의학박사 학위논문

Discovery of novel amplified genes in primary breast cancer with copy number and gene expression analysis of whole exome and transcriptome sequencing data

유방암에서 엑솜 및 전사체 분석을 통한 새로운 치료대상 유전자의 발굴

2016년 2월

서울대학교 대학원

의학과 외과학 전공

임 우 성

Abstract

Discovery of novel amplified genes in primary breast cancer with copy number and gene expression analysis of whole exome and transcriptome sequencing data

Woosung Lim, M.D.
College of Medicine, Surgery
The Graduate School
Seoul National University

Introduction: Copy number alteration (CNA) of genome is common in breast cancer and tend to have more driver role than single point mutations. Traditionally, genome-wide analysis of DNA copy number changes were done by array CGH or SNP array method. Here, we did DNA whole exome sequencing (WES) and RNA-seq using Next Generation Sequencing (NGS) technology to find common genes or chromosomal regions of which DNA copy was highly amplified and at the same time RNA expression was also upregulated.

Materials and Methods: RNA and DNA were extracted from fresh frozen tissues of 93 breast cancer patients. WES and RNA-seq were done using

NGS technology (Illumina HiSeq 2000). As a control, normal DNA from all matched patients were also sequenced. GATK was used to gain mean depth and coverage data for targeted regions. CNAs were calculated with Exome CNA, a statistical method to detect somatic CNAs using depth-of-coverage information from mapped short sequence reads. To estimate expression levels, the relative transcript abundances were measured in FPKM using Cufflinks. FISH was performed for amplification validation of target genes and siRNA transfection for target genes' role in breast cancer cell lines.

After siRNA transfection, stem cell marker assay and mammosphere formation count were used for the target genes' effect on stem cell.

Results: DNA of 1,737 genes were highly amplified (log R>1.0) in two or more samples. The two most commonly amplified chromosomes were chromosome 8 and 17. We applied a cut-off for higher gene expression as relative FPKM >1.5. ERBB2 amplifications and high expression were most common (21.5%) of all genes and it was in agreement with HER-2 IHC and FISH result. Among previously reported amplified genes, FGFR1 (5.4%) and PVT1 (8.6%) in chromosome 8, CCND1, PAK1 (3.2%) and EMSY (4.3%) in chromosome 11, CCNE1 (4.3%) in chromosome 19 were also identified in this study. IGF1R high amplification and expression was found in two samples, and ESR1, MDM2, KIT was found in only one sample each. We

found uncommon but novel and recurrent highly amplified and expressed genes: CLK4 in 5q (3.2%), AHI/MYB in 6q (3.2%), MMP7 (2.2%) and MALAT1 in 11q (1.1%), and NEK8 in 17q (4.3%) We designed FISH probe for this 5 new genes and confirmed the high amplifications in each sample with FISH. Real-time PCR results showed that mRNA expression of 4 genes was significantly down regulated after interfered with siRNA for each of 4 genes((NEK8, CLK4, MMP7 and MALAT1). Proliferation, migration and invasion were inhibited in various breast cancer cell lines after knock-down by siRNA for 4 genes and inhibition of NEK8 down-regulated stem cell properties in MDA-MB-157 cell line.

Conclusion: Our results revealed the amplification and the role in breast cancer of 4 new genes. These genes will be candidates for new therapeutic targets for breast cancer.

Keywords : breast cancer, NGS, exome, copy number alterations, Gene amplification, therapeutic target

Student Number : 2009-31093

List of Tables

List of Figures

Contents

1. Introduction

Copy number variations (CNVs) of DNA are important in genetic variation, influencing greater than single nucleotide polymorphisms (SNPs). CNVs are structurally variant regions in which copy number changes between two or more genomes (1). The DNA sequencing began earnestly in 1970s with development of the Sanger method. Termed the chain-termination method, it involves a reaction where chain-terminator nucleotides are labeled with fluorescent dyes, combined with fragmented DNA, DNA sequencing primers and DNA polymerase. Each nucleotide in the DNA sequence is labeled with a different dye color and a chromatogram is produced, with each color representing a different letter in the DNA code - A, T, C, or G (2). Relatively fast and cost efficient DNA sequencing were allowed by advances in sequencing technology and computer programming. The

Human Genome Project was performed with first-generation sequencing, known as Sanger sequencing. However, sequencing of entire genomes of organisms was difficult, costly and time consuming. Sanger sequencing had considered the gold standard for DNA sequencing but demands faster and cheaper sequencing methods has increased. This demand has driven the development of second-generation sequencing methods, or next-generation sequencing (NGS). NGS platforms perform massively parallel sequencing, during which millions of fragments of DNA from a single sample are sequenced in unison. Massively parallel sequencing technology facilitates high-throughput sequencing, which allows an entire genome to be sequenced in less than one day, therefore NGS technique have been used in analysis of nucleic acid sequencing nowadays.

Somatically acquired copy number alteration (CNA) of genome is

common in breast cancer. The Cancer Genome Atlas Network reported a study of breast tumors at the molecular level, including epigenetics, miRNA, gene expression profiling, copy number changes, sequencing, and protein analysis (3). This study identified variations in the somatic mutations present of 4 breast cancer subtypes (luminal A, luminal B, HER2-enriched, and basal-like), and suggested that putative therapeutic targets for luminal cancers might include highly mutated genes in this subtype, such as PIK3CA or AKT1, FGF receptors, cyclin D1, and CDK4 and 6. For basal-like tumors, they suggested targeting genes influenced by copy number changes, such as PIK3CA, KRAS, BRAF, EGFR, PTEN, or INPP4B. However, the driver events that are selected for during tumorigenesis are difficult to elucidate as they co-occur alongside a much larger landscape of random non-pathogenic passenger alterations (4,5). Large scale analysis by METABRIC (Molecular Taxonomy of Breast Cancer

International Consortium) showed that acquired somatic CNAs were associated with expression in 40% of genes. And they identified putative cancer genes by delineating expression outlier genes driven in *cis* by CNAs (6). In breast cancer, the somatic point mutation of single nucleotide is relatively uncommon compared with other cancers such as, lung, head and neck, or colon cancer (7). So the CNAs might have a more role as a driver in breast cancer.

Traditionally, genome-wide analysis of DNA copy number changes were done by array comparative genomic hybridization (CGH) or SNP array method. The combination of two or more genome-wide microarray-based expression profiling resulted in the identification of genetic profiles for predicting cancer progression and treatment response (8-11).

Analysis of CNAs investigated the use of genetic variations as prognostic markers for cancer patients (12-16). Only 12% of gene

expression variation can be explained by differences in CNAs in relation between CNAs and RNA expression in breast cancer (17). Genes involved in molecular processes may be targeted by RNA expression changes resulting from CNAs. By combining analysis of CNAs and RNA expression, genes involved in cancer processes can be better distinguished. However, a few of studies have used analysis of data combining CNAs and RNA expression in cancer prognosis. Several studies used oligonucleotide CGH arrays, and demonstrated a correlation between CNAs and mRNA expression levels (18-20).

In this study, we did DNA whole exome sequencing (WES) and RNA-seq using Next Generation Sequencing (NGS) technology to find common genes or chromosomal regions of which DNA copy was highly amplified and at the same time RNA expression was also upregulated. We hypothesized that the outlying RNA expression

coincident with high-level DNA amplification could be a putative driver gene in the individual breast cancer and be a candidate of potential therapeutic target in breast cancer.

2. Materials and Methods

2.1 Copy number analysis

RNA and DNA were extracted from fresh frozen tissues of 93 breast cancer patients. WES and RNA-seq were done using NGS technology (Illumina HiSeq 2000). As a control, normal DNA from all matched patients were also sequenced. We used GATK (Genome Analysis Toolkit) Depth of Coverage to convert bam file into coverage files. We got exome.interval_list with a format "chr#:start-end" for individual normal and cancer samples. CNVs were calculated with Exome CNV, a statistical method to detect somatic CNVs using

depth-of-coverage information from mapped short sequence reads. Briefly, Exome CNV were run with the following steps: 1) calculation of log coverage ratio, 2) calling CNV for each exon (l=101), 3) combining exome CNV into larger segments. With the result of these steps, locations in chromosomes and the CNVs were calculated. Genes included in the chromosome loci with CNVs were annotated.

2.2 Gene expression analysis with RNA-seq data

For RNA-Seq, total RNA quality and quantity were verified spectrophotometrically (NanoDrop 1000 spectrometer; Thermo Scientific, Wilmington, DE, USA) and electrophoretically (Bioanalyzer 2100; Agilent Technologies, Palo Alto, CA, USA). To construct Illumina-compatible libraries, a TruSeq RNA library preparation kit (Illumina, San Diego, CA, USA) was used according to the manufacturer's instructions. Sequencing was performed using a

HiSeq™ 2000 platform (Illumina) to obtain 100-bp paired-end reads. To estimate expression levels and to find alternative spliced transcripts, the RNA-Seq reads were mapped to the human genome using TopHat (version 1.3.3), which is capable of reporting split-read alignments across splice junctions and determined using Cufflinks software (version 1.2.1) in default options. The reference genome sequence (hg19, Genome Reference Consortium GRCh37) and annotation data were downloaded from the UCSC website (http://genome.uscs.edu). The transcript counts in isoform level and gene level were calculated, and the relative transcript abundances were measured in FPKM using Cufflinks.

2.3 Fluorescence in situ hybridization (FISH)

FISH was performed on 4um thick tissue sections of formalin-fixed, paraffin-embedded (FFPE) samples. MacProbe CLK4, AHI1-MYB,

IGF1R, NEK8 specific probes (Macrogen, Seoul, Korea) and SureFISH MALAT1, MMP7 specific probes (Agilent, Santa Clara, USA) were used. Briefly, FFPE tissue sections were deparaffinized and rehydrated in 100, 85 and 70% ethanol, followed by incubation in 0.2N HCl for 20min at room temperature. The tissue slides were incubated 8% Sodium thiocyanate for 30min at 80℃, followed by treatment with pepsin (0.05% pepsin in 0.01N HCl) for 30min at 37℃. The hydrolysis of tissues were stopped with 1% formaldehyde in PBS, followed by dehydration in 70, 85 and 100% ethanol. The slides were denatured with MacProbes or SureFISH probes for 5min at 75℃ and hybridized for overnight at 37℃ in a humidity chamber according to manufacturer's protocol, respectively. The slides were washed and countstained with 4',6-diamidino-2-phenylindole (Vector Laboratories Inc.). FISH images were observed using a Leica DMRXA2 (Leica Microsystems, Wetzlar, Germany). Images were

captured by a CoolSNAP cf digital camera (Roper Scientific photometrics, Tucson, USA), and analyzed using a Leica CW4000 (Leica Microsystems, Wetzlar, Germany).

2.4 Cell Culture

MDA-MB-231 and MDA-MB-157 cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). Hs578T, BT20, HCC1937, HCC38, and MDA-MB-468 cell lines were obtained from Korean Cell Line Bank (KCLB, Seoul, Korea). These cells were cultured as described (www.atcc.org). In brief, MDA-MB-231, MDA-MB-436, MDA-MB-157, MDA-MB-468, and Hs578T cell lines were cultured in DMEM (Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Invitrogen-Life Technologies, USA) and 1% penicillin/streptomycin (Gibco). All other cell lines were grown in RPMI 1640 with 10% FBS and 1% penicillin/streptomycin. Cells were

maintained at 37°C in a humidified atmosphere of 95% air and 5% CO_2. Cells were screened periodically for mycoplasma contamination.

2.5 RNA extraction and RT-PCR

Total RNA was extracted from cell lines using Trizol reagent (Invitrogen) as per the manufacturer's protocol. Total RNA was isolated from the frozen tissues disrupted using a pellet pestles cordless motor (sigma-Aldrich) and using Trizol. The concentration was quantified using NanoDrop Specthophotometer (NanoDrop Technologies) and cDNA was synthesized using the PrimeScript 1st strand cDNA synthesis kit. PCR was performed using EmeraldAmp Master Mix kit according to the manufacturer's instructions.

The sequences of specific primers were as follows: MALAT1 5′-GCTCTGTGGTGTGGGATTGA-3′ (sense) and 5′-GTGGCAAAATGGCGGACTTT-3′ (antisense); NEK8, 5′-

CAGTCGGGTGTGACCATCA-3′ (sense) and 5′-AACTCGCTGAGCTTCCTGTC-3′ (antisense); MMP7, 5′-GTCTCTGGACGGCAGCTATG -3′ (sense) and 5′-TAGTCCTGAGCCTGTTCCCA-3′ (antisense); CLK4, 5′-CTCCCGAGGTCATTTTGGCT-3′ (sense) and 5′-TTCCATCATTGCCAGGTGCT-3′ (antisense) and CK18, 5′-CAACTACATGGTTTACATGTTC-3′ (sense) and 5′-GCCAGTGGACTCCACGAC-3′ (antisense) as a control. The PCR products were separated by 2% agarose gel and visualized by Ecodye (Biofact, Korea).

2.6 siRNA transfection

ON-TARGET Plus Smart pool siRNA targeting MALAT1, NEK8, MMP7, CLK4 and non-targeting siRNA were purchased from Thermo Scientific (USA). RNA transfections were done in 6-well plate

formats. For 6-well plate, the day before transfection, cells were plated at a density of 2 ~5 x 10^5 cells per well in antibiotic-free media. The next day, Lipofectamine RNAiMax (Invitrogen) was used to transfect cell, as per manufacture's protocol. Briefly, 3 μL Lipofectamine RNAiMax was combined with 10 ~ 20 nM siRNA in a volume of 250 μL Opti-mem media (Gibco) and incubated for 20 min; the complexes of Lipofectamine RNAiMax and siRNAs were then added directly to each well and the cells were incubated until nearly confluent, 48 ~ 72 h later depending on growth conditions.

2.7 Cell viability assay

siRNA transfected cells were plated onto 96-well assay plates at different cell densities, depending on cell type, in 100 μL media per well. Cells were incubated overnight and assayed the following day using the CellTiter-Glo® Luminescent Cell Viability Assay Kit

(Promega, USA). Cells were allowed to equilibrate to room temperature at which time 100 µl of a 1:1 mixture of Cell Titer-Glo reagent. Cells were placed on rocking shaker for 2 min and incubated for an additional 10 min on the bench top. Luminescent measurements were done on a PerkinElmer's VICTOR light luminescence counter for microplate (1420-061, USA). All values are represented graphically as mean ± SD for three independent samples.

2.8 Cell migration, invasion and stem cell properties assay

Migration of desired siRNA transfected cells was assayed using chamber with 8 µm pore filters (6.5 mm in diameter, Corning). Cells (1 ~3 × 10^5) in 100 µl of serum-free medium were placed in the upper chamber, whereas the lower chamber was loaded with 700 µL of medium containing 5% FBS. Cells were incubated for 24 h at 37 °C, and then non-migrating cells were removed with cotton

swabs. Cells that migrated to the bottom of the membrane were then fixed with 4% paraformaldehyde and stained with 0.5% crystal violet for 20 min and washed twice with PBS. Then stained cells were visualized under a microscope and counted. Experiments were performed in triplicated. For invasion assay, cells with a siRNA transfection in serum-free medium were added to the upper chamber precoated with 2 mg/ml Matrigel. Then 700 μL of medium containing 1 ~ 5% FBS was added to the lower chamber. Cells were incubated for 24 h at 37 °C, and then non-invading cells were removed with cotton swabs. Invasive cells were fixed, stained and counted. Experiments were performed in duplicate.

For stem cell properties assay, cells were treated with the indicated compounds for 7 days, after which the numbers of mammospheres were counted at 100X and 200X magnification.

3. Results

3.1 Gene expression analysis with RNA-seq data and validation with FISH

DNA of 1,737 genes were highly amplified (log R>1.0) in two or more samples. The two most commonly amplified chromosomes were chromosome 8 and 17 (Figure 1). We applied a cut-off for higher gene expression as relative FPKM (FPKM of the gene in index sample/mean FPKM of the gene in all samples) >1.5. All the lists of genes in each sample according to chromosomes are shown in supplementary Table 1. Patients' age, molecular subtype, and recurrence sites are also shown.

ERBB2 amplifications and high expression were most common (21.5%) in chromosome 17q of all genes and it was in agreement with HER-2 IHC and FISH result (data not shown). Among previously

reported amplified genes, FGFR1 (5.4%) and PVT1 (8.6%) in chromosome 8, CCND1, PAK1 (3.2%) and EMSY (C11orf30) (4.3%) in chromosome 11, CCNE1 (4.3%) in chromosome 19 were identified. Figure 2A shows the correlation plots between copy number (log R) and expression (FPKM) of each gene. IGF1R high amplification in chromosome 15q and expression was found in two samples, and ESR1 in chromosome 6q, MDM2 in chromosome 12q, KIT in chromosome 4q high amplification and expression was found in only one sample each.

We found uncommon but novel and recurrent highly amplified and expressed genes: CLK4 in chromosome 5q (3.2%), AHI/MYB in chromosome 6q (3.2%), MMP7 (2.2%) and MALAT1 in chromosome 11q (1.1%), and NEK8 in chromosome 17q (4.3%). Figure 2B shows the correlation plots between copy number (log R) and expression (FPKM) of these novel genes. It shows the good correlation between

outlier copy number gain and expression of the genes.

We designed FISH probe for this 4 novel genes (CLK4, MMP7, MALAT1, NEK8) and confirmed the high amplifications in each sample with FISH (Figure 3).

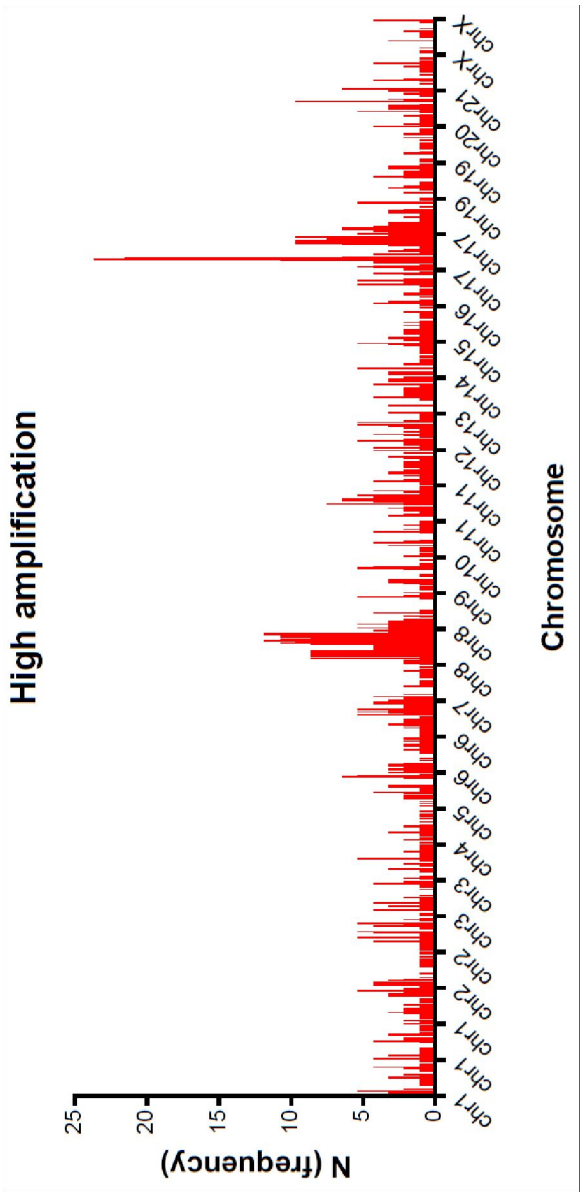

Figure 1. Distribution of gene amplifications according to chromosomes

DNA of 1,737 genes were highly amplified in two or more samples. The two most commonly amplified chromosomes were 8 and 17

Table 1. Genes with copy number gain in tumor of patients with metastasis during follow-up

chr_name	gene	AVG	Amplification	count		count (>log2 3)	non mets	
chr5	CLK4	0.3002	6.451613	5	17.2%	0	1	1.6%
chr8	TNFRSF11B	0.4261	10.75269	5	17.2%	4	5	7.8%
chr8	COLEC10	0.4261	10.75269	5	17.2%	4	5	7.8%
chr8	SAMD12	0.4261	10.75269	5	17.2%	4	5	7.8%
chr8	POU5F1B	0.3593	10.75269	5	17.2%	3	5	7.8%
chr8	MYC	0.3912	11.82796	5	17.2%	3	6	9.4%
chr8	MIR1204	0.39005	10.75269	5	17.2%	3	5	7.8%
chr8	CASC8	0.3912	11.82796	5	17.2%	3	6	9.4%
chr8	PVT1	0.3912	11.82796	5	17.2%	4	6	9.4%
chr9	NFIB	0.2582	5.376344	3	10.3%	1	2	3.1%
chr11	MALAT1	0.338574	7.526882	3	10.3%	3	4	6.3%
chr11	LRRC32	-0.2455	4.301075	3	10.3%	3	1	1.6%
chr11	B3GNT6	0.1807	4.301075	3	10.3%	3	1	1.6%
chr11	TSKU	-0.289	4.301075	3	10.3%	3	1	1.6%
chr11	MYO7A	-0.3847	4.301075	3	10.3%	3	1	1.6%
chr11	AQP11	0.2139	4.301075	3	10.3%	3	1	1.6%
chr11	AAMDC	0.2121	4.301075	3	10.3%	3	1	1.6%
chr11	GUCY2EP	-0.175	4.301075	3	10.3%	3	1	1.6%
chr11	ACER3	-0.1742	4.301075	3	10.3%	3	1	1.6%
chr11	CAPN5	-0.3847	4.301075	3	10.3%	3	1	1.6%
chr11	GDPD4	0.1731	5.376344	3	10.3%	4	2	3.1%
chr11	PAK1	0.2156	5.376344	3	10.3%	4	2	3.1%
chr11	CLNS1A	0.2156	4.301075	3	10.3%	3	1	1.6%

chr11	RSF1	0.2156	4.301075	3	10.3%	3	1	1.6%
chr11	INTS4	0.2156	4.301075	3	10.3%	3	1	1.6%
chr17	ARL5C	-0.2769	10.75269	4	13.8%	9	6	9.4%
chr17	RPL19	-0.2769	10.75269	4	13.8%	9	6	9.4%
chr17	LRRC37A11P	-0.2598	6.451613	3	10.3%	6	3	4.7%
chr17	CACNB1	-0.2769	10.75269	4	13.8%	9	6	9.4%
chr17	LOC100131347	-0.2663	7.526882	3	10.3%	7	4	6.3%
chr17	PTRH2	0.332	9.677419	3	10.3%	6	6	9.4%
chr17	CLTC	0.3562	9.677419	3	10.3%	5	6	9.4%
chr20	PCMTD2	0.4933	9.677419	5	17.2%	2	4	6.3%

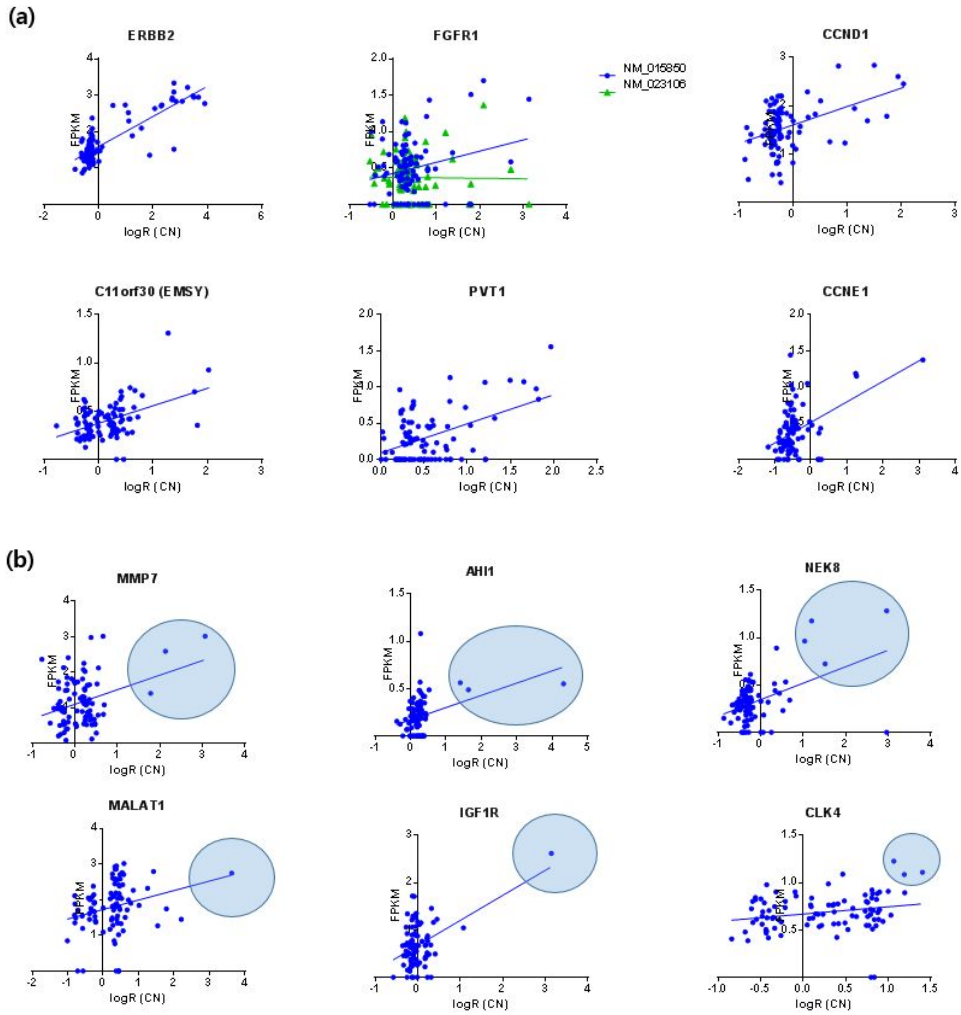

Figure 2. Correlation between expression and copy number variation

(A) for known 6 genes, (B) for novel 6 genes

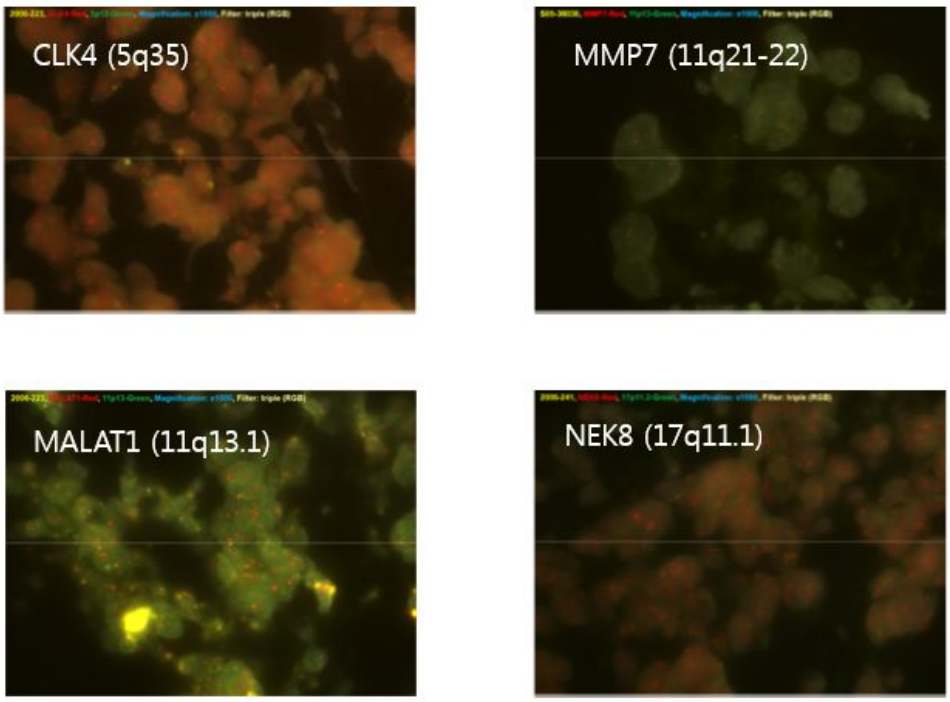

Figure 3. FISH validation for the amplification of 4 genes

3.2 Expression of NEK8, CLK4, MMP7 and MALAT-1 in various breast cancer cell lines

Real time PCR was performed to investigate the expression of NEK8, CLK4, MMP7 and MALAT-1 in various breast cancer cell lines (ER positive : T47D, ZR-75-1 / HER-2 positive : BT-474, SK-BR-3, MDA-MB-453, HCC1957, Jimt-1 / Basal A : HCC70, HCC1937, BT20, MDA-MB-468 / Basal B : MDA-MB-463, MDA-MB-157, HCC38, Hs578T, MDA-MB-231, BT549).

Compared with normal breast cell (MCF10A), NEK8 showed higher expression of mRNA in most cancer cell lines. MALAT1 was also highly expressed in most cancer cell lines except some of basal B cell lines. CLK4 was highly expressed in basal A and basal B cancer cell lines, but weakly expressed in ER+ and HER2+ cell lines. MMP7(especially in basal A) mRNA was highly expressed in some of triple negative breast cancer (Basal A and Basal B) cell lines (Figure 4).

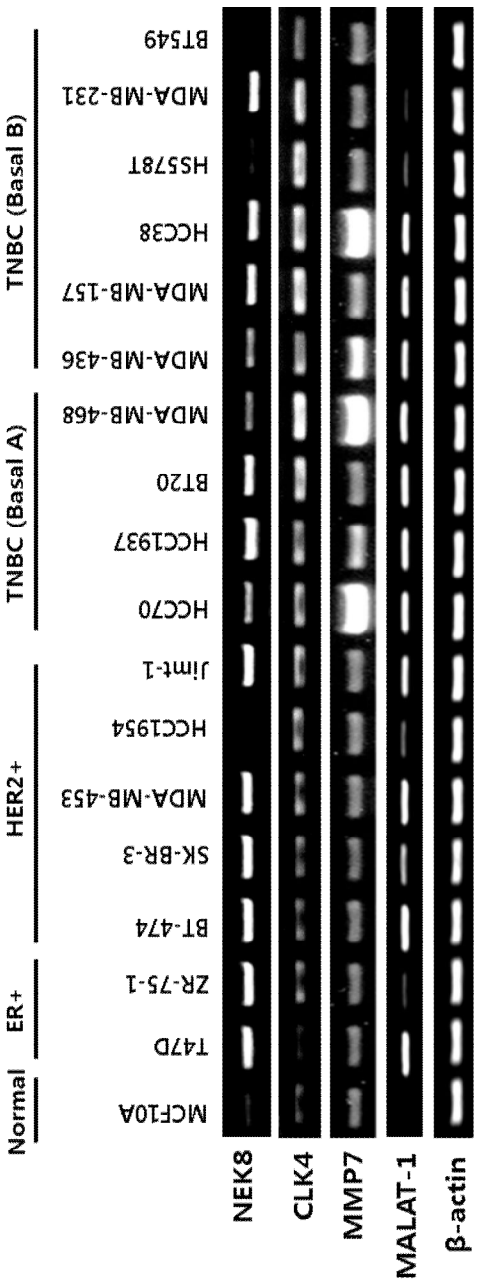

Figure 4. mRNA expression of NEK8, CLK4, MMP7 and MALAT-1 in various cancer cell lines

3.3 Functional study of NEK8, CLK4, MALAT-1 and MMP7 in various breast cancer cell lines

To investigate the role of target genes in breast cancer, we compared changes in proliferation, migration and invasion of various cell lines after treating with siRNA for each of 4 genes (NEK8, CLK4, MALAT1 and MMP7).

3.3.1 Effect of NEK8 knockdown on MDA-MB-157 cell proliferation, migration, invasion and stem cell properties

After knockdown of NEK8 with siRNA, inhibition of mRNA expression was confirmed by RT-PCR (Figure 5A). Relative growth rate was calculated after the MDA-MB-157 cells with siRNA were incubated during 72 hours. Compared with control siRNA, MDA-MB-157 cell proliferation was significantly inhibited by siRNA at 72 hours (Figure 5B) and also colony number and optical density (OD) value

were significantly inhibited ($p<0.05$) (Figure 5C).

Compared with control, MDA-MB-157 cells with NEK8 siRNA showed decreased migration and invasion (Figure 6Aand 6B) on H & E-stain and on electromicroscopic pictures (Figure 6C), and expression of N-cadherin, vimentin, ß-catenine decreased and expression of E-cadherin, p27 increased (Figure 6D). These results suggested that NEK8 may play an important role in proliferation, migration and invasion of cancer cells.

To investigate the effect of NEK8 inhibition on stem cell properties in MDA-MB-157 cells, sphere formation was observed after transfection with siRNA. Figure 7 showed that the number of spheres of stem cell was significantly inhibited. We also investigated the change of breast cancer stem cell markers by NEK8 knockdown. Real-time PCR experiment result demonstrated that expression of stemness markers Sox2, Oct4b and CD44 was downregulated by

inhibition of NEK8 (Figure 8).

3.3.2 Effect of NEK8 knockdown on MDA-MB-231 cell proliferation, migration and invasion

Downregulating with NEK8 siRNA, relative growth rate of MDA-MB-231 cells were not changed till 72 hours (Figure 9A).

The role of NEK8 knockdown by siRNA in MDA-MB-231 cell migration was investigated in transwell migration assay. It was found that NEK8 knockdown inhibited cancer cell migration and invasion, which suggested that NEK8 may play an important role in cancer cell migration and invasion (Figure 9B) and electromicroscopic findings also showed decreased capability of invasion and migration (Figure 9C).

After knockdown of NEK8, expression of Vimentin and N-cadherin was down-regulated (Figure 9D).

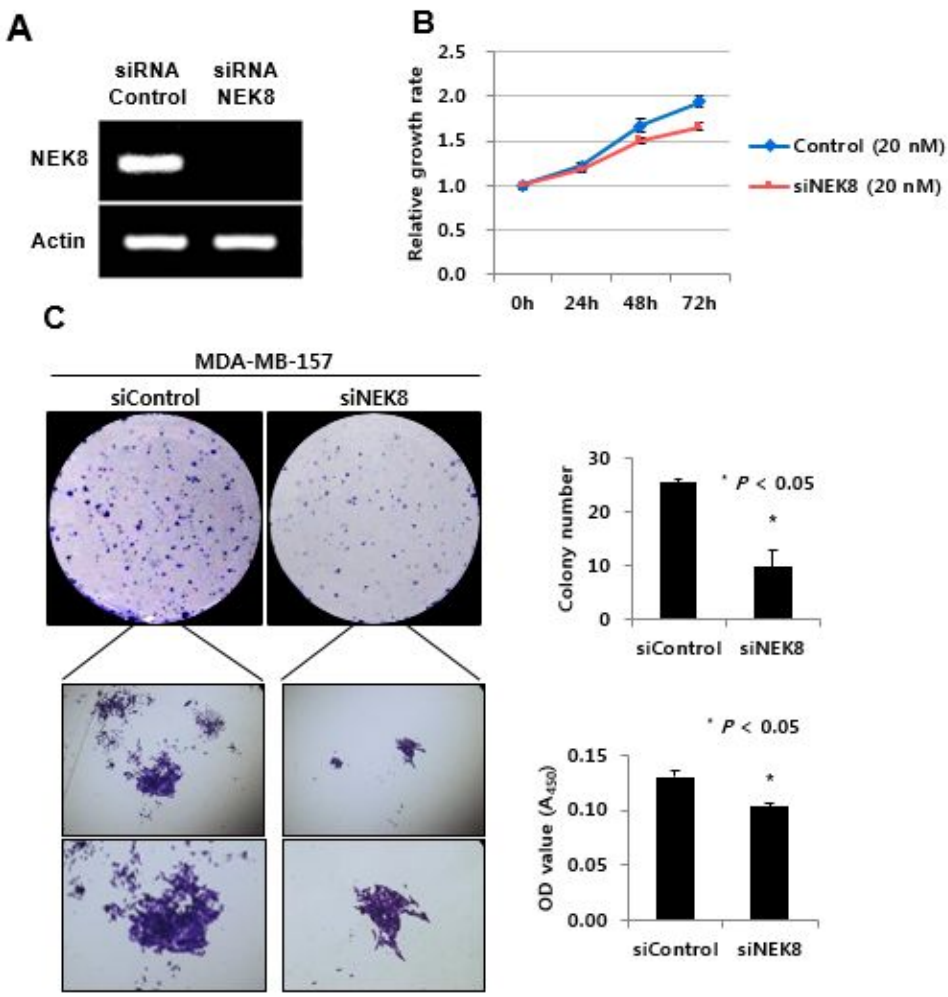

Figure 5. The effect of NEK8 knockdown on MDA-MB-157 cell Proliferation

(A) NEK8 knockdown by siRNA (B) growth rate was significantly inhibited by siRNA (C) Number of colony number was decreased. (D) Proliferation was inhibited by NEK8 siRNA (p,0.05)

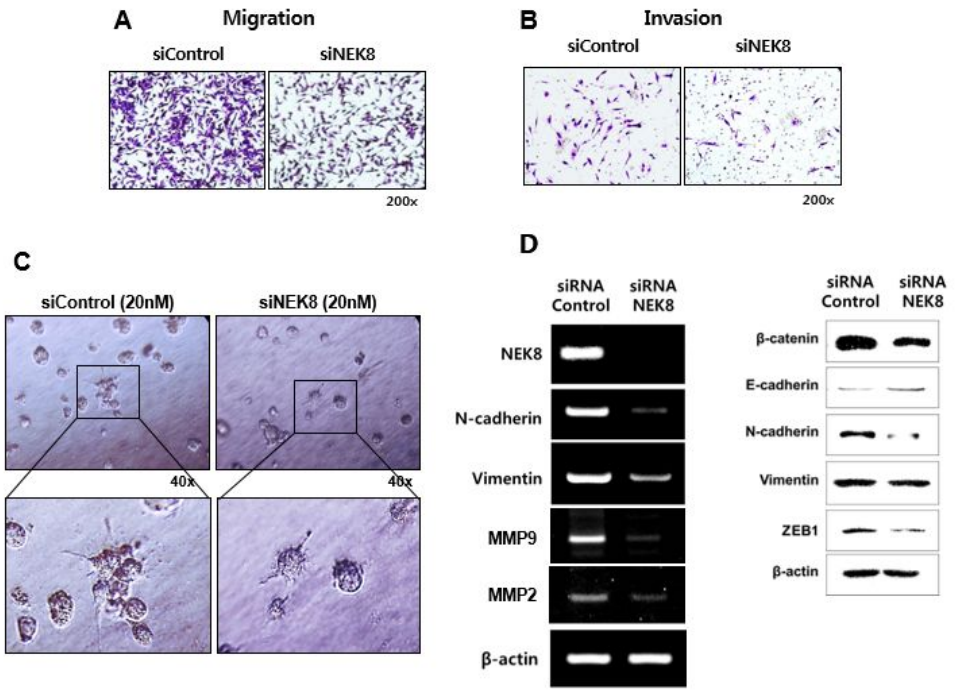

Figure 6. The effect of NEK8 knockdown on migration and invasion of MDA-MB-157 cells

(A,B) Inhibition of migration and invasion was observed on H&E stain by NEK8 siRNA (C) on electro-microscopic findings (D) Expression of genes associated with cancer cell migration and invasion

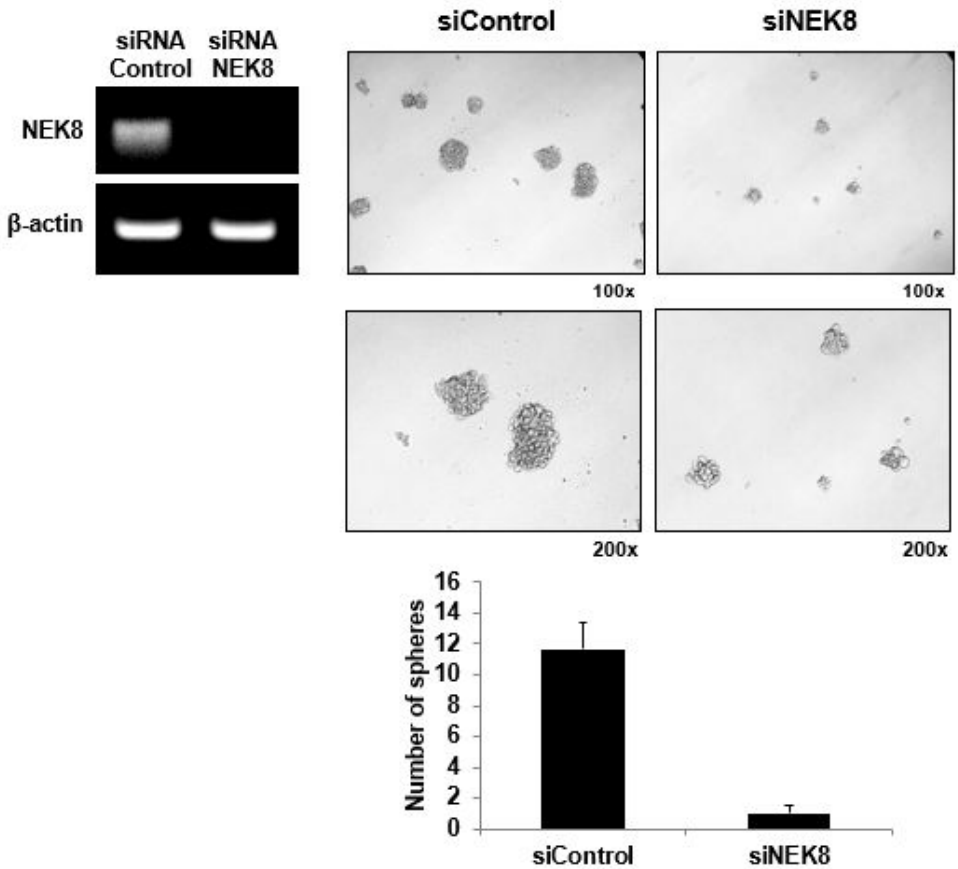

Figure 7. Inhibition of NEK8 suppresses stem cell properties in MDA-MB-157 cells

Number of mammosphere formation was significantly inhibited by NEK8 siRNA

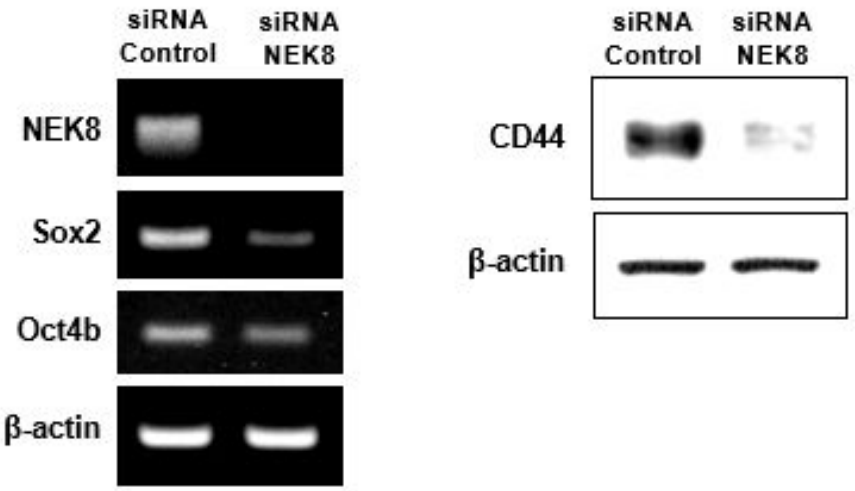

Figure 8. Inhibition of NEK8 down-regulates stem cell properties in MDA-MB-157 cells

Expression of breast cancer stem cell markers (Sox2, Oct4b, CD44) was decreased by NEK8 siRNA

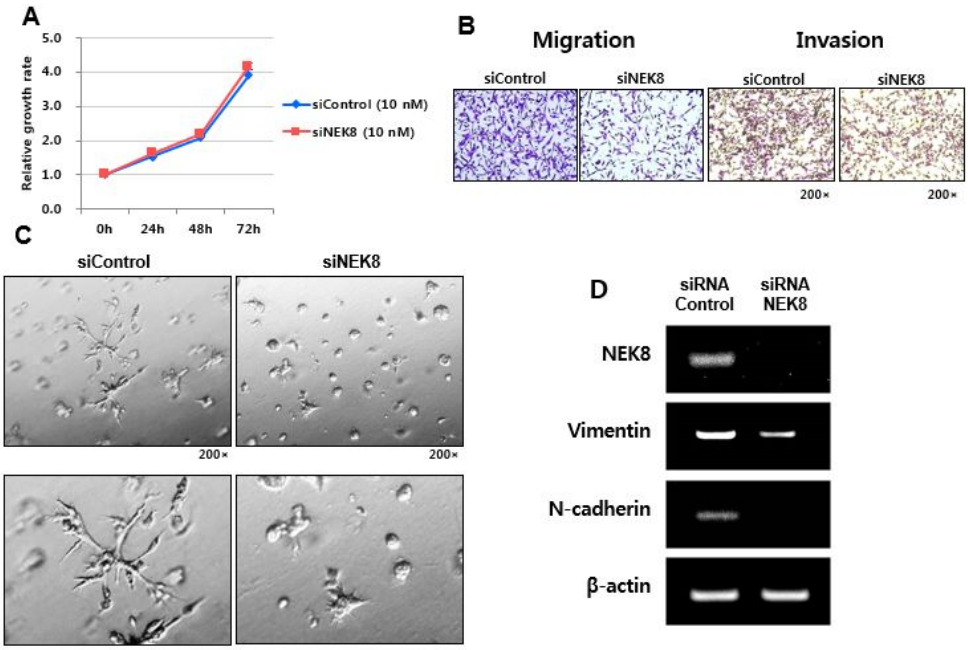

Figure 9. The effect of NEK8 knockdown on MDA-MB-231 cell proliferation, migration and invasion

After NEK8 downregulated, (A) proliferation was not changed (B) inhibition of migration and invasion was observed on H&E stain (C) on electromicroscopic findings (D) Expression of Vimentin and N-cadherin was decreased

3.3.3 Effect of NEK8 knockdown on HCC38 cell proliferation, migration and invasion

After knockdown of NEK8 with siRNA, inhibition of mRNA expression was confirmed by RT-PCR (Figure 10A). Compared with control siRNA, the effect of downregulated NEK8 on proliferation of HCC38 cells was determined by MTT assay. It was found that proliferation of HCC38 cells was inhibited by NEK8 knockdown at 48 hours (Figure 10B). The effect of NEK8 downregulated by siRNA in HCC38 cell migration was investigated in transwell migration assay. It was found that downregulated NEK8 inhibited the migration of cancer cell, which suggested that NEK8 may play an important role in migration of HCC38 cells. The result of invasion assay demonstrated that downregulated HCC38 significantly inhibited invasion of HCC38 cells (Figure 10C

3.3.4 Effect of CLK4 knockdown on MDA-MB-231 cell proliferation, migration and invasion

In CLK4, when siRNA was transfected to MDA-MB-231 cell, mRNA of CLK4 decreased and the effect of downregulated CLK4 on proliferation of MDA-MB-231 cells was determined by MTT assay. It was found proliferation of MDA-MB-231 cells was not changed by CLK4 which was downregulated by siRNA compared with control The effect of CLK4 downregulated by siRNA in MDA-MB-231 cell migration was investigated in transwell migration assay. It was found that downregulated CLK4 inhibited the migration of cancer cell, which suggested that CLK4 may play an important role in migration of breast cancer cells. The result of invasion assay demonstrated that downregulated CLK4 significantly inhibited invasion of MDA-MB-231 cells, suggesting that CLK4 may play a role in invasion of cancer cells (Figure 11).

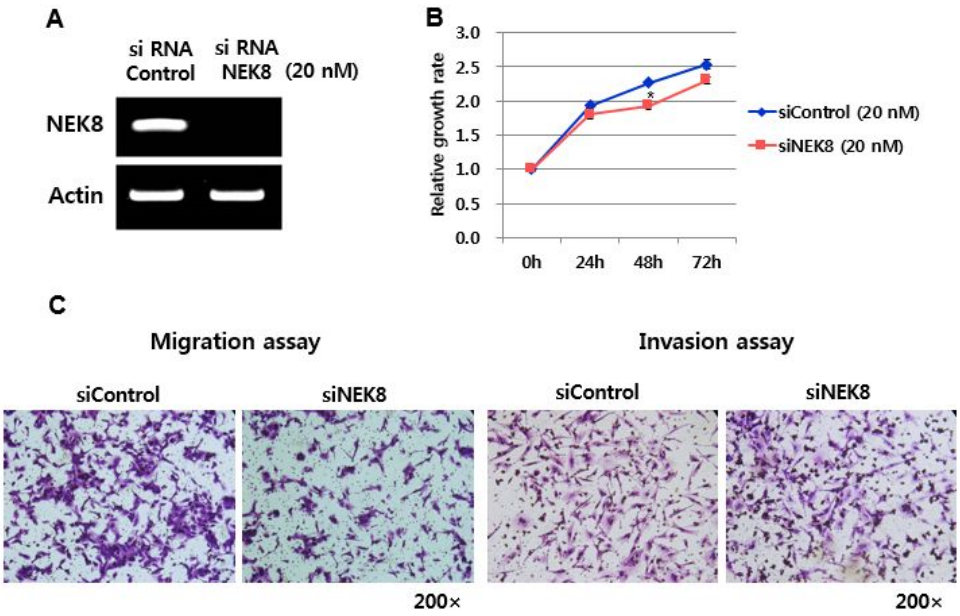

Figure 10. The effect of NEK8 knockdown on HCC38 cell proliferation, migration and invasion

(A) NEK8 was downregulated by siRNA (B) Growth rate was significantly inhibited by NEK8 siRNA at 48hours (p,0.05) (C) Inhibition of migration and invasion was observed on H&E stain

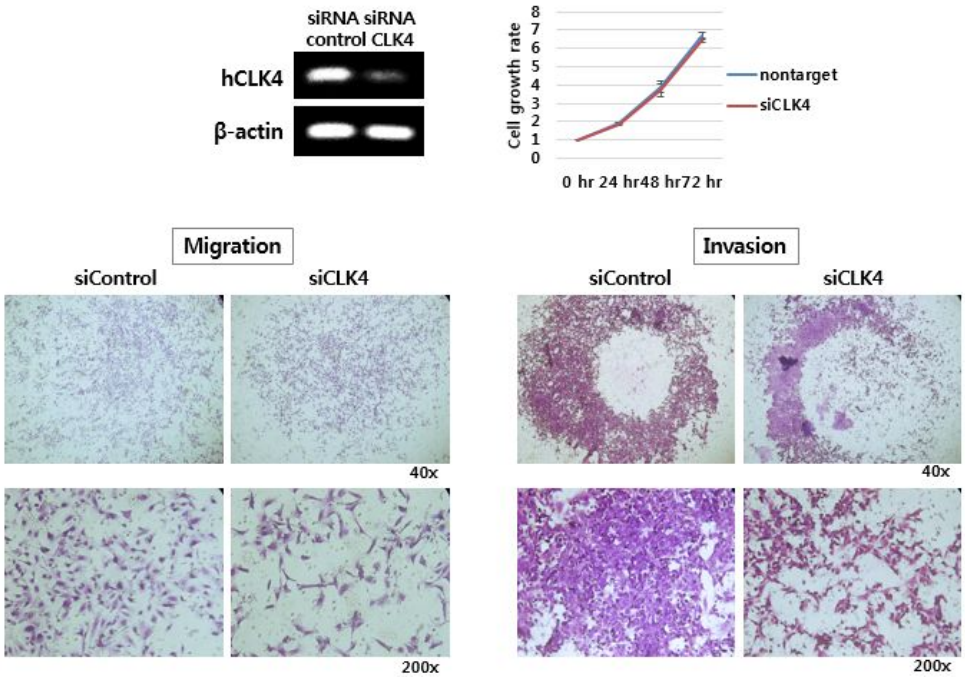

Figure 11. The effect of CLK4 knockdown on MDA-MB-231 cell proliferation, migration and invasion

(A) CLK4 was downregulated by siRNA (B) Growth rate was not changed by CLK4 siRNA (C) Inhibition of migration and invasion was observed on H&E stain

3.3.5 Effect of CLK4 knockdown on Hs578T cell proliferation, migration and invasion

After inhibition of CLK4, cell proliferation of Hs578T was not changed till 72 hours. The role of CLK4 downregulated by siRNA in Hs578T cell migration was investigated in transwell migration assay. It was found that downregulated CLK4 inhibited the migration of cancer cell, which suggested that CLK4 may play an important role in migration of Hs578T cells. The result of invasion assay demonstrated that downregulated CLK4 significantly inhibited invasion of Hs578T cells, suggesting that CLK4 may play a role in invasion of cancer cells (Figure 12).

3.3.6 Effect of CLK4 knockdown on HCC38 cell proliferation, migration and invasion

There was no change in proliferation of HCC38 cells after

transfection with CLK4 siRNA. The role of CLK4 downregulated by siRNA in HCC38 cell migration was investigated in transwell migration assay. It was found that downregulated CLK4 facilitated the migration of cancer cell, which suggested that CLK4 may play an inhibiting role in migration of breast cancer cells. The result of invasion assay demonstrated that downregulated CLK4 also significantly facilitated invasion of HCC38 cells. These results contrast with those in other cell lines (MDA-MB-231 and Hs578T) (Figure 13).

3.3.7 Effect of MALAT1 knockdown on MDA-MB-157 cell proliferation, migration and invasion

PCR result showed that MALAT1 expression was significantly down-regulated after transfection with siRNA (Figure 14A). Compared with control siRNA, MDA-MB-157 cell proliferation was

significantly inhibited by siRNA at 72 hours (Figure 14B) and also colony number and OD value were significantly inhibited ($p<0.05$) (Figure 14C). The result of flow cytometry, after knockdown of MALAT1, showed increased G0/G1, G2/M phase arrest and decreased S phase proportion (Figure 14D). Compared with control, MDA-MB-157 cells with MALAT1 siRNA showed decreased migration and invasion (Figure 15A and 15B). Cell to Cell adhesion was decreased (Figure 15C) and expression of N-cadherin, vimentin, ß-catenine decreased but expression of E-cadherin, p27 increased (Figure 15D).

3.3.8 Effect of MALAT1 knockdown on MDA-MB-436 cell proliferation, migration and invasion

After knockdown of MALAT1 with siRNA, inhibition of mRNA expression was confirmed by RT-PCR (Figure 16A). Relative growth

rate was significantly inhibited at 48 hours ($p<0.05$) and 72 hours ($p,0.001$) (Figure 16B), and the number of colonies was significantly decreased in MALAT1-deleted MDA-MB-468 cells compared with control cells (Figure 16C). The effect of MALAT1 downregulated by siRNA in MDA-MB-468 cell migration was investigated in transwell migration assay. It was found that downregulated MALAT1 inhibited the migration of cancer cell, which suggested that MALAT1 may play an important role in migration of breast cancer cells. The result of invasion assay showed that downregulated MALAT1 significantly inhibited invasion of MDA-MB-468 cells, suggesting that MALAT1 may play a role in invasion of MDA-MB-468 cells (Figure 17A and 17B). To know about the role of MALAT1 in downstream signaling pathway, whether or not the expression of Vimentin and N-cadherin were affected by MALAT1 and expression of Vimentin and N-cadherin decreased by downregulated MALAT1 (Figure 17C).

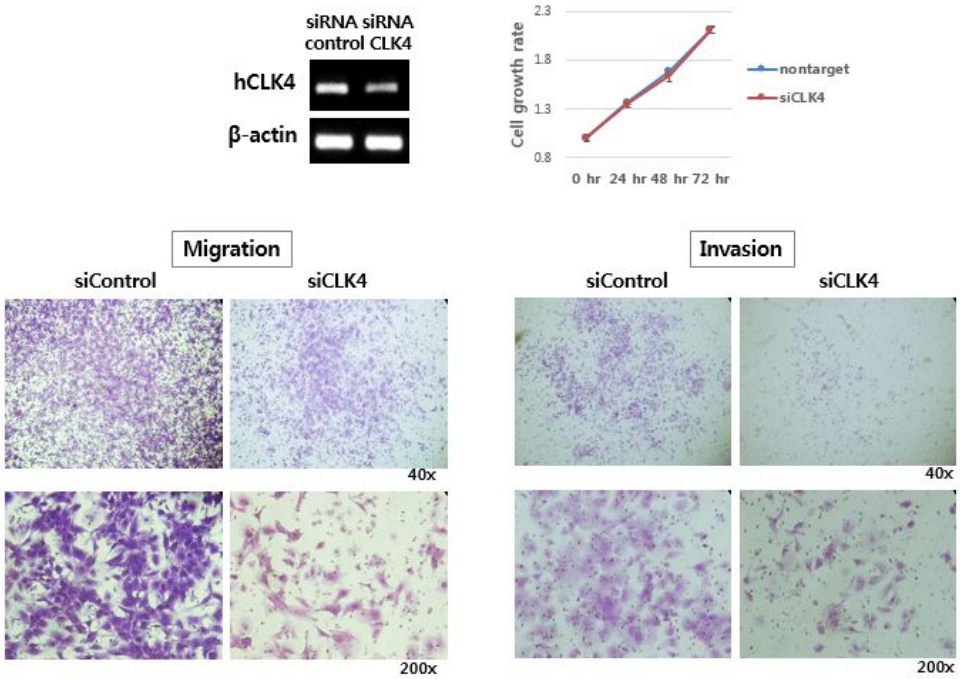

Figure 12. The effect of CLK4 knockdown on Hs578T cell proliferation, migration and invasion

(A) CLK4 was downregulated by siRNA (B) Growth rate was not changed by CLK4 siRNA (C) Inhibition of migration and invasion was observed on H&E stain

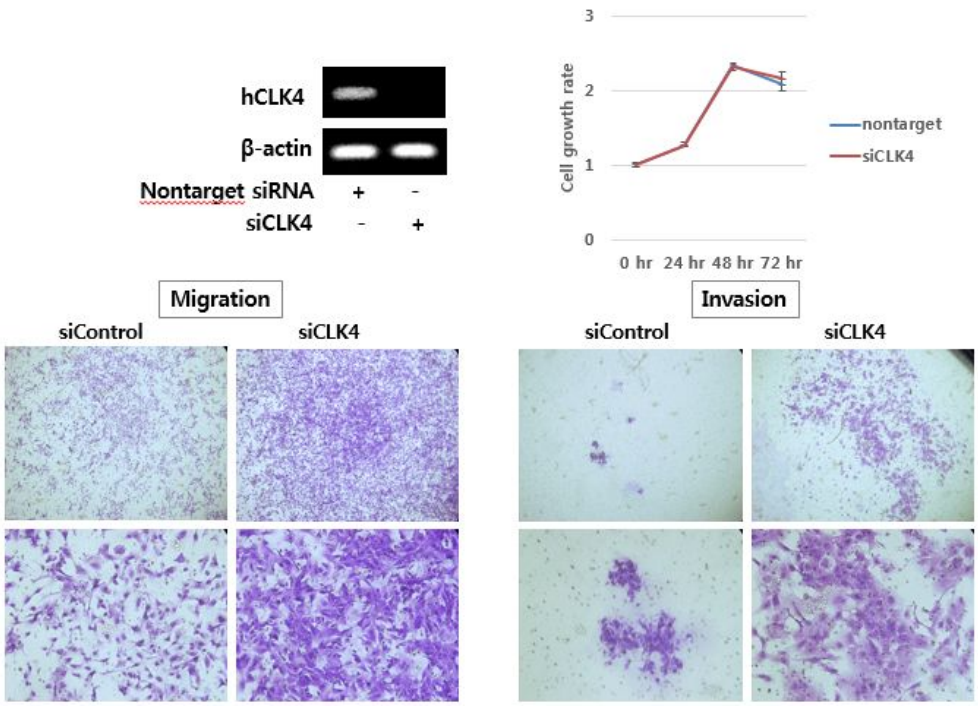

Figure 13. The effect of CLK4 knockdown on HCC38 cell proliferation, migration and invasion

(A) CLK4 was downregulated by siRNA (B) Growth rate was not changed by CLK4 siRNA (C) Facilitation of migration and invasion was observed on H&E stain

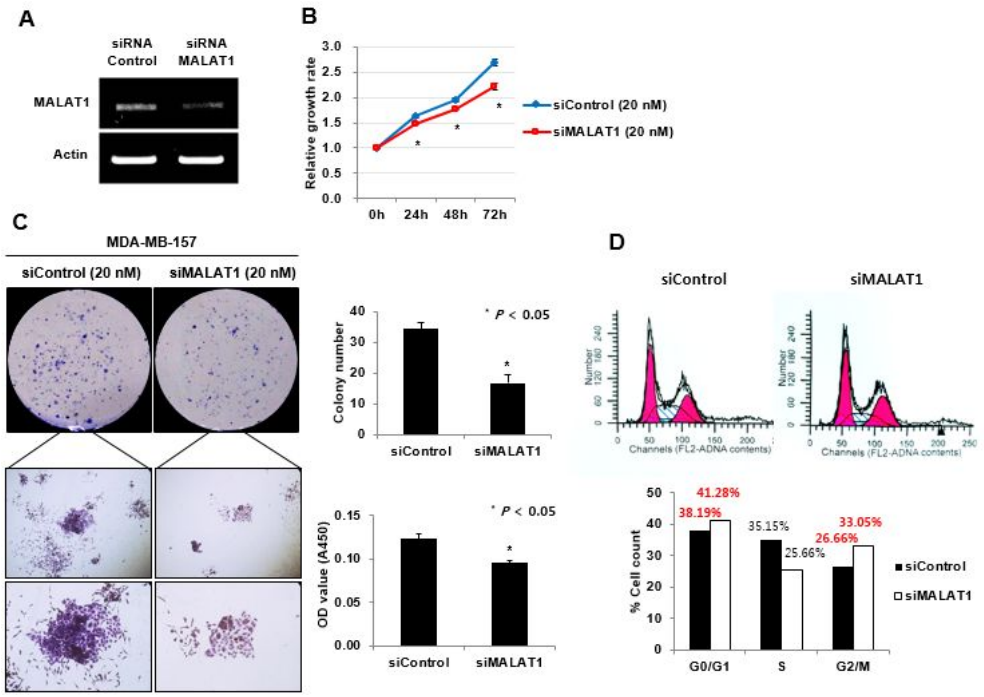

Figure 14. The effect of MALAT1 knockdown on MDA-MB-157 cell Proliferation

(A) MALAT1 was downregulated by siRNA (B) Growth rate was inhibited by siRNA (C) Number of colony formation was decreased (D) Cell cycle arrest in G0/G1 and G2/M phases was proved by flowcytometry

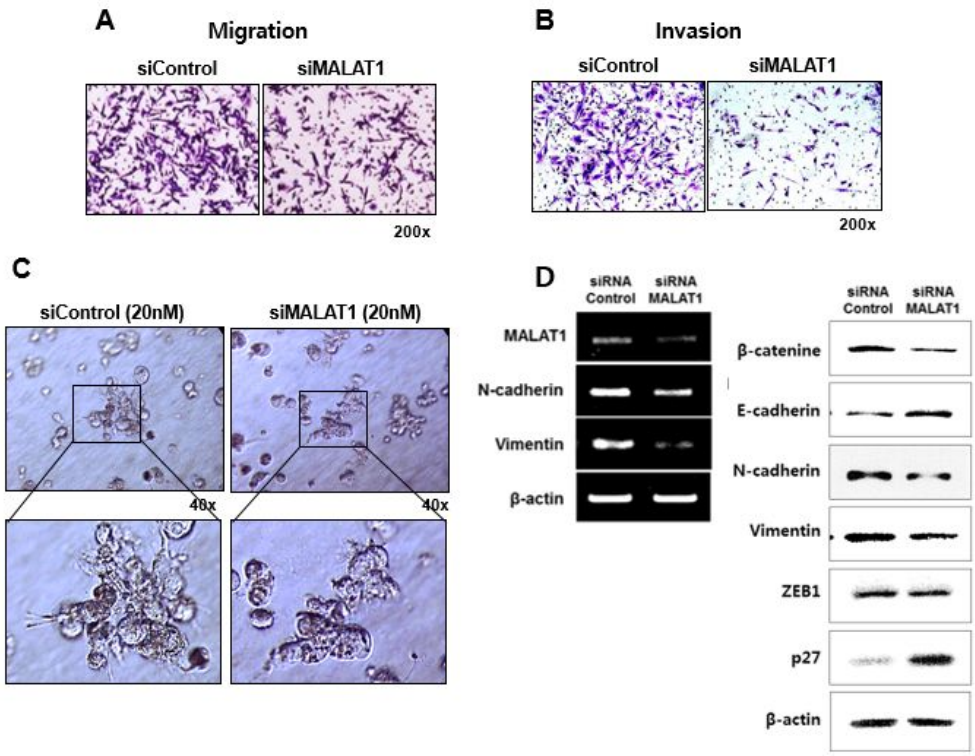

Figure 15. The effect of MALAT1 knockdown on the migration and invasion of MDA-MB-157 cells

(A,B) Inhibition of migration and invasion was observed on H&E stain by siRNA (B) on electromicroscopic findings (D) Expression of genes associated with migration and invasion was showed

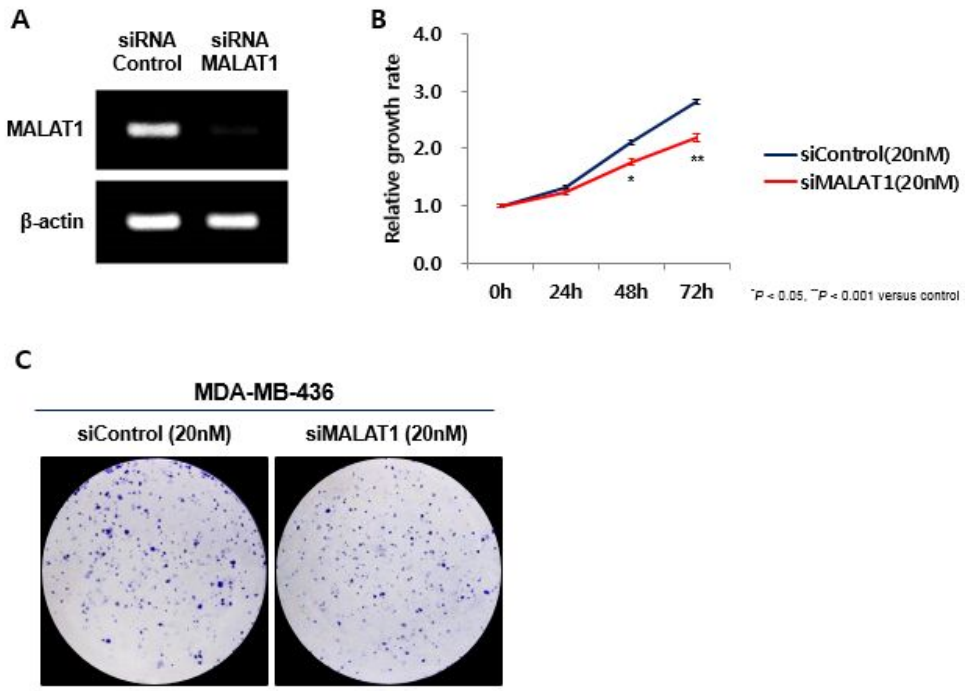

Figure 16. The effect of MALAT1 knockdown on MDA-MB-436 cell Proliferation

(A) MALAT1 was downregulated by siRNA (B) Proliferation was inhibited by MALAT1 siRNA (C) Number of colony formation was decreased

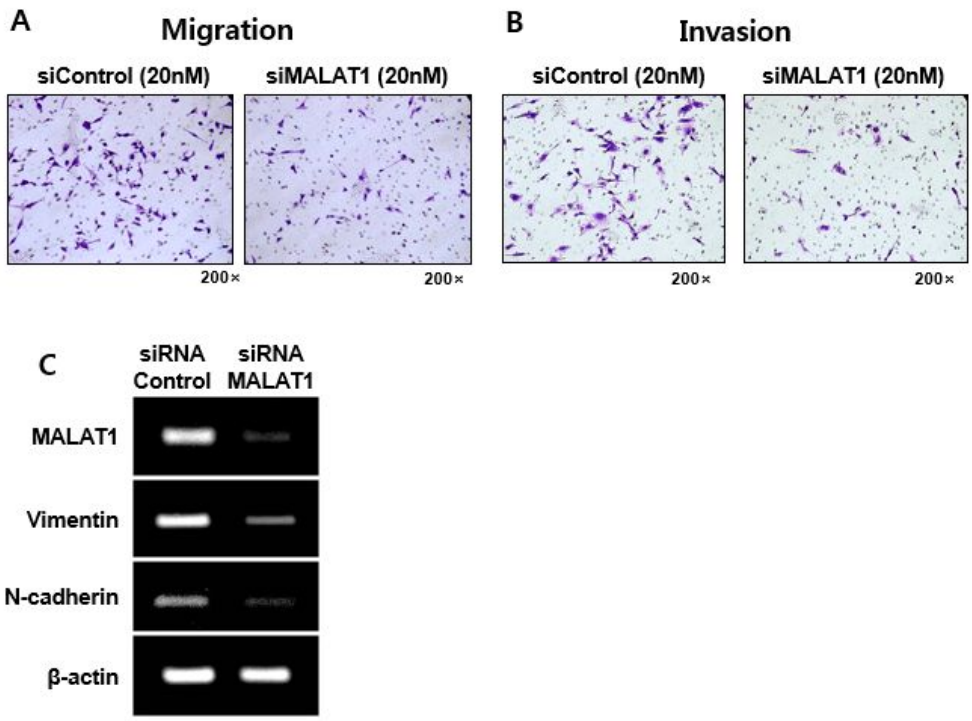

Figure 17. The effect of MALAT1 knockdown on MDA-MB-436 cell migration and invasion

(A,B) Inhibition of migration and invasion was observed on H&E stain

(C) Expression of Vimentin and N-cadherin was decreased

3.3.9 Effect of MALAT1 knockdown on HCC38 cell proliferation, migration and invasion

The change of proliferation was not observed after MALAT1 siRNA transfection to HCC38 till 72 hours. However, transwell migration and invasion assays showed migration and invasion ability decreased after inhibition of MALAT1 (Figure 18)

3.3.10 Effect of MALAT1 knockdown on HCC1937 cell proliferation, migration and invasion

Compared with control siRNA transfection, the effect of proliferation of HCC1937 cells was decreased after inhibition of MALAT1. The knockdown of MALAT1 also inhibited migration and invasion in HCC1937 cells (Figure 19).

3.3.11 Effect of MMP7 knockdown on MDA-MB-231, MDA-MB-468, HCC38 and BT20 cell proliferation, migration and invasion

Inhibition of proliferation after treating with MMP7 siRNA was observed at 72 hours in BT20 cell line, but there was no change of relative growth rate in other three cell lines(MDA-MB-231, MDA-MB-468 and HCC38) (Figure 20). The effect of MMP7 downregulated by siRNA in BT20, MDA-MB-231, MDA-MB-468 and HCC38 cell migration was investigated in transwell migration assay. We found that downregulated MMP7 inhibited the migration of cancer cell, which suggested that MMP7 may play an important role in migration of breast cancer cells. The result of invasion assay demonstrated that downregulated MMP7 significantly inhibited invasion of MDA-MB-231 cells, suggesting that MMP7 may play a role in invasion of triple negative breast cancer cells (Figure 21).

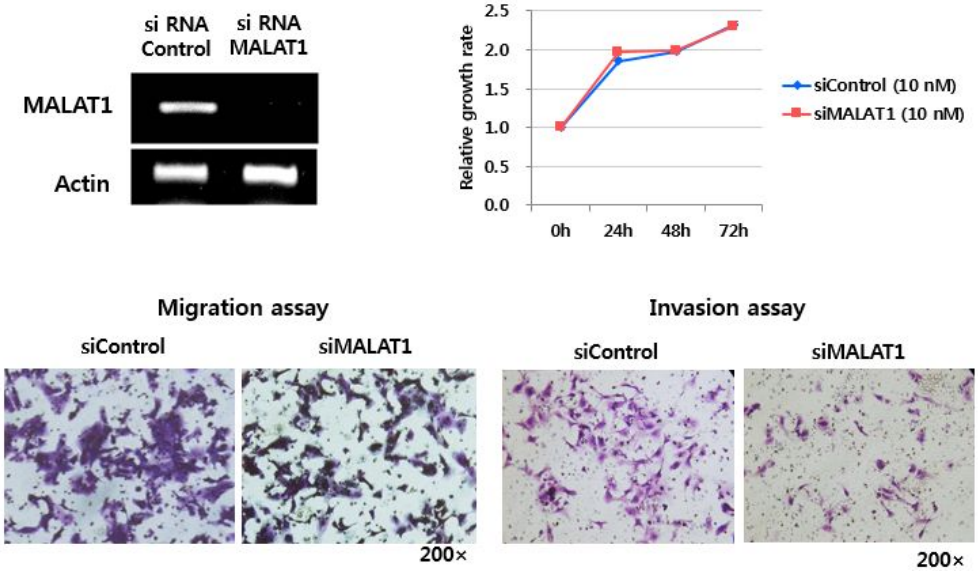

Figure 18. The effects of MALAT1 knockdown on HCC38 cell proliferation, migration and invasion

(A) MALAT1 was downregulated by siRNA (B) Proliferation was not changed by MALAT1 siRNA (C) Migration and invasion were inhibited by siRNA

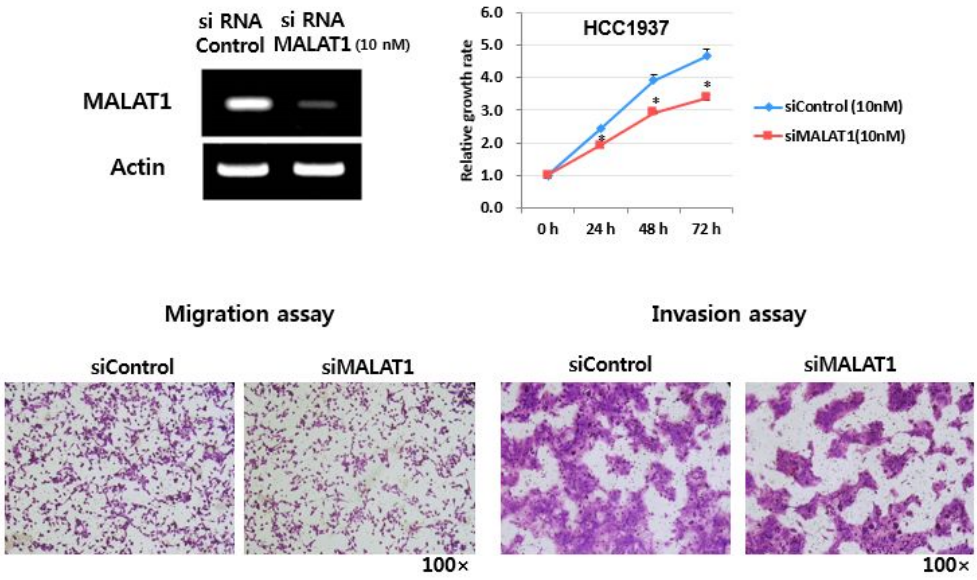

Figure 19. The effect of MALAT1 knockdown on HCC1937 cell proliferation, migration and invasion

(A) MALAT1 was downregulated by siRNA (B) Proliferation was significantly inhibited by MALAT1 siRNA (p,0.05) (C) Migration and invasion were inhibited by siRNA

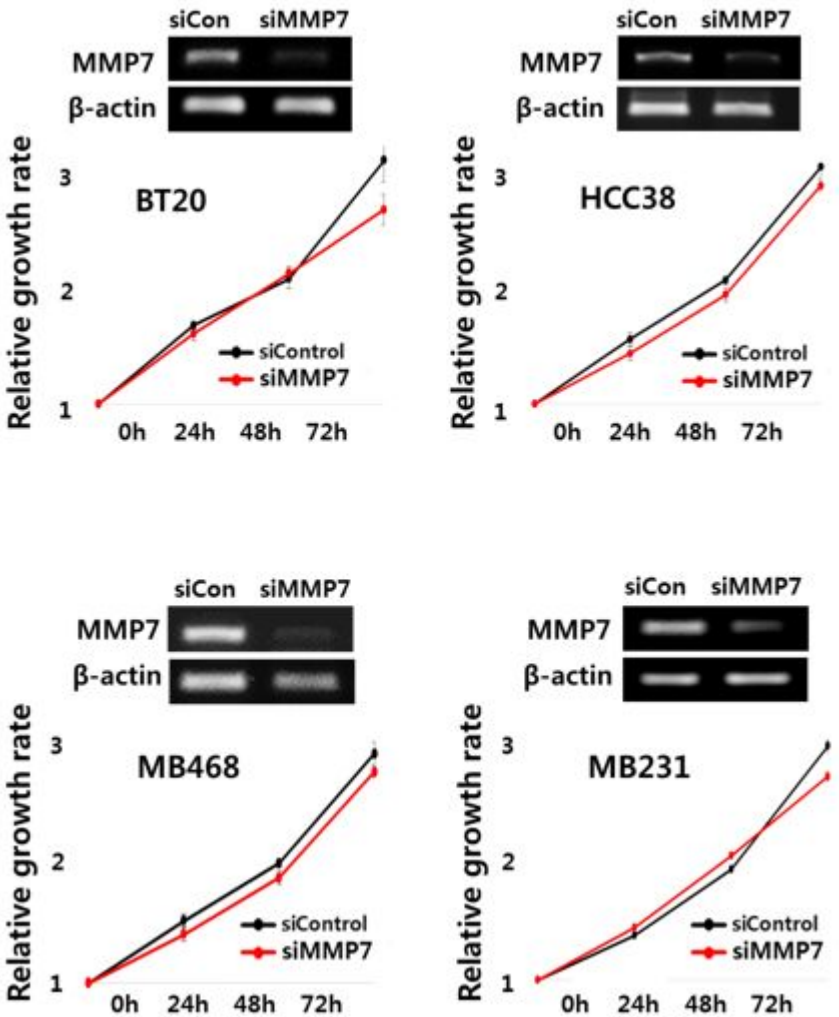

Figure 20. The effect of MMP7 knockdown on proliferation of BT20, HCC38, MDA-MB-468, and MDA-MB-231 cells

Proliferation of triple negative cancer cells was not changed by MMP7 siRNA

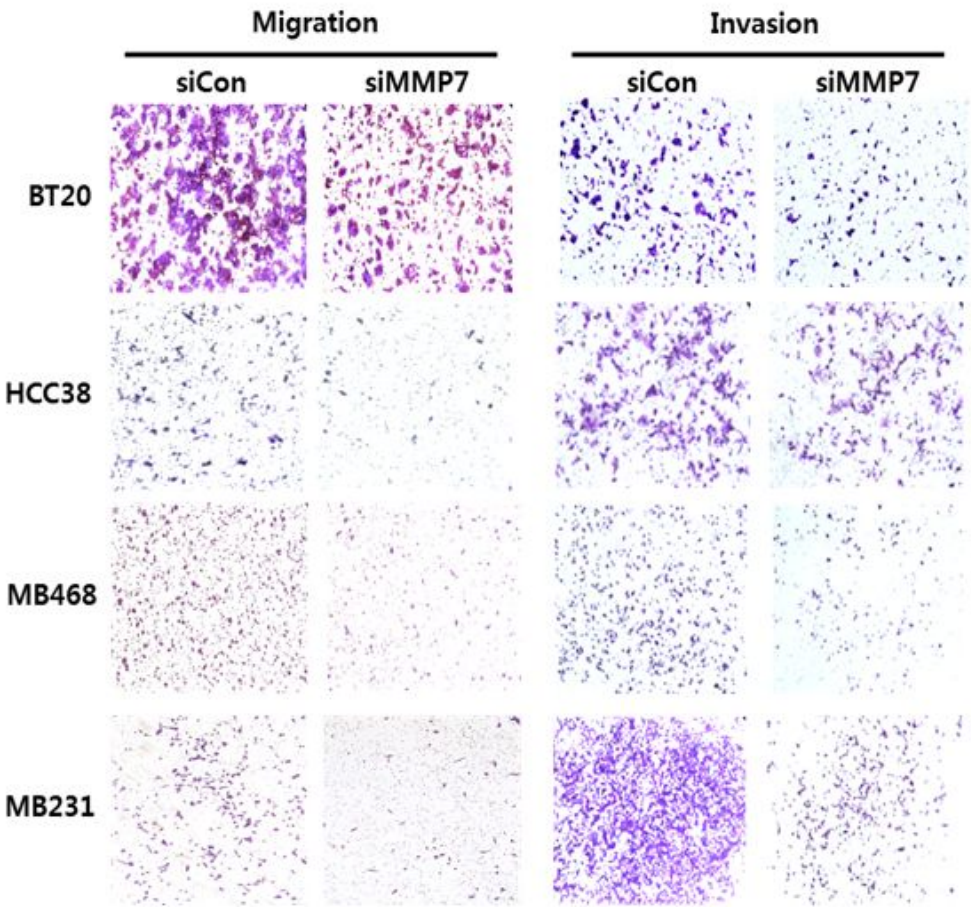

Figure 21. The effect of MMP7 knockdown on migration and invasion of BT20, HCC38, MDA-MB-468, and MDA-MB-231 cells

Inhibition of migration and invasion was observed on H&E stain by MMP7 siRNA

4. Discussion

NGS technologies allow for the generation of whole exome or whole genome sequencing data, which can be used to identify novel genetic alterations associated with defined phenotypes or to expedite discovery of functional variants for improved patient care. NGS-based tests are more sensitive, faster, easier to use, and less expensive than the conventional Sanger method, which is the most widely test currently used and is considered the gold standard test (21).

Traditionally, array CGH and FISH are used to detect CNVs. However, because of their low resolutions (about 5～10 Mbp for FISH, and 10～25 kbp with 1 million probes for aCGH), hybridization noise, limited coverage for genome, and difficulty in detecting short CNVs (22-24). NGS technology was employed for the detection of

CNVs with high resolution (<10 kbp) (25).

NGS has been used for genotyping and has included comprehensive characterization of CNVs by generating hundreds of millions of short reads in a single run (26). The advantages of the NGS include higher resolution and depth of coverage, more accuracy of copy numbers, more detailed detection of breakpoints region, and higher ability to detect novel CNVs (27,28).

Analyzing with NGS, we found DNA of 1,737 genetic alterations with high amplification. Although some have been shown to be driver mutation, such as oncogenic ERBB2 mutations (29), there is still much to be learned about the genetic alterations and their function in breast cancer.

As previous reports, our CNAs analysis showed highly amplified known drivers in relatively large number of sample, such as ERBB2 (21.5%) in chromosome 17, FGFR1(5.4%) and PVT1(8.6%) in

chromosome 8, CCND1 (3.2%), EMSY(4.3%), PAK1(3.2%) in chromosome 11 and CCNE1 (4.3%) in chromosome. Compared with above genes, ESR1 (1.1%) in chromosome 6, MDM2 (1.1%) in chromosome 12 and KIT (1.1%) in chromosome 4 in small number of samples. We also demonstrated positive correlation between known drivers and RNA expression via WES and RNA-seq with NGS.

Among several novel and highly amplified genes, we chose 4 genes (NEK8, CLK4, MALAT1 and MMP7), and demonstrated their amplification in breast cancer tissues and correlation with RNA expression. To our knowledge, there are little reports about expression of these genes and their function in breast cancer.

NEK8 is one of a family of serine/threonine kinases named the Nek (NIMA Related Kinase) kinases. Nek family of protein kinases were identified as participating in the control of the cell cycle and the overexpression of a kinase domain mutation of NEK8 leads to

reduced actin protein levels and an increase in cdk1/CyclinB1 protein levels, indicating a role for NEK8 in cell cycle progression from G2 to M phase. Bowers et al found that normal human tissue expression of NEK8 is restricted, while primary human breast tumors overexpress Nek8. Although their study demonstrated the overexpression of NEK8 in breast tumor, they observed the effect of NEK8 overexpression in osteosarcoma cell line (30).

In our study, the overexpression of NEK8 in various breast cancer cell lines as well as breast cancer tissue and this is the first report about the role of NEK8 associated with cancer cell migration and invasion in breast cancer.

Previous report showed NEK8 participates in the control of cell cycle progression from G2 to M phase, and our study showed the association between NEK8 and cancer cell proliferation firstly.

We also reveal the knockdown of NEK8 downregulated breast cancer stem cell marker (Sox2, Cot4b and CD44), hence, anti-NEK8 could be putative therapeutic target for refractory triple negative breast cancer.

CDC-like kinase 4(CLK4) belongs to Serine/arginine-rich (SR) protein kinases that are known to control alternative splicing processes via modulation of SR protein phosphorylation (31). Hypoxia was also shown to modulate alternative splicing in normal tissues as well as in cancer cells and the regulation of alternative splicing plays an essential role in breast cancer (32).

The role of CLK4 in cancer is largely unknown and amplification of CLK4 in breast cancer tissue is not reported. Our study demonstrated amplification and expression in breast cancer tissues and triple negative breast cancer cell lines. We observed CLK4 regulates migration and invasion via in vitro study, but were not able

to reveal the association between CLK4 and cancer cell proliferation.

The metastasis associated lung adenocarcinoma transcript 1 (MALAT1) is a long noncoding RNA transcribed from the nuclear-enriched transcript 2 (NEAT2). As a highly abundant nucleus-restricted RNA, MALAT1 can modulate the activity of serine/arginine (SR) splicing factor including its pre-mRNA splicing and mRNA export (33,34), and MALAT1 was reported to enhance the proliferation and invasion of nasopharyngeal carcinoma cell line CNE-1 (35). Several *in vivo* studies have shown that overexpression of MALAT1 enhances carcinogenesis of gastric cancer (36), gallbladder cancer (37), and lung cancer (38) in nude mice while depletion of MALAT1 in tumor cells reduces tumorigenicity (39). Although MALAT1 is considered as an independent prognostic marker and a potential target for metastatic lung cancer and plays an important role in carcinogenesis (39), the role of MALAT1 in breast

cancer remains largely unknown.

In our patients analysis, MALAT1 was associated with breast cancer cell proliferation, migration and invasion, and 2.3 fold higher expressed in patient with distant metastasis compared with patients without distant metastasis(p=0.01, data not shown).

To our knowledge, we demonstrated that inhibition of MALAT1 enhanced cell cycle arrest in G0/G1 and G2/M phases for the first time.

Matrix Metalloproteinases(MMPs) have been considered regarded as critical role in cancer metastasis via extracellular matrix degradation and cancer cell invasion. Many studies showed that upregulation of MMPs led to increased cancer cell migration, invasion and cancer cell-mediated tissue remodeling (40). MMP7 is a member of MMPs and structural-related zinc-dependent endopepidase. Its primary role is to break down extracellular matrix

by degrading molecules such as casein, gelatins, fibronectin and proteoglycan (41). Compared with other 3 genes, there are relatively many reports concerning about MMP7 in breast cancer. Several reports showed that MMP7 is overexpressed in breast cancer tissue and is associated with triple negative subtype of breast cancer. Overexpression of MMP7 is related to poor survival and pathological complete response in breast cancer patients (42).

There are no report that MMP7 enhances migration and invasion in various breast cancer cell lines, directly. We demonstrated not only mRNA expression of MMP7 in various cancer cell lines but also implication of MMP7 as a potential therapeutic target for the treatment of triple negative breast cancer.

Metastasis of cancer is a complex process by cancer cells disseminate from primary cancer to distant organs. Multistep process includes escaping, migration and invasion to surrounding tissues,

intravasation, extravasation and proliferation in metastatic focus (43,44). Migration and invasion are very important in metastasis and inhibition of these process could be therapeutic option in breast cancer.

Several studies reported the relationship between DNA CNAs and RNA expression in cancer. Phillips *et al.* have shown that with the acquisition of tumorigenicity in prostate cancer cell line, new chromosomal gains and losses resulted in a statistically significant increase and decrease in the average expression (45). In contrast, Platzer et al. reported that in metastatic colon cancers only ~4% of genes within amplified regions were found more highly expressed (46).

Analysis of combined DNA CNAs and RNA expression in breast cancer has several important implications.

Pollack JR et al. suggested high degree of copy number-dependent gene expression in tumors and elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. They also suggested that analyzing the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (17).

The representative study using analysis of CNAs and mRNA expression by NGS is large scale analysis by METABRIC.

They investigated germline CNVs, SNPs and somatic CNAs, and described an integrated genomic/transcriptomic analysis of breast cancer with long-term clinical outcomes composed of a discovery set of 997 primary tumors and a validation set of 995 tumors.

This study found that putative drivers with amplification (MDM1,

MDM4, CDK3, CDK4, CAMK1D, PI4KB, NCOR1) and also with deletion (PPP2R2A, MTAP, MAP2K4). Using integrative clustering analysis, 10 subgroups were established and clinical outcomes was calculated according to subgroups (6).

Compared with METABRIC study, our study also analyzed somatic CNAs combined with RNA expression and reported association between target genes aberration and distant metastasis, but we did not analyze germline CNVs and genes with deletion.

However, we found 4 novel drivers previously not reported in breast cancer and revealed the function of each genes.

Our study have some limitations. Four novel genes are amplified in very small number of patients and to identify the function of these genes, it is needed to observe whether proliferation, migration and invasion of breast cancer are changed or not after amplification of these genes. Further study is necessary to demonstrate alterations

of target genes in larger number of patients and the effect of target genes amplification on various cancer cell lines.

In conclusion, our result demonstrated that amplification of NEK8, CLK4, MALAT1 was found in breast cancer patients and the role of 4 genes (NEK8, CLK4, MALAT1 and MMP7) modulated cancer proliferation, migration, invasion and the effect of NEK8 on breast cancer stem cell in various breast cancer cell lines systematically for the first time. Further study is needed to confirm these genes as a prognostic factor in breast cancer patients and discovery of 4 new genes might be used as a therapeutic target especially for triple negative breast cancer in the future.

5. References

1. Feuk L, Carson AR, scherer SW. Structural variation in the human genome. Nat Rev Genet 2006;7:85-97.

2. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975;94:441-8.

3. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.

4. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci USA 2008;105:16224-9.

5. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM et al. Signatures of mutation and selection in the cancer genome.

Nature 2010;893-8.

6. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ et al. The genomic and transcriptomic architecture of 2,000 breast tumors reveals novel subgroups. Nature 2012;486:346-52.

7. Knadoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C et al. Mutational landscape and significance across 12 major cancer type. Nature 2013;502:333-9.

8. Yuan E, Haghighi F, White S, Costa R, McMinn J, Chun K et al. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. Cancer research 2008;66(7): 3443-51.

9. Juric D, Bredel C, Sikic BI, Bredel M. Integrated high-resolution genome-wide analysis of gene dosage and gene expression in human brain tumors. Methods Mol Biol 2007;377:187-202.

10. Chi B, deLeeuw RJ, Coe BP, Ng RT, MacAulay C, Lam WL. MD-

SeeGH: a platform for integrative analysis of multi-dimensional genomic data. BMC Bioinformatics 2008;9:243.

11. Andrews J, Kennette W, Pilon J, Hodgson A, Tuck A B, Chambers AF et al. Multi-platform whole-genome microarray analyses refine the epigenetic signature of breast cancer metastasis with gene expression and copy number. PLoS One 2010;5(1);e8665. doi:10.1371/journal.pone.0008665.

12. Balsara B R, Testa J R. Chromosomal imbalances in human lung cancer. Oncogene 2002;21(45): 6877-83.

13. Schlomm T, Kirstein P, Iwers L, Daniel B, Steuber T, Walz J et al. Clinical significance of epidermal growth factor receptor protein overexpression and gene copy number gains in prostate cancer. Clin Cancer Res 2007;13(22): 6579-84.

14. Cheng I, Levin AM, Tai YC, Plummer S, Chen GK, Neslund-Dudas C et al. Copy number alterations in prostate tumors and disease

aggressiveness. Genes Chromosomes Cancer 2012;51(1): 66–76.

15. Johnson CE, Gorringe KL, Thompson ER, Opeskin K, Boyle SE, Wang Y et al. Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. Breast Cancer Res Treat 2012;133(3): 889–98.

16. Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg Å, Karhu R et al. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. Cancer Res 2013;63(24): 8861–8.

17. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Nat Acad Sci U S A 2002;99(20): 12963–8.

18. Callagy G, Pharoah P, Chin S F, Sangan T, Daigo Y, Jackson L et al.

Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. J Pathol 2005;205: 388-96.

19. Chin S F, Teschendorff A E, Marioni J C, Wang Y, Barbosa-Morais N L, Thorne NP et al. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 2007;8(10): R215.

20. Lahti L, Schäfer M, Klein H U, Bicciato S, Dugas M. Cancer gene prioritization by integrative analysis of mRNA expression and DNA copy number data: a comparative review. Brief Bioinform 2013;14(1): 27-35.

21. Previati M, Manfrini M, Galasso M, Zerbinati C, Palatini J, Gasparini P et al. Next generation analysis of breast cancer genomes for precision medicine. Cancer Lett 2013;339:1-7..

22. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate

detection of copy number variants using read depth of coverage. Genome Res 2009;19: 1586-92.

23. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, et al.: Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 2001; 29:263-4.

24. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008;26:1135-45.

25. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007;318: 420-6.

26. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11:31-46.

27. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet 2011;12:363-376.

28. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer

genomes through second-generation sequencing. Nat Rev Genet 2010;11:685-696.

29. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC et al. Activating HER2 mutations in HER2 gene amplification-negative breast cancer. Cancer Discov 2013; 3: 224-37.

30. Bowers AJ, Boylan JF. Nek8, a NIMA family kinase member, is overexpressed in primary human breast tumors. Gene 2004;328:135-42.

31. Eisenreich A, Malz R, Pepke W, Ayral Y, Poller W, Schultheiss HP et al. Role of the phosphatidylinositol 3-kinase/protein kinase B pathway in regulating alternative splicing of tissue factor mRNA in human endothelial cells. Circ J 2009;73:1746-1752.

32. Hirschfeld M, zur Hausen A, Bettendorf H, Jäger M, Stickeler E. Alternative splicing of Cyr61 is regulated by hypoxia and significantly changed in breast cancer. Cancer Res 2009;69:2082-

2090.

33. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 2010;39:925-938.

34. Wilusz JE, Freier SM, Spector DL. 3`end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 2008;135:919-932.

35. Xie L, Hu Z, Wang X, et al. Expression of long noncoding RNA MALAT1 gene in human nasopharyngeal carcinoma cell lines and its biological significance. Nan Fang Yi Ke Da Xue Xue Bao 2013;33:692-7.

36. Okugawa Y, Toiyama Y, Hur K, et al. Metastasis-associated long non-coding RNA drives gastric cancer development and promotes peritoneal metastasis. Carcinogenesis 2014;35:2731-9.

37. Wu XS, Wang XA, Wu WG, et al. MALAT1 promotes the proliferation and metastasis of gallbladder cancer cells by activating the ERK/MAPK pathway. Cancer Biol Ther 2014;15:806-14.

38. Gutschner T, Hammerle M, Eissmann M, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res 2013;73:1180-9.

39. Schmidt LH, Spieker T, Koschmieder S, et al. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol 2001;6:1984-92.

40. Liotta LA. Tumor invasion and metastasis—role of the extracellular matrix: Rhoads Memorial Award lecture. Cancer Res 1986;46:1-7.

41. Miyazaki K, Hattori Y, Umenishi F, et al. Purification and characterization of extracellular matrix-degrading metalloproteinase, matrin (pump-1), secreted from human rectal carcinoma cell line.

Cancer Res 1990;50:7758-64.

42. Sizemore ST, Sizemore GM, Booth CN, et al. Hypomethylation of the MMP7 promoter and increased expression of MMP7 distinguishes the basal-like breast cancer subtype from other triple-negative tumors. Breast Cancer Res Treat 2014;146:25-40.

43. Chambers AF, Groom AC, MacDonald IC. Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2002;2:563-72.

44. Geho DH, Bandle RW, Clair T, et al. Physiological mechanisms of tumor-cell invasion and migration. Physiology 2005;20;194-200.

45. Phillips JR, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H et al. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res 2001;61:8143-9.

46. Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A et al. Silence of chromosomal amplifications in colon cancer. Cancer

Res 2002;62:1134-8.

국문 초록

배경: 유방암에서 유전자의 single point mutation은 다른 암종에 비해 상대적으로 드물게 관찰되며 유전자의 복제수 변이 (Copy Number Alterations; CNAs)는 single point mutation보다 유방암의 발생 및 진행에 중요한 역할은 하는 변화일 가능성이 높다. 최근에 보고된 METABRIC(Molecular Taxonomy of Breast Cancer International Consortium) 연구결과를 보면 유방암에서 acquired CNAs가 유전자의 40%에서 연관이 있으며, 전통적으로 DNA의 CNAs를 분석하는 방법으로 CGH나 SNP array가 사용되어 왔지만 본 연구에서는 차세대 염기서열 분석 기술(Next Generation Sequencing; NGS)을 이용하여 높게 증폭되는 DNA와 DNA의 chromosomal regions, 각 DNA에 대한 RNA의 과발현을 알아보기 위해 whole exome sequencing(WES)과 RNA-seq을 분석하였다.

방법: 93명의 유방암 환자에게 얻은 신선 동결조직을 이용하여 DNA와 RNA를 추출하였다. WES와 RNA-seq은 Illumina HiSeq 2000을 이용하여 NGS를 시행하였으며 대조군으로 각 환자에서 얻은 정상 유방조직의 DNA도 염기서열을 분석하여 비교하였다. 해당 유전자 위치의 mean depth와 coverage는 Genome Analysis ToolKit(GATK)를 이용하여 분석하였다. NGS에서 증폭을 보인 4개에 유전자에 대해 유방암 조직내에서의 증폭을 확인하기 위해 probe를 제작하여 FISH를 시행하였다. 해당 유전자의 기능을 규명하기 위해 8개의 유방암 세포주

를 사용하였으며 siRNA로 knock-down 시킨 후 세포주의 증식, 이동, 침윤의 변화를 관찰하였다. 또한 유방암 줄기세포에 대한 영향을 관찰하기 위해 줄기세포 마커와 mammosphere 형성의 변화를 관찰하였다.

결과: 두 개 이상의 샘플에서 1,737 유전자의 DNA가 증폭을 보였으며(log R>1.0), 가장 흔한 증폭을 보인 염색체는 8번과 17번이었다. 유전자 증폭의 cut-off로 FPKM>1.5을 기준으로 삼았으며, 모든 유전자 중에 가장 흔한 증폭을 보인 유전자는 ERBB2이고(21.5%) HER-2 IHC와 FISH에서도 동일한 결과를 확인했다. 높은 증폭을 보인 유전자 중에는 이미 유방암에서의 증폭이 보고되었던 8번 염색체에 위치한 FGGR1(5.4%), PVT1(8.6%), 11번 염색체의 CCND1, PAK1(3.2%), EMSY(4.3%), 19번 염색체의 CCNE1(4.3%)의 증폭을 관찰할 수 있었다. 기존에 알려진 유전자 이외에도 드물지만 반복적으로 증폭을 보이는 새로운 유전자 관찰되었으며 5번 염색체의 CLK4(3.2%), 6번 염색체의 AHI/MYB(3.2%), MMP7(2.2%), 11번 염색체의 MALAT1(1.1%), 17번 염색체의 NEK8(4.3%)였다. 이 중 4개의 유전자(CLK4, MMP7, MALAT1, NEK8)에 대해 specific probe를 제작하여 FISH를 통해 조직내에서의 유전자 증폭을 확인하였다. 해당 유전자의 mRNA의 과발현을 확인하기 위해 Real-time PCR을 시행하였으며 siRNA를 이용하여 억제시킨 후 유방암 세포주의 변화를 관찰하였다. Knock-down 후 삼중음성유방암 세포주에서 세포의 및 세포 이동과 침윤의 감소가 관찰되었으며 NEK8을 억제하였을 때 줄기세포 마커의 발현 감소 및 mammosphere 형성이 억제되었다.

결론: 본 연구의 결과는 유방암에서 증폭을 보이는 새로운 4개의 유전자를 발굴했으며 유방암에서의 역할을 규명했다. 유방암 세포주의 변화에서 얻은 결과를 바탕으로 해당 유전자들은 향후 추가적인 연구를 통해 유방암에서 새로운 치료의 표적으로 이용될 수 있을 것으로 판단된다.

주요어: 유방암, NGS, exome, 복제수 변이, 유전자 증폭, 치료 표적

학 번: 2009-31093