

Discovery of a Pathogenic Variant rs139379666 (p. P2974L) in *ATM* for Breast Cancer Risk in Chinese Populations



Xingyi Guo¹, Weiqiang Lin^{2,3}, Mengqiu Bai^{2,3}, Hongzhi Li⁴, Wanqing Wen¹, Chenjie Zeng¹, Zhishan Chen¹, Jing He¹, Jianghua Chen³, Qiuyin Cai¹, Jirong Long¹, Wei-Hua Jia⁵, Xiao-Ou Shu¹, and Wei Zheng¹

Abstract

Background: Pathogenic variants in susceptibility genes lead to increased breast cancer risk.

Methods: To identify coding variants associated with breast cancer risk, we conducted whole-exome sequencing in genomic DNA samples from 831 breast cancer cases and 839 controls of Chinese women. We also genotyped samples, including 4,580 breast cancer cases and 6,695 controls, using whole exome-chip arrays. We further performed a replication study using a Multi-Ethnic Global Array in samples from 1,793 breast cases and 2,059 controls. A single marker analysis was performed using the Fisher exact test.

Results: We identified a missense variant (rs139379666, P2974L; AF = 0.09% for breast cancer cases, but none for controls) in the *ATM* gene for breast cancer risk using combing data from 7,204 breast cancer cases and 9,593 controls ($P = 1.7$

$\times 10^{-5}$). To investigate the functionality of the variant, we first silenced *ATM* and then transfected the overexpression vectors of *ATM* containing the risk alleles (TT) or reference alleles (CC) of the variant in U2OS and breast cancer SK-BR3 cells, respectively. Our results showed that compared with the reference allele, the risk allele significantly disrupts the activity of homologous recombination-mediated double-strand breaks repair efficiency. Our results further showed that the risk allele may play a defected regulation role in the activity of the *ATM* structure.

Conclusions: Our findings identified a novel mutation that disrupts *ATM* function, conferring to breast cancer risk.

Impact: Functional investigation of genetic association findings is necessary to discover a pathogenic variant for breast cancer risk.

Introduction

Breast cancer is one of the leading causes of cancer morbidity and mortality in the world. Genetic factors contribute to the pathogenesis of both sporadic and familial breast cancer (1). The genetic abnormalities of cancer predisposition genes have been known to contribute to breast cancer (2, 3). Over the past

few decades, multiple breast cancer susceptibility genes harboring deleterious or protein-truncating variants have been identified. Those include *BRCA1*, *BRCA2*, *ATM*, *TP53*, *CHEK2*, *PALB2*, *CDH1*, *STK11*, *NF1*, and *PTEN* (3–16). Of these, the *ATM* gene has been identified as a moderate penetrance gene that contributes to breast and other cancer types (14–19). The *ATM* gene, functioning as a protein kinase, plays an essential role in regulating DNA double-stranded break repair pathways and cell-cycle checkpoints in breast cancer (20, 21). Using gene-panel or whole-exome sequencing (WES) data from large cohorts of patients with breast cancer, previous case-control studies have characterized more than 50 putative pathogenic variants in *ATM* in European populations, conferring approximately 3-fold to breast cancer risk (14–16). Further identification of novel variants associated with breast cancer risk is essential, especially in non-European populations, given different genetic architectures and environmental exposures. Although previously reported variants are statistically associated with breast cancer risk, in-depth functional investigation is necessary to identify putative pathogenic variants with biological evidence, providing strong candidates to develop genetic biomarkers for clinical utilities of personalized prevention or therapeutic targets.

In this study, we utilized the data resources from three study populations to search for coding variants associated with breast cancer risk. In the discovery stage, we conducted WES in blood samples from a total of 1,670 Chinese women, including 831 breast cancer cases and 839 controls in population I (see Materials

¹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee. ²The Kidney Disease Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ³Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China. ⁴Department of Bioinformatics, Beckman Research Institute of City of Hope, Duarte, California. ⁵Department of Experimental Research, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, Guangzhou, Guangdong, China.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Authors: Xingyi Guo, Vanderbilt University Medical Center, 2525 West End Ave., Ste. 330, Nashville, TN 37203. Phone: 615-936-3471; Fax: 615-332-0502; E-mail: xingyi.guo@vumc.org; and Weiqiang Lin, Zhejiang University School of Medicine, Hangzhou 310029, China. Phone: 868-697-1990; E-mail: wlin@zju.edu.cn

Cancer Epidemiol Biomarkers Prev 2019;28:1308–15

doi: 10.1158/1055-9965.EPI-18-1294

©2019 American Association for Cancer Research.

and Methods). We also genotyped 247,870 coding variants using the Illumina HumanExome BeadChip (or whole exome-chip arrays) in an independent set of samples from a total of 11,275 women, including 4,580 breast cancer cases and 6,695 controls in population II. In the replication stage, we conducted genotyping using the Multi-Ethnic Global Array (MEGA) in an independent set of samples from a total of 3,852 Chinese women, including 1,793 breast cancer cases and 2,059 controls in population III (Supplementary Table S1).

Materials and Methods

Study populations

Participants of this study comprise 7,204 breast cancer cases and 9,593 controls from three case-control studies, the Shanghai Breast Cancer Study, Shanghai Endometrial Cancer Study (controls only), and Guangzhou Colorectal Cancer Study (female controls only), and two cohort studies, the Shanghai Breast Cancer Survival Study (breast cancer cases only) and the Shanghai Women's Health Study. The two cohort studies provided nested case-control samples including 1,183 breast cancer cases and 5,464 controls matched by age at blood collection. Detailed descriptions of these cohort studies have been described in previous literature (22). Detailed descriptions of these participating studies are presented in Supplementary Table S1. Genomic DNA for all included participants was extracted using commercial DNA purification kits. All participants provided written informed consent, and the institutional review boards of all relevant institutes in both China and the United States approved the study protocols. In our association analyses (see below sections), we used both sequencing and chip-based genotyping for the independent samples from three study populations including study population I: WES for 831 cases and 839 controls; study population II: whole exome-chip arrays for 4,580 breast cancer cases and 6,695 controls; and MEGA for 1,793 breast cases and 2,059 controls.

WES data analysis

For study population I, we performed WES using the ILLUMINA GAII sequencing platform with paired-end reads in a length of 2×90 bp (average read depth ~ 50 M), which target >1 M genetic variants in all coding regions. The sequencing reads for each sample were mapped to the human reference genome (hg19) using the Burrows-Wheeler Aligner BWA program (version 0.75; ref. 23). Aligned reads were marked as duplicates to be removed using PICARD MarkDuplicates (<http://picard.sourceforge.net/>). The remaining aligned reads were further processed using the Genome Analysis ToolKit (GATKv3.2; ref. 24). We further performed multiple data processing, including local realignment (GATK RealignerTargetCreator and IndelRealigner) and base qualities recalibration (GATK TableRecalibration), following the GATK procedure. We evaluated the sequencing mapping quality, including mapping rate and coverage for each sample, using the QPLOT tool (25). Variant calling was performed individually on each sample using the GATK HaplotypeCaller tool. We next performed GenotypeGVCFs on variants for all samples together to create complete raw SNPs and indel Variant Call Format (VCF) files. Variant Quality Score Recalibration was further applied to filter variants of low quality that were possibly false positives. In addition, we removed variants with low depth of coverage (average < 8 per sample) and high missingness ($>2\%$). Principal

component analyses were conducted using EIGENSTRAT (26) to identify population outliers. We also estimated the pair-wise proportion of identity-by-descent to identify potentially genetically identical samples, unexpected duplicate samples, or close relatives. After filtering eight samples from the above quality control (QC) steps, we included a total of 1,670 samples from study population I for downstream analysis. Using the 1,000 Genomes Project data as reference, no population outliers were observed for these samples (Supplementary Fig. S1).

Genotyping in whole exome-chip arrays and MEGA

For study population II, we performed genotyping using the Illumina Exome array, an expanded Illumina HumanExome-12v1_A Beadchip, which included approximately 250 K genetic variants focused on protein-coding regions (details described in previous literature; ref. 27). For population study III, we performed genotyping using the Illumina MEGA-Expanded Array (Illumina Inc.), which covers >2.5 M genetic variants including common and additional promising variants discovered from study population I.

In both whole exome-chip arrays and MEGA, genotype calling was carried out using Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Raw data were imported into GenomeStudio. Genotypes were called using cluster definitions provided by Illumina. Cluster boundaries were determined using study samples. We further conducted QC using PLINK (28), and repeated the QC procedures conducted in WES. In addition, samples were excluded if: (i) the call rate $<98\%$, (ii) the consistence rates between the HapMap samples with 1000 Genomes data $<99\%$, (iii) they were a heterozygosity outlier, (iv) they were an ethnic outlier, (v) the samples were in close relationship, (vi) the consistency rates among duplicated samples $<99\%$, or (vii) the samples were of the wrong sex. From the above QC steps, we filtered approximately 1,000 samples using exome-chip platform from study population II, and 500 samples using MEGA platform from study population III, respectively. No population outliers were observed for the remaining samples (Supplementary Fig. S1).

Variant annotation, bioinformatics, and statistical analyses

The ANNOVAR tool was used to annotate coding variants (29). Five protein prediction algorithms, including Polyphen-2 HumDiv, Polyphen HumVar, Sorting Intolerant From Tolerant, logistic regression test scores and MutationTaster was applied to predict the possible impact of an amino acid substitution on the structure/function. For statistical analysis, we evaluated single mark association with breast cancer risk using the Fisher exact test from the additive regression model. The analysis was implemented in the R package.

Cell culture, plasmids, and antibody

U2OS cells containing the DR-GFP reporter [the gift from Dr. Yi Sun (Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China)], 293T and SK-BR3 cells were purchased from ATCC and maintained in DMEM supplemented with 10%. *Mycoplasma* testing is conducted routinely every month using a Mycoplasma Stain Assay Kit (Beyotime). FLAG-ATM and HA-ATM plasmid were gifts from Dr. Anyong Xie (Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China) and Yingli Sun (Beijing Institute

of Genomics, Chinese Academy of Sciences, Beijing, China). P2974L was generated by PCR-based site-directed mutagenesis. The PCR primers for mutagenesis were as follows: sense, AGAGGCTGGAAGATGAACTGAGCTTCAC; antisense, CATCTCCAGCCTCTGCTGTAATACAAAG. The antibodies used in this study include anti-ATM (Bethyl Laboratories A300-299A), anti-phospho-ATM (S1981; Abcam ab81292), anti-phospho-H2AX (S139; Cell Signaling Technology 9718), anti-phospho-CHK2 (T68; Cell Signaling Technology 2197), anti-phospho-p53 (S15; Cell Signaling Technology 9286), and anti-GAPDH (ProteinTech 60004-1-Ig).

shRNA endogenous *ATM*, semiquantitative RT-PCR, and immunoblot

The short hairpin RNA (shRNA) sequences for *ATM* were as follows: sense, CCGGGCTAAGTCACTGACCCATATTCTCGAGA-ATATGGGTCAGTACTTAGCTTTTGG; antisense, AATTCAAAAA-GCTAAGTCACTGACCCATATTCTCGAGAATATGGGTCAGTACTTAGC and cloned into pLKO.1 puro plasmid (Addgene 8453). The plasmid construction, viral production, and transfection were instructed by pLKO.1 protocol (<http://www.addgene.org/tools/protocols/plko/>). Briefly, the construction of the *ATM* shRNA plasmid and pLKO.1-TRC control plasmid (Addgene 10879), together with lentiviral packaging, envelope, and transfer plasmid psPAX2 and (Addgene 12260) and pMD2.G (Addgene 12259), were transfected to 293T cells using the Lipofectamine 3000 transfection reagent (Invitrogen L3000075). After two to three days of transfection, a culture supernatant of mtecs was collected. The viruses were concentrated by high-speed ultracentrifugation. The collected virus was further used for retroviral infection on a U2OS/DR reporter cell line. To confirm the knockdown efficiency for *ATM*, we performed semiquantitative RT-PCR and immunoblot. qPCR primers for *ATM* and GRAPDH were listed below: TGGGCATTACGGG-TGTTGAA; antisense primer for *ATM*, CTTCCGGCCTCTG-CTGAAA; sense primer for GRAPDH, ACCACAGTCCATGCCAT-CAC; antisense primer for GRAPDH: TCCACCACCCTGTGCTGT.

Homologous recombination assay

The silenced *ATM* in U2OS/DR reporter cell lines was transfected by pCDNA3.1, FLAG-*ATM*, and FLAG-*ATM*-PL plasmids. Control cells were transfected by pCDNA3.1. The I-SceI expression vector was transfected in all of the above cells at the same time. After they were transfected for three days, we collected cells and performed flow cytometry to analyze the efficiency of HR. The result was normalized by independently transfected cells from pEGFP-N1 (Clontech).

ATM knockout experiment in SK-BR3

The specific gRNA primers targeting *ATM* exon 9 were designed as follows: sense for *ATM* gRNA: CACCGAATGGAGACAGCTCA-CAGTT; antisense for *ATM* gRNA: AAACAACCTGTGAGCTGTCTC-CATTC. They were cloned into lentiCRISPR v2 plasmid (Addgene 52961). The procedures for viral production and transfection were described above. SK-BR3 cells were transfected with lentivirus, and selected with a specific amount of puromycin. Individual colonies were subjected to PCR-based sequencing using sense primer (GCTTACCCAGCTAGCCAAACG) and antisense primer (ATGGCTCCAAGTAAGCCAAAGT) to test whether the mutations occurred at the targeting region.

Immunofluorescence staining

The SK-BR3 cells were subjected to X-ray irradiation (PXI X-RAD 160), and the total IR amount reached 10 Gy every time. At 16 hours after IR, the cells were fixed with 4% paraformaldehyde for 20 minutes. The cells were then incubated with a rabbit polyclonal antibody against γ H2AX (CST 2577), following the protocol as described previously (Lin and colleagues, 2013). Nuclei were stained with DAPI. The slide was examined under a fluorescence microscope (Leica DM4000). The cell, which has at least 8 γ H2AX foci, was to be counted as γ H2AX positive (30).

The analysis of *ATM* protein structure

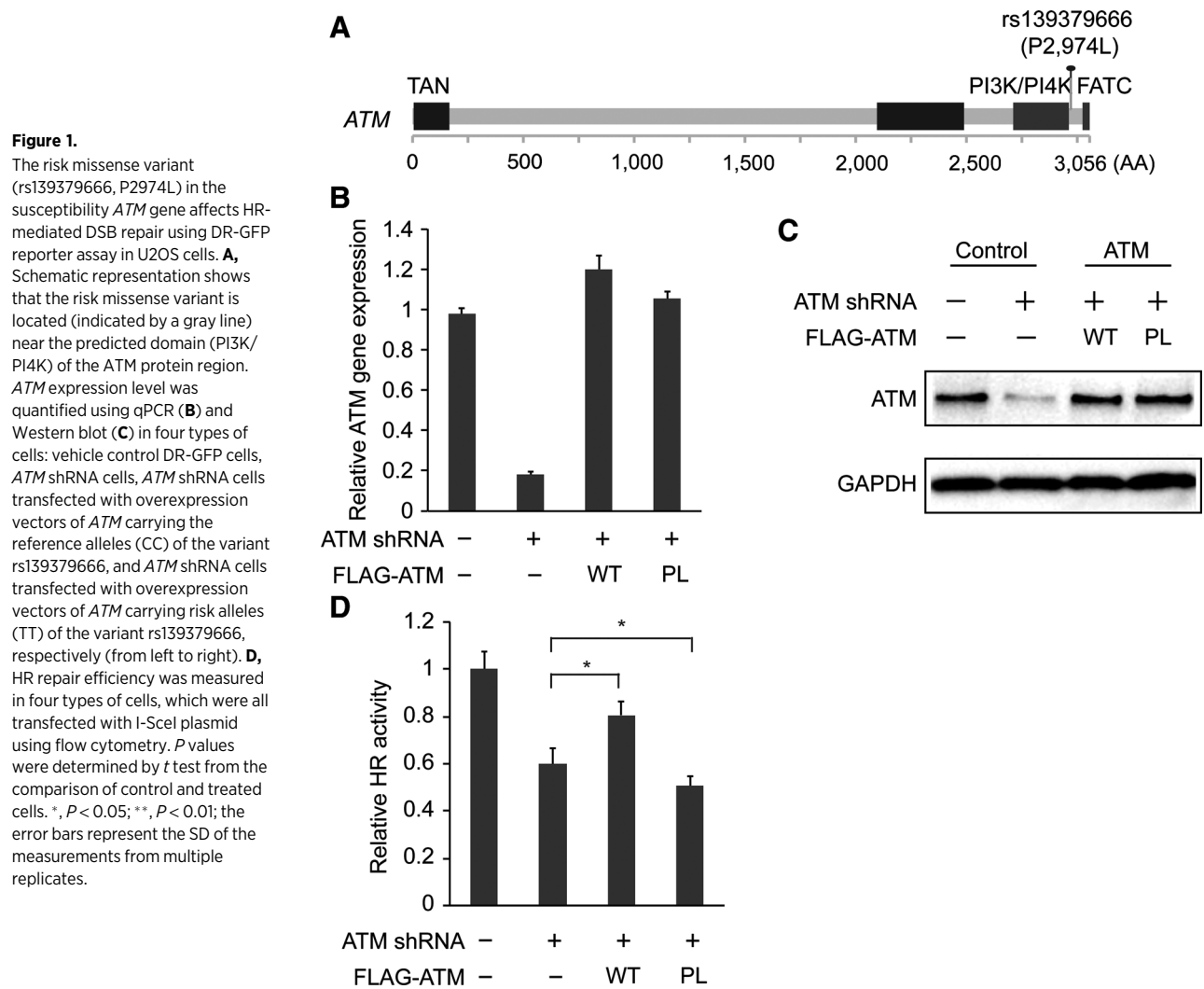
The structure of the close state of the *ATM* dimer was downloaded from RCSB PDB [id 5np0 (31), which is produced via electron microscopy at 5.7Å resolution]. The missing side-chains of the *ATM* structure were added by using a protein preparation wizard from Schrödinger software. The peptide substrate and the ATP molecule to display the kinase ATP-binding pocket were adapted from kinase CDK2 (PDB id 3qhw; ref. 32) by using the CE alignment algorithm in PyMol software (33).

Results

Discovering a missense variant (rs139379666, P2974L) in the *ATM* gene for breast cancer risk

After QC processes, we performed single marker analyses on rare coding variants for breast cancer risk. A missense variant (rs139379666, C/T, P2,974L) in the susceptibility *ATM* gene was observed to be associated with breast cancer risk. This missense variant was located near the PI3K/PI4K domain and predicted to be functionally deleterious by the MutationTaster tool (Fig. 1A). The risk allele (T) of this SNP was found in two of 831 breast cases, but in none of the 839 controls from the analysis of WES data in population I (Supplementary Fig. S2). We further examined the variant using whole exome-chip data from 4,580 breast cancer cases and 6,695 controls of Chinese women in population II (see Materials and Methods). The risk allele of this SNP was observed in eight of 4,580 breast cases, but in none of the 6,695 controls (Supplementary Fig. S2). Using a Fisher exact test, we estimated an association with $P = 1.6 \times 10^{-4}$ for breast cancer risk, after combining both data from population I and II (Table 1). We further replicated the finding using an independent set of samples from 1,793 breast cancer cases and 2,059 controls in population III. We observed that the risk allele (T) of this SNP was carried in three of the 1,794 breast cases, but in none of the 2,059 controls (Supplementary Fig. S2). We estimated the association with $P = 1.7 \times 10^{-5}$ for breast cancer risk, after combining data from all three populations (Table 1). In addition, we evaluated the association for the variant rs139379666 with breast cancer risk by using data only from early-onset patients with breast cancer (<45 years old). No significant statistical enrichment of this variant in *ATM* in early-onset patients with breast cancer was observed, compared with the observation in all breast cancer cases (Supplementary Table S2).

In addition, we examined allele frequency (AF) of this variant using the data from the reference controls from the genome Aggregation Database (gnomAD, <http://gnomad.broadinstitute.org/>) and the SG10K Consortium (bioRxiv 390070). In gnomAD, we observed that there is an AF = 8.3×10^{-5} of the variant



rs139379666 in all combined populations ($N = 138,574$), including Europeans, Asians, and Africans. Notably, we observed only two carriers in 4,932 females ($AF = 2.0 \times 10^{-4}$) from East Asian populations. We also examined the AF based on the whole genome sequencing data from SG10K (median read depth of $13.7\times$). We observed that this variant was not present in the healthy set of samples from 4,810 individuals of Chinese ($N = 2,780$), Malay, and Indian populations. Taken together, our results indicated that this variant may confer a risk of moderate to high for breast cancer.

The effects of the *ATM* missense variant (rs139379666, P2974L) on the activity of homologous recombination-mediated double-strand break repair efficiency

To explore the molecular mechanism of rs139379666 for breast cancer risk, we applied the well-established Direct Repeat (DR)-GFP reporter system to measure the efficiency of homologous recombination (HR) in the U2OS reporter cells. We applied the shRNA technique to knockdown the expression of endogenous *ATM* in U2OS cells (Fig. 1B). We observed that approximately 80% of the transcribed *ATM* was repressed in the cells (Fig. 1B).

Table 1. Association of a missense variant (rs139379666, P2974L) in *ATM* with breast cancer risk

Chr	Position(hg19)	Allele ^a	Frequency in cases ^b	Frequency in controls ^b	<i>P</i> ^c
Combing samples from population I & II (cases/controls: 5,411/7,534)					
11	108235879	C/T	0.09 (10/5,411)	0 (0/7,534)	1.6×10^{-4}
Samples from population III (cases/controls: 1,793/2,059)					
11	108235879	C/T	0.08 (3/1,793)	0 (0/2,059)	0.10
Combing samples from population I, II, & III (cases/controls: 7,204/9,593)					
11	108235879	C/T	0.09 (13/7,204)	0 (0/9,593)	1.7×10^{-5}

^aReference/risk allele; risk alleles are shown in bold.

^bRisk allele frequency (%).

^c*P* value derived from Fisher exact test under additive model.

We next introduced a fragment of the *ATM* gene containing the risk alleles (TT) or reference alleles (CC) of the variant into the *ATM* knockdown U2OS cells. Both qPCR and Western blot results showed that the *ATM* knockdown U2OS cells, with both exogenous *ATM* vector transfections, can restore the *ATM* expression to similar levels in the wild-type U2OS reporter cells (Fig. 1B and C). Using the DR-GFP reporter system, we found that the cells with transformed reference alleles can significantly recover the HR DNA repair, compared with the *ATM* knockdown cells (Fig. 1C). However, the HR DNA repair could not be recovered by the *ATM* knockdown cells with transformed risk alleles, and its repair efficiency was observed to be even lower than for the *ATM* knockdown cells (Fig. 1C). This observation is consistent with a previous study of mixed lineage leukemia (MLL) that there may be a dominant-negative effect of the repair ability of the pathogenic P2974L mutation in *ATM* (34).

To explore how the risk variant rs139379666 affects DNA repair ability, we conducted a CRISPR/CAS9 *ATM* knockout experiment on the breast cancer SK-BR3 cell line. The sequencing in a CRISPR/CAS9 *ATM* knockout colony showed that these colony cells carried three frameshift mutation patterns with one, eight, and 19 base-pairs in the ninth exon regions of *ATM* (Fig. 2A). Our Western blot results confirmed that the protein expression signal of *ATM* largely disappeared in these colony cells, when compared with the signal in the wild-type SK-BR3 cells (Fig. 2B). We next transfected a fragment of *ATM* containing the risk alleles (TT) or reference alleles (CC) of the variant into these colony cells. The Western blot results showed that the cells with both exogenous *ATM* vector transfections can restore *ATM* protein expression. These cells, together with wild-type SK-BR3 cells, were further subjected to irradiation (10 Gy), and later γ H2AX staining (phosphorylated form of H2AX immunofluorescence assay) was conducted in the cells at 16 hours after IR. Our γ H2AX staining experiments showed that *ATM* knockout dramatically increased the percentage of γ H2AX-positive cells (defined as the cell which had more than eight foci count), from 32% in wild-type cells to 52% in *ATM* knockout cells ($P < 0.01$; Fig. 2C and D), while transfected wild-type *ATM* expression in *ATM* knockout cells can more significantly decrease the percentage of γ H2AX-positive cells (33%), compared with *ATM* knockout cells ($P < 0.01$; Fig. 2C and D). However, transfected *ATM* expression containing the risk alleles in *ATM* knockout cells showed a significantly higher percentage of γ H2AX-positive cells than exogenous *ATM* expression containing the reference alleles in *ATM* knockout cells (43% vs. 33%, $P < 0.05$; Fig. 2C and D). This indicates that the risk alleles of *ATM* can lead to low repair efficiency for damaged DNA.

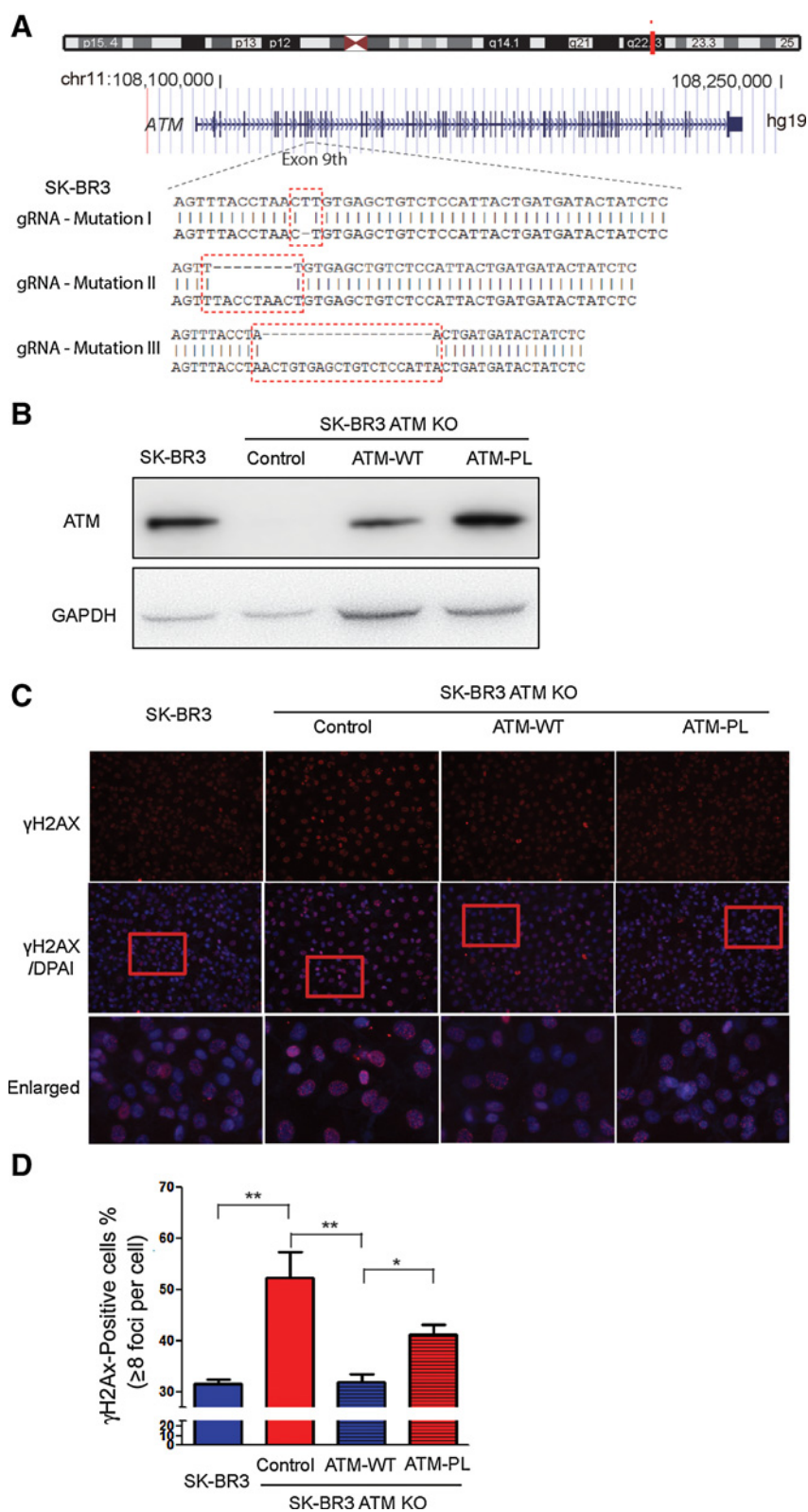
Discussion

In this study, we conducted WES, whole exome-chip arrays, and MEGA of 7,204 patients with breast cancer and 9,593 controls to search for additional coding variants associated with breast cancer risk. A putative pathogenic variant in the known susceptibility *ATM* gene, rs139379666, was discovered with strong statistical evidence for association with breast cancer risk. This newly identified variant may be further used to develop clinical utility for personalized prevention, as well as future potential therapeutic targets of breast cancer.

Baretic and colleagues determined the dimer structures of *ATM* in both closed state (substrate binding is blocked, PDB id 5np0)

and open state (substrate binding is allowed, PDB id 5np1) through electron microscopy at a resolution of 5.7Å (31). The key difference between closed and open state is the conformational change of the PIKK regulatory domain (PRD, with three helices, $\alpha 9b$, $\alpha 9c$, and $\alpha 9d$). In the closed state of the *ATM* dimer, the helices $\alpha 9c$ (2966-2974a.a.) and $\alpha 9d$ (2975-2979a.a.) on PRD is in closed conformation to block the peptide substrate-binding site, and thus, *ATM* is in an inactive state. In the dimer of open state, helix $\alpha 9d$ and most of $\alpha 9c$ are disordered to open conformation (Fig. 3A). Therefore, the peptide substrate can access the kinase ATP-binding pocket for phosphorylation. This is in agreement with the experimental data from previous literature that active *ATMs* do not necessarily require a complete dissociation of dimer to bind downstream kinases (35). As depicted in Fig. 3B, P2974 locates at the PRD, which is near the protein-protein interaction surface of the *ATM* dimer and is one of key regions for the regulation of *ATM* (also see Supplementary Video). The special cyclic structure of the proline amino acid of P2974 forms a kink in between the helices $\alpha 9c$ and $\alpha 9d$ in the closed state of the *ATM* dimer. In the closed state, the helices $\alpha 9c$ and $\alpha 9d$ on PRD, of which P2974 is in the middle, occupy the peptide substrate-binding site to disrupt the *ATM* regulation function. In the open state, the helices $\alpha 9d$ and $\alpha 9c$ are open to allow peptide substrate to access the kinase active site. In the closed state, P2974 is near a hydrophobic core that contains L2900 in helix α -AL on an activation loop and F3049 in helix α 12b of the FAT C-terminal. The hydrophobic interaction of L2900 of the A-loop and F3049 of FATC stabilizes the dimer interface and kinase active pocket. When P2974 is mutated from a weak hydrophobic proline to a strong hydrophobic leucine, it will not only change the conformation of the kink helices $\alpha 9c$ and $\alpha 9d$ to affect dimer interface, but most importantly, it will also enhance the binding of PRD to the L2900/F3049 pair via the hydrophobic interaction of mutated leucine. Therefore, a P2974L mutation may cause more difficulty to the *ATM* protein to leave the closed state than wild-type. Consequently, P2974L-mutated *ATMs* may lose part of its phosphorylation activity to regulate downstream kinases, as most of them are trapped in the closed states. This can explain the dominant negative effect of the P2974L mutant: the mutated *ATM* has defected regulation activity, although it may still form a dimer with other *ATM* proteins. To verify this, we performed additional experiments by transfecting them with a fragment of *ATM* containing the risk alleles (TT) and reference alleles (CC) of the variant into 293T cells, individually. At three days after transfection, the cells subjected to IR (5 Gy) were collected for Western blot assay at 20 and 50 minutes after IR. The Western blot results showed that transfected wild-type *ATM* expression can increase the phosphorylation of *ATM* downstream substrates (*ATM* S1981, γ H2AX, CHK2 T68, and TP53 S15) at two different times. However, transfected *ATM* expression containing the risk alleles can dramatically reduce the phosphorylation of *ATM* downstream substrates compared with control cells at both timepoints. When the P2974L-mutant *ATM* proteins are introduced from out-resources, the downstream substrate, such as p53 proteins, will have more of a chance to bind with the defected *ATMs* and cannot be phosphorylated (Fig. 3C). Thus, the overall activities of downstream substrates will be dropped, as observed in a previous experiment (34).

It should be noted that the *in vitro* silencing and knockout experiments were quite challenging due to multiple copies of the *ATM* gene in the breast cancer cells. For example, we observed

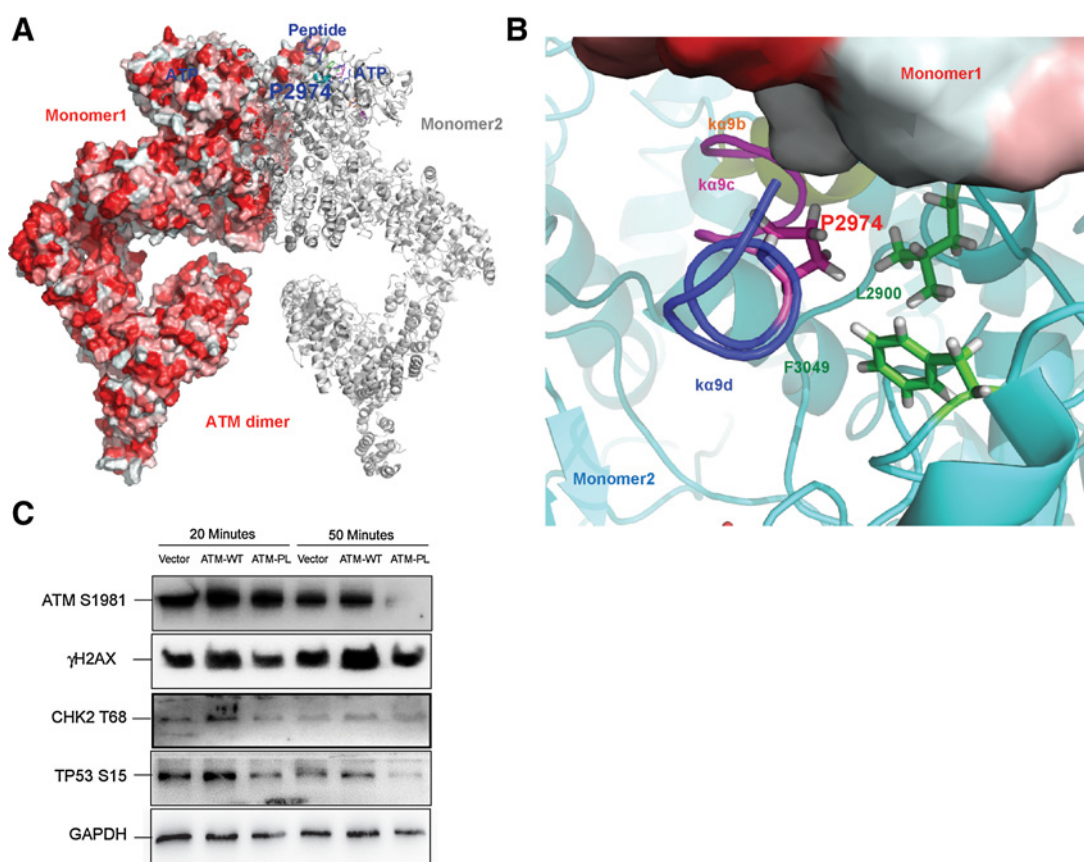
**Figure 2.**

The effects of the ATM missense variant on the DNA damage repair efficiency using an *in vitro* Gamma-H2AX assay in a breast cancer SK-BR3 cell line. **A**, Establishment of *ATM* knockout breast cell line. The #6 colony from the single cell transfected with a specific *ATM* gRNA sequence, targeted to the ninth Exon, was extracted, and a specific PCR was conducted using the genomic DNA as the template. The results of the sequencing of the subclones are listed. The *ATM* expression level was quantified by Western blot (**B**) in four types of cells (**C**): vehicle control SK-BR3 cells, *ATM* knockout cells, *ATM* knockout cells transfected with overexpression vectors of *ATM* carrying the reference alleles (CC) of the variant rs139379666, and *ATM* knockout cells transfected with overexpression vectors of *ATM* carrying risk alleles (TT) of the variant rs139379666, respectively (from left to right). **C**, Immunofluorescence staining of γ H2AX was conducted in four different types of cells at 16 hour after IR. **D**, Quantity of γ H2AX-positive cells. The cell that carries at least eight γ H2AX foci was counted as a γ H2AX-positive cell. *P* values were determined by *t* test from the comparison of control and treated cells. *, *P* < 0.05; **, *P* < 0.01; The error bars represent the standard deviation of the measurements from multiple replicates.

approximately four copies existing in SK-BR3 cells, based on PCR-based DNA sequencing for the *ATM* genomic sequence (Fig. 2A). We actually performed the *ATM* knockdown and knockout in the

breast cancer cells MCF7. Unfortunately, we failed to generate high efficiency of silencing and knockout *ATM* in the MCF7 cells. Using *ATM* knockout experiments based on CRISPR in SK-BR3

Guo et al.

**Figure 3.**

The risk missense variant (rs139379666, P2974L) in the susceptibility *ATM* gene predicted to affect the ATM activity from the examined protein structure. **A**, Dimer structure of ATM. One monomer is shown as a hydrophobic surface, with the hydrophobic surface displayed in red. The other monomer is shown as a gray cartoon. The substrate peptides and P2974 residues are on top and near the dimer interface. **B**, P2974 in close state of ATM dimer. P2974 is at a kink point of helices $\alpha 9c$ (pink) and $\alpha 9d$ (blue). It is near the hydrophobic core of L2900 in helix α -AL on an activation loop and F3049 in helix α 12b of FAT C-terminal. **C**, Protein expressions signals measured by Western blot analysis with conditions: vehicle control SK-BR3 cells, *ATM* knockout cells transfected with overexpression vectors of *ATM* carrying the reference alleles (CC) of the variant rs139379666, and *ATM* knockout cells transfected with overexpression vectors of *ATM* carrying risk alleles (TT) of the variant rs139379666, in 20 and 50 minutes.

cells and exogenous *ATM* introduction, we showed that the risk allele significantly disrupts the activity of HR-mediated double-strand breaks (DSB) repair efficiency when compared with the reference allele. Nevertheless, further *in vitro* in additional breast cancer cells or *in vivo* functional assays are needed to further explore the functional consequences of this particular variant in *ATM* that confer breast cancer risk.

Together with data from both population genetic studies and functional assays, we provide strong evidence that the variant rs139379666 in the *ATM* gene may be a novel pathogenic mutation for breast cancer. In line with previous findings, this missense variant has also been reported to be pathogenic in contributing to childhood acute leukemia (34). The study also demonstrated that this pathogenic variant can lead to the defective function of ATM. Notably, the variant was very rare or not detected in other populations (i.e., AF = 0.02% and AF = 0.005% for Africans and Europeans, respectively) using data from genomAD, likely indicating its ubiquitous pathogenicity in all populations. It is important to note that because most exomic sequencing platforms would not routinely detect this particular variant in a low frequency, the true prevalence of the variant contributing to other

diseases remains unknown. These results together indicate that this pathogenic variant may ubiquitously contribute to human cancers and diseases.

In conclusion, we identified a novel pathogenic variant, rs139379666, in a known susceptibility *ATM* gene, conferring to breast cancer risk. Results from our functional experiments further showed that this pathogenic variant can lead to the loss of HR activities and DNA damage repair abilities. Our analysis of the *ATM* protein structure further suggests that this pathogenic variant may play a defected regulation in the activity of ATM structure. Our findings provide additional insight into the pathogenesis of breast cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: X. Guo, W. Lin, W. Zheng
 Development of methodology: X. Guo, M. Bai, W. Wen, W. Zheng
 Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X. Guo, Q. Cai, J. Long, W.-H. Jia, X.-O. Shu, W. Zheng

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Guo, W. Lin, H. Li, W. Wen, C. Zeng, J. He, J. Chen, Q. Cai, W. Zheng

Writing, review, and/or revision of the manuscript: X. Guo, W. Lin, C. Zeng, Q. Cai, J. Long, X.-O. Shu, W. Zheng

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Guo, Z. Chen, W. Zheng

Study supervision: X. Guo, W. Lin, X.-O. Shu, W. Zheng

Acknowledgments

The authors thank the study participants and research staff for their contributions and support to this project. We thank Regina Courtney and Jie Wu for laboratory assistance, and Marshal Younger for assistance with editing and manuscript preparation. The data analyses were conducted using the Advanced

Computing Center for Research and Education (ACCRE) at Vanderbilt University. This research is supported by the U.S. NIH grant R01CA158473 (to W. Zhang) and the research development fund from Vanderbilt University Medical Center (to X. Guo). This research was also partly supported by grants from the National Key R&D Program of China (no. 2016YFC0902700) and the National Natural Science Foundation (nos. 31470776 and 1670651, to W. Lin).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 30, 2018; revised March 20, 2019; accepted May 28, 2019; published first June 3, 2019.

References

- Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med* 2001;7:552–6.
- Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014;505:302–8.
- Apostolou P, Fostira F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* 2013;2013:747318.
- Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res* 2012;18:400–7.
- Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 2010;42:410–4.
- Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, et al. Beyond Li Fraumeni syndrome: clinical characteristics of families with p53 germline mutations. *J Clin Oncol* 2009;27:1250–6.
- Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008;40:17–22.
- Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* 2007;39:1338–49.
- Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 2007;39:165–7.
- Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 2006;38:1239–41.
- Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* 2006;38:873–5.
- Pharoah PD, Guilford P, Caldas C, International Gastric Cancer Linkage Consortium. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology* 2001;121:1348–53.
- Guo X, Shi J, Cai Q, Shu XO, He J, Wen W, et al. Use of deep whole-genome sequencing data to identify structure risk variants in breast cancer susceptibility genes. *Hum Mol Genet* 2018;27:853–9.
- Couch FJ, Shimelis H, Hu CL, Hart SN, Polley EC, Na J, et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol* 2017;3:1190–6.
- Lu HM, Li S, Black MH, Lee S, Hoiness R, Wu S, et al. Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA Oncol* 2019;5:51–7.
- Easton DF, Pharoah PD, 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.
- Angele S, Hall J. The ATM gene and breast cancer: is it really a risk factor? *Mutat Res* 2000;462:167–78.
- Concannon P. ATM heterozygosity and cancer risk. *Nat Genet* 2002;32:89–90.
- Helgason H, Rafnar T, Olafsdottir HS, Jonasson JG, Sigurdsson A, Stacey SN, et al. Loss-of-function variants in ATM confer risk of gastric cancer. *Nat Genet* 2015;47:906–10.
- Li S, Ting NS, Zheng L, Chen PL, Ziv Y, Shiloh Y, et al. Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* 2000;406:210–5.
- Nielsen FC, van Overeem Hansen T, Sorensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* 2016;16:599–612.
- Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* 2014;46:886–90.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- Li B, Zhan X, Wing MK, Anderson P, Kang HM, Abecasis GR. QPLOT: a quality assessment tool for next generation sequencing data. *BioMed Res Int* 2013;2013:865181.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
- Zhang Y, Long J, Lu W, Shu XO, Cai Q, Zheng Y, et al. Rare coding variants and breast cancer risk: evaluation of susceptibility Loci identified in genome-wide association studies. *Cancer Epidemiol Biomarkers Prev* 2014;23:622–8.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
- Lin W, Sampathi S, Dai H, Liu C, Zhou M, Hu J, et al. Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J* 2013;32:1425–39.
- Baretic D, Pollard HK, Fisher DI, Johnson CM, Santhanam B, Truman CM, et al. Structures of closed and open conformations of dimeric human ATM. *Sci Adv* 2017;3:e1700933.
- Bao ZQ, Jacobsen DM, Young MA. Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis. *Structure* 2011;19:675–90.
- DeLano WL. Use of PYMOL as a communications tool for molecular science. *Abstr Pap Am Chem S* 2004;228:U313–U4.
- Oguchi K, Takagi M, Tsuchida R, Taya Y, Ito E, Isoyama K, et al. Missense mutation and defective function of ATM in a childhood acute leukemia patient with MLL gene rearrangement. *Blood* 2003;101:3622–7.
- Sawicka M, Wanrooij PH, Darbari VC, Tannous E, Hailemariam S, Bose D, et al. The dimeric architecture of checkpoint kinases Mec1ATR and Tel1-ATM reveal a common structural organization. *J Biol Chem* 2016;291:13436–47.

Cancer Epidemiology, Biomarkers & Prevention

Discovery of a Pathogenic Variant rs139379666 (p. P2974L) in *ATM* for Breast Cancer Risk in Chinese Populations

Xingyi Guo, Weiqiang Lin, Mengqiu Bai, et al.

Cancer Epidemiol Biomarkers Prev 2019;28:1308-1315. Published OnlineFirst June 3, 2019.

Updated version Access the most recent version of this article at:
doi:[10.1158/1055-9965.EPI-18-1294](https://doi.org/10.1158/1055-9965.EPI-18-1294)

Supplementary Material Access the most recent supplemental material at:
<http://cebp.aacrjournals.org/content/suppl/2019/06/01/1055-9965.EPI-18-1294.DC1>

Cited articles This article cites 35 articles, 7 of which you can access for free at:
<http://cebp.aacrjournals.org/content/28/8/1308.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/28/8/1308>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.