

RESEARCH

Open Access



Association of cyclin-dependent kinase inhibitor 2B antisense RNA 1 gene expression and rs2383207 variant with breast cancer risk and survival

Shahad W. Kattan¹, Yahya H. Hobani², Sameerah Shaheen³, Sara H. Mokhtar⁴, Mohammad H. Hussein⁵, Eman A. Toraih^{5,6}, Manal S. Fawzy^{7,8*}  and Hussein Abdelaziz Abdalla^{9,10}

*Correspondence:
manal2_khashana@gmail.com
⁷ Department of Medical
Biochemistry and Molecular
Biology, Faculty of Medicine,
Suez Canal University, Ismailia,
Egypt
Full list of author information
is available at the end of the
article

Abstract

Background: The expression signature of deregulated long non-coding RNAs (lncRNAs) and related genetic variants is implicated in every stage of tumorigenesis, progression, and recurrence. This study aimed to explore the association of lncRNA cyclin-dependent kinase inhibitor 2B antisense RNA 1 (*CDKN2B-AS1*) gene expression and the rs2383207A>G intronic variant with breast cancer (BC) risk and prognosis and to verify the molecular role and networks of this lncRNA in BC by bioinformatics gene analysis.

Methods: Serum *CDKN2B-AS1* relative expression and rs2383207 genotypes were determined in 214 unrelated women (104 primary BC and 110 controls) using real-time PCR. Sixteen BC studies from The Cancer Genome Atlas (TCGA) including 8925 patients were also retrieved for validation of results.

Results: *CDKN2B-AS1* serum levels were upregulated in the BC patients relative to controls. A/A genotype carriers were three times more likely to develop BC under homozygous (OR = 3.27, 95% CI 1.20–8.88, $P = 0.044$) and recessive (OR = 3.17, 95% CI 1.20–8.34, $P = 0.013$) models. G/G homozygous patients had a higher expression level [median and quartile values were 3.14 (1.52–4.25)] than A/G [1.42 (0.93–2.35)] and A/A [1.62 (1.33–2.51)] cohorts ($P = 0.006$). The Kaplan–Meier curve also revealed a higher mean survival duration of G/G cohorts (20.6 months) compared to their counterparts (A/A: 15.8 and A/G: 17.2 months) ($P < 0.001$). Consistently, BC data sets revealed better survival in cohorts with high expression levels ($P = 0.003$). Principal component analysis (PCA) showed a deviation of patients who had shorter survival towards A/A and A/G genotypes, multiple lesions, advanced stage, lymphovascular invasion, and HER2⁺ receptor staining. Ingenuity Pathway Analysis (IPA) showed key genes highly enriched in BC with *CDKN2B-AS1*.

Conclusions: The findings support the putative role of *CDKN2B-AS1* as an epigenetic marker in BC and open a new avenue for its potential use as a therapeutic molecular target in this type of cancer.



Keywords: Breast cancer, CDKN2B-AS1, Gene expression, Long non-coding RNA, Single nucleotide polymorphism, Survival

Introduction

According to recent cancer statistics, an estimated one-third of newly diagnosed female cancers will be breast cancers (BC) in 2020 [1]. Although the progress in diagnosis, surgical techniques, and targeted therapy has improved BC prognosis, it still represents one of the leading causes of cancer death among females aged 20–59 years [1, 2]. Hence, it is essential to explore the role of the newly emerged (epi)genetic molecular players in the BC scenario to improve disease prognosis and survival [3].

Extensive mammalian genomic and transcriptomic analyses have revealed a cluster of long non-coding RNAs (lncRNAs; >200 nucleotides in length), which are implicated in several genetic and epigenetic regulations of the cells [4–6]. This family of non-coding RNAs can be classified into five classes: sense, antisense, bidirectional, intronic, and intergenic lncRNAs [7]. A deregulated lncRNA expression signature was found to be implied in every stage of BC tumorigenesis, progression, and recurrence [8] (Additional file 1: Table S1). Given the stability of serum/plasma lncRNAs and their RNase resistance, many studies have explored their potential use as putative epigenetic markers for cancer diagnosis/prognosis, with promising results [9–11].

The lncRNA cyclin-dependent kinase inhibitor 2B antisense RNA 1; CDKN2B-AS1, NONCODE Gene ID: NONHSAG051899.2, also known as ANRIL (antisense non-coding RNA in the INK4 locus) was identified in a genomic hotspot region at chromosome 9p21.3 (Fig. 1a). This region is well known to be associated with human cancers [12] and was chosen based on screening of the previous studies and our detailed in silico analysis. The gene includes 19–21 exons spanning around 134 kb in the genome (Fig. 1b, c), and the coded lncRNA was reported to interact with the polycomb repressive complex-1 (PRC1) and -2 (PRC2) proteins in invasive BC [13]. These latter protein complexes play essential roles in cell differentiation and carcinogenesis [14].

CDKN2B-AS1 deregulation has been implicated in several pathological processes, including increased cellular proliferation, deregulated metabolic activity, inflammation, and decreased apoptosis [15–17]. Silencing this lncRNA was reported to prevent fibroblast and smooth muscle cell proliferation [15, 16]. Furthermore, increased CDKN2B-AS1 levels in several cancers, including BC tissues, suggest its putative role in promoting tumorigenesis [15, 18–20].

Accumulating evidence suggests that lncRNAs gene polymorphisms are also associated with cancer risk [21] and can influence lncRNA gene expression and/or function [22]. Additionally, studies have revealed that lncRNA variants may impact mRNAs splicing and stability, with subsequent changes in their cellular behavior and interacting partners [21, 23].

Given the association of *CDKN2B-AS1* expression and variants with cancer and the limitation of related studies in our region, the current study aimed to (1) investigate the impact of circulating *CDKN2B-AS1* expression and the intronic variant rs2383207 A>G (which has not been studied before) on BC risk and prognosis, (2) execute bioinformatics analysis for the specified lncRNA and (3) verify the present results against data in The Cancer Genomic Atlas (TCGA).

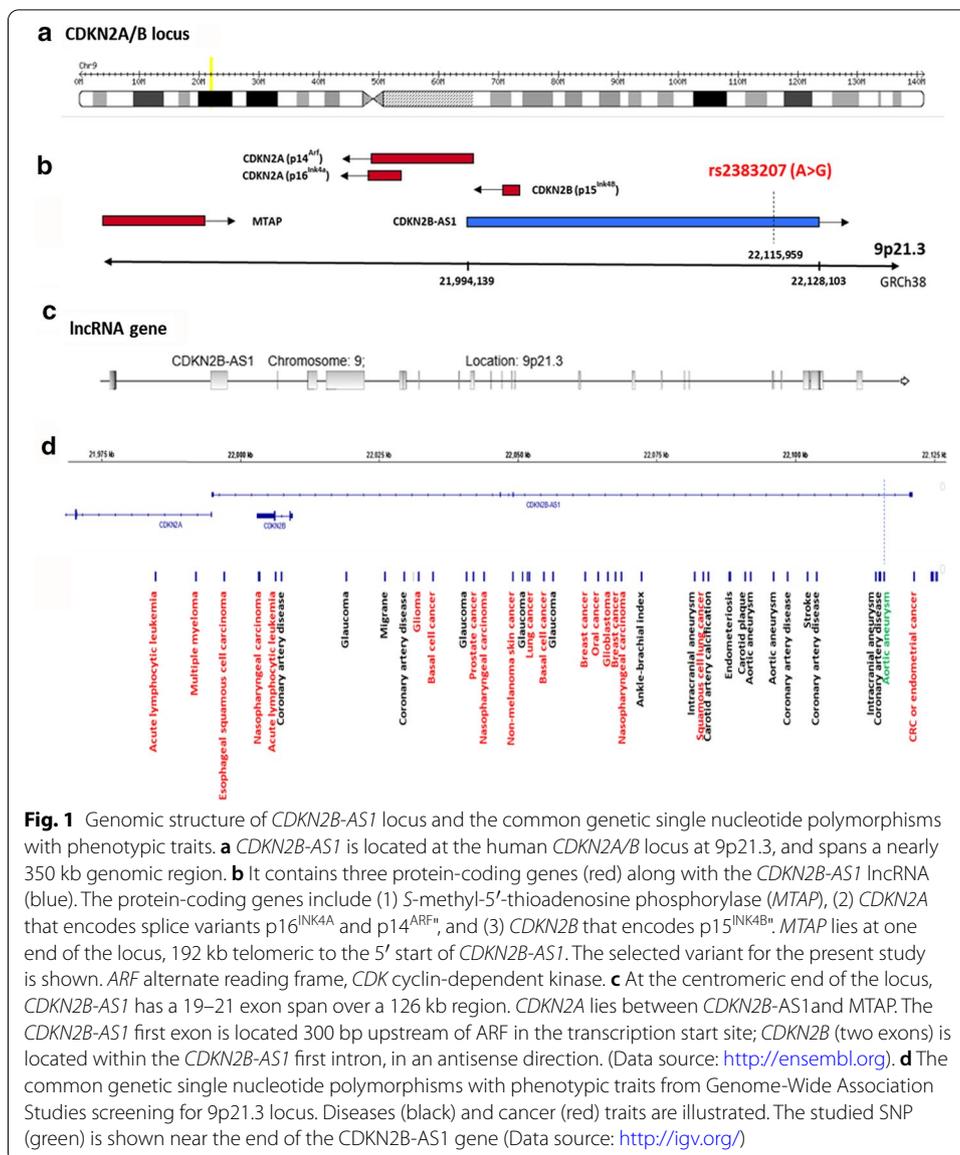


Fig. 1 Genomic structure of *CDKN2B-AS1* locus and the common genetic single nucleotide polymorphisms with phenotypic traits. **a** *CDKN2B-AS1* is located at the human *CDKN2A/B* locus at 9p21.3, and spans a nearly 350 kb genomic region. **b** It contains three protein-coding genes (red) along with the *CDKN2B-AS1* lncRNA (blue). The protein-coding genes include (1) *S*-methyl-5'-thioadenosine phosphorylase (*MTAP*), (2) *CDKN2A* that encodes splice variants p16^{INK4A} and p14^{ARF}, and (3) *CDKN2B* that encodes p15^{INK4B}. *MTAP* lies at one end of the locus, 192 kb telomeric to the 5' start of *CDKN2B-AS1*. The selected variant for the present study is shown. *ARF* alternate reading frame, *CDK* cyclin-dependent kinase. **c** At the centromeric end of the locus, *CDKN2B-AS1* has a 19–21 exon span over a 126 kb region. *CDKN2A* lies between *CDKN2B-AS1* and *MTAP*. The *CDKN2B-AS1* first exon is located 300 bp upstream of *ARF* in the transcription start site; *CDKN2B* (two exons) is located within the *CDKN2B-AS1* first intron, in an antisense direction. (Data source: <http://ensembl.org>). **d** The common genetic single nucleotide polymorphisms with phenotypic traits from Genome-Wide Association Studies screening for 9p21.3 locus. Diseases (black) and cancer (red) traits are illustrated. The studied SNP (green) is shown near the end of the *CDKN2B-AS1* gene (Data source: <http://igv.org/>)

Subjects and methods

Study population

The institutional research ethics committee of the Faculty of Medicine, Suez Canal University, Ismailia, Egypt, approved the current work (Approval no. 3960). Before the start of the study, written informed consent was obtained from all participants. A total of 214 unrelated women (110 consecutive primary BC and 104 controls) were included in this study. The patients were recruited from the General Surgery Department and Oncology Diagnostic Unit, Suez Canal University Hospitals, Ismailia, Egypt. They were diagnosed clinically, radiologically, and confirmed by biopsy [24]. The patients did not have a history of receiving any treatments such as chemotherapy, radiotherapy, hormonal therapy, or immunotherapy before blood sampling. Patients with other malignancies or chronic diseases were excluded. Healthy blood donors without any signs of chronic diseases,

recent pregnancy, or lactation in the last 2 years and/or associated malignancy were included as controls.

Histopathological and immunohistochemical assessment

Specimens from BC tissue were histopathologically analyzed after the operations. Assessment of pathological grade and clinical stage were performed using the Elston and Ellis modification of the Scarff–Bloom–Richardson system and the American Joint Committee on Cancer (AJCC) tumor-lymph node-metastasis (TNM) staging system [25]. Assessment of hormonal receptors for molecular subtyping was investigated [26]. Estrogen receptor (ER) and progesterone receptor (PR) evident nuclear staining in $\geq 10\%$ of the tumor cells was reported as positive (+ve), and if absent or $< 10\%$ staining was recorded as negative (–ve) [27]. At the same time, human epidermal growth factor receptor 2 (HER2)/neu expression was semi-quantified by the following membrane scoring system: 0, no staining or membrane staining in $< 10\%$ of tumor cells; 1+, faint or partly stained membranes in $\geq 10\%$ of tumor cells; 2+, weak to moderate complete membrane staining in $\geq 10\%$ of tumor cells; 3+, strong complete membrane staining in $\geq 10\%$ of tumor cells; and accordingly, samples were classified into different molecular subtypes [22]. Then patients were subdivided into four molecular subgroups: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2⁺), and triple-negative BC, as previously detailed [28, 29].

Clinical assessment and prognostic evaluation

Clinical features, risk factor assessment, and investigations were evaluated. The Nottingham Prognostic Index (NPI) and the Immunohistochemical Prognostic Index (IHPI) were applied as previously described [28], and patients were classified accordingly to have a good, moderate, and poor prognosis. The follow-up period was extended for up to 3 years for overall survival (OS) and disease-free survival (DFS) assessment.

Sample collection

Venous blood (5 ml) samples were collected into plain and ethylenediamine tetra-acetic acid (EDTA) Vacutainers. The samples were preserved after collection for 30 min to 2 h in the refrigerator to allow blood clotting. After that, all samples were centrifuged for 10 min. The obtained serum from plain tubes was divided into aliquots and stored at $-80\text{ }^{\circ}\text{C}$. One freeze–thaw cycle was carried out for the samples. The EDTA tubes were used for genomic DNA analysis.

Nucleic acids extraction

From the serum samples, total RNAs were isolated via the Qiagen miRNeasy Serum/Plasma Kit (Qiagen, Clinilab Co., Catalog no. 217184), and from the buffy coat of EDTA blood samples, the genomic DNA was isolated via the QIAamp DNA Blood Mini kit (Catalog No. 51104; Qiagen) following the instructions of the vendors. The purity/concentrations of nucleic acids were assessed using the NanoDrop ND-1000 spectrophotometer (NanoDrop Tech., Inc. Wilmington, DE, USA). Furthermore, the integrity of the nucleic acids was tested by agarose gel electrophoresis.

Reverse-transcription and quantitative polymerase chain reaction analysis

The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guideline was followed during the quantitative real-time reverse-transcription polymerase chain reaction (qRT-PCR) runs [30]. The first step of the amplification process was the reverse transcription (RT) of 1 µg of total RNA to yield complementary DNA (cDNA) using a High-Capacity cDNA RT Kit (P/N 4368814, Applied Biosystems, Foster City, California, USA). The second step is the real-time PCR reaction using a specific TaqMan probe for the lncRNA *CDKN2B-AS1* (Assay no. Hs04406279_m1), compared to the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) for normalization of the data. In each run, appropriate negative controls were applied: no template and no reverse transcriptase controls. The final volume of the reaction (20 µl) contained RT products (1.33 µl), 2× TaqMan Universal PCR Master Mix (10 µl), and TaqMan RNA assay (1 µl). The StepOne Real-Time PCR machine (Applied Biosystems) was set up as follows: 10 min at 95 °C, followed by 40 cycles of 15 s at 92 °C and 60 s at 60 °C [31].

CDKN2B-AS1 expression data analysis

The quantification cycle (C_q) is the cycle number at which the fluorescence passed a fixed threshold [30]. The relative amount of the study lncRNA to *GAPDH* in patients compared to controls was calculated using the formula: $2^{-\Delta\Delta C_q}$; where $\Delta\Delta C_q = (C_q \text{ *CDKN2B-AS1* - } C_q \text{ *GAPDH*)}_{BC} - (C_q \text{ *CDKN2B-AS1* - } C_q \text{ *GAPDH*)}_{\text{mean controls}}$ [32].

Allelic discrimination analysis

TaqMan Genotyping PCR Master Mix, No. UNG (4440043), and TaqMan single nucleotide polymorphism (SNP) Genotyping Assay Mix (assay IDC__15789010_20, Catalog# 4351379, Thermo Fisher Scientific) were used for the rs2383207 real-time allele discrimination assay. Briefly, the extracted DNA (20 ng) was diluted to 11.25 µl with DNase-free water, then added to the reaction mix containing TaqMan Master mix (12.5 µl) and TaqMan SNP Genotyping Assay (20×) Mix (1.25 µl). The PCR program was described previously [33]. The investigators were not aware of the sample status (patient versus control) during the genotyping. In each run, non-template and TaqMan enzyme negative controls were included. Ten percent of samples were tested in duplicate with a 100% concordance rate for genotype calls. The SDS software version 1.3.1 (Applied Biosystems) was applied for genotype calling.

Genomic and functional analysis of *CDKN2B-AS1* gene

Gene locus analysis was performed in the Ensembl Genome Browser (<http://ensembl.org>), a database that annotates genes, transcripts, and genomic variations, computes multiple alignments, predicts regulatory function, and collects disease-related data. Variant analysis of *CDKN2B-AS1* and the impact of each were determined in the Varsome web application (<https://varsome.com/>), a search engine, aggregator, and impact analysis tool for human genetic variation. It displays a detailed annotation of the queried variant, including multiple notations, predicted pathogenicity status from various tools, and genomic context. Genetic variants of the *CDKN2B-AS1* gene with known pathogenicity were confirmed in ClinVar and PubMed. The predicted and tested *CDKN2B-AS1*-related

SNPs' impact on gene expression was determined from the literature and the LDHap project, a haplotype map project of the human genome that describes the common genetic patterns' variation associated with human disease. Ensembl.org and varsome.com were used to analyze the structural and functional impact of the selected rs2383207 intron variant. The Integrative Genomics Viewer (<http://igv.org/>) web application was run to view the locations and Genome-Wide Association Studies (GWAS)-associated phenotypic traits in the *CDKN2A/B* locus.

The Gene-Tissue Expression (GTEx) Portal (<http://gtexportal.org>) and BioGPS database (<http://biogps.org>) were used to identify the gene expression pattern of the gene across diverse normal human tissues from the U133plus2 Affymetrix microarray experiments. The expression level of *CDKN2B-AS1* in BC tissues with different molecular subtypes was determined from Expression Atlas (www.ebi.ac.uk/), an open resource that assists in finding information about gene and protein expression. Localization of the lncRNA was detected in the Compartment subcellular localization database (www.compartments.jensenlab.org/).

To retrieve high-throughput experiments in public repositories, 16 BC studies (total number of patients=8925) in The Cancer Genomic Atlas (TCGA) were downloaded from the cBioPortal for Cancer Genomics database (<http://cBioportal.org>) to identify the rate of genetic alterations in the *CDKN2B-AS1* genome and aberrant expression of the gene in BC. The online Kaplan–Meier plotter program (<http://kmplot.com>) was utilized to plot Kaplan–Meier survival curves for *CDKN-AS1* expression from the datasets stored in this database [34]. Next, to identify the predictive role of the *CDKN2B-AS1* gene in the therapeutic response, a receiver operator characteristics (ROC) plotter tool (<http://rocplot.org/>) was used to link gene expression and response to therapy using transcriptome-level data of 3104 BC patients [35]. Lnc2Cancer version 2.0, a manually curated database, was used to identify experimentally supported associations between lncRNA and human cancer. It provides information on circulating, drug-resistant, prognostic lncRNAs in cancer [36].

The Gene Ontology Annotation (GOA) database was used to define the gene ontology terms related to the *CDKN2B-AS1* gene. Next, the Automatic Cancer Hallmarks Analytics Tool (CHAT) was used for classification of the PubMed literature according to the cancer taxonomy hallmarks. Breast cancer and *CDKN2B-AS1* (data are shown as NPMI; normalized pointwise mutual information) were compared. The LncMAP lncRNA Modulator Atlas in Pan-cancer (<http://bio-bigdata.hrbmu.edu.cn/LncMAP/index.jsp>) was used to determine putative transcription factors (TFs). NetworkAnalyst (www.networkanalyst.ca) version 3.0 was used to construct gene regulatory networks (GRN) for the upstream regulators of the *CDKN2B-AS1* gene. The TF targets derived from the JASPAR TF binding site profile database and inferred from integrating the literature curated Chip-X database were identified. Also, the literature curated regulatory interaction information was collected from the RegNetwork repository [37]. IPA was employed to identify the most relevant spatial molecular interactions within the cell, predict the direction of the downstream effect in BC, and the activation and inhibition of upstream TFs. Finally, ENCORI for the RNA interactome (<http://starbase.sysu.edu.cn/>) with default settings presented the ceRNA (competing endogenous RNA) networks from multiple interactions of miRNA targets in

pan-cancer supported by CLIP-seq data. The results were examined at a P -value ≤ 0.01 and the false discovery rate (FDR) ≤ 0.01 [38].

Statistical analysis

When G power-3 software (<http://www.gpower.hhu.de/>) was used for study power estimation with a medium effect size, specified study design, and allowable error rate (alpha error=0.05), it yielded 95% study power. Data were expressed as number and percentage or mean \pm standard deviation for categorical and continuous data. Data were checked for normality, outliers, and skewness and were transformed in case of violation of the normality assumption. Appropriate non-parameter tests were then applied and presented as median and quartiles. Two-sided chi-square, Fisher's exact, Student t-test, one-way ANOVA, Mann–Whitney U, or Kruskal–Wallis tests were used. Construction of contingency tables was performed for genotype distribution and allelic frequency comparison. The Hardy–Weinberg equilibrium (HWE) was applied using the χ^2 test to assess genotype distribution in all study subjects. Six genetic association models were analyzed as previously described [39]. Odds ratios (ORs) with 95% confidence intervals (CIs) and corresponding P -values were computed by logistic regression analysis and adjusted by the patients' demographic, clinical, and pathological characteristics. The statistical significance was set at P -value < 0.05 . All statistical analyses were executed using SPSS version 26.0 and GraphPad Prism version 8.0. SNPStats (<https://www.snptest.net/start.htm>) was applied for genetic analysis. Kaplan–Meier curves were plotted for *CDKN-AS1* genotypes with overall survival. In RStudio, correlation plots were constructed using the corrplot package, while principal component analysis (PCA) was performed using the factoextra and FactoMineR packages.

Results

Baseline characteristics of the study participants

The mean age of BC women was 44.3 ± 12.2 years. Only one-third (30.9%) had a positive family history of cancer. Two-thirds (63.6%) were overweight/obese. Most of the patients had a sedentary lifestyle (90%); nevertheless, only 11.8% had night shifts at work, and 7.3% of women had a prior history of breast problems. About 64.5% had early onset of menarche. A few of them had late first gravida (4.5%), nullipara (12.7%), and breastfed (16.4%), while one-third of patients were at the post-menopausal state at the time of diagnosis.

CDKN2B-AS1 expression profile

Based on screening 52 cases (who had available extracted RNA matched with the DNA samples for further genotype-expression correlation analysis) and 104 controls, circulatory *CDKN2B-AS1* was upregulated in cancer patients compared to controls ($P < 0.001$). The median log fold change of *CDKN2B-AS1* in the BC cohort was 1.82 (IQR: 1.38–3.71) relative to the controls.

Allelic discrimination analysis

Allele and genotype frequencies of CDKN2B-AS1(rs2383207) variant

A total of 110 cases and 104 controls were genotyped. Genotype frequencies followed the Hardy–Weinberg equilibrium in controls ($P = 0.80$) and cases ($P = 0.06$). Minor allele frequency (A allele) was 31% in the overall population and 26% in controls. Apart from

the African population, allele frequencies were similar to those of different populations in the 1000 Genome Project Phase 3 (<http://ensembl.org>) (Fig. 2a). In our study population, half of the patients carried G/G genotype (107 cohorts), while A/G heterozygotes exceed a third (83 patients, 39%). A higher frequency of the A allele was observed in BC patients (77, 35%) compared to controls (54, 26%), $P=0.042$ (Fig. 2b). Similarly, A/A homozygosity was the most frequent genotype among patients, representing 16% (18 patients) compared to 6% (6 patients), $P=0.047$ (Fig. 2c).

CDKN2B-AS1 (rs2383207) genotypes and disease risk

According to the genetic association models, A allele (A/A) homozygotes were three times more liable to develop BC under homozygote comparison (OR=3.27, 95% CI 1.20–8.88, $P=0.044$) and the recessive model (OR=3.17, 95% CI 1.20–8.34, $P=0.013$) (Table 1).

CDKN2B-AS1 genotypes modulate lncRNA expression in the breast cancer cohort

Although different CDKN2B-AS1 genotypes in the control group showed no significant difference in gene relative expression ($P=0.069$) (Fig. 2d), the G/G genotype was associated with a higher expression level than A/G in the BC cohort; median and quartile values were 3.14 (1.52–4.25) in G/G vs. 1.42 (0.93–2.35) in A/G and 1.62 (1.33–2.51) in A/A ($P=0.006$) (Fig. 2e).

Association of CDKN2B-AS1 with clinicopathological features

Apart from body weight, in which overweight and obesity were more prevalent among G/G cohorts ($P=0.004$), no significant differences in BC risk factors among patients with different genotypes were found (Table 2).

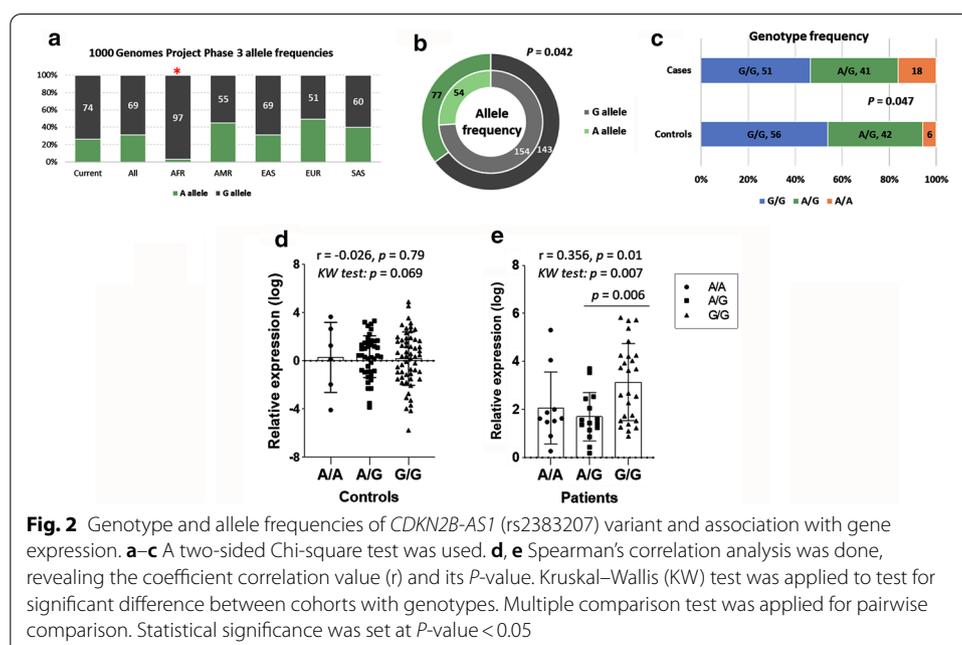


Table 1 Genetic association models and disease risk

Model	Genotypes	Controls	Patients	Adjusted OR (95% CI)	P-value
Codominant	G/G	56 (53.9%)	51 (46.4%)	1.00	0.044
	A/G	42 (40.4%)	41 (37.3%)	1.07 (0.60–1.90)	
	A/A	6 (5.8%)	18 (16.4%)	3.27 (1.20–8.88)	
Dominant	G/G	56 (53.9%)	51 (46.4%)	1.00	0.28
	A/G–A/A	48 (46.1%)	59 (53.6%)	1.35 (0.79–2.30)	
Recessive	G/G–A/G	98 (94.2%)	92 (83.6%)	1.00	0.013
	A/A	6 (5.8%)	18 (16.4%)	3.17 (1.20–8.34)	
Over-dominant	G/G–A/A	62 (59.6%)	69 (62.7%)	1.00	0.64
	A/G	42 (40.4%)	41 (37.3%)	0.88 (0.51–1.52)	
Log-additive	–	–	–	1.48 (0.99–2.22)	0.053

Data are presented as number (percentage). Adjusted odds ratio (OR) by age. Bold values indicate statistical significance at P -value < 0.05

CI confidence interval

Immunohistochemical analysis was conducted for tumor samples. Selected samples with positive estrogen, progesterone, and HER2⁺ receptors are illustrated in Fig. 3.

Regarding the pathological characteristics of BC tumors, no significant difference was observed among patients with various genotypes. However, the rs2383207*A variant was associated with poor survival. Higher frequencies of A/A and A/G genotypes were found in cohorts with short disease-free survival ($P = 0.009$) and overall survival ($P < 0.001$) (Table 2). Consistently, the principal component analysis showed a deviation of patients who had shorter survival towards A/A and A/G genotypes, multiple lesions, advanced stage, lymphovascular invasion, and HER2⁺ receptor staining (Fig. 4a). Kaplan–Meier curve analysis also revealed a higher mean survival duration of G/G cohorts (20.6 months) compared to their counterparts (A/A: 15.8 months and A/G: 17.2 months), $P < 0.001$ (Fig. 4b).

In an association of the transcriptomic signature of *CDKN2B-AS1* with clinicopathological features, the lower expression level (1.64 ± 1.85) was associated with poor pathological grade compared to well-differentiated and intermediately differentiated tumors (2.65 ± 1.60 , $P = 0.004$) in breast cancer patients. However, there were no other significant associations with any other clinical or pathological parameters (Fig. 5).

In silico data analysis

Structural analysis of *CDKN2B-AS1* gene

Although the *CDKN2B-AS1* gene locus includes other protein-coding genes, i.e. *CDKN2B* (encoding p15ink4b) and *CDKN2A* (encoding p16ink4a and p14ARF) (Fig. 1b) with a key role in cell cycle inhibition, senescence, and stress-induced apoptosis, prior studies highlighted that SNPs in this locus act through effects on *CDKN2B-AS1* itself. The gene has 28 splice variant linear and circular transcripts ranging from 602 bp up to 7173 bp and includes LINE, SINE, and Alu repetitive sequences. Some are tissue-specific, while others have distinct roles in cell physiology.

Table 2 Association between clinicopathological characteristics of breast cancer patients and *CDKN2B-AS1* genotypes

Characteristics	Total	A/A	A/G	G/G	P-value
Demographics and risk factors					
Age (years)					
≤ 45	98 (48.3)	8 (44.4)	24 (58.5)	26 (51)	0.38
> 45	105 (51.7)	10 (55.6)	17 (41.5)	25 (49)	
Smoking					
Negative	99 (90)	18 (100)	37 (90.2)	44 (86.3)	0.248
Positive	11 (10)	0 (0)	4 (9.8)	7 (13.7)	
Weight					
Underweight	12 (10.9)	4 (22.2)	4 (9.8)	4 (7.8)	0.004
Normal	28 (25.5)	7 (38.9)	13 (31.7)	8 (15.7)	
OW/Obese	70 (63.6)	7 (38.9)	24 (58.5)	39 (76.5)	
FH cancer					
Negative	76 (69.1)	13 (72.2)	28 (68.3)	35 (68.6)	0.951
Positive	34 (30.9)	5 (27.8)	13 (31.7)	16 (31.4)	
Breast problems					
Negative	102 (92.7)	18 (100)	38 (92.7)	46 (90.2)	0.387
Positive	8 (7.3)	0 (0)	3 (7.3)	5 (9.8)	
OCP intake					
Negative	87 (79.1)	17 (94.4)	29 (70.7)	41 (80.4)	0.114
Positive	23 (20.9)	1 (5.6)	12 (29.3)	10 (19.6)	
Menarche					
Late onset	39 (35.5)	4 (22.2)	11 (26.8)	24 (47.1)	0.058
Early onset	71 (64.5)	14 (77.8)	30 (73.2)	27 (52.9)	
Parity					
Multipara	96 (87.3)	14 (77.8)	37 (90.2)	45 (88.2)	0.401
Nullipara	14 (12.7)	4 (22.2)	4 (9.8)	6 (11.8)	
Gravida					
Early G1	105 (95.5)	18 (100)	38 (92.7)	49 (96.1)	0.443
Late G1	5 (4.5)	0 (0)	3 (7.3)	2 (3.9)	
Menopause					
Early	98 (89.1)	16 (88.9)	37 (90.2)	45 (88.2)	0.953
Late	12 (10.9)	2 (11.1)	4 (9.8)	6 (11.8)	
Breast feeding					
Negative	92 (83.6)	13 (72.2)	34 (82.9)	45 (88.2)	0.284
Positive	18 (16.4)	5 (27.8)	7 (17.1)	6 (11.8)	
Menopausal status					
Pre	80 (72.7)	14 (77.8)	28 (68.3)	38 (74.5)	0.698
Post	30 (27.3)	4 (22.2)	13 (31.7)	13 (25.5)	
Night work					
Negative	97 (88.2)	16 (88.9)	36 (87.8)	45 (88.2)	0.993
Positive	13 (11.8)	2 (11.1)	5 (12.2)	6 (11.8)	
Sedentary life style					
Negative	11 (10)	1 (5.6)	3 (7.3)	7 (13.7)	0.470
Positive	99 (90)	17 (94.4)	38 (92.7)	44 (86.3)	
Pathological data					
Side					
Right	70 (63.6)	11 (61.1)	27 (65.9)	32 (62.7)	0.92
Left	40 (36.4)	7 (38.9)	14 (34.1)	19 (37.3)	

Table 2 (continued)

Characteristics	Total	A/A	A/G	G/G	P-value
Site					
Outer quadrants	48 (43.6)	5 (27.8)	16 (39)	27 (52.9)	0.13
Others	62 (56.4)	13 (72.2)	25 (61)	24 (47.1)	
No masses					
Single	85 (77.3)	12 (66.7)	30 (73.2)	43 (84.3)	0.22
Multiple	25 (22.7)	6 (33.3)	11 (26.8)	8 (15.7)	
Grade					
≤ 2	90 (81.8)	16 (88.9)	32 (78)	42 (82.4)	0.60
> 2	20 (18.2)	2 (11.1)	9 (22)	9 (17.6)	
T stage					
≤ 3	79 (71.8)	12 (66.7)	29 (70.7)	38 (74.5)	0.80
> 3	31 (28.2)	6 (33.3)	12 (29.3)	13 (25.5)	
N stage					
N0	30 (27.3)	8 (44.4)	10 (24.4)	12 (23.5)	0.20
N1–3	80 (72.7)	10 (55.6)	31 (75.6)	39 (76.5)	
M stage					
M0	50 (45.5)	8 (44.4)	17 (41.5)	25 (49)	0.76
M1	60 (54.5)	10 (55.6)	24 (58.5)	26 (51)	
LVI					
Negative	58 (52.7)	10 (55.6)	18 (43.9)	30 (58.8)	0.35
Positive	52 (47.3)	8 (44.4)	23 (56.1)	21 (41.2)	
Skin involvement					
Negative	89 (80.9)	14 (77.8)	31 (75.6)	44 (86.3)	0.40
Positive	21 (19.1)	4 (22.2)	10 (24.4)	7 (13.7)	
Clinical stage					
≤ 2	48 (43.6)	8 (44.4)	17 (41.5)	23 (45.1)	0.93
> 2	62 (56.4)	10 (55.6)	24 (58.5)	28 (54.9)	
NPI					
Good	56 (50.9)	12 (66.7)	19 (46.3)	25 (49)	0.332
Poor	54 (49.1)	6 (33.3)	22 (53.7)	26 (51)	
Molecular subtype					
Luminal A	52 (47.3)	9 (50)	16 (39)	27 (52.9)	0.45
Luminal B	14 (12.7)	3 (16.7)	3 (7.3)	8 (15.7)	
HER2 ⁺	7 (6.4)	1 (5.6)	4 (9.8)	2 (3.9)	
Basal	37 (33.6)	5 (27.8)	18 (43.9)	14 (27.5)	
IHPI					
Good	66 (181.6)	12 (66.7)	19 (46.3)	35 (68.6)	0.25
Moderate	37 (33.6)	5 (27.8)	18 (43.9)	14 (27.5)	
Poor	7 (6.4)	1 (5.6)	4 (9.8)	2 (3.9)	
Follow-up					
ESMO					
Low risk	41 (37.3)	9 (50)	16 (39)	16 (31.4)	0.35
High risk	69 (62.7)	9 (50)	25 (61)	35 (68.6)	
DFS					
Prolonged	52 (47.3)	7 (38.9)	13 (31.7)	32 (62.7)	0.009
Short	58 (52.7)	11 (61.1)	28 (68.3)	19 (37.3)	
Recurrence					
Negative	55 (50)	9 (50)	19 (46.3)	27 (52.9)	0.82
Positive	55 (50)	9 (50)	22 (53.7)	24 (47.1)	

Table 2 (continued)

Characteristics	Total	A/A	A/G	G/G	P-value
OS					
Prolonged	53 (48.2)	2 (11.1)	8 (19.5)	43 (84.3)	< 0.001
Short	57 (51.8)	16 (88.9)	33 (80.5)	8 (15.7)	

Data are presented as numbers (percentage). A two-sided Chi-square test was used. Short survival was set at 12 months. Bold values indicate significance at P -value < 0.05

OW over-weight, FH family history, OCP oral contraceptive pills, G1 first pregnancy, LVI lymphovascular invasion, NPI Nottingham prognostic index, IHPI immunohistochemistry prognostic index, ESMO European Society of Medical Oncology for risk estimation of recurrence, DFS disease-free survival, OS overall survival

Variant analysis of *CDKN2B-AS1* gene

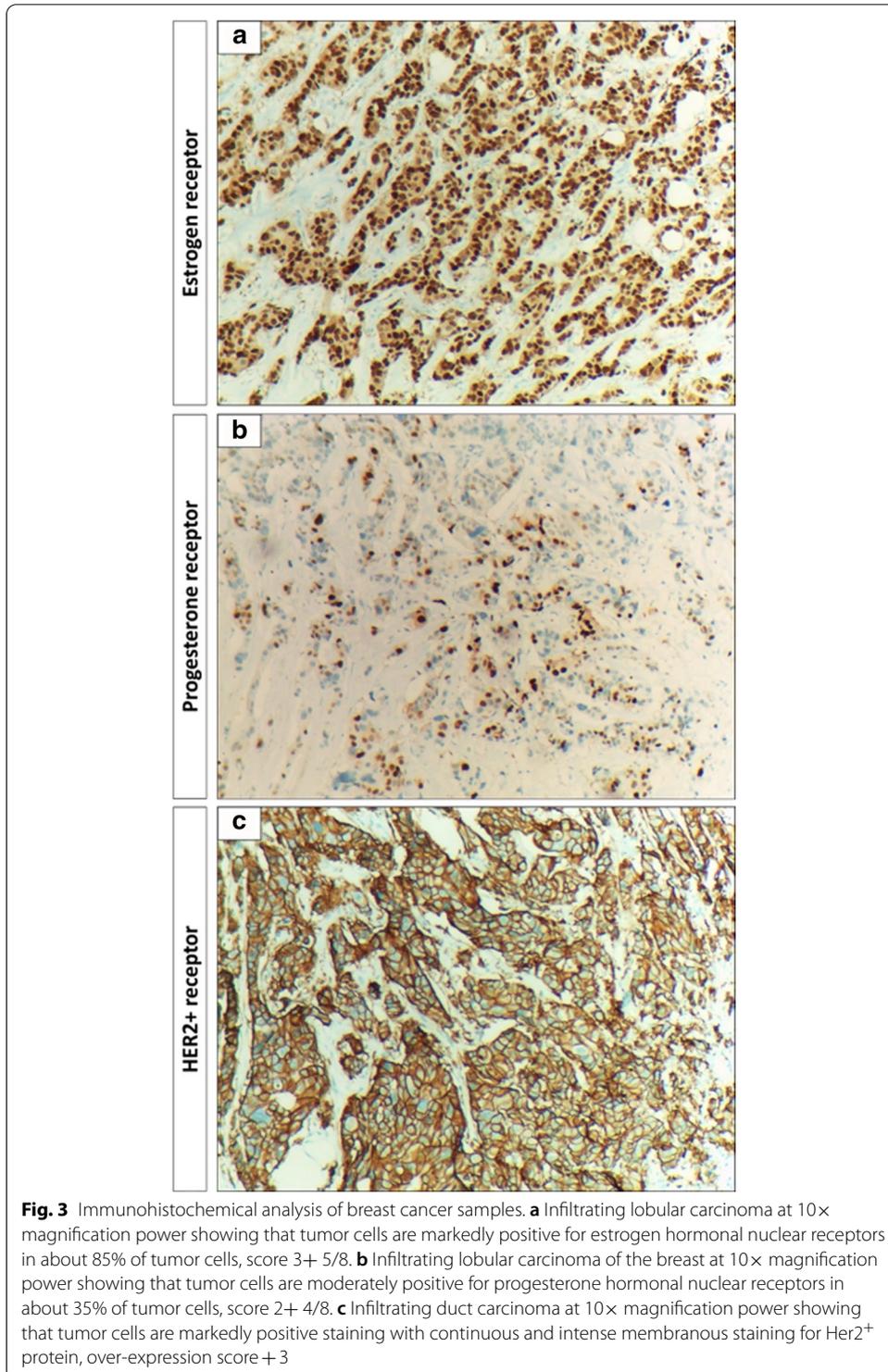
The studied intronic variant rs2383207A>G in the *CDKN2B-AS1* gene is located at chromosome 9: 22115959 (forward strand) 2161 bp away from the splice site with a minor allele frequency (G allele) of 0.31. This SNP overlaps 15 transcripts of the lncRNA and is predicted to be benign (Additional file 1: Table S2). Common disease genome-wide association studies (GWAS) have identified the *CDKN2B-AS1* gene as a shared locus for genetic susceptibility to multiple cancers (Fig. 1d).

Transcriptomic profile of *CDKN2B-AS1* gene

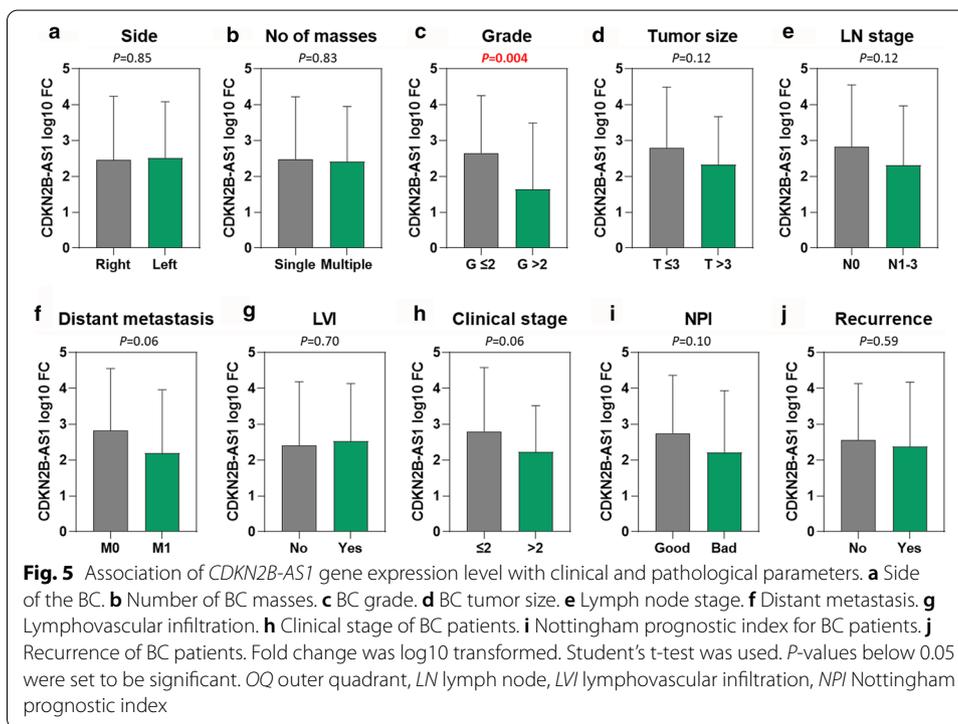
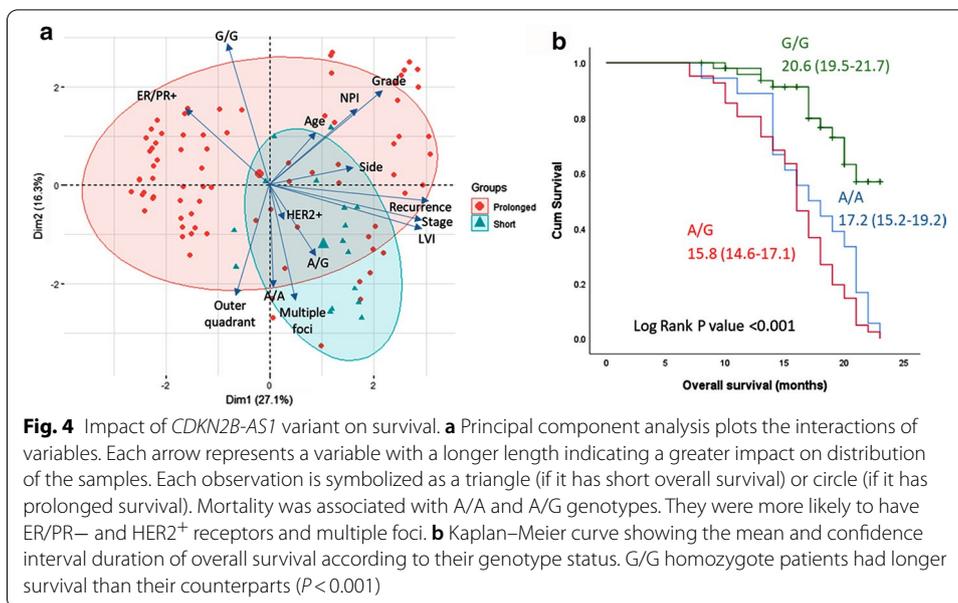
RNAseq analysis of 27 different normal tissues showed higher expression of the colon and small intestine and down-regulation in normal breast tissues (Fig. 6a). However, in BC tissues, overexpression was observed in all molecular subtypes (Fig. 6b). *CDKN2B-AS1* accumulates in both the nucleus and cytoplasm. The family of linear *CDKN2B-AS1* contains proximal (exon 1) and distal (exon 13b, 19) exons, and is enriched in the nucleus. However, the circular isoforms usually contain the middle exons (exons 5, 6, and 7), are enriched in the cytoplasm, and they differed markedly in their stability. Circular lncRNAs were reported to be associated with ribosome biogenesis and nucleolar stress, while nuclear isoforms are more likely to be involved in regulating gene transcription via chromatin modulation (Fig. 6c). Sixteen BC studies from the TCGA including 8925 patients were retrieved. Of these, 146 patients (2.1%) had a genetic alteration in the *CDKN2B-AS1* gene (44 amplification and 102 deep deletions). Somatic mutations were not studied in these datasets. Higher expression levels were encountered in patients with more copies of the genes and vice versa. Patients with altered *CDKN2B-AS1* had poorer survival than their counterparts ($P=0.031$) (Fig. 6d–f).

Prognostic and predictive role of *CDKN2B-AS1* gene in breast cancer

Despite the up-regulation of the lncRNA in BC tissues and blood, a higher expression level of the gene was associated with better overall survival (Fig. 7a–e). The prognostic performance of lncRNA expression, represented as a ROC curve, is shown in Fig. 7f–h. The utility of *CDKN2B-AS1* as a diagnostic and prognostic biomarker is illustrated in Table 3. Articles showing the association of *CDKN2B-AS1* up-regulation with treatment resistance are shown in Table 4. The top 30 drug-lncRNA pairs were used to build the Drug-lncRNA Network. Of the drugs associated with *CDKN2B-AS1*, Lapatinib, a dual inhibitor of EGFR and HER2, is indicated for patients with advanced or metastatic BC

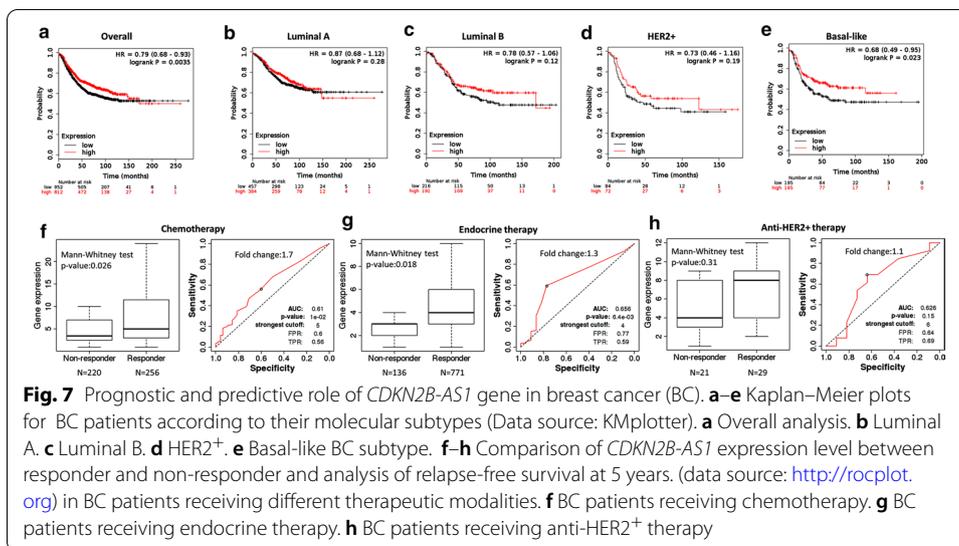
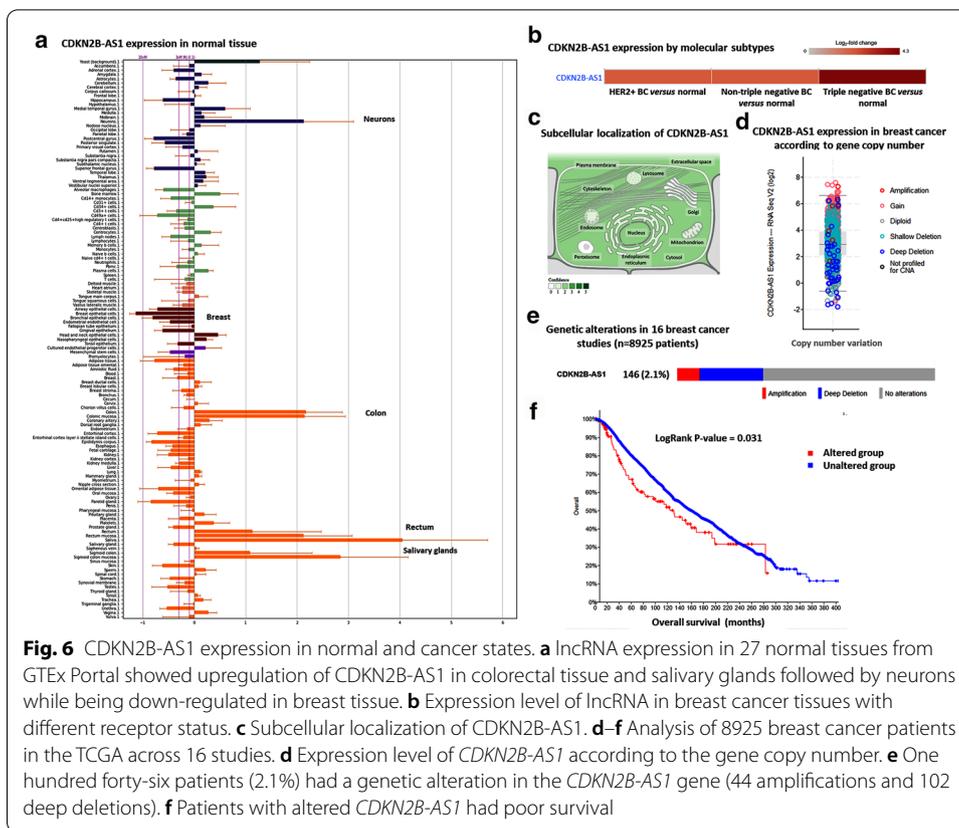


treatment whose tumors overexpress HER2. Other tyrosine kinase inhibitors such as erlotinib and sorafenib were also connected to the lncRNAs. The single drug topotecan was also indicated in patients with BC (Fig. 8).



Functional enrichment analysis of *CDKN2B-AS1* gene

Functional enrichment analysis revealed *CDKN2B-AS1* as a key player in regulating gene expression and cell migration. Being within the *INK4b-ARF-INK4a* gene family, which encodes the p15, p14, and p16 tumor-suppressor proteins, respectively, it is transcriptionally silenced or homozygously deleted in various human cancers. These three proteins are implicated in apoptosis, senescence, and stem cell renewal via promoting anti-proliferative and pro-apoptotic activities of Rb1 and p53 (Fig. 9a).



We used CHAT to analyze PubMed literature on BC and our studied gene. Cell invasion and metastasis was the most common hallmark associated with BC literature, followed by sustaining proliferative signaling. *CDKN2B-AS1* studies have two main hallmarks similar to those of the BC profile: sustaining proliferative signaling and genome instability and mutation (Fig. 9b). Upstream regulators enhancing transcription of the

Table 3 CDKN2AB-AS1 quantification in cancer patients as a diagnostic and prognostic biomarker. Data source: Lnc2Cancer version 2.0

Cancer type	Methods	Expression pattern	PubMed ID
Circulatory diagnostic biomarker			
Breast cancer	Microarray, qPCR	Up-regulated	28248879
Intraductal Papillary Mucinous Neoplasms of the Pancreas	qPCR	Up-regulated	28874676
Acute lymphoblastic leukemia	MassARRAY assay	Up-regulated	21414664
Neurofibromatosis type 1	Microarray, qPCR	Up-regulated	22034633
Non-small cell lung cancer	qPCR	Up-regulated	29504701
Oral cancer	qPCR, a Luciferase reporter assay, in vitro knockdown	Up-regulated	29635126
Prognostic biomarker			
Breast cancer	Microarray, qPCR	Up-regulated	28248879
Triple-negative breast cancer	qPCR, Luciferase reporter assay, RIP	Up-regulated	28961506
Non-small cell lung cancer	qPCR	Up-regulated	29504701
Oral cancer	qPCR, a Luciferase reporter assay, in vitro knockdown	Up-regulated	29635126
Gastric cancer	qPCR, RNAi	Up-regulated	27121324
Lung adenocarcinoma	qPCR, Western blot	Up-regulated	28402932
Bladder cancer	qPCR, RNAi, Western blot	Up-regulated	26800519
Cervical cancer	qPCR, RNAi, Western blot, Cell proliferation assay	Up-regulated	27899255
Colorectal cancer	qPCR, RNAi, Western blot, Northern blot	Up-regulated	27286457
Colorectal cancer	qPCR, RNAi	Up-regulated	27314206
Esophageal squamous cell cancer	qPCR, RNAi, Western blot	Up-regulated	24747824
Gallbladder cancer	qPCR, Western blot	Down-regulated	26812694
Gastric cancer	qPCR, RNAi, Western blot, RIP	Up-regulated	24810364
Gastric cancer	qPCR, RNAi, Western blot, a Luciferase reporter assay, Cell proliferation assay	Up-regulated	27027260
Glioblastoma	qPCR	Up-regulated	23046790
Hepatocellular carcinoma	qPCR, Western blot	Up-regulated	29029488
Hepatocellular carcinoma	qPCR, RNAi	Up-regulated	26045820
Lung cancer	qPCR, RNAi, Western blot, Cell apoptosis assay	Up-regulated	25964559
Nasopharyngeal cancer	RT-qPCR, Western blot, a Luciferase reporter assay, in vitro knockdown, RIP	Up-regulated	29463902
Nasopharyngeal cancer	qPCR, RNAi, Western blot, MTT assay	Up-regulated	27557514
Non-small cell lung cancer	qPCR, RNAi, Western blot, Luciferase reporter assay	Up-regulated	27307748
Non-small cell lung cancer	qPCR, RNAi, Cell proliferation assay	Up-regulated	25889788
Non-small cell lung cancer	qPCR, Western blot	Up-regulated	25504755
Osteosarcoma	qRT-PCR, Western blot, in vitro knockdown	Up-regulated	29520337
Ovarian cancer	qPCR, Western blot	Up-regulated	27095571
Pancreatic cancer	qPCR, Cell transfection, Western blot, cell migration, and invasion assay	Up-regulated	28344092
Renal cell carcinoma	qPCR, RNAi, Western blot, Cell migration and invasion assay, CCK-8 assay	Up-regulated	28251886
Serous ovarian cancer	qPCR, RNAi, Western blot	Up-regulated	25845387
Thyroid cancer	qPCR, RNAi, Western blot	Up-regulated	27507052

Breast cancer-related articles are shown in bold

Table 4 Publications on the role of *CDKN2B-AS1* with treatment resistance in cancer

Cancer site	Study type	Treatment	Putative function	Putative mechanism	PubMed ID
Osteosarcoma	In vitro	Cisplatin	Chemoresistance	Through STAT3 and miR-125a-5p	30777616
Colorectal cancer	In vitro	5-Fluorouracil	Chemoresistance	by regulating ATP-binding cassette subfamily C member 1 through binding Let-7a	30279206
Lung adenocarcinoma	In vitro	Paclitaxel	Chemoresistance	Through the mitochondrial pathway by modulating the expression of apoptosis-related protein cleaved-PARP and Bcl-2	28402932
Nasopharyngeal carcinoma	In vitro	Radiation	Radioresistance	Via functioning as a miR-125a sponge	28402230
	In vitro	Cisplatin	Chemoresistance	Via regulating microRNA let-7a	28117929
Multiple myeloma	Patients	Hematopoietic stem cell transplantation	Relapse	By modulating p14ARF-MDM2-p53 axis	28150872
Gastric cancer	In vitro	Cisplatin and 5-fluorouracil	Multidrug resistance	Decreased the expression of MDR1 and MRP1	27121324
Bladder cancer		Gemcitabine	Resistance	Through the Wnt signaling pathway	29937935
laryngeal squamous cell cancer		Cisplatin and paclitaxel		CDKN2B-AS1 expression decrease	25257554
Oral squamous cancer		Cisplatin	Resistance	Via impairment of the drug transporters MRP1 and ABCC2	29176691

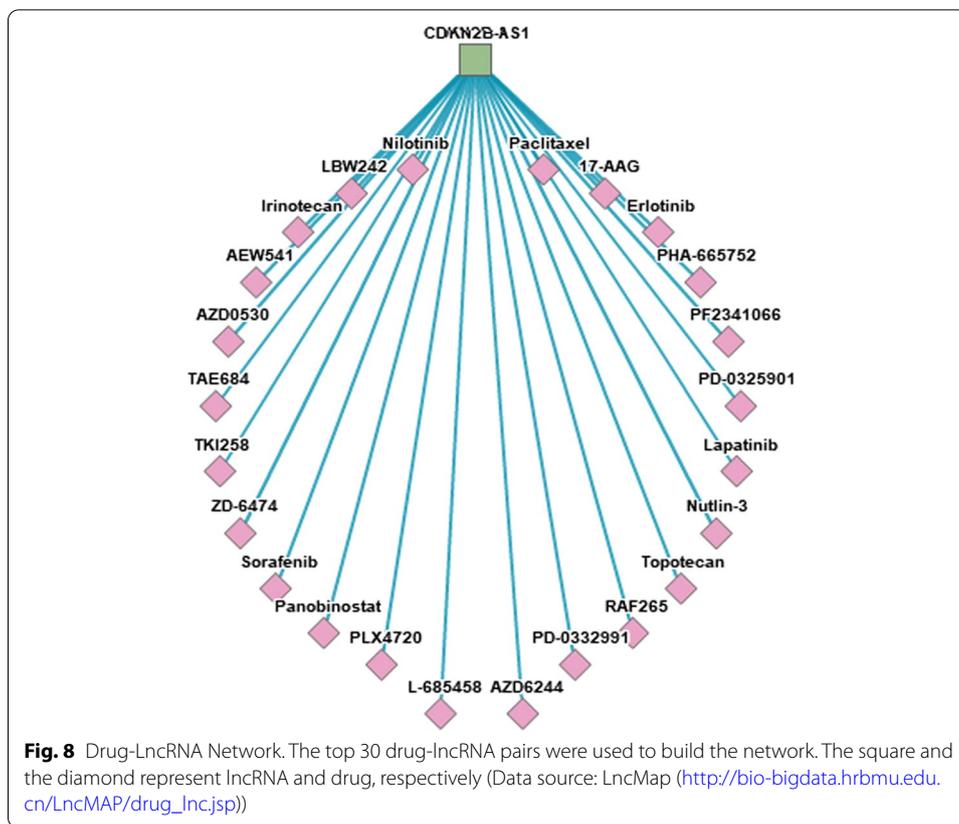
CDKN2B-AS1 gene are presented in Fig. 9c. The LncMAP revealed that *CDKN2B-AS1* could also bind to 12 TFs: E2F4, REST, BCL11A, RARA, ETS1, ESR2, GATA6, FLI1, EBF1, CTNBN1, STAT2, and ESRRA.

As depicted in Fig. 9d, Ingenuity Pathway Analysis showed key genes highly enriched in BC with *CDKN2B-AS1*, namely (1) *RELA* proto-oncogene, (2) prostaglandin-endoperoxide synthase 2 (*PTGS2*), (3) interleukin 6 (*IL6*), (4) vascular endothelial growth factor A (*VEGFA*), and (5) tumor necrosis factor-alpha (*TNF*).

Discussion

The last two decades have witnessed remarkable developments in the era of non-coding RNAs [40]. Given their high stability in different storage and handling conditions, being easily monitored by repeated sampling and their circulatory levels mirroring those in tissue cancer, serum/plasma lncRNAs demonstrated features of particular relevance for ideal biomarkers [41].

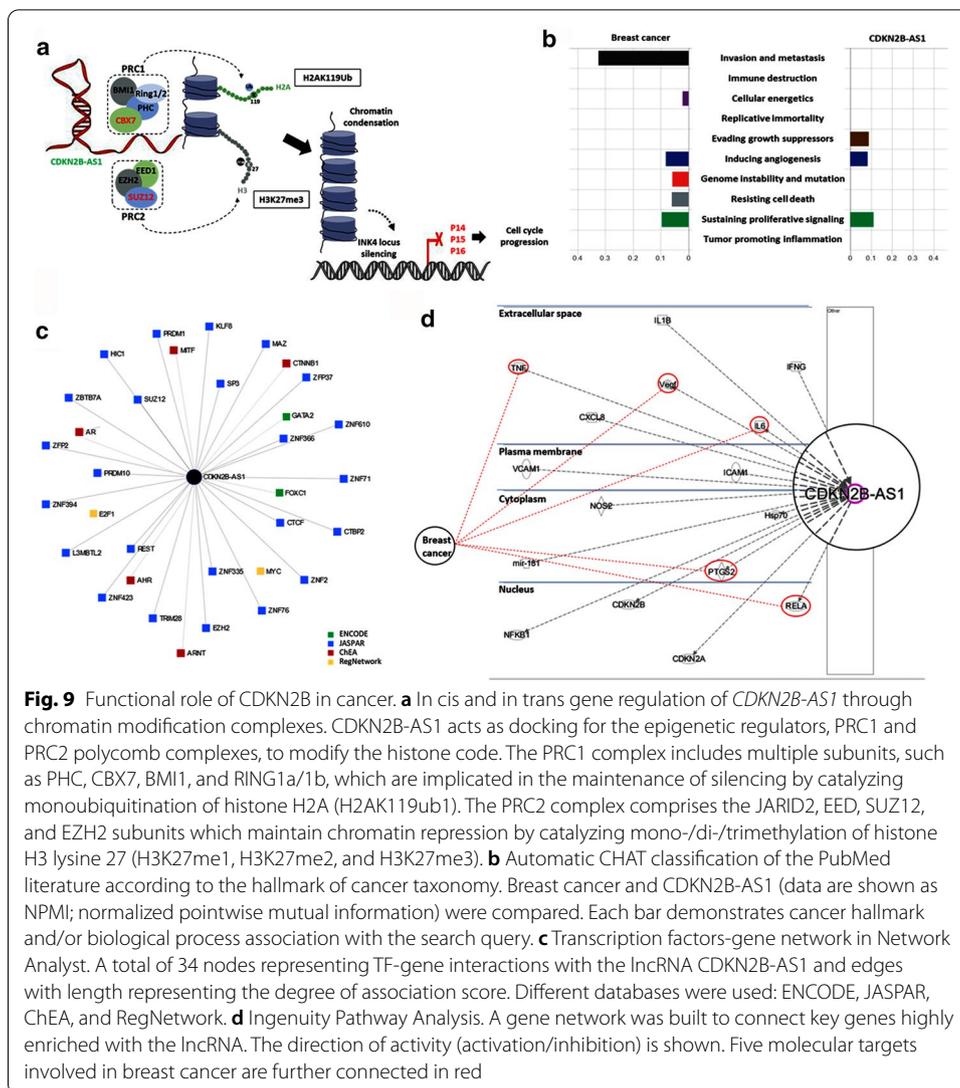
In this study, we evaluated the impact of serum lncRNA *CDKN2B-AS1* expression and variant signature in BC patients supported by in silico analyses followed by verifying the



results on samples from TCGA. We found that *CDKN2B-AS1* is upregulated in sera of BC patients compared with controls. This observation was in line with the oncogenic role this lncRNA plays in several cancers (Additional file 1: Table S3), which can support the potential role this type of lncRNA can play as a universal biomarker differentiating cancer from non-cancer patients.

CDKN2B-AS1 is known to be involved in transcriptional repression through forming chromatin modification complexes that execute histone modifications at specific sites. It can interact with both chromobox 7 (CBX7), a polycomb repressor component within PRC-1, and SUZ12, a subunit of PRC2 [42]. Next, PRC2 downregulates *INK4* expression (Fig. 9a) by inducing H3K27 tri-methylation. Meanwhile, PRC1 maintains the repressive chromatin structure by H2AK119 mono-ubiquitination [15, 43].

CDKN2B-AS1 is mainly upregulated by the E2F1 in an ATM-dependent manner after DNA damage, leading to cell cycle arrest to allow for DNA repair [44]. Several known potent oncogenes that regulate *CDKN2B-AS1* expression in various cancers have been reported in the literature, including MYC, RELA, and ERBB2 [12, 45]. *CDKN2B-AS1* expression is also regulated by interferon-gamma (IFN γ), tumor necrotic factor (TNF α) [46], and interleukin 1 beta (IL1B). These inflammatory mediators activate the nuclear factor kappa-B (NF- κ B), a pro-proliferation and survival factor, pathway and form a complex with the YY1 transcription factor to create transcriptional regulatory loops [47]. *CDKN2B-AS1* abundance can also be induced by exposure to hypoxia (a well-known phenomenon associated with the tumor microenvironment). It can bind the aryl



hydrocarbon receptor nuclear translocator (ARNT), which functions as a transcriptional regulator of the adaptive response to hypoxia [48]. Apart from the potential oncogenic role of *CDKN2B-AS1* in BC, all the above mechanisms could support *CDKN2B-AS1* upregulation in breast cancer in part as an adaptive response to DNA and cellular damage during the tumorigenesis process, which could explain the association of high expression of this type of lncRNA with survival and cancer grade; the findings were validated by the same results from the TCGA data set. Interestingly, using RNAScope, a recent study underscored that presence of *CDKN2B-AS1* in different subcellular locations in breast tumors may affect its functionality in cancer progression [49].

Our in-silico analysis, including the Ingenuity Pathway Analysis, showed several key genes highly enriched with *CDKN2B-AS1* in BC and could mediate part of its oncogenic role or impact the tumor microenvironment in this type of cancer. These genes include (1) the *RELA* (v-rel avian reticuloendotheliosis viral oncogene homolog A) gene encoding the NF-κB-p65 subunit (a ubiquitous TF held in the cytoplasm in

an inactive state by a specific inhibitor), upon degradation of which the NF- κ B translocates to the nucleus and activates specific gene expression, (2) *PTGS2* encoding a major enzyme in prostaglandin (PTG) biosynthesis implicated in biosynthesis of prostanoids that are involved in inflammation and mitogenesis, (3) *IL6*, which functions in inflammation and B-cell maturation, (4) *VEGFA*, which stimulates proliferation and migration of vascular endothelial cells and is a key player in physiological/pathological angiogenesis, and (5) the *TNF* gene that regulates cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation.

It is worth noting that the transcriptomic abundance of *CDKN2B-AS1* is not only modulated by epigenetic control through promoter transcriptional activity, promoter methylation, and splicing [50] but also post-transcriptionally regulated by miRNA sponging and RNA stability [12].

Using the ENCORI database, we identified the crosstalk of *CDKN2B-AS1* lncRNA with mRNAs and transcribed pseudogenes in cancer. Putative lncRNA–ceRNA interactions in cancer represent a new layer of gene regulation (Table 5). Mounting data from the literature suggest that deregulated *CDKN2B-AS1*–miRNA interacting networks have also been implicated in carcinogenesis, thus representing another molecular mechanism of the lncRNA. *CDKN2B-AS1* contains miRNA-binding domains in its sequence and therefore acts as a "sponge" to sequester miRNAs away from its mRNA targets. Through modulation of microRNA pathways, *CDKN2B-AS1* can also be involved in post-transcriptional regulation (Table 6). miRNA sponge mechanism was reported in BC patients with triple-negative receptors through interaction with microRNA-199a [51]. All the aforementioned cellular and genetic/epigenetic mechanisms support *CDKN2B-AS1* involvement in breast carcinogenesis and suggest that it could be a biomarker for BC detection with other molecular panels.

As one of the well-known lncRNAs that occur in genomic loci that harbor many cancer-associated SNPs [12], the *CDKN2B-AS1*-related variant rs2383207 showed, for the first time, an association with BC risk in the present study. The A/A homozygous carriers were three times more liable to develop BC under dominant and recessive genetic models. They were associated with short disease-free/overall survival compared to their counterpart G/G homozygous carriers. Intriguingly, the biallelic study variant was also associated with the gene expression, as G/G genotype was associated with a higher expression level than other genotypes (A/G and A/A).

Studies indicate that SNPs in *CDKN2B-AS1* can impact its expression and function with potential risk modification for diseases, including cancers [18, 52]. As most of the polymorphisms in this region do not impact any protein sequence, they likely act by affecting the expression of a nearby gene in cis, as proposed by Cunningham et al. [18]. Furthermore, as our in silico analysis showed, the existence of key genes highly enriched in BC with *CDKN2B-AS1* and upstream regulators enhancing the transcription of the *CDKN2B-AS1* gene (as detailed in Fig. 8) collectively suggests a wide "regulatory panorama" for this type of lncRNA [53]. Notably, it is not proved that the risky allele (the A allele in our case) is associated with high gene expression to increase BC susceptibility. As demonstrated in our in silico analysis and reported previously, the *CDKN2B-AS1* variants could act through "independent mechanisms" to increase/decrease susceptibility to diseases [53], and more than one functional variant as well

Table 5 Crosstalk between the lncRNA (*CDKN2B-AS1*), mRNAs, and pseudogenes in competing endogenous RNA (ceRNA) network. Data source: ENCORI for RNA interactomes supported by CLIP-seq data (<http://starbase.sysu.edu.cn/>)

ceRNA symbol	ceRNA name	Molecule type	Location	ceRNA gene Type	FDR	Pan-cancer number
AL160290.2	Cytochrome C Oxidase Subunit VIIc (COX7C)	–	–	LincRNA	2.00E–03	14
RPL39P26	Ribosomal Protein L39 Pseudogene 26	–	–	Processed pseudogene	2.00E–03	8
CLCN1	Chloride Voltage-Gated Channel 1	Ion channel	Plasma membrane	Protein-coding	7.94E–04	17
ARHGAP44	Rho GTPase Activating Protein 44	Other	Cytoplasm	Protein-coding	7.94E–04	13
RBM6	RNA Binding Motif Protein 6	Other	Nucleus	Protein-coding	7.94E–04	19
AK9	Adenylate Kinase 9	Kinase	Nucleus	Protein-coding	8.97E–04	18
USP2	Ubiquitin Specific Peptidase 2	Peptidase	Cytoplasm	Protein-coding	8.78E–04	17
SPRYD7	SPRY Domain Containing 7	Other	Other	Protein-coding	7.94E–04	10
CUL3	Cullin 3	Enzyme	Nucleus	Protein-coding	7.97E–04	9
EPHA7	EPH Receptor A7	Kinase	Plasma membrane	Protein-coding	1.30E–03	10
TLK1	Tousled Like Kinase 1	Kinase	Nucleus	Protein-coding	7.94E–04	10
ICMT	Isoprenylcysteine carboxyl methyltransferase	Enzyme	Cytoplasm	Protein-coding	1.83E–03	11
GPD2	Glycerol-3-phosphate dehydrogenase 2	Enzyme	Cytoplasm	Protein-coding	2.26E–03	15
LRRC58	Leucine-Rich Repeat Containing 58	Other	Other	Protein-coding	1.49E–03	14

FDR false discovery rate

as others in disequilibrium with the specified study variant, collectively, could impact *CDKN2B-AS1* expression [18].

In agreement with our results, the T allele of the rs2151280 variant (one of the *CDKN2B-AS1* SNPs) was related to increased susceptibility to neurofibromatosis type 1, associated with gene downregulation [54]. Also, the melanoma-associated rs1011970-T variant was reported to be associated with *CDKN2B-AS1* downregulation [18]. In contrast, the glioma-associated risky rs1063192-C allele was associated with increased *CDKN2B-AS1* expression [16, 18]. The presence of multiple regulatory elements and binding sites in the region of the studied variant suggests that its transcript, *CDKN2B-AS1*, with other genes in the same locus as *CDKN2A/B*, is subject to intricate "temporal and tissue-specific regulation" underlying the gene associations [53].

Table 6 Role of CDKN2B-AS1 in microRNA sponging in cancer. *Data source:* PubMed and NCBI

Cancer site	Sponged microRNA	Genes/pathway	Putative function	PubMed ID
Breast (TNBC)	miR-199a		Proliferation and apoptosis	28961506
Cervical cancer	miR-186		Cancer development	28550682
Oral cancer	miR-125a		Cell proliferation, migration, and invasion	29635126
Hepatocellular carcinoma	miR-122-5p		Cell proliferation, invasion, and metastasis	29127494
	miR-153-5p	ARHGAP18	Metastasis	30510148
	miR-384	STAT3		31679275
	miR-191	NF- κ B and Wnt/ β -catenin	Apoptosis, proliferation, metastasis, and invasion	30249208
	let-7c-5p	NAP1L1		30165194
	miR-122-5p		Cell proliferation, metastasis, and invasion	29127494
Gastric cancer	miR-99a	BMI1, Bcl-2	Cancer development	30156609
Colorectal	let-7a	ATP-binding cassette subfamily C member 1		30279206
Prostate cancer	let-7a	TGF- β 1/Smad	Proliferation and migration	29278879
Glioma	miR-34a	Sirt1, PI3K/AKT, and mTOR	Cell proliferation, migration, and invasion	29057547
Retinoblastoma	miR-99a	c-MYC	Apoptosis, proliferation, metastasis and invasion	31184221
	miR-24	c-MYC, MEK/ERK, and Wnt/ β -catenin	Viability, migration, and invasion	30703428
Osteosarcoma	miR-125a-5p	STAT3	Cisplatin chemoresistance	30777616
Nasopharyngeal carcinoma	miR-125a			28402230
	let-7a		Cisplatin chemoresistance	28117929
Medulloblastoma	miR-323	BRI3 and CDK6, p38 MAPK, ERK, and AKT, Wnt signaling	Cell proliferation and migration through miR-323-mediated regulation of BRI3	28513871

Based on the limited study cohort, future large-scale studies in other populations are warranted to confirm the studied variant's association with BC risk. Also, supporting in vitro studies are needed to confirm the impact of different rs2383207 genotypes on *CDKN2B-AS1* expression. Finally, follow-up studies are warranted to assess the specified lncRNA's validity as a diagnostic/prognostic epigenetic marker and the response to chemotherapy in BC patients.

Conclusion

In summary, the study findings for the first time support the association of the *CDKN2B-AS1* rs2383207 (A>G) variant with gene expression and BC risk in the study population. It might provide new insight into the molecular stratification of BC patients. Also, high levels of circulating CDKN2B-AS1 could discriminate BC

patients from controls and had prognostic value in patients' survival, supporting its use in future epigenetic personalized BC treatment.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s11658-021-00258-9>.

Additional file 1: Table S1. Several types of long non-coding RNAs are involved in breast cancer. **Table S2.** Overlapping transcripts for rs2383207 variant according to NCBI and Varsome.com. **Table S3.** Role of CDKN2B-AS1 in cancer as a diagnostic and prognostic biomarker and therapeutic target.

Authors' contributions

EAT, SWK, MSF: study design; SWK, YHH, SS, HAA: study identification and data extraction; MHH, EAT: statistical and bioinformatic analyses; SWK, YHH, SS, HAA, EAT, MSF: contributed the reagents and materials; EAT, HAA, MSF: draft preparation. All authors revised the manuscript. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The data that support the findings of this study are available in the manuscript and the additional files.

Declarations

Ethics approval and consent to participate

All procedures related to human participants conducted in the present study followed the ethical standards of the institutional research ethics committee of the Faculty of Medicine, Suez Canal University (Approval no. 3960). All study subjects gave written informed consent before the start of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Medical Laboratory, College of Applied Medical Sciences, Taibah University, Yanbu, Saudi Arabia.

² Department of Medical Laboratory Technology, College of Applied Medical Sciences, Jazan University, Jazan, Kingdom of Saudi Arabia.

³ Anatomy Department and Stem Cell Unit, College of Medicine, King Saud University, Riyadh, Saudi Arabia.

⁴ Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

⁵ Department of Surgery, Tulane University, School of Medicine, New Orleans, LA, USA.

⁶ Genetics Unit, Department of Histology and Cell Biology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt.

⁷ Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt.

⁸ Department of Biochemistry, College of Medicine, Northern Border University, Arar, Saudi Arabia.

⁹ Department of Medical Biochemistry, Faculty of Medicine, Taibah University, Medina, Saudi Arabia.

¹⁰ Department of Medical Biochemistry, Faculty of Medicine, Mansoura University, Mansoura, Egypt.

Received: 5 November 2020 Accepted: 3 April 2021

Published online: 13 April 2021

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34. <https://doi.org/10.3322/caac.21551>.
2. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223–7. <https://doi.org/10.1038/nature07672>.
3. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: moving forward. *PLoS Genet*. 2018;14(6):e1007362. <https://doi.org/10.1371/journal.pgen.1007362>.
4. Wei JW, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics (Review). *Oncol Rep*. 2017;37(1):3–9. <https://doi.org/10.3892/or.2016.5236>.
5. Bartoszewski R, Sikorski AF. Editorial focus: entering into the non-coding RNA era. *Cell Mol Biol Lett*. 2018;23(1):45. <https://doi.org/10.1186/s11658-018-0111-3>.
6. Wu Z, Liu X, Liu L, Deng H, Zhang J, Xu Q, et al. Regulation of lncRNA expression. *Cell Mol Biol Lett*. 2014;19(4):561–75. <https://doi.org/10.2478/s11658-014-0212-6>.
7. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136(4):629–41. <https://doi.org/10.1016/j.cell.2009.02.006>.

8. Youness RA, Gad MZ. Long non-coding RNAs: functional regulatory players in breast cancer. *Noncoding RNA Res.* 2019;4(1):36–44. <https://doi.org/10.1016/j.ncrna.2019.01.003>.
9. Shi T, Gao G, Cao Y. Long noncoding RNAs as novel biomarkers have a promising future in cancer diagnostics. *Dis Markers.* 2016;2016:9085195. <https://doi.org/10.1155/2016/9085195>.
10. Fawzy MS, Toraih EA, Abdallah HY. Long noncoding RNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1): a molecular predictor of poor survival in glioblastoma multiforme in Egyptian patients. *Egypt J Med Hum Genet.* 2017;18:231–9. <https://doi.org/10.1016/j.ejmhg.2016.08.003>.
11. Ma J, Qi G, Li L. LncRNA NNT-AS1 promotes lung squamous cell carcinoma progression by regulating the miR-22/FOXM1 axis. *Cell Mol Biol Lett.* 2020;25(1):34. <https://doi.org/10.1186/s11658-020-00227-8>.
12. Kong Y, Hsieh CH, Alonso LC. A lncRNA at the CDKN2A/B locus with roles in cancer and metabolic disease. *Front Endocrinol (Lausanne).* 2018;9:405. <https://doi.org/10.3389/fendo.2018.00405>.
13. Meseure D, Vacher S, Alsibai KD, Nicolas A, Chemlali W, Cally M, et al. Expression of ANRIL-polycomb complexes-CDKN2A/B/ARF genes in breast tumors: identification of a two-gene (EZH2/CBX7) signature with independent prognostic value. *Mol Cancer Res.* 2016;14(7):623–33. <https://doi.org/10.1158/1541-7786.MCR-15-0418>.
14. Laugesen A, Højfeldt JW, Helin K. Role of the polycomb repressive complex 2 (PRC2) in transcriptional regulation and cancer. *Cold Spring Harb Perspect Med.* 2016;6(9):a026575. <https://doi.org/10.1101/cshperspect.a026575>.
15. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell.* 2010;38(5):662–74. <https://doi.org/10.1016/j.molcel.2010.03.021>.
16. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, et al. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis.* 2012;220(2):449–55. <https://doi.org/10.1016/j.atherosclerosis.2011.11.017>.
17. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, et al. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. *PLoS Genet.* 2013;9(7):e1003588. <https://doi.org/10.1371/journal.pgen.1003588>.
18. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet.* 2010;6(4):e1000899. <https://doi.org/10.1371/journal.pgen.1000899>.
19. Iranpour M, Soudyab M, Geranpayeh L, Mirfakhraie R, Azargashb E, Movafagh A, et al. Expression analysis of four long noncoding RNAs in breast cancer. *Tumour Biol.* 2016;37(3):2933–40. <https://doi.org/10.1007/s13277-015-4135-2>.
20. Roys JA, Pilbrow AP, Ahn A, Morrin HR, Frampton C, Russell IA, et al. The rs11515 polymorphism is more frequent and associated with aggressive breast tumors with increased ANRIL and decreased p16 (INK4a) expression. *Front Oncol.* 2015;5:306. <https://doi.org/10.3389/fonc.2015.00306>.
21. Huang X, Zhang W, Shao Z. Association between long non-coding RNA polymorphisms and cancer risk: a meta-analysis. *Biosci Rep.* 2018;6:46:227.
22. Duan F, Jiang J, Song C, Wang P, Ye H, Dai L, et al. Functional long non-coding RNAs associated with gastric cancer susceptibility and evaluation of the epidemiological efficacy in a central Chinese population. *Gene.* 2018;646:227–33. <https://doi.org/10.1016/j.gene.2017.12.063>.
23. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* 2010;6(12):e1001233. <https://doi.org/10.1371/journal.pgen.1001233>.
24. Kataja V, Castiglione M, Group EGW. Primary breast cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol.* 2009;20(Suppl 4):10–4. <https://doi.org/10.1093/annonc/mdp114>.
25. Meyer JS, Alvarez C, Milikowski C, Olson N, Russo I, Russo J, et al. Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Mod Pathol.* 2005;18(8):1067–78. <https://doi.org/10.1038/modpathol.3800388>.
26. Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, et al. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res.* 2009;11(3):R27. <https://doi.org/10.1186/bcr2257>.
27. Toraih EA, El-Wazir A, Ageeli EA, Hussein MH, Eltoukhy MM, Killackey MT, et al. Unleash multifunctional role of long noncoding RNAs biomarker panel in breast cancer: a predictor classification model. *Epigenomics.* 2020;12(14):1215–37. <https://doi.org/10.2217/epi-2019-0291>.
28. Toraih EA, Mohammed EA, Farrag S, Ramsis N, Hosny S. Pilot study of serum MicroRNA-21 as a diagnostic and prognostic biomarker in Egyptian breast cancer patients. *Mol Diagn Ther.* 2015;19(3):179–90. <https://doi.org/10.1007/s40291-015-0143-6>.
29. Babteen NA, Fawzy MS, Alelwani W, Alharbi RA, Alruwetei AM, Toraih EA, et al. Signal peptide missense variant in cancer-brake gene CTLA4 and breast cancer outcomes. *Gene.* 2020;737:144435. <https://doi.org/10.1016/j.gene.2020.144435>.
30. Bustin SA, Benes V, Garson JA, Hellems J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55(4):611–22. <https://doi.org/10.1373/clinchem.2008.112797>.
31. Fawzy MS, Toraih EA, Abdallah HY. Long noncoding RNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1): a molecular predictor of poor survival in glioblastoma multiforme in Egyptian patients. *Egypt J Med Hum Genet.* 2016. <https://doi.org/10.1016/j.ejmhg.2016.08.003>.
32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001;25(4):402–8. <https://doi.org/10.1006/meth.2001.1262>.
33. Toraih EA, Fawzy MS, Elgazzaz MG, Hussein MH, Shehata RH, Daoud HG. Combined genotype analyses of precursor miRNA196a2 and 499a variants with hepatic and renal cancer susceptibility: a preliminary study. *Asian Pac J Cancer Prev.* 2016;17(7):3369–75.

34. Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients. *Breast Cancer Res Treat.* 2010;123(3):725–31. <https://doi.org/10.1007/s10549-009-0674-9>.
35. Fekete JT, Györfy B. ROCplot.org: validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3104 breast cancer patients. *Int J Cancer.* 2019;145(11):3140–51. <https://doi.org/10.1002/ijc.32369>.
36. Gao Y, Wang P, Wang Y, Ma X, Zhi H, Zhou D, et al. Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.* 2019;47(D1):D1028–33. <https://doi.org/10.1093/nar/gky1096>.
37. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 2019;47(W1):W234–41. <https://doi.org/10.1093/nar/gkz240>.
38. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42(Database issue):D92–7. <https://doi.org/10.1093/nar/gkt1248>.
39. Fawzy MS, Hussein MH, Abdelaziz EZ, Yamany HA, Ismail HM, Toraih EA. Association of MicroRNA-196a2 variant with response to short-acting beta 2-agonist in COPD: an Egyptian pilot study. *PLoS ONE.* 2016;11(4):e0152834. <https://doi.org/10.1371/journal.pone.0152834>.
40. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics.* 2013;193(3):651–69. <https://doi.org/10.1534/genetics.112.146704>.
41. Pardini B, Sabo AA, Birolo G, Calin GA. Noncoding RNAs in extracellular fluids as cancer biomarkers: the new frontier of liquid biopsies. *Cancers (Basel).* 2019;11(8):1170. <https://doi.org/10.3390/cancers11081170>.
42. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene.* 2011;30(16):1956–62. <https://doi.org/10.1038/onc.2010.568>.
43. Latgé G, Poulet C, Bours V, Josse C, Jerusalem G. Natural antisense transcripts: molecular mechanisms and implications in breast cancers. *Int J Mol Sci.* 2018;19(1):123. <https://doi.org/10.3390/ijms19010123>.
44. Wan G, Mathur R, Hu X, Liu Y, Zhang X, Peng G, et al. Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. *Cell Signal.* 2013;25(5):1086–95. <https://doi.org/10.1016/j.cellsig.2013.02.006>.
45. Lu Y, Zhou X, Xu L, Rong C, Shen C, Bian W. Long noncoding RNA ANRIL could be transactivated by c-Myc and promote tumor progression of non-small-cell lung cancer. *Onco Targets Ther.* 2016;9:3077–84. <https://doi.org/10.2147/OTT.S102658>.
46. Zhou X, Han X, Wittfeldt A, Sun J, Liu C, Wang X, et al. Long non-coding RNA ANRIL regulates inflammatory responses as a novel component of NF- κ B pathway. *RNA Biol.* 2016;13(1):98–108. <https://doi.org/10.1080/15476286.2015.1122164>.
47. Aguilo F, Di Cecilia S, Walsh MJ. Long non-coding RNA ANRIL and polycomb in human cancers and cardiovascular disease. *Curr Top Microbiol Immunol.* 2016;394:29–39. https://doi.org/10.1007/82_2015_455.
48. Kuo TC, Kung HJ, Shih JW. Signaling in and out: long-noncoding RNAs in tumor hypoxia. *J Biomed Sci.* 2020;27(1):59. <https://doi.org/10.1186/s12929-020-00654-x>.
49. Mehta-Mujoo PM, Cunliffe HE, Hung NA, Slatter TL. Long non-coding RNA ANRIL in the nucleus associates with periostin expression in breast cancer. *Front Oncol.* 2019;9:885.
50. Popov N, Gil J. Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics.* 2010;5(8):685–90. <https://doi.org/10.4161/epi.5.8.12996>.
51. Xu ST, Xu JH, Zheng ZR, Zhao QQ, Zeng XS, Cheng SX, et al. Long non-coding RNA ANRIL promotes carcinogenesis via sponging miR-199a in triple-negative breast cancer. *Biomed Pharmacother.* 2017;96:14–21. <https://doi.org/10.1016/j.biopha.2017.09.107>.
52. Shastry BS. SNPs: impact on gene function and phenotype. *Methods Mol Biol.* 2009;578:3–22. https://doi.org/10.1007/978-1-60327-411-1_1.
53. Congrains A, Kamide K, Ohishi M, Rakugi H. ANRIL: molecular mechanisms and implications in human health. *Int J Mol Sci.* 2013;14(1):1278–92. <https://doi.org/10.3390/ijms14011278>.
54. Pasmant E, Sabbagh A, Masliah-Planchon J, Ortonne N, Laurendeau I, Melin L, et al. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. *J Natl Cancer Inst.* 2011;103(22):1713–22. <https://doi.org/10.1093/jnci/djr416>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.