

# Die genomische Struktur des ATM-Gens und Hinweise auf die Möglichkeit umfangreicher Deletionen

Matthias Platzer, Thomas Wiehe, Klaus Rohde, Natascha Sandoval, André Rosenthal

## Zusammenfassung

Mit dem Ziel, die Mutationssuche im AT Gen (ATM) und die Erforschung der Transkriptionsregulation des Gens zu unterstützen, haben wir 184 kb der chromosomalen Region 11q22.3 bestimmt. Das ATM-Gen besteht aus 66 Exons und erstreckt sich über 146 kb. Der Start der Translation liegt in Exon 4. Die strukturelle Diversität der nicht-translatierten Regionen (UTR) weisen auf eine komplexe post-transkriptionale Regulation des Gens hin, die eine schnelle Anpassung des Protein-spiegels an Faktoren der Umwelt oder Stadien des Zellzyklus ermöglichen könnte.

Die Expression des ATM-Gens wird durch einen bidirektionalen Promotor gesteuert, der gleichzeitig die Transkription des benachbarten E14/NPAT Gens in entgegengesetzter Richtung betreibt. Zusätzlich identifizieren verschiedene Vorhersageprogramme ein zweites ATM-Promotor im Intron 3, unmittelbar vor dem ersten kodierenden Exon. Die gesamte 148 kb große genomische Region wird zu 85% transkribiert und zu 5% translatiert. Genomweite repetitive Elemente stellen 37,2% des Bereiches dar. Drei nahezu vollständige LINE-1 Elemente in den Introns 18 und 63, sowie 4,5 kb nach dem Gen sind in gleicher Orientierung angeordnet und stellen mögliche Rekombinationsorte für die Entstehung großer Deletionen des ATM dar.

## Schlüsselwörter

Genstruktur, repetitive Elemente, Einzelbasen-Polymorphismus (SNP)

## Summary

*Genomic structure of the ATM gene reveals potential hotspots for large deletions*

*To facilitate the search for mutations of the AT gene (ATM) and to establish the basis for the identification of transcriptional regulatory elements, we have sequenced 184 kb of the human 11q22.3 chromosomal region. The ATM gene consists of 66 exons and spans 146 kb. The initiation codon falls within exon 4. Structural diversity of untranslated regions (UTR) of the ATM gene suggest complex post-transcriptional regulation of the ATM gene, enabling rapid modulation of the protein level in response to environmental stimuli or alterations of the cellular physiological states. The ATM shares a bidirectional promoter with E14/NPAT which is transcribed in the opposite direction. Moreover, promoter recognition algorithms identified a putative second ATM promoter located within intron 3, immediately upstream of the first coding exon. The entire 148 kb region is transcribed to about 85% and translated to 5%. Genome-wide repeats were found to constitute 37,2%. Three full-length LINE-1 elements in introns 18 and 63, as well as 4.5 kb downstream of ATM are arranged in the same orientation and may represent hotspots for the generation of large deletions.*

## Keywords

*Gene structure, repetitive elements, single nucleotide polymorphism (SNP)*

## Einleitung

Das bei Patienten mit Ataxia teleangiectatica mutierte Gen (ATM) wurde auf Chromosom 11q22-23 lokalisiert (1) und durch positionelle Klonierung identifiziert (2). Es enthält einen offenen Leserahmen (ORF) von 9168 Nucleotiden. Das daraus abgeleitete Protein mit 3056 Aminosäuren (aa) gehört zu einer Familie großer Proteine mit Sequenzhomologien zu der katalytischen Domäne von Phosphatidylinositol-3 (PI-3)-Kinasen (3).

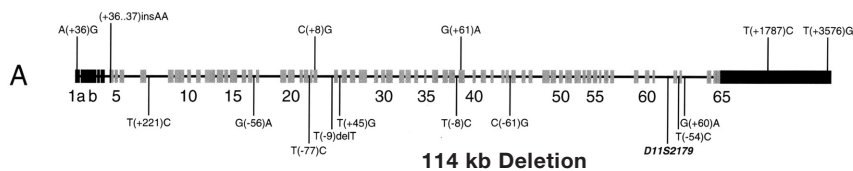
Um die Identifizierung von AT-Mutationen in genomischer DNA als Ausgangsmaterial möglich zu machen, haben wir die gesamte genomische Organisation des ATM-Gens bestimmt. Es besteht aus 66 Exons (4). Andere Arbeitsgruppen haben unsere Ergebnisse bestätigt (5, 6).

Wegen der Bedeutung des ATM-Gens in der biomedizinischen Forschung führten wir die Sequenzierung des gesamten ATM-Lokus durch. Durch Sequenzierung fünf überlappender Cosmide, die den größten Teil des D11S384-D11S1818 Intervalls überspannen (3), bestimmten wir einen zusammenhängenden genomischen Sequenzbereich von 184.490 bp (GenBank access. no. U82828), der das gesamte ATM-Gen sowie die 5'-Region des E14/NPAT-Gens enthält (7).

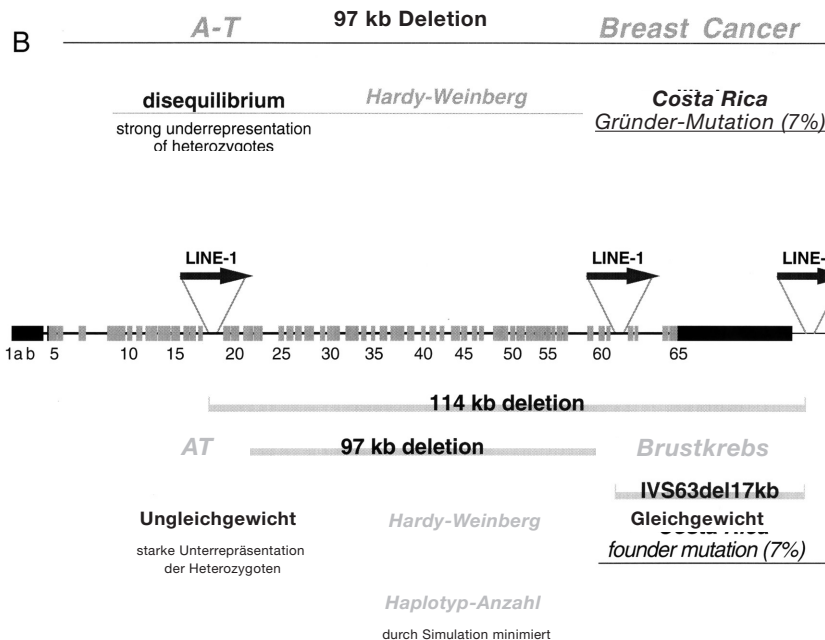
## Exon-Intron-Struktur

Ein Vergleich der 184 kb-Genesequenz mit der Sequenz der ATM-mRNA führte uns zur Exon-Intron-Struktur des ATM-Gens mit einer Auflösung im Nucleotidbereich. Das ATM-Gen hat 66 Exons (Abb 1), die über einen genomischen Bereich von 146.182 bp verteilt sind. Die Introns variieren beträchtlich in ihrer Größe zwischen 77 und 10.207 bp. Das erste Exon des E14/NPAT-Gens beginnt 468 bp in 5'-Richtung von dem ersten Exon des ATM-Gens und setzt sich in entgegengesetzter Richtung fort.

Während bisher kein alternatives Spleißen in der kodierenden Region gefunden wurde, unterliegen die ersten vier Exons innerhalb der 5'-untranslatierten Region (UTR) extensivem alternativem Spleißen. 12 verschiedene 5'UTRs weisen beträchtli-



**Abb 1**  
**Genomische Organisation**  
**des ATM-Gens**  
 Exons sind durch Boxen, Introns durch Linien dargestellt. Graue Boxen, translattiert, schwarze Boxen, Introns.  
 Maßstab Exon:Intron = 10:1



**Abb 2**  
 A) Schematische Darstellung von SNPs in 66 AT- und 36 Brustkrebspatienten  
 B) Schätzung von Genotyp- und Haplotypfrequenzen

che Längenvariationen (150-884 bp) und Sequenzunterschiede auf. Der längste 5'UTR enthält insgesamt 18 AUG-Codons strangaufwärts des Translationsinitiationscodons. Innerhalb des 3590 bp großen 3'UTR wurden mindestens vier aktive Polyadenylierungsstellen identifiziert (8).

Mit seinen 66 Exons, einschließlich der beiden alternativ genutzten „leader“-Exons 1a and 1b, enthält das ATM-Gen eine der größten Exonzahlen, die bisher für menschliche Gene berichtet wurden. In dieser Hinsicht ist die genomische Organisation des ATM-Gens vergleichbar mit der des Huntingtin-Gens, dessen 67 Exons über 180 kb verteilt sind, jedoch verschieden von der einiger riesiger Gene, wie beispielsweise des Dystrophin-Gens, dessen 79 Exons über 2,4 Mb verteilt sind.

**Analyse des lokalen Sequenzgehalts**

Die 184 kb-Sequenz des ATM-Genlokus weist einen geringen G+C-Gehalt von 38,1% auf. Dies ist konsistent mit der Lage des ATM-Genlokus in der chromosomalen Region 11q22.3, einem spät replizierenden G-Band. Wenn alle genomweit verbreiteten repetitiven Sequenzen innerhalb des 184 kb ATM-Lokus ausgenommen werden, sinkt der G+C-Gehalt der verbleibenden 115.767 bp auf 34,0%. Daher kann vermutet werden, daß ein früher Vorläufer des menschlichen

ATM-Gens einen noch niedrigeren G+C-Gehalt aufwies, bevor sich bei Säugern repetitive Sequenzen ausbreiteten.

**Promotorvorhersage**

Wir benutzten zwei unabhängige Algorithmen, um potentielle Promotorbereiche auf beiden Strängen der 184 kb-Sequenz zu identifizieren. Zwei Promotorbereiche, einer in Vorwärts- und einer in reverser Richtung, liegen innerhalb der CpG-Insel, die das jeweils erste Exon des ATM-Gens und des E14/NPAT-Gens überspannt. Die Sequenz zwischen beiden Genen beträgt nur 468 bp. Die Nachbarschaft beider Gene legt die Möglichkeit einer koordinierten Regulation der Genexpression nahe. Experimente mit Reporter-Gen-Konstrukten zeigten, daß die CpG-Insel als ein bidirektionaler Promotor wirkt und daß die Stimulation der Genexpression in Richtung des ATM-Gens dreifach höher ist als in Richtung des E14/NPAT-Gens (9).

Die Transkripte des ATM-Gens gehören zu jenen 5–10% der Vertebraten-mRNAs, die lange und hoch strukturierte 5'UTRs aufweisen. Solche Gene nutzen oft alternative Promotoren, um ergänzend Transkripte mit kurzen „leader“-Sequenzen zu erzeugen. Interessanterweise fanden wir einen zusätzlichen putativen ATM-Promotor, mit einer „TATA-Box“, innerhalb des Introns 3 unmittelbar vor dem ersten kodierenden Exon. Die vorhergesagte

Promotorregion überlappt mit zwei repetitiven Einheiten der „Alu“-Familie. Eine dieser Wiederholungseinheiten, AluSg, weist eine hohe Homologie zu einem funktionell aktiven östrogen-responsiven Element auf (10). Dieser vermeintliche interne Promotor könnte für die kurze 5'UTR-Sequenz sorgen, die eine basale ATM-Translation erlaubt, wogegen der strangaufwärts gelegene Promotor regulative UTRs ermöglichen würde.

**Analyse repetitiver Sequenzen**

Wir führten eine Analyse repetitiver Sequenzen durch, um verschiedene Typen einfacher Sequenzwiederholungen, Mikrosatelliten oder genomweit vorkommender repetitiver Einheiten zu charakterisieren. Wiederholungen von fünf und mehr ununterbrochener Di- und Trinukleotide, die zur Etablierung polymorpher Marker genutzt werden konnten, sind in Tabelle 1 aufgeführt. Die Dinukleotideinheit CA/TG wurde an zwölf Stellen gefunden, einschließlich des Markerlokus D11S2179, 180 bp strangabwärts des Exons 62.

Genomweit verteilte Sequenzwiederholungen machen zusammen 37,2% des ATM-Genlokus aus, hierzu leisten Alu (14,6%) und LINE-1 (17,1%) Einheiten die Hauptbeiträge. Drei LINE-1 Elemente voller Länge (6017 bp, 6031 bp und 6116 bp) befinden sich in den Introns 18 und 63 sowie 4,5 kb strangabwärts der Polyadenylierungsstelle. Sie sind untereinander höchst

Tab 1

## Sequenzfolgen mit mehr als fünf ununterbrochenen Di- und Trinukleotidwiederholungen

Sequenz	Kopien	Position	Lokalisation	relative Position*
TA	5	5254	NPAT intron 1	+4947
TA	9	37828	ATMintron 11	-555
CA	5	38803	intron 12	+49
CA	5	56733	intron 20	+444
TG	7	58327	"	-418
TG	6	62026	intron 24	+1493
TG	8	68202	intron 25	-473
TA	13	81568	intron 33	+416
TG	5	81594	"	+442
TG	5	81620	"	+468
TC	5	86981	intron 35	-407
TG	5	103030	intron 42	-463
TA	15	105288	intron 45	+95
CA	13	105318	"	+125
TA	6	105345	"	+152
TA	5	105385	"	+192
GA	5	105422	"	+229
TA	5	135887	intron 61	+858
TA	5	136070	"	+1041
TA	5	136255	"	+1226
CA	26	141909	intron 62	+180
TA	5	158201	ATM3'	+1248
CA	6	159092	"	+2139
TC	8	169782	"	+12829
TG	9	178989	"	+22036
CAA	6	330	NPAT intron 1	+9871
CAA	8	7439	"	+2762
GAA	5	123978	ATMintron 58	+353
CAA	5	131465	intron 59	+430

\* bezüglich Anfang (+) oder Ende (-) des Exons/Introns

konserviert, in derselben Orientierung arrangiert und könnten „Brennpunkte“ für homologe Rekombinationsereignisse sein, die zu großen Deletionen führen (Abb 1). Eine entsprechende 16 kb-Deletion, IVS63 del16kb, ist kürzlich in der Bevölkerung von Costa Rica identifiziert worden (11).

Hinweise für das häufige Vorkommen großer Deletionen innerhalb des ATM-Genorts können auch durch Anwendung einer Methode zur Rekonstruktion von Haplotypfrequenzen gewonnen werden (12). Wir haben die Genotyphäufigkeiten für 15 SNP-Marker bei 66 AT- und 36 Brustkrebspatienten bestimmt. Im Gegensatz zur Gruppe der Brustkrebspatientinnen zeigen die Marker bei AT-Patienten eine deutliche Abweichung vom Hardy-Weinberg-Gleichgewicht aufgrund einer Unterzahl an Heterozygoten (Abb 2). Der Heterozygotenmangel könnte damit erklärt werden, daß Genotypen mit großen Deletionen (Hemizygoten) als Homozygoten fehlklassifiziert werden.

### Zusammenfassung

Die vorgestellte 184.490 bp-Sequenz genomischer DNA einschließlich des gesamten menschlichen ATM-Gens stellt eine wichtige Ressource bereit für weitere Untersuchungen der ATM-Regulation, für die Detektion von Mutationen und Polymorphismen in diesem Gen und für die Entwicklung diagnostischer Verfahren. Die Analyse dieser Genregion illustriert die Mög-

lichkeiten der gegenwärtig in großem Maßstab durchgeführten Sequenzierprojekte, Fragen der Organisation und Evolution menschlicher Gene und chromosomaler Regionen zu beleuchten. Vergleichende Sequenzierungen in Modellorganismen werden weitere Einblicke in diese Prozesse gewähren.

### Danksagung

Wir danken Diana Wiedemann, Hella Ludewig, Dreen Sötzer und Nicole Steinlein für die ausgezeichnete technische Mitarbeit und dem Deutschen ATM-Konsortium für den Zugang zu SNP-Daten.

### Literatur

- Gatti RA, Berkel I, Boder E, Braedt G, Charney P, Concannon P, Ersoy F, Foroud T, Jaspers NG, Lange K et al. (1988) Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. *Nature* 336, 577-80.
- Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S et al. (1995a) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 268, 1749-1753.
- Savitsky K, Sfez S, Tagle DA, Ziv Y, Sartiel A, Collins FS, Shiloh Y, Rotman G (1995b) The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Hum Mol Genet* 4, 2025-2032.
- Uziel T, Savitsky K, Platzer M, Ziv Y, Helbitz T, Nehls M, Boehm T, Rosenthal A, Shiloh Y, Rotman G (1996) Genomic Organization of the ATM gene. *Genomics* 33, 317-320.
- Rasio D, Schichman, SA, Negrini, M, Canaani, E, and Croce, CM (1996) Complete exon structure of the ALL1 gene. *Cancer Res* 56, 1766-1769.
- Vorechovsky, I, Rasio D, Luo L, Monaco C, Hammarstrom L, Webster AD, Zaloudik J, Barbanti-Brodani G, James M, Russo G, Croce CM, Negrini M (1996) The ATM gene and susceptibi-

lity to breast cancer: analysis of 38 breast tumors reveals no evidence for mutation. *Cancer Res* 56, 2726-2732.

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

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

9. Byrd PJ, Cooper PR, Stankovic T, Kullar HS, Watts GD, Robinson PJ, Taylor MR (1996) A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies. *Hum Mol Genet* 5, 1785-1791.

10. Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Iglehart JD, Deininger PL, McDonnell DP (1995) Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J Biol Chem* 270, 22777-22782.

11. Telatar M, Wang S, Castellvi-Bel S, Tai LQ, Sheikhavandi S, Regueiro JR, Porras O, Gatti RA (1998) 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.

12. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 12, 921-927.