelencia, Junta de Castilla y León (Ref. CLU-2019-2002). EB-M is a postdoctoral researcher funded by the University of Valladolid (POSTDOC-UVA05, 2022-2025). AV-P and IL-B are supported by predoctoral fellowships from the Consejería de Educación, Junta de Castilla y León. LS-M is supported by a predoctoral fellowship from the AECC-Scientific Foundation, Sede Provincial de Valladolid (2019–2023). AE-S is supported through the Operational Program for Youth Employment and Youth Employment Initiative (YEI), called by the Community of Madrid in 2020, and co-financed by the European Social Fund.

Author contributions statement

MdlH and EAV-S conceptualized and designed the study. Variant data were curated by EB-M, LS-M, AV-P, JA and DFE. Bioinformatics analysis was performed by EB-M, LS-M, AV-P, MdlH and EAV-S. PD, DFE, MPGV, MdlH and EAV-S obtained funds for this study. All authors contributed to all the experiments and the analysis and interpretation of data. This study was supervised by EAV-S. EB-M, MdlH and EAV-S wrote the original draft. EB-M, PD, DFE, MPGV, MdlH and EAV-S reviewed and edited the article. All authors approved the final version of the article.

Data availability statement

All sequencing and fragment analysis data are available at Digital.CSIC (http://hdl.handle.net/10261/265669; https://doi.org/10.20350/digitalCSIC/145).

References

1. Savitsky K, Bar-Shira A, Gilad S, *et al*. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 1995; **268:** 1749–1753.
2. Paull TT. Mechanisms of ATM activation. *Annu Rev Biochem* 2015; **84:** 711–738.
3. Choy KR, Watters DJ. Neurodegeneration in ataxia-telangiectasia: multiple roles of ATM kinase in cellular homeostasis. *Dev Dyn* 2018; **247:** 33–46.
4. Reiman A, Srinivasan V, Barone G, *et al*. Lymphoid tumours and breast cancer in ataxia telangiectasia; substantial protective effect of residual ATM kinase activity against childhood tumours. *Br J Cancer* 2011; **105:** 586–591.
5. Suarez F, Mahlaoui N, Canioni D, *et al*. Incidence, presentation, and prognosis of malignancies in ataxia-telangiectasia: a report from the french national registry of primary immune deficiencies. *J Clin Oncol* 2015; **33:** 202–208.
6. Dorling L, Carvalho S, Breast Cancer Association Consortium, *et al*. Breast cancer risk genes - association analysis in more than 113,000 women. *N Engl J Med* 2021; **384:** 428–439.
7. Hu C, Hart SN, Gnanaolivu R, *et al*. A population-based study of genes previously implicated in breast cancer. *N Engl J Med* 2021; **384:** 440–451.
8. Narod SA. Which genes for hereditary breast cancer? *N Engl J Med* 2021; **384:** 471–473.
9. Hsu FC, Roberts NJ, Childs E, *et al*. Risk of pancreatic cancer among individuals with pathogenic variants in the ATM gene. *JAMA Oncol* 2021; **7:** 1664–1668.
10. Lesueur F, Easton DF, Renault AL, *et al*. First international workshop of the ATM and cancer risk group (4-5 December 2019). *Fam Cancer* 2022; **21:** 211–227.
11. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; **8:** 749–761.
12. Schafer S, Miao K, Benson CC, *et al*. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr Protoc Hum Genet* 2015; **87:** 11.16.1–11.16.14.
13. Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods* 2005; **37:** 331–340.
14. Rhine CL, Neil C, Glidden DT, *et al*. Future directions for high-throughput splicing assays in precision medicine. *Hum Mutat* 2019; **40:** 1225–1234.
15. Sanoguera-Miralles L, Valenzuela-Palomo A, Bueno-Martínez E, *et al*. Comprehensive functional characterization and clinical interpretation of 20 splice-site variants of the *RAD51C* gene. *Cancers (Basel)* 2020; **12:** 3771.
16. Bueno-Martínez E, Sanoguera-Miralles L, Valenzuela-Palomo A, *et al*. RAD51D aberrant splicing in breast cancer: identification of splicing regulatory elements and Minigene-based evaluation of 53 DNA variants. *Cancers (Basel)* 2021; **13:** 2845.
17. Valenzuela-Palomo A, Bueno-Martínez E, Sanoguera-Miralles L, *et al*. Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants. *J Pathol* 2022; **256:** 321–334.
18. Lopez-Perolio I, Leman R, Behar R, *et al*. Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. *J Med Genet* 2019; **56:** 453–460.
19. Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, *et al*. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol* 2019; **248:** 409–420.
20. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11:** 377–394.
21. Houdayer C, Caux-Moncoutier V, Krieger S, *et al*. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012; **33:** 1228–1238.
22. Moles-Fernández A, Duran-Lozano L, Montalban G, *et al*. Computational tools for splicing defect prediction in breast/ovarian cancer genes: how efficient are they at predicting RNA alterations? *Front Genet* 2018; **9:** 366.
23. de Garibay GR, Acedo A, García-Casado Z, *et al*. Capillary electrophoresis analysis of conventional splicing assays: IARC analytical and clinical classification of 31 BRCA2 genetic variants. *Hum Mutat* 2014; **35:** 53–57.
24. Acedo A, Hernández-Moro C, Curiel-García Á, *et al*. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 2015; **36:** 210–221.
25. Tavtigian SV, Greenblatt MS, Harrison SM, *et al*. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 2018; **20:** 1054–1060.
26. Tavtigian SV, Harrison SM, Boucher KM, *et al*. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* 2020; **41:** 1734–1737.

27. Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; **17:** 405–424.
28. Feliubadaló L, Moles-Fernández A, Santamariña-Pena M, *et al*. A collaborative effort to define classification criteria for ATM variants in hereditary cancer patients. *Clin Chem* 2020; **67:** 518–533.
29. Parada GE, Munita R, Cerda CA, *et al*. A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res* 2014; **42:** 10564–10578.
30. LaDuca H, Stuenkel AJ, Dolinsky JS, *et al*. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med* 2014; **16:** 830–837.
31. van Marcke C, Collard A, Vikkula M, *et al*. Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: a systematic review and meta-analysis of gene-panel data. *Crit Rev Oncol Hematol* 2018; **132:** 138–144.
32. Landrith T, Li B, Cass AA, *et al*. Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *NPJ Precis Oncol* 2020; **4:** 4.
33. Baralle D, Lucassen A, Buratti E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 2009; **10:** 810–816.
34. Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, *et al*. Functional classification of DNA variants by hybrid minigenes: identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet* 2017; **13:** e1006691.
35. Montalban G, Fraile-Bethencourt E, López-Perolio I, *et al*. Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: BRCA2 c.7976+5G > T as a case study. *Hum Mutat* 2018; **39:** 1155–1160.
36. Dörk T, Bendix-Waltes R, Wegner RD, *et al*. Slow progression of ataxia-telangiectasia with double missense and in frame splice mutations. *Am J Med Genet A* 2004; **126A:** 272–277.
37. Laake K, Jansen L, Hahnemann JM, *et al*. Characterization of ATM mutations in 41 Nordic families with ataxia telangiectasia. *Hum Mutat* 2000; **16:** 232–246.
38. Teraoka SN, Telatar M, Becker-Catania S, *et al*. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am J Hum Genet* 1999; **64:** 1617–1631.
39. Soukupova J, Dundr P, Kliebl Z, *et al*. Contribution of mutations in ATM to breast cancer development in the Czech population. *Oncol Rep* 2008; **19:** 1505–1510.
40. Stankovic T, Kidd AM, Sutcliffe A, *et al*. 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* 1998; **62:** 334–345.
41. Cavaciuti E, Laugé A, Janin N, *et al*. Cancer risk according to type and location of ATM mutation in ataxia-telangiectasia families. *Genes Chromosomes Cancer* 2005; **42:** 1–9.
42. Demuth I, Dutrannoy V, Marques W Jr, *et al*. New mutations in the ATM gene and clinical data of 25 AT patients. *Neurogenetics* 2011; **12:** 273–282.
43. Sandoval N, Platzer M, Rosenthal A, *et al*. Characterization of ATM gene mutations in 66 ataxia telangiectasia families. *Hum Mol Genet* 1999; **8:** 69–79.
44. Broeks A, de Klein A, Floore AN, *et al*. ATM germline mutations in classical ataxia-telangiectasia patients in the Dutch population. *Hum Mutat* 1998; **12:** 330–337.
45. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, *et al*. Predicting splicing from primary sequence with deep learning. *Cell* 2019; **176:** 535–548.e24.
46. Ha C, Kim JW, Jang JH. Performance evaluation of SpliceAI for the prediction of splicing of NF1 variants. *Genes (Basel)* 2021; **12:** 1308.
47. Moles-Fernández A, Domènech-Vivó J, Tenés A, *et al*. Role of splicing regulatory elements and in silico tools usage in the identification of deep intronic splicing variants in hereditary breast/ovarian cancer genes. *Cancers (Basel)* 2021; **13:** 3341.
48. Riepe TV, Khan M, Roosing S, *et al*. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum Mutat* 2021; **42:** 799–810.
49. Wai HA, Lord J, Lyon M, *et al*. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med* 2020; **22:** 1005–1014.
50. Thanaraj TA, Clark F. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res* 2001; **29:** 2581–2593.
51. Churbanov A, Winters-Hilt S, Koonin EV, *et al*. Accumulation of GC donor splice signals in mammals. *Biol Direct* 2008; **3:** 30.
52. Ding H, Mao C, Li SM, *et al*. Lack of association between ATM C.1066-6T > G mutation and breast cancer risk: a meta-analysis of 8,831 cases and 4,957 controls. *Breast Cancer Res Treat* 2011; **125:** 473–477.
53. Brnich SE, Abou Tayoun AN, Couch FJ, *et al*. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* 2019; **12:** 3.
54. Abou Tayoun AN, Pesaran T, DiStefano MT, *et al*. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018; **39:** 1517–1524.
55. Lee JH, Paull TT. Cellular functions of the protein kinase ATM and their relevance to human disease. *Nat Rev Mol Cell Biol* 2021; **22:** 796–814.
56. Jansma M, Linke-Winnebeck C, Eustermann S, *et al*. Near-complete structure and model of Tel1ATM from Chaetomium thermophilum reveals a robust autoinhibited ATP state. *Structure* 2020; **28:** 83–95.e5.
57. Stakyte K, Rotheneder M, Lammens K, *et al*. Molecular basis of human ATM kinase inhibition. *Nat Struct Mol Biol* 2021; **28:** 789–798.
58. Baretić D, Pollard HK, Fisher DI, *et al*. Structures of closed and open conformations of dimeric human ATM. *Sci Adv* 2017; **3:** e1700933.
59. Xiao J, Liu M, Qi Y, *et al*. Structural insights into the activation of ATM kinase. *Cell Res* 2019; **29:** 683–685.
60. Xin J, Xu Z, Wang X, *et al*. Structural basis of allosteric regulation of Tel1/ATM kinase. *Cell Res* 2019; **29:** 655–665.
61. Yates LA, Williams RM, Hailemariam S, *et al*. Cryo-EM structure of nucleotide-bound Tel1ATM unravels the molecular basis of inhibition and structural rationale for disease-associated mutations. *Structure* 2020; **28:** 96–104.e3.
62. Kozlov SV, Graham ME, Peng C, *et al*. Involvement of novel autophosphorylation sites in ATM activation. *EMBO J* 2006; **25:** 3504–3514.
63. Cavalieri S, Funaro A, Porcedda P, *et al*. ATM mutations in Italian families with ataxia telangiectasia include two distinct large genomic deletions. *Hum Mutat* 2006; **27:** 1061.
64. Sun X, Becker-Catania SG, Chun HH, *et al*. Early diagnosis of ataxia-telangiectasia using radiosensitivity testing. *J Pediatr* 2002; **140:** 724–731.
65. Buzin CH, Gatti RA, Nguyen VQ, *et al*. Comprehensive scanning of the ATM gene with DOVAM-S. *Hum Mutat* 2003; **21:** 123–131.
66. Soukupova J, Pohlreich P, Seemanova E. Characterisation of ATM mutations in Slavic ataxia telangiectasia patients. *Neuromolecular Med* 2011; **13:** 204–211.
67. Pietrucha BM, Heropolitańska-Pliszka E, Wakulińska A, *et al*. Ataxia-telangiectasia with hyper-IgM and wilms tumor: fatal reaction to irradiation. *J Pediatr Hematol Oncol* 2010; **32:** e28–e30.
68. Allinen M, Launonen V, Laake K, *et al*. ATM mutations in Finnish breast cancer patients. *J Med Genet* 2002; **39:** 192–196.

69. Jacquemin V, Rieunier G, Jacob S, *et al*. Underexpression and abnormal localization of ATM products in ataxia telangiectasia patients bearing ATM missense mutations. *Eur J Hum Genet* 2012; **20:** 305–312.
70. Pylkäs K, Tommiska J, Syrjäkoski K, *et al*. Evaluation of the role of Finnish ataxia-telangiectasia mutations in hereditary predisposition to breast cancer. *Carcinogenesis* 2007; **28:** 1040–1045.
71. Lee JH, Mand MR, Kao CH, *et al*. ATM directs DNA damage responses and proteostasis via genetically separable pathways. *Sci Signal* 2018; **11:** eaan5598.
72. Tian Y, Pesaran T, Chamberlin A, *et al*. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci Rep* 2019; **9:** 12752.
73. Verhagen MMM, Abdo WF, Willemsen MAAP, *et al*. Clinical spectrum of ataxia-telangiectasia in adulthood. *Neurology* 2009; **73:** 430–437.
74. Tavtigian SV, Oefner PJ, Babikyan D, *et al*. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet* 2009; **85:** 427–446.
75. Mitui M, Bernatowska E, Pietrucha B, *et al*. ATM gene founder haplotypes and associated mutations in Polish families with ataxia-telangiectasia. *Ann Hum Genet* 2005; **69:** 657–664.
76. Cavalieri S, Funaro A, Pappi P, *et al*. Large genomic mutations within the ATM gene detected by MLPA, including a duplication of 41 kb from exon 4 to 20. *Ann Hum Genet* 2008; **72:** 10–18.
77. Davis MY, Keene CD, Swanson PD, *et al*. Novel mutations in ataxia telangiectasia and AOA2 associated with prolonged survival. *J Neurol Sci* 2013; **335:** 134–138.
78. Verhagen MM, Last JI, Hogervorst FB, *et al*. Presence of ATM protein and residual kinase activity correlates with the phenotype in ataxia-telangiectasia: a genotype-phenotype study. *Hum Mutat* 2012; **33:** 561–571.
79. Mitui M, Campbell C, Coutinho G, *et al*. Independent mutational events are rare in the ATM gene: haplotype prescreening enhances mutation detection rate. *Hum Mutat* 2003; **22:** 43–50.
80. Birrell GW, Kneebone K, Nefedov M, *et al*. ATM mutations, haplotype analysis, and immunological status of Russian patients with ataxia telangiectasia. *Hum Mutat* 2005; **25:** 593.
81. Nespoli L, Verri A, Tajè S, *et al*. A precocious cerebellar ataxia and frequent fever episodes in a 16-month-old infant revealing ataxia-telangiectasia syndrome. *Case Reports Immunol* 2013; **2013:** 296827.
82. Telatar M, Teraoka S, Wang Z, *et al*. Ataxia-telangiectasia: identification and detection of founder-effect mutations in the ATM gene in ethnic populations. *Am J Hum Genet* 1998; **62:** 86–97.
83. Aygün FD, Nepesov S, Çokuğraş H, *et al*. Bladder wall telangiectasia in a patient with ataxia-telangiectasia and how to manage? *Case Rep Pediatr* 2015; **2015:** 615368.
84. Ghosh R, Harrison SM, Rehm HL, *et al*. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat* 2018; **39:** 1525–1530.
85. Whiffin N, Roberts AM, Minikel E, *et al*. Using high-resolution variant frequencies empowers clinical genome interpretation and enables investigation of genetic architecture. *Am J Hum Genet* 2019; **104:** 187–190.
86. Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med* 2018; **20:** 1687–1688.
87. Dörk T, Bendix R, Bremer M, *et al*. Spectrum of ATM gene mutations in a hospital-based series of unselected breast cancer patients. *Cancer Res* 2001; **61:** 7608–7615.

References [53–87] are cited only in the supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1.

S1-A. Insert sequence of minigene mgATM_ex4–9

S1-B. Insert sequence of minigene mgATM_ex11–17

S1-C. Insert sequence of minigene mgATM_ex17–22

S1-D. Insert sequence of minigene mgATM_ex25–29

S1-E. Insert sequence of minigene mgATM_ex49–52

Figure S2. Workflow of the minigene protocol

Figure S3.

S3-A. Proposed decision tree assigning a PVS1_O/BP7_O code strength to mgATM minigene readouts

S3-B. Pathogenic/Benign code strengths applicable to individual transcripts produced by mgATM minigenes

S3-C. Pathogenic/Benign annotation of *ATM* transcripts

Figure S4. Splicing functional assays of four selected splice-site variants and WT minigenes in MDA-MB-231 (green) and MCF-7 (blue) cells

Figure S5. Proposed 'dosage-sensitive expression model' and tentative integration into the classification scheme to assigning a PVS1_O/BP7_O code strength to *ATM* leaky variants

Figure S6. Minigene mgATM_ex17–22. (A) Minigene structure. Exons are indicated by boxes. (B) Fluorescent fragment electrophoresis of the wildtype minigene mgATM_ex17-22 in MCF-7 cells. FAM-labelled products (blue peaks) were run with LIZ1200 (orange peaks) as size standard. FL, Full-length transcript

Figure S7. Structures and functional assays of the novel minigenes mgATM_ex19–22 and _ex21–22. (A) Structure of the minigenes mgATM_ex19–22 (left) and _ex21–22 (right). Exons are indicated by boxes. (B) Agarose gel electrophoresis of RT-PCR products of both minigenes

Figure S8. Agarose gel electrophoresis of RT-PCR products produced by minigene mgATM_ex49–54 (in duplicate)

Figure S9.

S9-A. Alignment and amino acid conservation of deleted in-frame sequences corresponding to the anomalous ATM transcripts Δ(E5) and Δ(E7_9)

S9-B. Alignment and amino acid conservation of deleted in-frame sequences corresponding to the anomalous ATM transcripts Δ(E11), Δ(E12), Δ(E15), Δ(E15_16), Δ(E16p3) and Δ(E16)

S9-C. Alignment and amino acid conservation of deleted in-frame sequences corresponding to the anomalous ATM transcripts Δ(E25_26), Δ(E25p159), Δ(E26q120) and Δ(E28_29)

S9-D. Alignment and amino acid conservation of deleted in-frame sequences corresponding to the anomalous ATM transcripts Δ(E51) and Δ(E52)

Table S1. Bioinformatics analysis of 381 BRIDGES *ATM* variants with Max Ent Scan

Table S2. Mutagenesis primers for *ATM* variants

Table S3 (S3.1-S3.5). ACMG/AMP-based tentative classification according to a Bayesian point system

Table S4. RNA and protein HGVS descriptions according to the reference sequence NM_000051.3

Table S5. Comparative classification of 21 *ATM* variants

Table S6. Impact of mgATM data on the classification of 56 *ATM* variants

Table S7. Comparative analysis of SpliceAI predictions, mgATM readouts, and experimental splicing data in carriers