Genome-wide long non-coding RNA association study on Han Chinese women identifies lncHSAT164 as a novel susceptibility gene for breast cancer

Jing-Kai Xu[1], Guo-Zheng Li[2,3], Zhi Li[2,3], Wen-Jing Li[2,3], Run-Sheng Chen[4], Bo Zhang[2,3], Xue-Jun Zhang[1]

[1]Department of Dermatology, The First Affiliated Hospital of Anhui Medical University, Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, Hefei, Anhui 230032, China;

[2]Department of Genetics, School of Life Sciences, Anhui Medical University, Hefei, Anhui 230032, China;

[3]Department of Oncology, No. 2 Hospital, Anhui Medical University, Hefei, Anhui 230032, China;

[4]Laboratory of Bioinformatics and Noncoding RNA, Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

Abstract

Background: Single-nucleotide polymorphisms (SNPs)-associated genes and long non-coding RNAs (lncRNAs) can contribute to human disease. To comprehensively investigate the contribution of lncRNAs to breast cancer, we performed the first genome-wide lncRNA association study on Han Chinese women.

Methods: We designed an lncRNA array containing >800,000 SNPs, which was incorporated into a 96-array plate by Affymetrix (CapitalBio Technology, China). Subsequently, we performed a two-stage genome-wide lncRNA association study on Han Chinese women covering 11,942 individuals (5634 breast cancer patients and 6308 healthy controls). Additionally, *in vitro* gain or loss of function strategies were performed to clarify the function of a novel SNP-associated gene.

Results: We identified a novel breast cancer-associated susceptibility SNP, rs11066150 ($P_{meta} = 2.34 \times 10^{-8}$), and a previously reported SNP, rs9397435 ($P_{meta} = 4.32 \times 10^{-38}$), in Han Chinese women. rs11066150 is located in NONHSAT164009.1 (lncHSAT164), which is highly expressed in breast cancer tissues and cell lines. lncHSAT164 overexpression promoted colony formation, whereas lncHSAT164 knockdown promoted cell apoptosis and reduced colony formation by regulating the cell cycle.

Conclusions: Based on our lncRNA array, we identified a novel breast cancer-associated lncRNA and found that lncHSAT164 may contribute to breast cancer by regulating the cell cycle. These findings suggest a potential therapeutic target in breast cancer.

Keywords: Breast cancer; lncHSAT164; Single-nucleotide polymorphisms

Introduction

Breast cancer is a leading cause of cancer-related deaths in women worldwide.[1] The 5-year prevalence statistics indicate that approximately 11% of breast cancers worldwide occur in China, and the incidence has increased rapidly in recent decades.[2] Genome-wide association studies and exome sequencing have been widely used to identify disease-associated susceptibility single-nucleotide polymorphisms (SNPs)/genes/loci,[3-10] and >180 breast cancer-associated SNPs have been identified in over 100 susceptibility genes/loci. The majority of identified SNPs fall outside of the protein-coding regions, which comprise only 2% of the whole genome.[11,12]

In the whole genome, the vast majority of transcribed sequences does not encode proteins and are thus called non-coding RNAs. Among these non-coding transcripts, transcripts of long non-coding RNAs (lncRNAs) can work together with other proteins and participate in various biological processes,[13-16] including carcinogenesis, tumor growth, and patient prognosis. Recent technological advances in high-throughput sequencing have created an opportunity to identify breast cancer-associated lncRNAs and generated compelling evidence that lncRNAs contribute to the metastasis of breast cancer.[17-20] In addition, the SNPs in lncRNA regions can also promote disease development.[21,22] However, investigations about lncRNAs are still limited due to small sample sizes and weak statistical power, rendering it imperative to thoroughly study the correlation between lncRNA SNPs and disease development.

To comprehensively investigate the contribution of lncRNAs to breast cancer, we designed a genome-wide

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000001429

Correspondence to: Prof. Xue-Jun Zhang, Department of Dermatology, The First Affiliated Hospital of Anhui Medical University, Key Laboratory of Dermatology, Ministry of Education, Hefei, Anhui 230032, China
E-Mail: ayzxj@vip.sina.com

Copyright © 2021 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2021;134(10)

Received: 22-12-2020 **Edited by:** Ning-Ning Wang

lncRNA array containing >800,000 SNPs and performed a genome-wide lncRNA association study on Han Chinese women. Furthermore, we investigated the function of SNP-associated genes in breast cancer.

Methods

Ethical approval

This study was approved by the Ethics Committee of Anhui Medical University (No. 20131301). All participants provided written informed consent. The current study was conducted according to the principles of the *Declaration of Helsinki*.

lncRNA array

Based on the NONCODE database, the Run-Sheng Chen team independently designed an lncRNA array,[12,23] which included >800,000 SNPs in total. Among these SNPs, 425,000 were located in the whole-genome non-coding region, 8187 were located in the non-coding gene regulatory region (promoter and enhancer regions), and 11,466 were located in the miRNA binding region. A total of 130,831 of these SNPs were recruited from RegulomeDB (https://www.regulomedb.org/regulome-search/), and 11,764 were associated with tumors, immune diseases, and cardiovascular diseases. This lncRNA array covers 150,000 SNPs in the Illumina Human OmniZhongHua array (https://www.illumina.com/products/by-type/microarray-kits/infinium-omni-zhonghua.html),[24] and >60% of SNPs in this lncRNA array were distributed in the Illumina Human OmniZhongHua array SNP genome linkage disequilibrium (LD) region. All SNP data were collected from the 1000 Genomes Project ($r^2 = 0.8$).

Study subjects

A total of 5634 cases and 6308 healthy controls were enrolled in our study through a collaborative consortium in China. All cases were diagnosed with breast cancer by at least two pathologists, and their clinical information was collected by professional investigators with a comprehensive clinical check-up. All of the healthy controls were clinically determined to be breast cancer free and to have no family history of breast cancer (including first- and second-degree relatives). Peripheral blood was collected from all participants with an anticoagulation tube and stored in a −80°C freezer for sequencing analysis. Specimens of cancer tissues and para-cancerous tissues were obtained with informed consent from patients (all Han Chinese women) at the No. 2 Hospital, Anhui Medical University (Hefei, China), between July 2011 and September 2016. For specimens of para-carcinoma tissues or healthy controls, we collected a small amount of adipose/skin tissue from breast cancer patients who underwent radical mastectomy. All tissue samples were stored in liquid nitrogen immediately after surgical resection. All participants provided written informed consent, and all investigators were blinded to the group allocation during the experiment when assessing outcomes.

Quality control (QC) and statistical analysis

PLINK 1.07 software (developed by Christopher Chang with support from NIH-NIDDK's Laboratory of Biological Modeling, the Purcell Lab and others; USA)[25] was used to examine potential genetic relatedness based on pairwise identity by state for all the successfully genotyped samples. A principal component analysis (PCA)-based approach was used to assess population outliers and stratification.[26] All human leukocyte antigen SNPs on chromosome 6 between 25 and 34 Mb and SNPs on non-autosomes were removed. We excluded SNPs with a call rate <99%, a minor allele frequency <0.0001, and/or a significant deviation from Hardy-Weinberg equilibrium in the controls ($P < 10^{-4}$) during each stage. LD pruning was carried out by using the PLINK option "-indep-pairwise 50 5 0.20."

Cell culture

MCF10A, MC7F, T47D, and HEK293T cell lines were purchased from the Institute of Basic Medical Sciences of the Chinese Academy of Medical Sciences. MCF10A cells were grown in Dulbecco's modified eagle medium/nutrient mixture F-12 (DMEM/F12; 1:1, Gibco, Life, USA) medium supplemented with 10% fetal bovine serum (FBS) (Gibco, Australia), 10 μg/mL insulin (Macklin, China), 20 ng/mL epidermal growth factor (EGF) (Peprotech, China), 0.5 μg/mL hydrocortisone Macklin, China), and 100 U/mL penicillin-streptomycin. MCF7 cells were maintained in Dulbecco's modified eagle medium (DMEM; Gibco, Life, USA) supplemented with 10% FBS (Gibco, Australia) and 100 U/mL penicillin-streptomycin. T47D cells were maintained in Roswell Park Memorial Institute-1640 (RPMI-1640) medium (Gibco, Life, USA) supplemented with 10% FBS (Gibco, Australia) and 100 U/mL penicillin-streptomycin. All cells were maintained at 37°C in a humidified atmosphere containing 5% CO_2 and confirmed to be mycoplasma free.

RNA extraction and quantitative real-time PCR (RT-qPCR) analysis

The total RNA from the cell lines, human breast cancer tissues, and para-cancerous tissues used in this study, was extracted with TRIzol reagent and incubated with DNase I (Thermo Fisher, USA) to remove genomic DNA. First-strand complementary DNA (cDNA) was synthesized by using the SuperScript III Reverse Transcriptase Kit (Thermo Fisher, USA). Relative RNA levels determined by RT-qPCR were measured on a Rotor-Gene Q real-time PCR machine (Qiagen, Germany). Glyceraldehyde-3-phosphate dehydrogenase was employed as an internal control. The relative expression of RNAs was calculated using the $2^{-\Delta\Delta Ct}$ method.

Subcellular fractionation

The separation of nuclear and cytosolic fractions was performed using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher, USA) according to the manufacturer's instructions. RNA was extracted and RT-qPCR was performed to assess lncHSAT164 expression in the nuclear and cytosolic fractions.

5′ and 3′ rapid amplification of cDNA ends (RACE) and northern blot analysis

5′ and 3′ RACE was performed with a SMARTer® RACE 5′/3′ Kit (Takara, China). PCR products were obtained by using SeqAmp™ DNA Polymerase (Takara, China) and separated on a 2% agarose gel. Gel bands were purified, cloned into the pEASY-T5 Zero vector (Transgen, China), and sequenced at Beijing Genomics Institute. Northern blot analysis was carried out with a NorthernMax-Gly Kit (Ambion, USA) according to the manufacturer's instructions. For this procedure, 30 μg of the total RNA was mixed with the glyoxal loading dye, denatured, and loaded onto a 1% agarose gel to detect lncHSAT164.

Plasmid construction, transfection, and lentivirus infection

Full-length lncHSAT164 was amplified according to RACE results and cloned into the pcDNA3.1 vector (Invitrogen, Waltham, MA, USA) according to the manufacturer's instructions. In total, 5×10^5 cells were seeded into each well of a six-well plate and transfected with plasmids upon reaching 80% to 90% confluence. The recombinant plasmid pcDNA3.1 (normal control [NC])-lncHSAT164+ (164+) and the control plasmid pcDNA3.1 (normal control [NC]) were constructed and transfected separately into cells using Lipofectamine 3000 (Invitrogen, USA) according to the instruction manual. Cells were harvested 48 h after transfection for further experiments.

Short hairpin RNAs against lncHSAT164 were designed and cloned into the enhanced green fluorescent protein-puromycin-pll3.7 plasmid. Based on the psPAX2-pMD2G lentiviral system (Addgene, Beijing, China), a lentivirus was constructed according to the manufacturer's instructions. After lentivirus infection, 1 μg/mL puromycin (Invitrogen, USA) was added for selection, and 48 to 72 h later, the cells were harvested for further experiments.

Colony formation assay

Cells (1×10^3 cells/well) were seeded in a six-well plate and incubated for 10 to 15 days at 37°C. Then, the cells were washed twice in phosphate buffered solution (PBS), fixed with 90% ethanol for 15 min, and stained with 0.1% crystal violet for 20 min. Images of colonies were taken with a digital camera, and the number of colonies was analyzed by ImageJ v1.8.0 software.

Flow cytometry analyses of apoptosis and the cell cycle

For apoptosis analysis, target cells were transferred to a 15 mL centrifuge tube, and annexin V binding buffer was added. After being centrifugated at 1000 rpm for 5 min at 4°C, the cells were washed three times in PBS. Then, 100 μL of binding buffer was added, and 5 μL of Annexin V-APC and 1 μL of 100 μg/mL propidium iodide (PI) stain (Thermo Fisher, USA) were added and incubated in the dark for 25 min. For cell cycle analysis, target cells were fixed with 75% ice-cold ethanol at 4°C overnight. Then, the cells were suspended in PBS supplemented with 100 mg/mL RNAse A for 30 min at 37°C and then stained with 50 μg/mL PI (Thermo Fisher, USA) in the dark at room temperature for 15 min. Finally, a total of 20,000 cells were analyzed on a FACSCalibur flow cytometer equipped with Cell Quest software (BD Biosciences, USA).

Results

Identification of breast cancer-associated SNPs

Based on our lncRNA array, 1496 cases and 1257 controls of Han Chinese women were genotyped in the first stage. After QC filtering and PCA, 1675 SNPs with P values < 10^{-4} were identified [Supplementary Table 1, http://links.lww.com/CM9/A495]. Based on the previously reported SNP selection criteria,[27,28] 70 SNPs with P values < 10^{-5} were selected for replication through the Sequenom MassARRAY system. An independent cohort including 4138 cases and 5051 controls was involved in the second replication stage, and after QC, 23 SNPs were replicated (P value < 0.05). A meta-analysis of the 70 SNPs in this lncRNA array discovery and replication stage study identified two variant SNPs and two suggestive SNPs [Figure 1 and Table 1]. All four SNPs were subjected to PCA [Supplementary Figure 1A–H, http://links.lww.com/

Figure 1: Flow chart of the genome-wide lncRNA analysis of breast cancer in Chinese Han women. CHB: Chinese Han Beijing; HWE: Hardy-Weinberg equilibrium; lncRNA: Long non-coding RNA; MAF: Minor allele frequency; PCA: Principal component analysis; SNPs: Single-nucleotide polymorphisms.

Table 1: Association of the two stages and meta-analyses based on logistic regression.

SNP	BP (hg19)	Non-coding RNA transcript annotation	Allele	lncRNA array					Genotyping validation					Meta		
				F_A	F_U	P value	OR (95% CI)	P-HWE	F_A	F_U	P value	OR (95% CI)	P-HWE	P value	HetISq	HetPVal
rs9397435	chr6_151951220	*N/A	G/A	0.3759	0.2825	4.34×10^{-13}	1.53 (1.36–1.72)	0.4452	0.3669	0.2914	4.05×10^{-27}	1.41 (1.32–1.50)	0.1397	4.32×10^{-38}	33.1	0.2216
rs11066150	chr12_112518803	NONHSAT164009.1: intron_1	A/G	0.4276	0.3721	3.50×10^{-5}	1.26 (1.13–1.41)	0.8564	0.4235	0.3933	3.91×10^{-5}	1.13 (1.07–1.20)	0.2234	2.34×10^{-8}	64.6	0.09297
rs12537	chr22_30423460	NONHSAT192799.1: exon_1	T/C	0.1922	0.2387	3.44×10^{-5}	0.76 (0.67–0.86)	0.3143	0.1772	0.1964	9.97×10^{-4}	0.88 (0.82–0.95)	0.3676	8.84×10^{-7}	73.5	0.05202
rs62112521	chr19_51162756	*N/A	A/G	0.4278	0.4865	1.72×10^{-5}	0.79 (0.71–0.88)	0.778	0.4305	0.4507	6.61×10^{-3}	0.92 (0.87–0.98)	0.8408	8.80×10^{-6}	83.6	0.01366

*N/A: No comment information. BP: Base position; CI: Confidence interval; HWE: Hardy-Weinberg equilibrium; lncRNA: Long non-coding RNA; OR: Odds ratio; SNP: Single-nucleotide polymorphism.

CM9/A492]. The novel breast cancer-associated SNP verified in this work, rs11066150 ($P_{meta} = 2.34 \times 10^{-8}$, odds ratio [OR] = 1.13), is located at 12q24.13. According to the NONCODE database, rs11066150 was located in the NONHSAT164009.1 transcript intron (lncHSAT164). A previously reported breast cancer-associated variant rs9397435 (CCDC170, $P_{meta} = 4.32 \times 10^{-38}$, OR = 1.41) was also confirmed by our data,[3] which suggests that the genotype data in this lncRNA array are highly reliable for future studies.

Characterization of lncHSAT164 in breast cancer

According to the NONCODE database, rs11066150 is located in the lncHSAT164 transcript, which is >867 bp. It is highly expressed in HepG2 (hepatocellular carcinoma cell line exosomes), invasive non-functional pituitary adenomas exosomes (NFPAs), non-invasive NFPAs, and tuberculosis patients serum (active tuberculosis patients serum exosomes), but did not express in A431 cell line (squamous cell carcinoma cell line exosomes) and MCF7 cell line (human breast cancer cell line exosomes) [Supplementary Figure 2, http://links.lww.com/CM9/A493]. We performed the qPCR analysis with four paired breast cancer tissues and para-cancerous tissues and found that lncHSAT164 was more highly expressed in the cancer tissues than in the para-cancerous tissues [Figure 2A]. Furthermore, the qPCR analysis was performed among the normal breast cell line MCF10A and the breast cancer cell lines MCF7 and T47D. The results demonstrated that lncHSAT164 was highly expressed in T47D cells [Figure 2B]. To determine the full length of lncHSAT164, 5′ and 3′ RACE [Supplementary Figure 3A, http://links.lww.com/CM9/A494] and Northern blotting were performed [Figure 2C]. Our results revealed that lncHSAT164 is an ~1700-nucleotide-long transcript with a polyadenylated tail.

Next, we examined the subcellular localization of lncHSAT164 in the MCF7 and T47D cell lines and discovered that lncHSAT164 predominately resided in the nucleus [Figure 2D]. Colony formation assays illustrated that clonogenic survival remarkably increased following lncHSAT164 overexpression [Figure 2F–G]. We reasoned that lncHSAT164 might play an important role in the nucleus and could promote the growth of breast cancer cells.

lncHSAT164 knockdown regulates cell proliferation and the cell cycle

To further investigate the function of lncHSAT164 in breast cancer, a lentivirus was constructed to knockdown lncHSAT164 in T47D cells [Figure 3A]. In addition, face shorting was used to examine the impact of lncHSAT164 knockdown on cell apoptosis and the cell cycle, and the results showed that the number of apoptotic cells increased after lncHSAT164 knockdown ($P < 0.01$) [Figures 3B and 3D]. Fewer cells remained in the G0/G1 phase, but more cells were arrested in the G2/M phase ($P < 0.05$) [Figures 3C and 3E]. Colony formation assays revealed that clonogenic survival greatly decreased after lncHSAT164 knockdown [Figures 3F and 3G], and this result was also

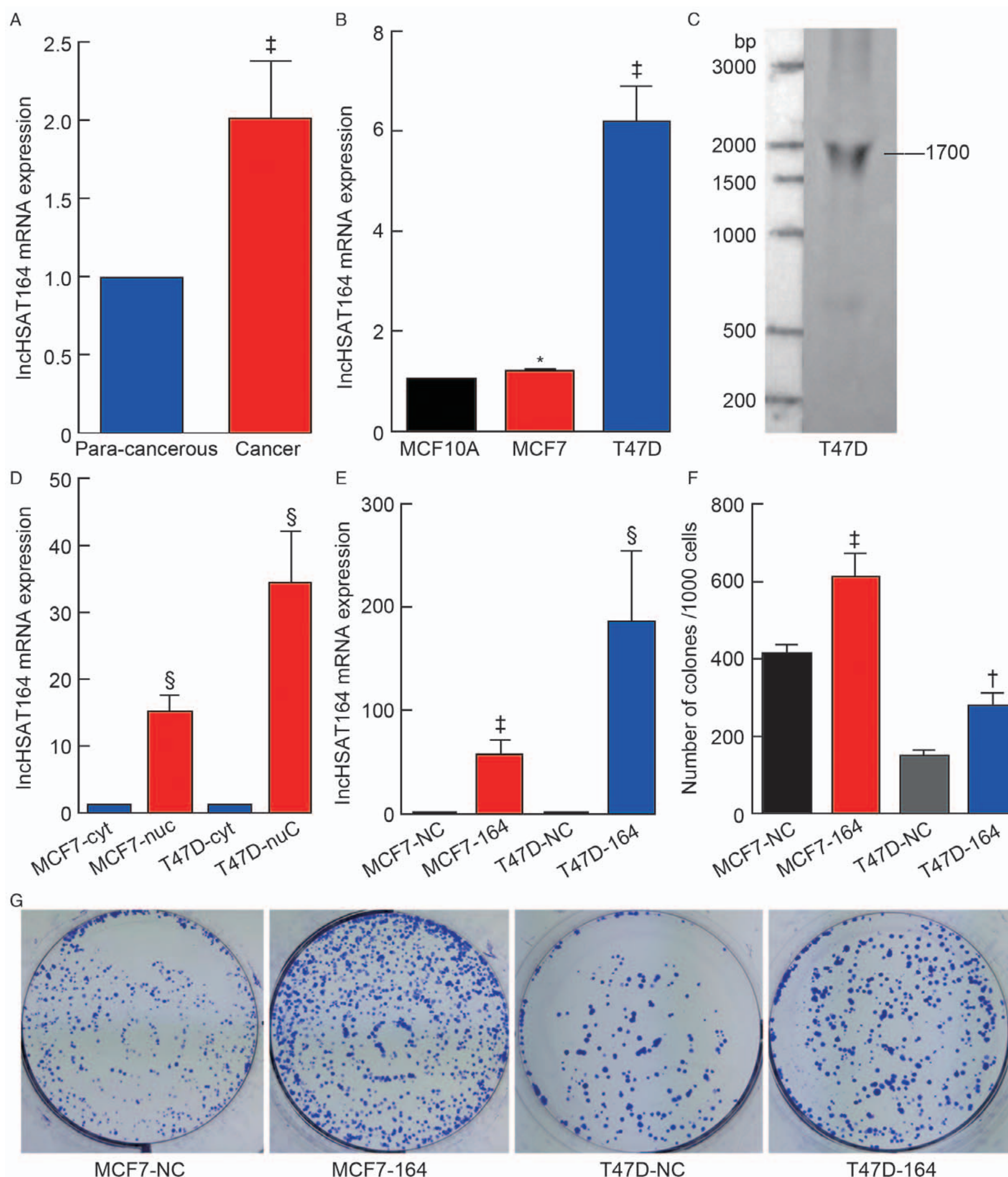

Figure 2: Characterization of lncHSAT164 in breast cancer. (A) lncHSAT164 mRNA expression in four paired human breast cancer tissues and para-cancerous tissues. (B) Relative lncHSAT164 expression levels in breast cancer cells. (C) Northern blot analysis of lncHSAT164 in T47D cells. (D) Subcellular localization of lncHSAT164 in breast cancer cells. Blue indicates cytoplasmic extracts. Red indicates nuclear extracts. (E) lncHSAT164 was overexpressed in MCF7 and T47D cells. (F–G) Overexpressed lncHSAT164 promotes the clonogenic potential of MCF7 and T47D cells. Data are presented as the mean $\pm$ standarad error (SE), and P values were computed by unpaired Student's t test. $^{*}P > 0.05$; $^{\dagger}P < 0.05$; $^{\ddagger}P < 0.01$; $^{\S}P < 0.001$.

found for MCF7 cells [Supplementary Figure 3B–D, http://links.lww.com/CM9/A494]. Subsequently, a pull-down analysis was applied to identify lncHSAT164-associated proteins, but unfortunately, we failed to identify any related proteins (data not shown). Based on the above analysis, it is reasonable to suggest that lncHSAT164 could promote tumor proliferation by regulating cell proliferation and the cell state.

Discussion

To our knowledge, this study was among the first attempts to perform a genome-wide lncRNA analysis based on an lncRNA array. In addition to the target sites in non-coding areas, this lncRNA array also includes thousands of SNPs in the Illumina Human OmniZhongHua array and its genome LD region. Despite the fact that only 1675 SNPs in

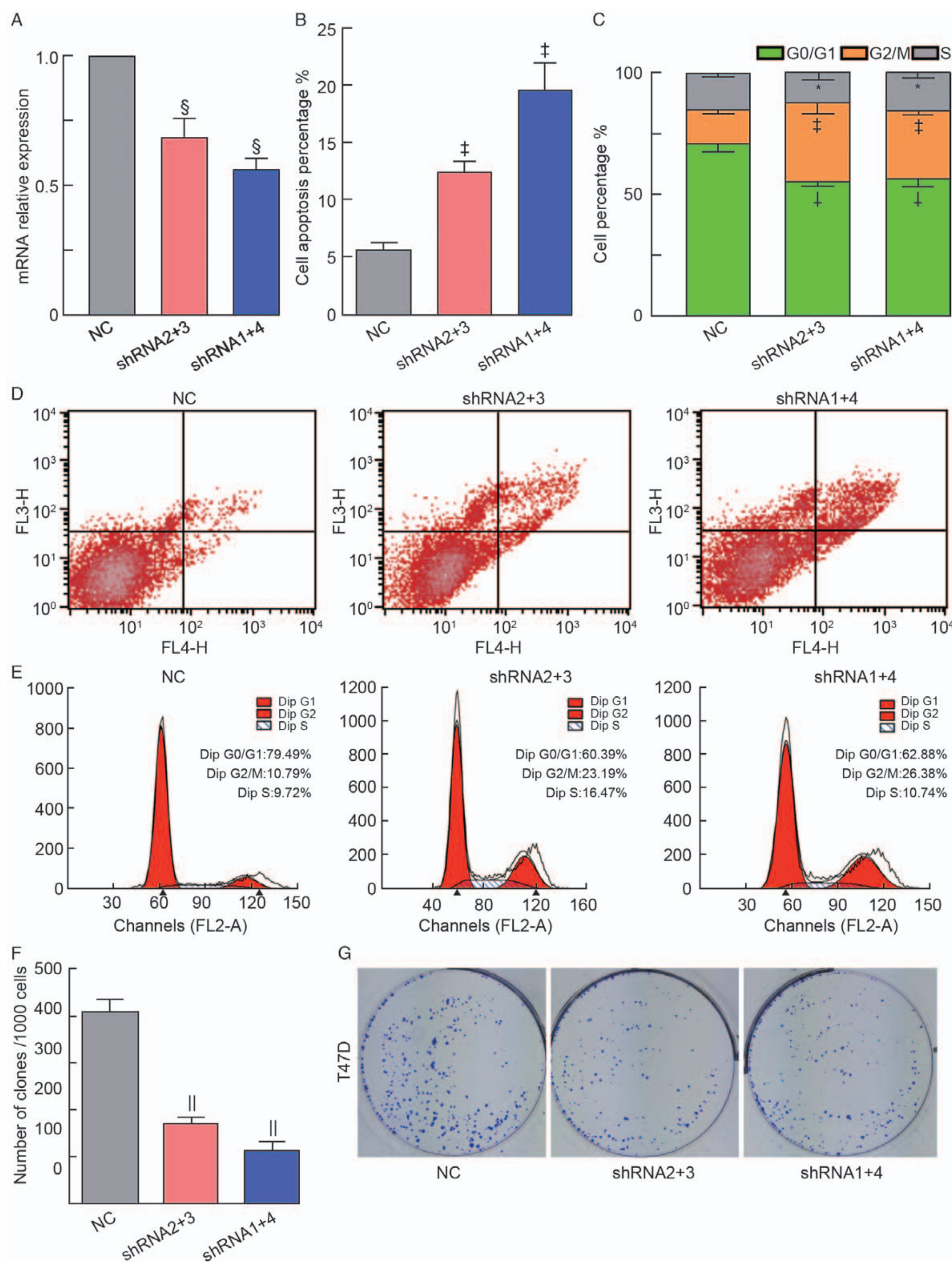

Figure 3: lncHSAT164 knockdown increases cell apoptosis and disturbs the cell cycle in breast cancer. (A) mRNA expression of lncHSAT164 after its knockdown in T47D cells. (B) Apoptosis analysis of the lncHSAT164-deficient T47D cell line. (C) Cell cycle analysis of the lncHSAT164-deficient T47D cell line. (D) and (E) Cell apoptosis and cell cycle analyses were assessed by flow cytometry assays after lncHSAT164 knockdown. (F) and (G) Downregulated lncHSAT164 reduced the clonogenic potential of T47D cells. Data are presented as the means ± standard error (SE), and P values were computed by an unpaired Student's t test. $^{*}P > 0.05$; $^{\dagger}P < 0.05$; $^{\ddagger}P < 0.01$; $^{\S}P < 0.001$; $^{\|}P < 0.0001$. NC: Normal control; shRNA: Short hairpin RNA.

our array had P values $< 10^{-4}$, resulting in lower array recovery than in the Illumina Human OmniZhongHua array, the current study still supports the existence of the breast cancer-associated SNP rs9397435 reported previously,[3,29] which also demonstrates that the data presented herein are reliable. The aforementioned results also suggest that this lncRNA array can be widely used to identify more disease-associated SNPs in future studies.

In this study, we identified a novel breast cancer-associated SNP, rs11066150 on Han Chinese. This SNP is an intron variant located on chr12q24.13. Based on the NONCODE database, rs11066150 was annotated to transcript lncHSAT164. It was first identified by Sharon *et al*[30] in 2013 based on the single-molecule long-read sequencing technology. However, they did not go further and conduct functional studies on lncHAST164. Due to the limitation of samples, we did not study the relationship between rs11066150 variant and lncHSAT164 expression, and further studies are needed to clarify their correlations.

Currently, lncRNAs have been verified to be involved in breast cancer progression by regulating cell proliferation and metastasis.[31] lncHSAT164, as a newly identified lncRNA, is highly expressed in breast cancer and other cancers. Downregulated lncHSAT164 can regulate the cell cycle and promote cell apoptosis in breast cancer cell lines. These findings suggest a potential therapeutic target in breast cancer. As previously reported, pull-down analysis or chromatin immunoprecipitation sequencing (ChiP-seq) analysis can be performed to identify lncRNA binding proteins or miRNAs,[19,32-34] but we did not obtain any positive data associated with lncHSAT164 in the current study. Therefore, the mechanisms by which lncHSAT164 contributes to breast cancer are still unclear, and further studies are required to fully understand how lncHSAT164 predisposes patients to breast cancer and other cancers.

Acknowledgements

The authors would like to thank all the participating patients and healthy controls, and they would also like to thank all the doctors and nurses in the Department of Oncology, No. 2 Hospital, Anhui Medical University.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81320108016) and the Natural Science Foundation of Higher Education of Anhui Province of China (No. KJ2017A181).

Conflicts of interest

None.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424. doi: 10.3322/caac.21492.
2. Li T, Mello-Thoms C, Brennan PC. Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence. Breast Cancer Res Treat 2016;159:395–406. doi: 10.1007/s10549-016-3947-0.
3. Antoniou AC, Kartsonaki C, Sinilnikova OM, Soucy P, McGuffog L, Healey S, *et al*. Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum Mol Genet 2011;20:3304–3321. doi: 10.1093/hmg/ddr226.
4. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet 2008;40:17–22. doi: 10.1038/ng.2007.53.
5. Zhang B, Chen MY, Shen YJ, Zhuo XB, Gao P, Zhou FS, *et al*. A large-scale, exome-wide association study of Han Chinese women identifies three novel loci predisposing to breast cancer. Cancer Res 2018;78:3087–3097. doi: 10.1158/0008-5472.CAN-17-1721.
6. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindström S, *et al*. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49:1767–1778. doi: 10.1038/ng.3785.
7. Kraft P, Haiman CA. GWAS identifies a common breast cancer risk allele among BRCA1 carriers. Nat Genet 2010;42:819–820. doi: 10.1038/ng1010-819.
8. Hoffman J, Fejerman L, Hu D, Huntsman S, Li M, John EM, *et al*. Identification of novel common breast cancer risk variants at the 6q25 locus among Latinas. Breast Cancer Res 2019;21:3. doi: 10.1186/s13058-018-1085-9.
9. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, *et al*. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 2010;42:504–507. doi: 10.1038/ng.586.
10. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, *et al*. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet 2012;44:312–318. doi: 10.1038/ng.1049.
11. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, *et al*. Landscape of transcription in human cells. Nature 2012; 489:101–108. doi: 10.1038/nature11233.
12. Fang SS, Zhang LL, Guo JC, Niu YW, Wu Y, Li H, *et al*. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res 2018;46:D308–D314. doi: 10.1093/nar/gkx1107.
13. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, *et al*. Cell 2010;142:409–419. doi: 10.1016/j.cell.2010.06.040.
14. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell 2013;154:26–46. doi: 10.1016/j.cell.2013.06.020.
15. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, *et al*. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet 2013;45:1392–1398. doi: 10.1038/ng.2771.
16. Kim J, Piao HL, Kim BJ, Yao F, Han Z, Wang Y, *et al*. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nat Genet 2018;50:1705–1715. doi: 10.1038/s41588-018-0252-3.
17. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, *et al*. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun 2016; 7:12791. doi: 10.1038/ncomms12791.
18. Betts JA, Moradi Marjaneh M, Al-Ejeh F, Lim YC, Shi W, Sivakumaran H, *et al*. Long noncoding RNAs CUPID1 and CUPID2 mediate breast cancer risk at 11q13 by modulating the response to DNA damage. Am J Hum Genet 2017;101:255–266. doi: 10.1016/j.ajhg.2017.07.007.
19. Lin A, Li C, Xing Z, Hu Q, Liang K, Han L, *et al*. The LINK-A lncRNA activates normoxic HIF1alpha signalling in triple-negative breast cancer. Nat Cell Biol 2016;18:213–224. doi: 10.1038/ncb3295.
20. Hu WL, Jin L, Xu A, Wang YF, Thorne RF, Zhang XD, *et al*. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. Nat Cell Biol 2018;20:492–502. doi: 10.1038/s41556-018-0066-7.
21. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, *et al*. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007;39:1181–1186. doi: 10.1038/ng1007-1181.
22. Fan Z, Chen X, Liu L, Zhu C, Xu J, Yin X, *et al*. The polymorphism rs13259960 in SLEAR predisposes to systemic lupus erythematosus. Arthritis Rheumatol 2020;72:985–996. doi: 10.1002/art.41200.
23. Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, *et al*. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res 2016;44:D203–D208. doi: 10.1093/nar/gkv1252.
24. He L, Wu WJ, Yang JK, Cheng H, Zuo XB, Lai W, *et al*. Two new susceptibility loci 1q24.2 and 11p11.2 confer risk to severe acne. Nat Commun 2014;5:2870. doi: 10.1038/ncomms3870.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575. doi: 10.1086/519795.
26. 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;38:904–909. doi: 10.1038/ng1847.

27. Wen L, Zhu C, Zhu Z, Yang C, Zheng X, Liu L, *et al*. Exome-wide association study identifies four novel loci for systemic lupus erythematosus in Han Chinese population. Ann Rheum Dis 2018;77:417. doi: 10.1136/annrheumdis-2017-211823.
28. Zuo X, Sun L, Yin X, Gao J, Sheng Y, Xu J, *et al*. Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. Nat Commun 2015;6:6793. doi: 10.1038/ncomms7793.
29. Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, *et al*. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet 2010;6: e1001029. doi: 10.1371/journal.pgen.1001029.
30. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol 2013; 31:1009–1014. doi: 10.1038/nbt.2705.
31. Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, *et al*. A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. Cancer Cell 2015;27:370–381. doi: 10.1016/j.ccell.2015.02.004.
32. Liu X, Xiao Z-D, Han L, Zhang J, Lee S-W, Wang W, *et al*. LncRNA NBR2 engages a metabolic checkpoint by regulating AMPK under energy stress. Nat Cell Biol 2016;18:431–442. doi: 10.1038/ncb3328.
33. Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, *et al*. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell 2014;159:1110–1125. doi: 10.1016/j.cell.2014.10.013.
34. Grelet S, Link LA, Howley B, Obellianne C, Palanisamy V, Gangaraju VK, *et al*. A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression. Nat Cell Biol 2017;19:1105–1115. doi: 10.1038/ncb3595.

How to cite this article: Xu JK, Li GZ, Li Z, Li WJ, Chen RS, Zhang B, Zhang XJ. Genome-wide long non-coding RNA association study on Han Chinese women identifies lncHSAT164 as a novel susceptibility gene for breast cancer. Chin Med J 2021;134:1138–1145. doi: 10.1097/CM9.0000000000001429