Breast Cancer Research

RESEARCH

Open Access

Breast tumors from *ATM* pathogenic variant carriers display a specific genome-wide DNA methylation profile

Nicolas M. Viart[1], Anne-Laure Renault[1,2], Séverine Eon-Marchais[1], Yue Jiao[1], Laetitia Fuhrmann[3], Sophia Murat El Houdigui[1], Dorothée Le Gal[1], Eve Cavaciuti[1], Marie-Gabrielle Dondon[1], Juana Beauvallet[1], Virginie Raynal[4], Dominique Stoppa-Lyonnet[5], Anne Vincent-Salomon[3], Nadine Andrieu[1†], Melissa C. Southey[2,6†] and Fabienne Lesueur[1*]

Abstract

Background The ataxia-telangiectasia mutated (ATM) kinase phosphorylates and activates several downstream targets that are essential for DNA damage repair, cell cycle inhibition and apoptosis. Germline biallelic inactivation of the *ATM* gene causes ataxia-telangiectasia (A-T), and heterozygous pathogenic variant (PV) carriers are at increased risk of cancer, notably breast cancer. This study aimed to investigate whether DNA methylation profiling can be useful as a biomarker to identify tumors arising in *ATM* PV carriers, which may help for the management and optimal tailoring of therapies of these patients.

Methods Breast tumor enriched DNA was prepared from 2 A-T patients, 27 patients carrying an *ATM* PV, 6 patients carrying a variant of uncertain clinical significance and 484 noncarriers enrolled in epidemiological studies conducted in France and Australia to investigate genetic and nongenetic factors involved in breast cancer susceptibility. Genome-wide DNA methylation analysis was performed using the Illumina Infinium HumanMethylation EPIC and 450K BeadChips. Correlation between promoter methylation and gene expression was assessed for 10 tumors for which transcriptomic data were available.

Results We found that the *ATM* promoter was hypermethylated in 62% of tumors of heterozygous PV carriers compared to the mean methylation level of *ATM* promoter in tumors of noncarriers. Gene set enrichment analyses identified 47 biological pathways enriched in hypermethylated genes involved in neoplastic, neurodegenerative and metabolic-related pathways in tumor of PV carriers. Among the 327 differentially methylated promoters, promoters of *ARHGAP40*, *SCGB3A1* (*HIN-1*), and *CYBRD1* (*DCYTB*) were hypermethylated and associated with a lower gene expression in these tumors. Moreover, using three different deep learning algorithms (logistic regression, random forest and

†Nadine Andrieu and Melissa C. Southey contributed equally to this work.

*Correspondence:
Fabienne Lesueur
fabienne.lesueur@curie.fr

Full list of author information is available at the end of the article

XGBoost), we identified a set of 27 additional biomarkers predictive of *ATM* status, which could be used in the future to provide evidence for or against pathogenicity in *ATM* variant classification strategies.

Conclusions We showed that breast tumors that arise in women who carry an *ATM* PV display a specific genome-wide DNA methylation profile. Specifically, the methylation pattern of 27 key gene promoters was predictive of *ATM* PV status of the women. These genes may also represent new medical prevention and therapeutic targets for these women.

Keywords Breast cancer, *ATM* gene, Epigenetics, DNA methylation, Biomarker, Molecular testing

Background

Germline biallelic inactivation of *ATM* is responsible for ataxia-telangiectasia (A-T), a rare autosomal recessive disorder affecting about one in 100,000 children in Europe. A-T is characterized by neuronal degeneration, immunological deficiency, cutaneous telangiectasias, genetic instability, radiosensitivity and predisposition to cancers [1–4]. Women related to an A-T child and heterozygous for the *ATM* familial pathogenic variant (PV) have a 2- to 3-fold increased risk of developing a cancer, and 5- to 9-fold increased risk of developing a breast cancer (BC) as compared to women from the general population [3–8]. Although A-T is a rare disorder, 0.5 to 1% of the general population is estimated to be heterozygous for such an *ATM* variant [9–12]. Moreover, *ATM* PV or predicted PV are identified in about 5% of index cases of families predisposed to breast and ovary cancers who undergo genetic testing [11–13]. In these families, heterozygous variant carriers have a 2- to 4-fold increased risk of BC compared to noncarriers, but reported risk estimates for BC and other cancers vary greatly according to the type of variant [9, 11, 12, 14, 15]. To inform clinical management of *ATM* variant carriers, accurate variant classification and precise age-specific cumulative risk of specific cancers of *ATM* variant carriers is a prerequisite, and current efforts aim to address these questions through large-scale genetic epidemiological studies.

At the somatic level, *ATM* mutations or deletions are also commonly found in lymphoid malignancies and in a variety of solid tumors including breast tumors [16] but as yet are not strong indicators for specific targeted therapies [17]. We showed that breast tumors developed by *ATM* PV carriers are more frequently estrogen receptor-positive (ER+; in 97% of the cases) [18], consistent with other studies [11, 12, 19–21], and of histological subtype luminal B or luminal B/HER2 + in 60.7% of the cases. Tumors arising in *ATM* PV carriers lack the homologous recombination deficiency (HRD)-related mutational signature (signature 3 from the COSMIC database [22]) commonly observed in *BRCA1*- and *BRCA2*-deficient tumors [18, 19, 23] but display specific copy number aberrations, including loss of heterozygosity (LOH) at the *ATM* locus on 11q22-23 (in 67% of tumors) and loss at the *RB1* locus on 13q14 (in 69.6% of tumors) [18]. However, a somatic genomic signature that predicts the *ATM* status of the tumor has not yet been reported. Given that ATM mediated the retinoblastoma protein pRB function to control the DNA methyltransferase DNMT1 stability and thus DNA methylation [24], we hypothesized that breast tumors arising in carriers of *ATM* PVs may have methylation aberrations. Our aim was therefore to describe the genome-wide DNA methylation profile of breast tumors of *ATM* PV carriers to further examine the role of *ATM* in oncogenesis and to identify potential therapeutic targets that could benefit this distinct subgroup of women with BC.

Materials and methods

Participants

Breast formalin-fixed, paraffin-embedded (FFPE) tumor samples were collected from patients enrolled in the French studies CoF-AT2 (French prospective cohort on families segregating an *ATM* variant) [25–27] and GENESIS (GENE SISters study) [28], and in the Australian studies ABCFS (Australian Breast Cancer Family Study) [29] and MCCS (Melbourne Collaborative Cohort Study) [30].

CoF-AT2 is an ongoing prospective cohort initiated in 2003 to follow women related to an A-T patient. Epidemiological data including detailed information on familial and clinical data, together with biological samples (blood, tumors) of participants are being collected.

GENESIS is a study on familial BC [28]. Index cases are women diagnosed with invasive breast carcinoma or in situ ductal carcinoma, having at least one sister affected with BC, and tested negative for PV in *BRCA1* and *BRCA2*. *ATM* variant carriers were identified through a case-control mutation-screening study thanks to a resequencing of 113 DNA repair genes [15].

ABCFS is a population-based case-control family study of BC with an emphasis on early-onset BC cases (age at diagnosis < 40 years), carried out in Melbourne and Sydney (Australia) [31]. Women affected with BC were identified using the Victorian and the New South Wales cancer registries and were invited to participate in the study between 1992 and 1999. *ATM* PV carriers were identified through a population-based case-control mutation-screening [32].

Table 1 Clinical characteristics of *ATM* variant carriers, and available tumor material used for analyses

Study	Sample Name	Nucleotide change	Effect on protein	Variant classification	LOH[§]	QC passed[¶]	Sex	Age at BC diagnosis	BC type	Grade	ER status	Subtype
CoF-AT	T0249‡[&]	c.2413C>T; c.7517_7520del	p.Arg805*; p.Arg2506Thrfs*3	PV; PV	No	Yes	Female	31	IDC	III	ER+	Luminal B
CoF-AT	T0252	c.2098C>T	p.Gln700*	PV	ND	Yes	Female	47	ILC	II	ER+	Luminal A
CoF-AT	T0072‡	c.2839–580_577del(-/-)	p.?	PV	Yes	Yes	Male	28	IDC	III	ER+	Luminal/HER2
CoF-AT	T0008	c.3085dup	p.Thr1029Asnfs*19	PV	ND	Yes	Female	52	Mixed ILC and IDC	II	ER+	Luminal A
CoF-AT	T0077R[&]	c.3754_3756delinsCA	p.Tyr1252Glnfs*4	PV	Yes	Yes	Female	67	Apocrine	III	ER+	Luminal B
CoF-AT	T0009[&]	c.3894dup	p.Ala1299Cysfs*3	PV	No	Yes	Female	36	IDC	II	ER+	Luminal B
CoF-AT	T0003	c.5644C>T	p.Arg1882*	PV	ND	Yes	Female	46	IDC	I	ER+	Luminal B
CoF-AT	T0181[&]	c.5644C>T	p.Arg1882*	PV	ND	Yes	Female	40	DCIS	NA	ER+	Luminal A
CoF-AT	T0015[&]	c.6404_6405insTT	p.Arg2136*	PV	No	Yes	Female	40	IDC	II	ER+	Luminal B
CoF-AT	T0076[&]	c.73–2A>G	p.?	PV	Yes	Yes	Female	47	IDC	II	ER+	Luminal B
CoF-AT	T0248[&]	c.7928–2A>C	p.?	PV	Yes	Yes	Female	35	IDC	II	ER+	Luminal A
CoF-AT	T0247[&]	c.8083G>A	p.Gly2695Ser	PV	No	Yes	Female	48	IDC	II	ER+	Luminal B
CoF-AT	T0078	c.8140C>T	p.Gln2714*	PV	Yes	Yes	Female	56	IDC	II	ER+	Luminal/HER2
GENESIS	T0192	c.1236–2A>T	p.?	PV	ND	No	Female	43	DCIS	NA	ER+	Luminal A
GENESIS	T0099[&]	c.3058dup	p.Thr1020Asnfs*28	PV	Yes	Yes	Female	40	IDC	III	ER+	Luminal B
GENESIS	T0111	c.5497–2A>C	p.?	PV	ND	Yes	Female	33	DCIS	NA	ER-	HER2
GENESIS	T0123[&]	c.5528del	p.Pro1843Hisfs*3	PV	Yes	Yes	Female	75	IDC	II	ER-	TNBC
GENESIS	T0141	c.6203T>C	p.Leu2068Ser	PV	ND	Yes	Female	46	DCIS	NA	ER+	Luminal A
GENESIS	T0220[&]	c.8494C>T	p.Arg2832Cys	PV	Yes	Yes	Female	42	IDC	III	ER+	Luminal B
GENESIS	T0091	c.8584+1G>A	p.?	PV	ND	Yes	Female	51	IDC	II	ER+	Luminal B
ABCFS	ABCFS5	c.1355delC	p.Thr452Asnfs*21	PV	ND	Yes	Female	39	IDC	III	ER+	ND
ABCFS	ABCFS8	c.1396C>T	p.Gln466*	PV	ND	Yes	Female	53	IDC	II	ER+	ND
ABCFS	ABCFS4	c.5156delA	p.Asn1719Ilefs*5	PV	ND	Yes	Female	55	IDC	II	ER+	ND
ABCFS	ABCFS7	c.7271T>G	p.Val2424Gly	PV	ND	No	Female	33	IDC	I	ER+	ND
ABCFS	ABCFS6	c.8098A>T	p.Lys2700*	PV	ND	Yes	Female	28	IDC	III	NA	ND
ABCFS	ABCFS1	c.8122G>A	p.Asp2708Asn	PV	ND	Yes	Female	47	IDC	III	ER+	ND
ABCFS	ABCFS2	c.8418+5_8418+8del	p.?	PV	ND	Yes	Female	55	Invasive Tubular Carcinoma	I	ER+	ND
ABCFS	ABCFS3	c.8418+5_8418+8del	p.?	PV	ND	Yes	Female	50	Mixed ILC and IDC	I	ER+	ND
ABCFS	ABCFS9	c.8672-6_8672-2del	p.?	PV	ND	Yes	Female	39	IDC	III	ER-	ND
GENESIS	T0106	c.1009C>T	p.Arg337Cys	VUS	ND	Yes	Female	58	IDC	I	ER+	Luminal A
GENESIS	T0120[&]	c.1464G>T	p.Trp488Cys	VUS	No	Yes	Female	66	IDC	II	ER+	Luminal A
GENESIS	T0191	c.5750G>C	p.Arg1917Thr	VUS	ND	No	Female	48	IDC	II	ER+	Luminal A/B
GENESIS	T0109	c.5882A>G	p.Tyr1961Cys	VUS	ND	Yes	Female	65	IDC	II	ER+	Luminal B

Table 1 (continued)

Study	Sample Name	Nucleotide change	Effect on protein	Variant classification	LOH[§]	QC passed[¶]	Sex	Age at BC diagnosis	BC type	Grade	ER status	Subtype
GENESIS	T0232	c.6059G>T	p.Gly2020Val	VUS	ND	Yes	Female	45	IDC	II	ER+	Luminal A/B
GENESIS	T0100	c.8624A>G	p.Asn2875Ser	VUS	ND	Yes	Female	41	ILC	II	NA	NA

‡A-T patient. [§]Samples with corresponding transcriptomic data available. [¶]Quality control on pre-processed methylation data. [§]As reported in Renault et al. 2017 [27]. QC: quality control; BC: breast cancer; ER: estrogen receptor; LOH: loss of heterozygosity; IDC: infiltrating ductal carcinoma; DCIS: ductal carcinoma in situ; ILC: infiltrating lobular carcinoma; NA: not applicable; ND: not determined; PV: pathogenic variant, VUS: variant of uncertain clinical significance. The combined annotation dependent depletion (CADD) phred scores for missense variants, according to CADD v1.6 are the following: p.Gly2695Ser: 30.0, p.Leu2068Ser: 28.2, p.Arg2832Cys: 32.0, p.Asn2875Ser: 25.7, p.Arg337Cys: 31.0, p.Tyr1961Cys: 28.1, p.Trp488Cys: 22.7, p.Arg1917Thr: 27.6, p.Gly2020Val: 27.3, p.Asp2708Asn: 28.5, p.Val2424Gly: 27.4

MCCS is a prospective cohort study of healthy adults recruited between 1990 and 1994 [30]. Incident BC cases were identified by linkage with the Victorian Cancer Registry, and FFPE tumor blocks related to each case were retrieved from the laboratory that made the cancer diagnosis. The *ATM* gene variant information was obtained via gene panel sequencing conducted by the Breast Cancer Risk after diagnostic Gene Sequencing (BRIDGES) project [21].

Selection of *ATM* variant carriers and noncarriers

Eligible *ATM* variant carriers were either A-T patients who developed BC ($N=2$), heterozygous carriers of a loss-of-function (LoF) variant or of a missense variant classified as pathogenic for A-T disorder ($N=27$). We also included heterozygous carriers of a variant of uncertain clinical significance (VUS) corresponding to missense variants predicted as being deleterious because of their low minor allele frequency (MAF) and high CADD (Combined Annotation Dependent Depletion) phred score [33]. Here we considered variants with a MAF < 0.0005 in GnomAD populations and a CADD phred score > 20 according to CADD v1.6 ($N=6$). Characteristics of *ATM* variant carriers with a description of the variant are provided in Table 1.

The control series was composed of 489 FFPE breast tumors from female noncarriers of *ATM* variants identified in the MCCS ($N=440$) and ABCFS ($N=49$) studies.

None of the *ATM* variant carriers and noncarriers included in the study carried a known pathogenic or likely pathogenic variant in the following BC susceptibility genes: *BRCA1*, *BRCA2*, *BARD1*, *BRIP1*, *CDH1*, *CHEK2*, *MRE11A*, *NBN*, *PALB2*, *PTEN*, *RAD50*, *RAD51C*, *RAD51D*, *STK11* and *TP53*.

DNA preparation

Tumor enriched DNA was prepared from each BC tumor as described in Wong et al. 2015 [34]. For each FFPE tumor sample, a hematoxylin and eosin (HE) or hematoxylin-eosin-safran (HES) stained slide marked up by a pathologist was used as a reference slide to delimit tumor-enriched areas. DNA was extracted from areas with ≥ 50% tumor content when possible. These areas were macro-dissected from at least two 8µm corresponding methyl green stained sections. Tumor tissues were first incubated with Proteinase K for 48h (replenished with 20µL at t = 24h), then DNA was extracted using the QIAamp DNA FFPE Tissue Kit following manufacturer's instructions (QIAGEN, Hilden, Germany). The tumor DNA was eluted twice in 15µL elution buffer to obtain a final volume of 30µL, and DNA concentration was assessed using the Qubit dsDNA BR assay (Life Technologies, Carlsbad, CA, USA).

Genome-wide methylation profiling

Genome-wide methylation profiling of all samples was performed in the Precision Medicine molecular genomics facility at Monash University. We used a previously described in-house experimental workflow to perform sodium bisulfite conversion and DNA restoration on FFPE tumor samples [34]. Briefly, FFPE tumor DNA quality was assessed by quantitative-PCR (qPCR) using the Infinium HD QC assay (Illumina, San Diego, CA, USA). All samples were assayed in duplicate. The non–bisulfite converted, unrestored U266 multiple myeloma cell line DNA was used as a negative control. The difference in quantification cycle (C_q) value (ΔC_q) was determined by subtracting the average C_q value of each DNA sample from the average C_q value of the negative control. All samples with a $\Delta C_q \geq 4$ were progressed through the sodium bisulfite conversion. Depending on the DNA concentration, a minimum of 50ng of DNA and a maximum of 750ng of DNA was sodium bisulphite converted using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine, CA, USA) and was restored using the Infinium HD FFPE DNA Restore kit (Illumina, San Diego, CA, USA). Successful conversion and restoration were verified by qPCR using a primer pair specific for bisulfite-converted DNA. Tumor DNA samples amplified at least 4 cycles earlier than the negative control (non-converted DNA) were progressed to the whole-genome amplification and the hybridization onto the Illumina Infinium HumanMethylationEPIC (EPIC) (all CoF-AT and GENESIS and 15 ABCFS samples) or the Illumina Infinium HumanMethylation450K (450K) BeadChips (43 ABCFS samples and all MCCS samples) (Illumina, San Diego, CA, USA).

The EPIC and 450K assays were performed on tumor DNA samples as per the manufacturer's instructions. The Freedom EVO automated liquid handler (TECAN, Månnedorf, Switzerland) was used for extension and staining steps. The BeadChips were scanned using the iScan machine (Illumina, San Diego, CA, USA). Raw methylation data, corresponding to the red and green signals measured for each probe, were stored in.IDAT files.

RNA sequencing

Tumor RNA was extracted from tumor-enriched areas delimited from the most representative HES slides of the FFPE block (with ≥ 50% tumor content when possible). Three 10-µm-thick sections from each block were macrodissected and tumor RNA was isolated using the NucleoSpin Tissue protocol which includes a DNA digestion step (Macherey-Nagel, Düren, Germany). RNA concentration and RNA purity were assessed using the Nanodrop instrument (Nanodrop, Indianapolis, In, USA). To assess RNA quality, the DV_{200} (percentage of RNA fragments ≥ 200 bp) was measured using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Capture-based libraries were prepared using the TruSeq RNA Exome kit (Illumina, San Diego, CA, USA) from an input of 100ng of tumor RNA as recommended by the manufacturer's instruction. Capture of coding RNA was then paired-end sequenced on the HiSeq2500 instrument (Illumina).

Methylation data preprocessing

Methylation data were pre-processed using the Bioconductor package minfi (version 1.40.0) [35]. The pipeline is detailed in Figure S1. Briefly, only probes present on both EPIC and 450K BeadChips were considered in the analyses. Samples with 10% of probes with a detection p-value > 1% and probes with a detection p-value > 1% in at least 10% of the control samples were excluded. Cross-reactive probes and polymorphic probes, mapping to cytosine or guanine with single nucleotide polymorphisms on either strand were removed [36, 37]. To normalize DNA methylation data across samples, the functional normalization [38] was then applied using the preprocessFunnorm function from *minfi* with default parameters. Probes on sexual chromosomes were excluded.

DNA methylation level of each CpG site was calculated as "β values" (corresponding to $\frac{\text{methylated signal}}{\text{methylated + unmethylated signals}}$). β values range from 0 to 1, where 0 indicates unmethylated and 1 fully methylated CpG. Because β values have severe heteroscedasticity for highly methylated or unmethylated CpGs sites [39], M values corresponding to $\log_2(\frac{\beta}{1-\beta})$, were used for the subsequent statistical analyses, after removal of probes with generated infinite values.

To obtain the mean level of methylation of each promoter, probes were mapped to promoters defined as between 1.5 kb upstream and 500 bp downstream the first base of the first exon of the genes according to the GRCh38.p5 human genome assembly using bedtools [40]. β values of probes belonging to same promoter were averaged (NB: for genes less than 500 bp long, only the upstream part was considered as promoter). Gene name symbols were recovered using the biomaRt R package [41].

Unsupervised clustering

Unsupervised clustering was performed with the R umap package [42] (version 0.2.7) using the 10,000 most variable probes based on interquartile ranges. The Pearson2 distance was used to calculate the distance between data points.

Identification of differentially methylated promoters

Differentially methylated (DM) promoters were identified by moderated t-statistics. A linear model was first fitted for each promoter using the lmFit function of the limma R package [43], with the number of probes mapped to each promoter being used as weight. T-statistics were then computed with the eBayes function and multiple testing correction was performed using the Benjamini and Hochberg method [44]. An adjusted p-value of 0.05 and an absolute $\log_2$(fold change) (log2FC) of 1 (corresponding to a 2-fold difference in methylation level) were used as thresholds to identify DM promoters, thus defining hypomethylated (log2FC < 1) and hypermethylated (log2FC > 1) promoters.

Gene set enrichment analyses

Gene Set Enrichment Analyses (GSEA) [45] were performed with the clusterProfiler R package (version 4.2.2) [46] using gene promoters ranked according to log2FC x the number of probes mapped to each promoter, and pathways annotation from the KEGG database (https://www.kegg.jp/brite/br08901) [47–49]. P-values were corr ected for false discovery rate (FDR) using the Benjamini-Hochberg correction, with the significance threshold set to 0.05. These analyses resulted for each significant pathway in an Enrichment Score (ES), a normalized ES (NES) and a core enrichment list of genes contributing the most to the pathway enrichment. The EnrichmentMap Cytoscape app [50, 51] was used for enrichment map visualization. Similarity between two gene sets was calculated using a composite score that integrated Jaccard and overlap metrics, resulting in a network visualized with the "prefused force directed layout".

Table 2 Comparison of clinical and histological characteristics of ATM and non-ATM tumors, after quality controls

Characteristics	ATM tumors ($N=32$)		Non-ATM tumors ($N=484$)
ATM variant type	**PV ($N=27$)**	**VUS ($N=5$)**	**–**
Age at diagnosis			
Mean [range]	45.6 [28–75]‡	55.0 [41–66]	61.5 [25–82]
SD	11.1	11.5	11.1
Year of diagnosis			
Mean [range]	2003 [1989–2017][&]	2007 [2004–2009]	2000 [1992–2016]
Median	2006	2007	2000
ER status			
Positive	23 (85.2%)[¶]	4 (80.0%)	350 (72.3%)
Negative	3 (11.1%)	0 (0%)	107 (22.1%)
Unknown	1 (3.7%)	1 (20%)	27 (5.6%)
Grade			
I	3 (11.1%)	1 (20%)	99 (20.5%)
II	12 (44.4%)[§]	4 (80%)	186 (38.4%)
III	9 (33.3%)[§]	0 (0%)	171 (35.3%)
Unknown	3 (11.1%)	0 (0%)	28 (5.8%)

‡Significantly different from non-ATM tumors (Wilcoxon test, adjusted p-value: 1.2×10^{-9}). [&] Significantly different from non-ATM tumors (Wilcoxon test, adjusted p-value: 1.8×10^{-2}). [¶]Proportion significantly different between ATM and non-ATM tumors (right-tailed two-proportions z-test, adjusted p-value: 0.12). [§]Proportion significantly different between ATM and non-ATM tumors (right-tailed two-proportions z-test, adjusted p-value: 0.21). SD: standard deviation; ER: estrogen receptors; PV: pathogenic variant; VUS: variant of uncertain clinical significance

Identification of biomarkers predictive of tumors arising in *ATM* PV carriers

A machine learning approach was used to identify a set of promoters whose methylation status was predictive of tumors arising in *ATM* PV carriers. Three models were used for classification of the tumors and features selection with a stratified 4-fold procedure: a logistic regression, a random forest and the XGBoost [52] model. Analyses were performed with Python version 3.9.12 and scikit-learn version 1.0.2. Datasets being imbalanced between tumors from patients carrying and not carrying *ATM* PVs, classification performances were reported using the Matthews Correlation Coefficient (MCC) score [53] (Supplementary Material).

Transcriptome analysis

RNA-Seq data were pre-processed using the RNA-Seq pipeline version 4.1.0 developed at Institut Curie, (https://zenodo.org/records/13744441). The main steps involved: (1) the identification and the suppression of ribosomal RNA reads with the bowtie1 aligner, (2) the alignment on the hg38 reference genome of the remaining reads with STAR, and (3) the generation of count tables with STAR. Transcripts Per Million (TPM) were then used to assess the correlation between gene expression and gene promoter methylation for the samples for which both types of data were available, and the Pearson correlation coefficient was calculated.

Results

Histopathological and clinical features of investigated tumor series

After pre-processing and quality control (QC) of DNA methylation data, 32 out of 35 tumors from *ATM* variant carriers (Table 1) and 484 out of 489 tumors from noncarriers were kept in the analyses. Clinical and histological characteristics of these tumors are presented in Table 2. The mean age at BC diagnosis was 45.6 years (range: 28–75) for *ATM* PV carriers and 61.5 years (range: 25–82) for noncarriers. The difference in age at diagnosis between the two groups was significant (Wilcoxon test, adjusted p-value: 1.2×10^{-9}). The mean time between tumor sampling and methylation measurement (age of the tumor blocks) was also significantly different between carriers and noncarriers. Indeed, BCs were diagnosed between 1989 and 2017 for *ATM* PV carriers and between 1992 and 2016 for noncarriers (Wilcoxon test,

adjusted *p*-value: 1.8×10^{-2}). Therefore, we first checked that age at diagnosis and age of the tumor block do not represent confounding factors when comparing methylation profile of ATM and non-ATM tumors. As no clustering of the tumors according to these two variables was observed, we concluded that these factors were unlikely to bias the subsequent analyses (Figure S2).

Among tumors with known ER/grade status, 88.4% of tumors of *ATM* PV carriers and 76.6% of tumors of noncarriers were ER+, 87.5% of tumors of *ATM* PV carriers and 78.3% of tumors of noncarriers were of grade II or III. These proportions were not significantly different (right-tailed two-proportions z-test, adjusted *p*-value: 0.12 for ER status and 0.21 for grade).

Because ER+ and ER- tumors showed different genome-wide methylation profiles (Figure S3), we focused the subsequent analyses on ER + tumors (25 ATM tumors and 359 non-ATM tumors). Of note, no batch effect was observed between ATM and non-ATM tumors nor between the French series and the Australian series (Figure S4).

ATM promoter is more frequently hypermethylated in ATM tumors than in non-ATM tumors

We first examined the methylation status of *ATM* promoter in the tumors, as this epigenetic event could be implicated in the inactivation of the second allele of *ATM* in tumors of heterozygous PV carriers, following the Knudson two-hit hypothesis for tumor suppressor genes [54]. In tumors of noncarriers, the mean methylation level of *ATM* promoter was -4.15 (standard deviation: 0.51). Using this mean methylation level as a reference, we found that *ATM* promoter was more frequently hypermethylated in tumors of heterozygous PV carriers (13/21 tumors) than in tumors of noncarriers (58/350), and that this difference was significant (Wilcoxon test, adjusted *p*-value: 8.6×10^{-7}, Fig. 1A). After exclusion of 6 tumors of heterozygous carriers showing LOH at the *ATM* locus (Table 1), *ATM* promoter was found hypermethylated in 53.3% (8/15) of tumors of *ATM* PV carriers, and the level of methylation of *ATM* promoter remained significantly higher than in non-ATM tumors (Wilcoxon test, adjusted *p*-value: 2.1×10^{-4}) (Fig. 1B). Interestingly, *ATM*

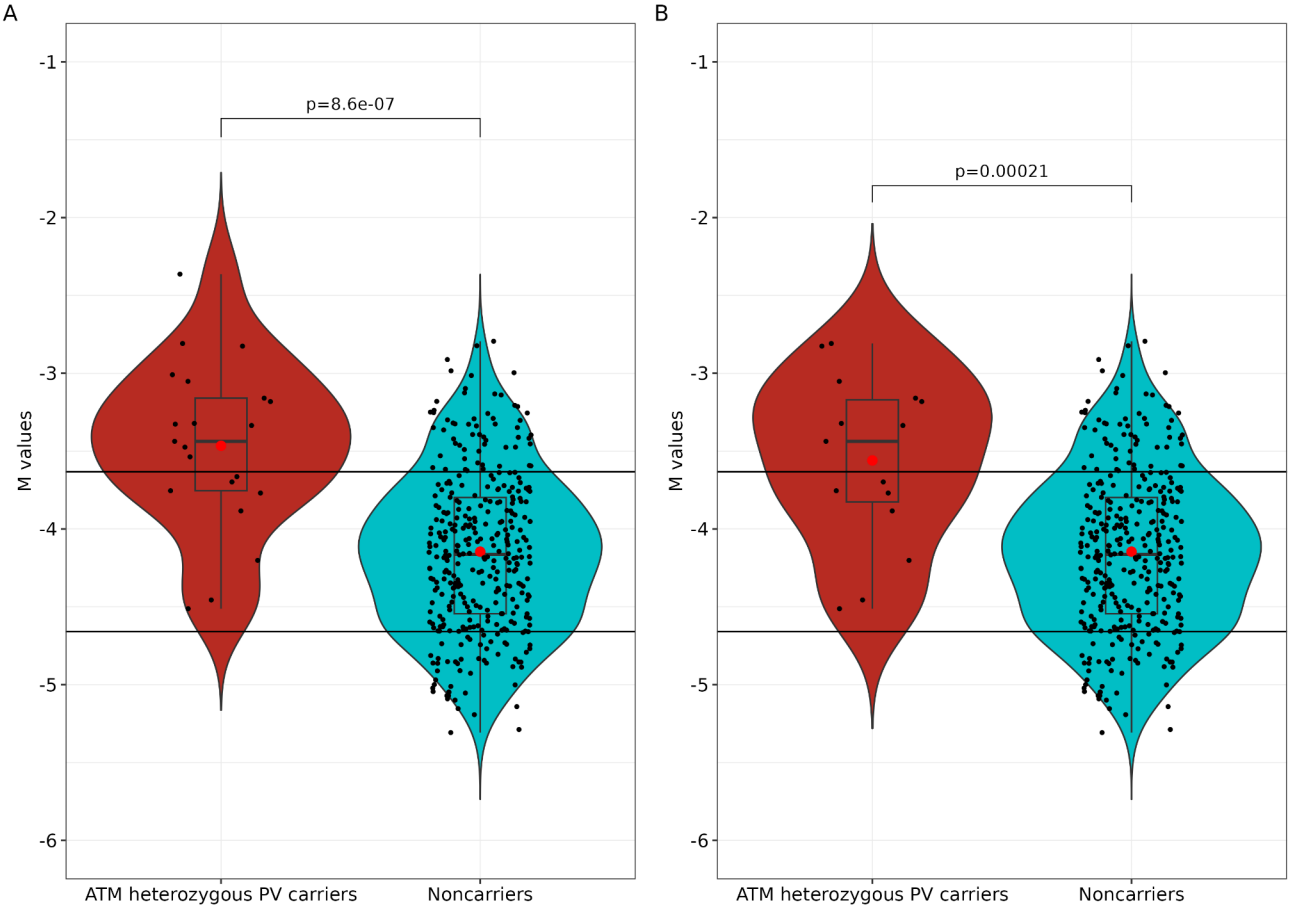

Fig. 1 *ATM* promoter is hypermethylated in ER+ tumors of *ATM* heterozygous pathogenic variant carriers. (**A**) All tumors of *ATM* heterozygous PV carriers were considered in the analysis ($N = 21$). (**B**) Only tumors where LOH at the *ATM* locus had not been reported were considered in the analysis ($N = 15$). Red dots indicate the mean of *ATM* promoter methylation for each tumor category; the horizontal bars indicate the mean ± the standard deviation of the *ATM* promoter methylation of non-carriers

promoter was also hypermethylated in breast tumors of carriers of VUS c.1009C>T, c.5882A>G and c.6059G>T, which could be used as evidence for pathogenicity of these variants (Table S1).

Because we could not verify in any of the tumors that the higher methylation level affected the wild-type allele, we conducted two types of analyses in the remaining part of the work: the main analyses compared ER+ tumors of PV carriers to ER+ tumors of noncarriers, and the secondary analyses compared only tumors of PV carriers with (probable) inactivation of the second *ATM* allele in the tumor, either because they were tumors of A-T patients, or tumors showing LOH at the *ATM* locus, or tumors showing hypermethylation of *ATM* promoter (Table S1).

Genome-wide DNA promoter methylation analysis

After preprocessing and QCs of the methylation data, 118,796 probes mapped to 22,337 gene promoters (Figure S1).

In the main analysis, we found 327 promoters differentially methylated between ER+ tumors of PV carriers and noncarriers, with 238 (72.8%) of them being hypomethylated in ATM tumors (Table S2). In this analysis, the log2FC for *ATM* promoter was 0.66, which means that *ATM* promoter is 1.6 times more methylated in tumors from *ATM* variant carriers than in tumors from noncarriers. The heatmap built using the 327 DM promoters allowed to cluster all ATM tumors. Tumors of carriers of the VUS c.1009C>T, c.5882A>G, c.1464G>T and c.6059G>T clustered among tumors of PV carriers, which would be also an indication in favor of pathogenicity (Fig. 2).

In the analysis restricted to ATM tumors with probable biallelic inactivation of *ATM*, the number of DM promoters increased to 773, which supports the Knudson second-hit hypothesis in ATM tumors. Among them, 315 (41%) promoters overlapped with those identified in the main analysis (Table S2).

Correlation between promoter methylation and gene expression in ATM tumors

Since DNA methylation is one of the main epigenetic mechanisms for the regulation of gene expression, we next investigated the potential correlation between promoter methylation level and TPM values of expressed genes in the 10 tumors of *ATM* PV carriers for which RNA-Seq data and methylation data were available (annotated in Table S1). In the main analysis, only DM promoters of genes *SCGB3A1*, *CYBRD1*, *ARHGAP40* and *GJA1* showed high negative correlation with gene expression in the tumor (absolute $r > 0.7$ and p-value < 0.05) (Figure S5.A). When restricting the analysis to tumors with probable biallelic inactivation of *ATM*, *SCGB3A1*, *CYBRD1* and 22 additional genes showed negative correlation with gene expression and five genes with positive correlation with gene expression (Figure S5.B). Hence these combinations of DNA methylation/gene expression markers may represent good candidate biomarkers of tumors arising in *ATM* PV carriers.

Enriched pathways in ATM tumors

To gain a biological systems-level understanding of the changes in methylation between tumors arising in *ATM* and non-*ATM* PV carriers, we next performed a GSEA to identify biological pathways enriched in DM genes. Out of the 357 tested KEGG pathways, 47 were significantly enriched in ER+ tumors of *ATM* PV carriers (Fig. 3 and Table S3). Forty-two pathways were significantly enriched when restricting the analysis to tumors with probable biallelic inactivation of *ATM*, including 39 pathways common two both analyses. Among enriched pathways in which *ATM* is part of the core enrichment, top pathways included *homologous recombination* (hsa03440), *cell cycle* (hsa04110), *platinum drug resistance* (hsa01524), *cellular senescence* (hsa04218), *p53 signaling* (hsa04115), *shigellosis* (hsa05131), *Human papillomavirus infection* (hsa05165) and *Human T-cell leukemia virus 1 infection* (hsa05166).

Enrichment maps built using the core enrichment promoters of these pathways highlighted two "master pathways" enriched in ER+ tumors of *ATM* PV carriers, the largest one regrouping 15 pathways involved in several cancer types and viral infection, the second one regrouping 8 pathways involved in neurodegenerative diseases or metabolic functions known to be involved in cancer development (Figure S6).

DNA methylation biomarker panels predictive of ATM tumors

We next employed three machine learning methods (logistic regression, random forest and XGBoost) to identify a set of promoters allowing to discriminate ATM tumors from non-ATM tumors (feature selection procedure detailed in Figure S7 and in Supplementary Material). For each repetition and classifier, identified promoters allowed to predict tumors arising in *ATM* PV carriers and non-*ATM* PV carriers of the validation set with precision, recall, f1 score, MCC and specificity equal or above 0.8 (Table S4). In the main analysis, the classifier based on logistic regression identified eight promoters (*INTS6P1*, *PTDSS2*, *RPL36AP30*, *SCAPER*, *ARF4*, *AMPD3*, *FLT4* and *POLR2L*). The classifier based on XGBoost identified, in addition to these eight promoters, promoters of *SNORA14A*, *RFX1*, *AADACL4* and *PDIA3P2*, while the classifier based on random forest identified a list of 21 promoters with only promoters of *INTS6P1*, *ARF4*, *PDIA3P2*, *RPL36AP30*, *SCAPER*,

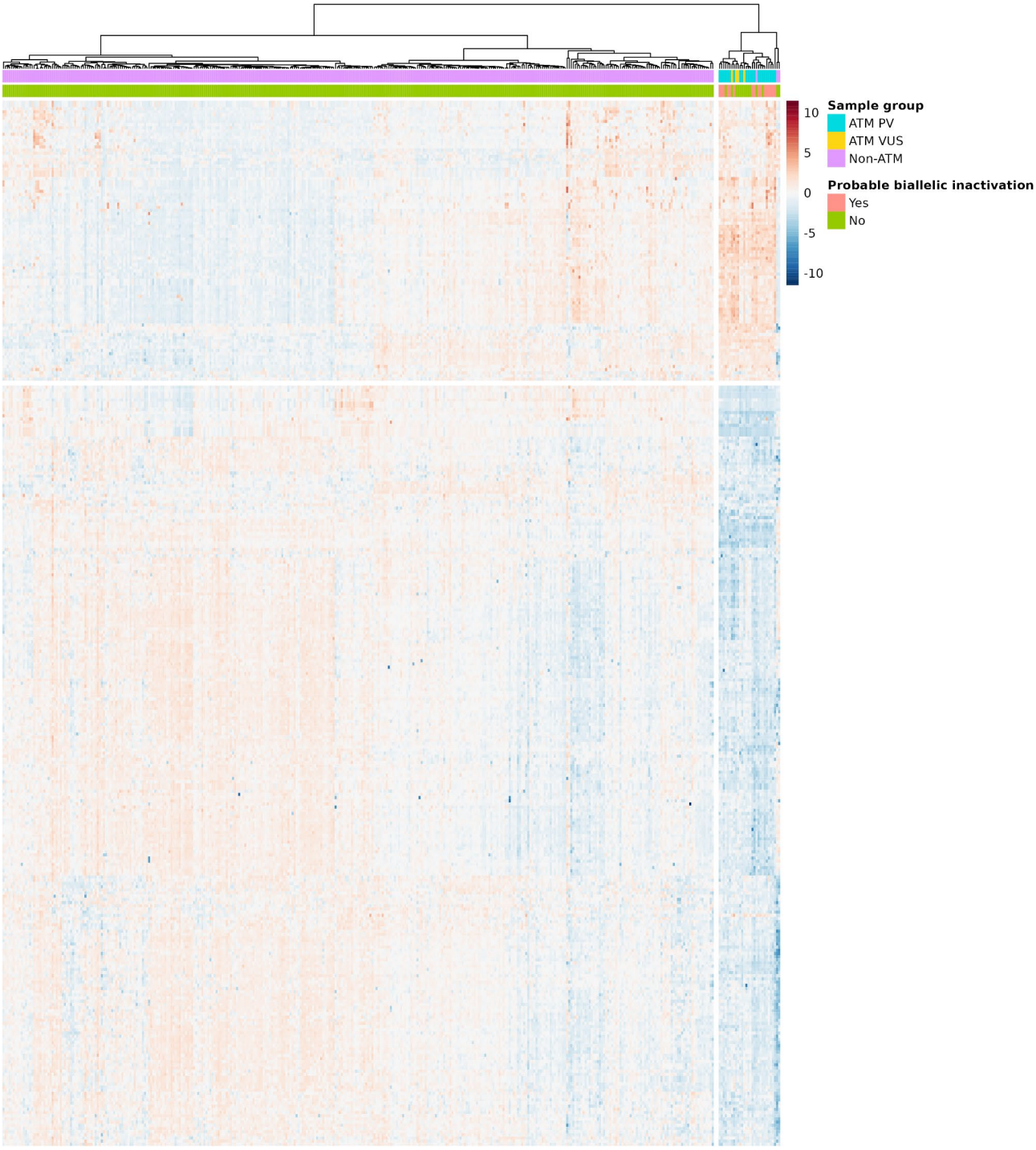

Fig. 2 Tumors from patients with and without *ATM* PVs display a different genome-wide methylation pattern

POLR2L present in the former lists. The good classification performances obtained with these three lists of genes are illustrated in Fig. 4.

To ensure that the identified biomarkers were not randomly selected by the models, we compared these results to the classification performance of 20 randomly selected promoters, repeating the analysis 1000 times. The MCC mean of all 1000 repetitions was below 0.7, confirming that the DM promoters selected by the three machine learning approaches are biomarkers specific to ATM tumors.

Using the respective three lists of genes identified by the three ML methods, the classifiers based on logistic regression and on random forest models classified

Fig. 3 Pathways enriched in genes showing aberrant methylation of promoters in ATM tumors

tumors from carriers of VUS c.1009C>T, c.5882A>G, c.1464G>T and c.6059G>T as ATM tumors, while the classifier based on XGBoost model classified the tumor of carrier of VUS c.6059G>T as a non-ATM tumor in 1 out of the 6 repetitions, and the tumor of carrier of VUS c.5882A>G as a non-ATM tumor in 5 out of the 6 repetitions (Table S5 and Table S6). However, heatmaps built with these biomarkers showed that tumors of the four VUS carriers clustered with tumors of PV carriers, which again provides evidence in favor of the pathogenicity of the variants (Fig. 4). Interestingly, in each of these visualizations, the non-ATM tumor ABCFS17 clustered with ATM tumors, and was also predicted as an ATM tumor when used in the validation set by logistic regression and random forest models (Table S6), which suggests a somatic biallelic inactivation of *ATM* in this tumor.

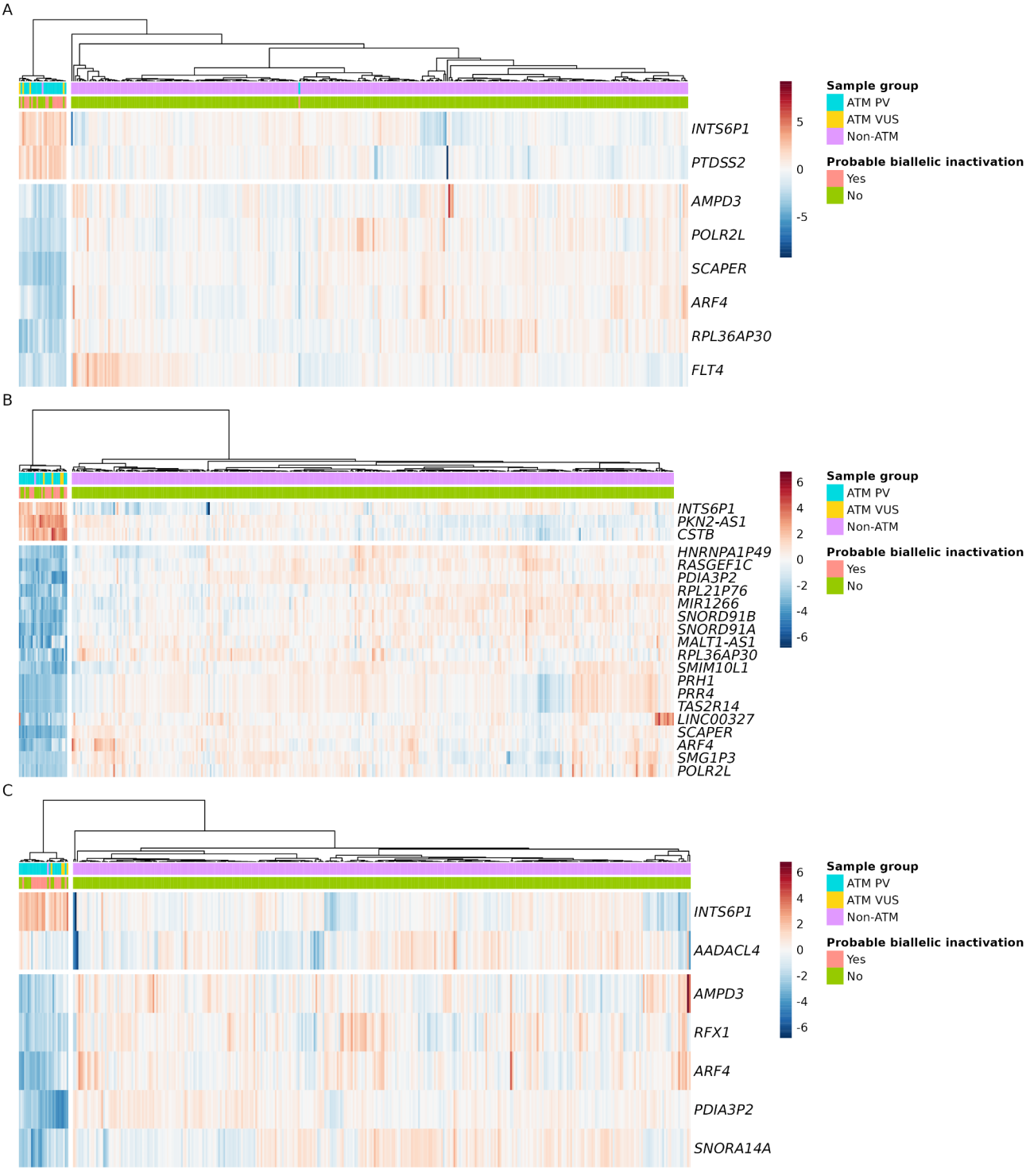

Fig. 4 Methylation of promoters selected by ML methods allow to cluster ER+ tumors from *ATM* variant carriers. (**A**) Logistic regression. (**B**) random forest. (**C**) XGBoost

Discussion

This study reports the first characterization of the DNA methylation profile of tumors developed by women with *ATM* PVs. These tumors were compared to tumors arising in women without *ATM* PVs participating in a population-based study of BC (ABCFS) and a prospective cohort (MCCS) with the same clinical and histopathological information.

Because 88% of the investigated tumors arising in women with *ATM* PVs were ER+ (in accordance with

previous findings [11, 12, 18–21]), and because we confirmed that the ER status is a confounding factor in the genome-wide DNA methylation analysis of breast tumors, we focused our analyses on this group of ER+ tumors. We showed that tumors of *ATM* PV carriers had a global hypermethylation of the *ATM* promoter as compared to tumors of noncarriers. Hence hypermethylation of *ATM* promoter may be useful as a potential new biomarker to identify BC tumors arising in *ATM* PV carriers. We could not further investigate here the correlation between *ATM* hypermethylation and *ATM* expression in ATM tumors because RNA-Seq data of the tumors without *ATM* PVs were not available to serve as control dataset. However, others reported an association between *ATM* promoter hypermethylation and lower expression of ATM mRNA in tumors from sporadic BC cases [55–57].

Although the use of DNA methylation arrays did not enable an allele-specific measurement of methylation, we can reasonably hypothesize that the observed hypermethylation of *ATM* promoter leads to the inactivation of the wild-type allele. Indeed, the secondary analyses performed on the subgroup of tumors with *ATM* PVs with a confirmed or probable biallelic inactivation of the gene identified 315 out of the 327 DM promoters (96.3%) of the main analysis and 458 new DM promoters, which is in agreement with our hypothesis that methylation alterations are accumulated in ATM-deficient breast tumors.

Activation of oncogenes and repression of tumor suppressor genes can be caused by aberrant hypo- or hypermethylation. In a meta-analysis investigating methylation profiles of normal and cancerous samples from multiple tissues from TCGA, a weak association between the hypermethylated signatures and gene expression repression was reported overall. High correlation between DNA methylation and gene expression variations was identified for a subset of genes in some specific cancer types making these methylation marks potentially important biomarkers for these cancers [58]. In line with these observations, among the genes with DM promoters between tumors with and without *ATM* PVs, those showing a high negative correlation between methylation and gene expression are of particular interest when searching for biomarkers of ATM deficiency. This was the case for *SCGB3A1* (*HIN-1*), *CYBRD1* (*DCYTB*), *ARHGAP40* identified as hypermethylated in both analyses, and *GJA1* identified only in the main analysis. *SCGB3A1* which encodes the Secretoglobin Family 3 A, Member 1 may play a tumor suppressor role in several cancers including breast cancer [59], prostate cancer [60], lung cancer [61] and non-small cell lung cancer [62], as its expression has been noted to be markedly lower in cancer tissues compared to normal tissue [63]. In BC, *SCGB3A1* gene methylation status has been proposed as a biomarker of prognosis and progression [64] and its promoter hypermethylation was found significantly associated with ER + and progesterone receptor (PR) + tumors [65, 66]. *CYBRD1* which encodes the cytochrome b reductase 1, was found to be a prognostic predictor for BC and may "retard cancer progression by reducing activation of FAK, a kinase that plays a central role in tumor cell adhesion and metastasis" [67]. Its promoter hypermethylation and expression inhibition may be connected to a faster development of BC. Little is known about the role of *ARHGAP40* encoding the Rho GTPase activating protein 40 in cancer, but the hypermethylation of its promoter has been correlated with a loss of *ARHGAP40* expression in basal cell carcinoma [68]. Finally, in ER+ BC, a high expression of GJA1 has been associated with a better prognosis [69]. A link with *GJA1* expression and other cancers has also been reported. It was correlated with a higher level of immune infiltrating cells and a good prognostic in colorectal cancer [70]. However, in cervical cancer, *GJA1* expression has been associated with a poor survival [71]. Of note, GJA1 is also a key regulator of the pathogenesis of Alzheimer's disease [72], a pathways found highly enriched in ATM tumors. Our study suggests that the hypermethylation of *SCGB3A1*, *CYBRD1*, *ARHGAP40* promoters and the hypomethylation of *GJA1* are biomarkers of ATM tumors.

Regarding the 47 KEGG pathways found to be dysregulated by aberrant methylation patterns, all of them were found enriched in hypermethylated (and not in hypomethylated) promoters, in agreement with the common hypothesis that promoter hypermethylation inactivates tumor suppressor genes during tumorigenesis. Interestingly, the homologous recombination pathway was highly enriched in DM gene promoters in ATM tumors while in a previous study the mutational signature 3 associated with defective homologous recombination DNA (HRD) repair was not highlighted in tumors of *ATM* PV carriers [19]. The other pathways linked to the DNA Damage Response (DDR) which were also enriched in hypermethylated promoters in tumors arising in *ATM* PV carriers in our study were: the Fanconi anemia pathway, involved in repairing inter-strand crosslinks in DNA and critical for genome stability, the Tumor Necrosis Factor signaling pathway, which can promote or inhibit cancer progression, and the p53 signaling pathway recognized as a tumor suppressor pathway due to its central role in DDR by activating various downstream cellular processes involved in DNA repair, cell cycle arrest and apoptosis. Furthermore, enrichment map based on these enriched pathways illustrates that the specific methylation pattern of ATM tumors leads to pleiotropic defects altering known biological functions involving ATM such as cancerous, neurodegenerative and metabolic functions. It has been long demonstrated that ATM is central in

cancer development [9, 73] through its role of sensing DNA damages and activating cascades of effectors of the DDR response [74]. Some viral infections being known to induce DNA double-strand breaks and several types of cancers [75], it is not surprising to find enriched pathways linked to viral infections. ATM is also activated by oxidative stress to maintain redox homeostasis through the regulation of central carbon metabolism. In neurological diseases, such as Alzheimer's or Huntington's diseases, the overactivation of the microglia releases reactive oxygen species which in turns activates ATM [76, 77]. Alterations of the neurodegenerative and metabolic/ROS functions are indeed related to symptoms of A-T patients [76].

Finally, ML approaches allowed to identify a set of 27 promoters that are highly discriminative of tumors developed by *ATM* PV carriers. Among those, *INTS6P1* and *ARF4* were found by the three ML methods employed. Four of the 27 DM genes (*AMPD3*, *ARF4*, *PTDSS2*, *TAS2R14*) are linked to metabolism and viral infections and 7 are linked to neoplastic mechanisms (*ARF4*, *CSTB*, *FLT4*, *INTS6P1*, *MIR1266*, *RFX1*, *TAS2R14*), thus related to the KEGG pathways found enriched in the GSEA. However, no correlation was observed between gene promoter methylation and gene expression for these 27 genes in our dataset. Furthermore, these biomarkers may help in the classification of VUS. Interestingly, one tumor that is not known to carry an *ATM* PV showed a methylation pattern resembling that of *ATM* PV carriers, suggesting a somatic biallelic alteration of *ATM*. These results will need to be replicated in an independent dataset or with functional analyses to determine their clinical relevance, some drugs being already available such azathioprine and mercaptopurine to target *AMPD3*, a member of nucleoside metabolic pathways.

The main limitation of this study is the lack of a replication dataset. We attempted to replicate our findings in the TCGA-BRCA [78] dataset, in which we identified 30 tumors from *ATM* variant carriers (9 PV and 21 VUS) and 478 tumors from noncarriers, but no promoters were differentially methylated between the two groups, which did not confirm our results. This non replication may be due to the protocols used for samples preparation: in our study, DNAs were prepared from tissue samples enriched in tumoral cells which was not the case for TCGA-BRCA samples. Thus, micro-environmental/normal tissue cells in TCGA-BRCA samples may have attenuated the signal of the epigenetic modifications due to ATM inactivation detected in tumoral cells in our study.

While we provide supplementary information for the characterization of breast tumors developed by *ATM* variant carriers, the genome-wide DNA methylation pattern of ATM tumors may not be sufficient to capture their full biological complexity, and other types of alterations, such as in gene expression and post-transcriptional modifications should be investigated. The integration of different biological omics data may help in understanding this biology and bring out stronger biological signals that may not be detectable with the analysis of a single omics layer or single genomics approach (i.e. gene panel sequencing) [79].

Conclusion

To conclude, breast tumors from *ATM* PV carriers have a recognizable genome-wide DNA methylation pattern which targets genes involved in neoplastic, neurodegenerative and metabolic-related pathways, in which *ATM* is also involved.

Three ML methods identified a panel of methylation biomarkers that can be helpful in the identification of ATM-deficient breast tumors outside a familial genetic context (A-T or HBOC families) because DNA methylation pattern alterations are one of the earliest modifications occurring in tumorigenesis. Moreover, such biomarkers may represent specific therapeutic targets. Additional functional analyses or replicating studies are needed to assess the relevance of these biomarkers and pathways as potential therapeutic targets.

Abbreviations

A-T	Ataxia-telangiectasia
ABCFS	Australian Breast Cancer Family Study
ATM	Ataxia-Telangiectasia Mutated
BC	Breast cancer
CoF-AT2	French prospective cohort on families segregating an *ATM* variant
CADD	Combined Annotation Dependent Depletion
C_q	Quantification cycle
DCIS	Ductal carcinoma in situ
DDR	DNA damage response
DM	Differentially methylated
ER+	Estrogen receptor-positive
FFPE	Formalin-fixed, paraffin-embedded
GENESIS	GENE SISters study
GSEA	Gene Set Enrichment Analysis
HRD	Homologous recombination deficiency
IDC	Infiltrating ductal carcinoma
ILC	Infiltrating lobular carcinoma
log2FC	Log_2(fold change)
LOH	Loss of heterozygosity
MCC	Matthews correlation coefficient
MCCS	Melbourne Collaborative Cohort Study
ML	Machine learning
PR+	Progesterone receptor-positive
PV	Pathogenic variant
TCGA	The Cancer Genome Atlas
TPM	Transcripts per million
VUS	Variant of uncertain clinical significance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-025-01988-w.

Supplementary Material 1: Procedure used to identify biomarkers of ATM PV carrying tumors.

Supplementary Material 2: **Table S1**: LOH at ***ATM*** locus and methylation status of ***ATM*** promoter in tumors of ***ATM*** variant carriers. **Table S2**:

Methylation level of promoters in ATM tumors as compared to non-ATM tumors (expressed as Log2FC) and correlation with expression of promoters. **Table S3**: KEGG pathways significantly enriched in the gene set enrichment analyses. **Table S4**: Classification performances of tumors. **Table S5**: Classification performances of tumors from *ATM* VUS carriers and noncarriers. **Table S6**: Predictions of tumors from *ATM* VUS carriers and noncarriers.

Supplementary Material 3: **Figure S1**: Preprocessing pipeline of methylation data using minfi. **Figure S2**: **Age and year of diagnosis of breast cancer do not represent confounding variables.**: Violin plots (A and C) showing the distributions of ages at diagnosis (A) and year of diagnosis (C) for *ATM* PV carriers and noncarriers. Statistical tests were performed with a Wilcoxon test (*p*-value indicated in each plot), the normality being rejected with a Shapiro test. UMAPs (B and D) were performed using the M values of the 10,000 most variables probes and colored according to the ages at diagnosis (B) and year of diagnosis of the tumors (C). **Figure S3**: **Genome-wide methylation profile of ER-positive and ER-negative tumors is different.**: Uniform manifold approximation and projection (UMAP) unsupervised clustering was performed using the 10.000 most variable probes based on interquartile ranges, between the 32 tumors from *ATM* variant carriers and the 488 tumors from noncarriers. **Figure S4**: **No batch effect exists between ATM and non-ATM tumors nor between French and Australian series.**: ATM tumors are circled while non-ATM tumors are not. **Figure S5**: **Correlation between promoter methylation and gene expression in tumors of PV carriers**: A. all ER+ tumors of ATM PV carriers. B. ER+ tumors of ATM PV carriers with a possible biallelic inactivation of ATM. **Figure S6**: Biological links between KEGG pathways enriched in tumors developed by ATM pathogenic variant carriers. **Figure S7**: Procedure used to identify biomarkers of ATM tumors.

Acknowledgements
We wish to thank all study participants. We acknowledge Mamy Andrianteranagna and Elodie Girard for technical expertise. We thank all GENESIS, CoF-AT2, ABCFS and MCCS investigators and participants without whom this research would not have been possible. We thank also Tú Nguyen-Dumont for the verification of variant carriers identified in the MCCS and ABCFS studies.

Author contributions
ALR, NMV, FL conceived and designed the study. FL supervised the study. SEM, LF, DLG, EC, MGD, JB, AVS, DSL, MCS, NA acquired the data (e.g. invited and managed patients, centralized and managed data, constructed databases). NMV, ALR, VR, SMEH, YJ, NA, FL analyzed and interpreted data (molecular analysis, statistical analysis, bioinformatics). NV, FL wrote the manuscript. ALR, YJ, SMEH, NA, MCS reviewed and revised the manuscript. All authors read and approved the final version of the manuscript.

Funding
This work was supported by Fondation ARC pour la recherche sur le cancer (grant PJA 20151203365), La Ligue comité de Paris (grant RS16/75 – 72) and Institut National du Cancer (grant INCa-9578). NMV was the recipient of a PhD fellowship from PSL Research University. MCS is supported by a L3 NHMRC investigator grant (GNT2017325). The Australian Breast Cancer Family Registry (ABCFR) is supported by the US National Institutes of Health (grant number 2U01CA164920). The ABCFR was supported in Australia by the National Health and Medical Research Council (NHMRC), the New South Wales Cancer Council, the Victorian Health Promotion Foundation, the Victorian Breast Cancer Research Consortium, Cancer Australia, and the National Breast Cancer Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry. RNA-Seq of ATM tumors was performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the ITMO-Cancer Aviesan (Plan Cancer III) and by the SiRIC-Curie program (SiRIC Grant INCa-DGOS-465 and INCa-DGOS-Inser-12554). Data management, quality control and primary analysis were performed by the Bioinformatics platform of the Institut Curie.

Data availability
Data presented in this report can be requested to the corresponding author for the French studies CoF-AT and GENESIS and via PEDIGREE (https://www.cancervic.org.au/research/epidemiology/pedigree) for the Australian studies MCCS and ABCFS. Results obtained on the TCGA dataset are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Declarations

Ethics approval and consent to participate
CoFAT2 and GENESIS: Written informed consent for genetic studies and use of medical records for the present analyses was obtained from all participants. The appropriate local ethics committee (Comité de Protection des Personnes [CCP] Ile-de-France III 2002/2006) and the French data protection authority (Commission Nationale de l'Informatique et des Libertés [CNIL]) approved the study protocols CoF-AT2 and GENESIS, the individual resource collections, and the specific study on tumor material. **MCCS**: Written informed consent was obtained from all participants to collect a blood sample and tumor pathology materials. The study protocols were approved by Human Research Ethics Committee at the Cancer Council Victoria (MCCS). **ABCFS**: The study was approved by the Human Research Ethics Committee of the University of Melbourne (12496). Informed consent was obtained from all participants involved in the study.

Competing interests
The authors declare no competing interests.

Author details
[1]Inserm, U1331, Institut Curie, PSL University, Mines ParisTech, Paris, France
[2]Monash University, Clayton, VIC; University of Melbourne, Parkville, VIC, Australia
[3]Service de Pathologie, Institut Curie, Paris, France
[4]ICGex Next-Generation Sequencing Platform, Institut Curie, PSL University, Paris, France
[5]Service de Génétique, Institut Curie, Université Paris Cité, Inserm, U830, Paris, France
[6]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia

Received: 18 November 2024 / Accepted: 27 February 2025
Published online: 11 March 2025

References

1. Rothblum-Oviatt C, Wright J, Lefton-Greif MA, McGrath-Morrow SA, Crawford TO, Lederman HM. Ataxia telangiectasia: a review. Orphanet J Rare Dis. 2016;11:159.
2. Amirifar P, Ranjouri MR, Yazdani R, Abolhassani H, Aghamohammadi A. Ataxia-telangiectasia: A review of clinical features and molecular pathology. Pediatr Allergy Immunol. 2019;30:277–88.
3. Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia–telangiectasia. N Engl J Med. 1991;325:1831–6.
4. Peterson RD, Funkhouser JD, Tuck-Muller CM, Gatti RA. Cancer susceptibility in ataxia-telangiectasia. Leukemia. 1992;6(Suppl 1):8–13.
5. Olsen JH, Hahnemann JM, Borresen-Dale A-L, Brondum-Nielsen K, Hammarstrom L, Kleinerman R, et al. Cancer in patients with Ataxia-Telangiectasia and in their relatives in the nordic countries. JNCI J Natl Cancer Inst. 2001;93:121–7.
6. Micol R, Ben Slama L, Suarez F, Le Mignot L, Beauté J, Mahlaoui N, et al. Morbidity and mortality from ataxia-telangiectasia are associated with ATM genotype. J Allergy Clin Immunol. 2011;128:382–e3891.
7. Suarez F, Mahlaoui N, Canioni D, Andriamanga C, Dubois d'Enghien C, Brousse N, 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;33:202–8.

8. Lindahl H, Svensson E, Danielsson A, Puschmann A, Svenningson P, Tesi B, et al. The clinical spectrum of ataxia telangiectasia in a cohort in Sweden. Heliyon. 2024;10:e26073.
9. The Breast Cancer Susceptibility Collaboration (UK), Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38:873–5.
10. Swift M, Morrell D, Cromartie E, Chamberlin AR, Skolnick MH, Bishop DT. The incidence and gene frequency of ataxia-telangiectasia in the united States. Am J Hum Genet. 1986;39:573–83.
11. Breast Cancer Association Consortium, Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, et al. Breast Cancer risk Genes — Association analysis in more than 113,000 women. N Engl J Med. 2021;384:428–39.
12. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A Population-Based study of genes previously implicated in breast Cancer. N Engl J Med. 2021;384:440–51.
13. Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-Panel sequencing and the prediction of Breast-Cancer risk. N Engl J Med. 2015;372:2243–57.
14. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast Cancer. Am J Hum Genet. 2009;85:427–46.
15. Girard E, Eon-Marchais S, Olaso R, Renault A, Damiola F, Dondon M, et al. Familial breast cancer and DNA repair genes: insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing. Int J Cancer. 2019;144:1962–74.
16. Choi M, Kipps T, Kurzrock R. ATM mutations in cancer: therapeutic implications. Mol Cancer Ther. 2016;15:1781–91.
17. Lesueur F, Easton DF, Renault A-L, T SV, B JL, K-J Z, et al. First international workshop of the ATM and cancer risk group (4–5 December 2019). Fam Cancer. 2022;21:211–27.
18. Renault A-L, Mebirouk N, Fuhrmann L, Bataillon G, Cavaciuti E, Le Gal D, et al. Morphology and genomic hallmarks of breast tumours developed by ATM deleterious variant carriers. Breast Cancer Res. 2018;20:1–18.
19. Weigelt B, Bi R, Kumar R, Blecua P, M DL, G FC, et al. The landscape of somatic genetic alterations in breast cancers from ATM germline mutation carriers. JNCI: J Natl Cancer Inst. 2018;110:1030–4.
20. Southey MC, Dowty JG, Riaz M, Steen JA, Renault A-L, Tucker K, et al. Population-based estimates of breast cancer risk for carriers of pathogenic variants identified by gene-panel testing. NPJ Breast Cancer. 2021;7:153.
21. Breast Cancer Association Consortium, Mavaddat N, Dorling L, Carvalho S, Allen J, González-Neira A, et al. Pathology of tumors associated with pathogenic germline variants in 9 breast Cancer susceptibility genes. JAMA Oncol. 2022;8:e216744.
22. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in Cancer. Nucleic Acids Res. 2019;47:D941–7.
23. Polak P, Kim J, Braunstein LZ, Tiao G, Karlic R, Rosebrock D, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49:1476–86.
24. Shamma A, Suzuki M, Hayashi N, Kobayashi M, Sasaki N, Nishiuchi T, et al. ATM mediates pRB function to control DNMT1 protein stability and DNA methylation. Mol Cell Biol. 2013;33:3113–24.
25. Janin N, Andrieu N, Ossian K, Lauge A, Croquette MF, Griscelli C, et al. Breast cancer risk in ataxia telangiectasia (AT) heterozygotes: haplotype study in French AT families. Br J Cancer. 1999;80:1042–5.
26. Cavaciuti E, Lauge A, Janin N, Ossian K, Hall J, Stoppa-Lyonnet D, et al. Cancer risk according to type and location of ATM mutation in ataxia-telangiectasia families. Genes Chromosom Cancer. 2005;42:1–9.
27. Renault A-L, Mebirouk N, Cavaciuti E, Le Gal D, Lecarpentier J, d'Enghien CD, et al. Telomere length, ATM mutation status and cancer risk in Ataxia-Telangiectasia families. Carcinogenesis. 2017;38:994–1003.
28. Sinilnikova OM, Dondon M-G, Eon-Marchais S, Damiola F, Barjhoux L, Marcou M, et al. GENESIS: a French National resource to study the missing heritability of breast cancer. BMC Cancer. 2016;16:13.
29. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The breast Cancer family registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res. 2004;6:R375–389.
30. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, et al. Cohort profile: the Melbourne collaborative cohort study (Health 2020). Int J Epidemiol. 2017;46:1757–i1757.
31. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MRE, et al. Familial risks, Early-Onset breast cancer, and BRCA1 and BRCA2 germline mutations. JNCI J Natl Cancer Inst. 2003;95:448–57.
32. Renault A-L, Dowty JG, Steen JA, Li S, Winship IM, Giles GG, et al. Population-based estimates of age-specific cumulative risk of breast cancer for pathogenic variants in ATM. Breast Cancer Res. 2022;24:24.
33. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–94.
34. Wong EM, Joo JE, McLean CA, Baglietto L, English DR, Severi G, et al. Tools for translational epigenetic studies involving formalin-fixed paraffin-embedded human tissue: applying the infinium HumanMethyation450 Beadchip assay to large population-based studies. BMC Res Notes. 2015;8:543.
35. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
36. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.
37. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the illumina methylationepic BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208.
38. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:1–17.
39. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11:587.
40. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
41. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart. Nat Protoc. 2009;4:1184–91.
42. Konopka T. umap: Uniform Manifold Approximation and Projection. 2020 [cited 2021 Jun 25]. Available from: https://CRAN.R-project.org/package=umap
43. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47–47.
44. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. J R Stat Soc B. 1995;57.
45. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102:15545–50.
46. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (N Y). 2021;2:100141.
47. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
48. Kanehisa M. Toward Understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947–51.
49. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587–92.
50. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE. 2010;5:e13984.
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
52. Chen T, Guestrin C, XGBoost:. A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM; 2016 [cited 2024 Aug 8]. pp. 785–94. Available from: https://doi.org/10.1145/2939672.2939785
53. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21:6.
54. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci. 1971;68:820–3.

55. Vo QN, Kim W-J, Cvitanovic L, Boudreau DA, Ginzinger DG, Brown KD. The ATM gene is a target for epigenetic Silencing in locally advanced breast cancer. Oncogene. 2004;23:9432–7.
56. Rondeau S, Vacher S, De Koning L, Briaux A, Schnitzler A, Chemlali W, et al. ATM has a major role in the double-strand break repair pathway dysregulation in sporadic breast carcinomas and is an independent prognostic marker at both mRNA and protein levels. Br J Cancer. 2015;112:1059–66.
57. Begam N, Jamil K, Raju SG. Promoter hypermethylation of the ATM gene as a novel biomarker for breast Cancer. Asian Pac J Cancer Prev. 2017;18:3003–9.
58. Moarii M, Boeva V, Vert J-P, Reyal F. Changes in correlation between promoter methylation and gene expression in cancer. BMC Genomics. 2015;16:873.
59. Haakensen VD, Bjøro T, Lüders T, Riis M, Bukholm IK, Kristensen VN, et al. Serum estradiol levels associated with specific gene expression patterns in normal breast tissue and in breast carcinomas. BMC Cancer. 2011;11:332.
60. Zhang P, Chen T, Yang M. Comparative analysis of prognosis and gene expression in prostate cancer patients with site-specific visceral metastases. Urologic Oncology: Seminars and Original Investigations. 2024;42:160.e1-160.e10.
61. Castro M, Grau L, Puerta P, Gimenez L, Venditti J, Quadrelli S, et al. Multiplexed methylation profiles of tumor suppressor genes and clinical outcome in lung cancer. J Transl Med. 2010;8:86.
62. Yu Y, Yin D, Hoque MO, Cao B, Jia Y, Yang Y, et al. AKT signaling pathway activated by HIN-1 methylation in non-small cell lung cancer. Tumour Biol. 2012;33:307–14.
63. Krop I, Player A, Tablante A, Taylor-Parker M, Lahti-Domenici J, Fukuoka J, et al. Frequent HIN-1 promoter methylation and lack of expression in multiple human tumor types. Mol Cancer Res. 2004;2:489–94.
64. Dai D, Dong X-H, Cheng S-T, Zhu G, Guo X-L. Aberrant promoter methylation of HIN-1 gene May contribute to the pathogenesis of breast cancer: a meta-analysis. Tumour Biol. 2014;35:8209–16.
65. Callahan CL, Wang Y, Marian C, Weng DY, Eng KH, Tao M-H, et al. DNA methylation and breast tumor clinicopathological features: the Western new York exposures and breast Cancer (WEB) study. Epigenetics. 2016;11:643–52.
66. Nomair AM, Ahmed SS, Mohammed AF, El Mansy H, Nomeir HM. SCGB3A1 gene DNA methylation status is associated with breast cancer in Egyptian female patients. Egypt J Med Hum Genet. 2021;22:66.
67. Lemler DJ, Lynch ML, Tesfay L, Deng Z, Paul BT, Wang X, et al. DCYTB is a predictor of outcome in breast cancer that functions via iron-independent mechanisms. Breast Cancer Res. 2017;19:25.
68. Yu N, Yuan B, Cai J, Liu J, Zhang W, Bao W, et al. Loss of ARHGAP40 expression in basal cell carcinoma via CpG Island hypermethylation. Exp Dermatol. 2023;32:2094–101.
69. Busby M, Hallett MT, Plante I. The complex Subtype-Dependent role of connexin 43 (GJA1) in breast Cancer. Int J Mol Sci. 2018;19:693.
70. Hu W, Li S, Zhang S, Xie B, Zheng M, Sun J, et al. GJA1 is a prognostic biomarker and correlated with immune infiltrates in colorectal Cancer. Cancer Manage Res. 2020;12:11649–61.
71. Meng S, Fan X, Zhang J, An R, Li S. GJA1 expression and its prognostic value in cervical Cancer. Biomed Res Int. 2020;2020:e8827920.
72. Kajiwara Y, Wang E, Wang M, Sin WC, Brennand KJ, Schadt E, et al. GJA1 (connexin43) is a key regulator of Alzheimer's disease pathogenesis. Acta Neuropathol Commun. 2018;6:144.
73. Ahmed M, Rahman N. ATM and breast cancer susceptibility. Oncogene. 2006;25:5906–11.
74. Blackford AN, Jackson SP. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. Mol Cell. 2017;66:801–17.
75. Schiller JT, Lowy DR. An introduction to virus infections and human Cancer. Recent Results Cancer Res. 2021;217:1–11.
76. Pizzamiglio L, Focchi E, Antonucci F. ATM protein kinase: old and new implications in neuronal pathways and brain circuitry. Cells. 2020;9:1969.
77. Choy KR, Watters DJ. Neurodegeneration in ataxia-telangiectasia: multiple roles of ATM kinase in cellular homeostasis. Dev Dyn. 2018;247:33–46.
78. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.
79. Sun YV, Hu Y-J. Integrative analysis of Multi-omics data for discovery and functional studies of complex human diseases. Adv Genet. 2016;93:147–90.

Publisher's note