

# Integrated analysis of copy number alteration and RNA expression profiles of cancer using a high-resolution whole-genome oligonucleotide array

Seung-Hyun Jung<sup>1,2</sup>, Seung-Hun Shin<sup>1,2</sup>,  
Seon-Hee Yim<sup>2</sup>, Hye-Sun Choi<sup>1,2</sup>,  
Sug-Hyung Lee<sup>3</sup> and Yeun-Jun Chung<sup>1,2,4</sup>

<sup>1</sup>Department of Microbiology

<sup>2</sup>Integrated Research Center for Genome Polymorphism

<sup>3</sup>Department of Pathology

The Catholic University of Korea  
Seoul 137-701, Korea

<sup>4</sup>Corresponding author: Tel, 82-2-590-1214;

Fax, 82-2-596-8969; E-mail, yejun@catholic.ac.kr

DOI 10.3858/emmm.2009.41.7.051

Accepted 2 February 2009

Abbreviations: array-CGH, microarray-based comparative genomic hybridization; CNA, copy number alteration; oligoarray, oligonucleotide microarrays; qPCR, quantitative PCR; RAR, repeatedly altered regions

## Abstract

Recently, microarray-based comparative genomic hybridization (array-CGH) has emerged as a very efficient technology with higher resolution for the genome-wide identification of copy number alterations (CNA). Although CNAs are thought to affect gene expression, there is no platform currently available for the integrated CNA-expression analysis. To achieve high-resolution copy number analysis integrated with expression profiles, we established human 30k oligoarray-based genome-wide copy number analysis system and explored the applicability of this system for integrated genome and transcriptome analysis using MDA-MB-231 cell line. We compared the CNAs detected by the oligoarray with those detected by the 3k BAC array for validation. The oligoarray identified the single copy difference more accurately and sensitively than the BAC array. Seventeen CNAs detected by both platforms in MDA-MB-231 such as gains of 5p15.33-13.1, 8q11.22-8q21.13, 17p11.2, and losses of 1p32.3, 8p23.3-8p11.21, and 9p21 were consistently identified in previous studies on breast cancer. There were 122 other small CNAs (mean size 1.79 mb) that were detected by oligoarray only, not by BAC-array. We performed genomic qPCR

targeting 7 CNA regions, detected by oligoarray only, and one non-CNA region to validate the oligoarray CNA detection. All qPCR results were consistent with the oligoarray-CGH results. When we explored the possibility of combined interpretation of both DNA copy number and RNA expression profiles, mean DNA copy number and RNA expression levels showed a significant correlation. In conclusion, this 30k oligoarray-CGH system can be a reasonable choice for analyzing whole genome CNAs and RNA expression profiles at a lower cost.

**Keywords:** cell line, tumor; gene dosage; gene expression profiling; oligonucleotide array sequence analysis

## Introduction

One of the hallmarks of cancer is copy number alteration (CNA) across the entire genome. CNA can affect the development or progression of human malignancies by altering the expression of cancer-related genes. Recently, microarray-based comparative genomic hybridization (array-CGH) has emerged as a very efficient technology with higher resolution for the genome-wide identification and characterization of CNAs (Pinkel *et al.*, 1998; Yim and Chung, 2004). Improved resolution enables the identification of submicroscopic chromosomal alterations which are less likely to be detected by conventional cytogenetics tools. These small-sized chromosomal changes including repeatedly altered regions (RAR) in various cancers are thought to contain cancer-related genes (Albertson and Pinkel, 2003; Kim *et al.*, 2006). In addition, whole-genome approach can help to understand the contribution of CNAs to tumorigenesis and tumor behaviors such as metastasis in comprehensive terms.

In spite of technical advancement, the precise identification of CNAs is still challenging and needs further improvement. Most currently available BAC arrays provide the resolution of ~1 mb, which makes it difficult to detect smaller than mb-sized CNAs (Brennan *et al.*, 2004; Carvalho *et al.*, 2004). Oligonucleotide microarrays (oligoarray) provide higher resolution and sensitivity in detecting sub-

microscopic CNAs. Especially long oligoarray (>60 mers) guarantees increased signal intensity thanks to improved hybridization kinetics, which enables more reliable copy number analysis (Ylstra *et al.*, 2006).

Since CNAs are thought to affect expression of genes by changing genomic dosages or gene structures, it is important to interpret copy number status together with gene expression profiles (Lynch, 2002). For example, genomic amplifications of 8q24, 11q13 and 17q12 are commonly observed to be associated with overexpression of *MYC*, *CCND1* and *ERBB1* in diverse cancers (Croce, 2008). Recently, integrated genome and transcriptome analysis results for breast and lung cancers have been reported, for which different platforms were used separately for copy number and expression profiling (Dehan *et al.*, 2007; Vincent-Salomon *et al.*, 2008). However, there is no array platform currently available for the integrated CNA-expression analysis.

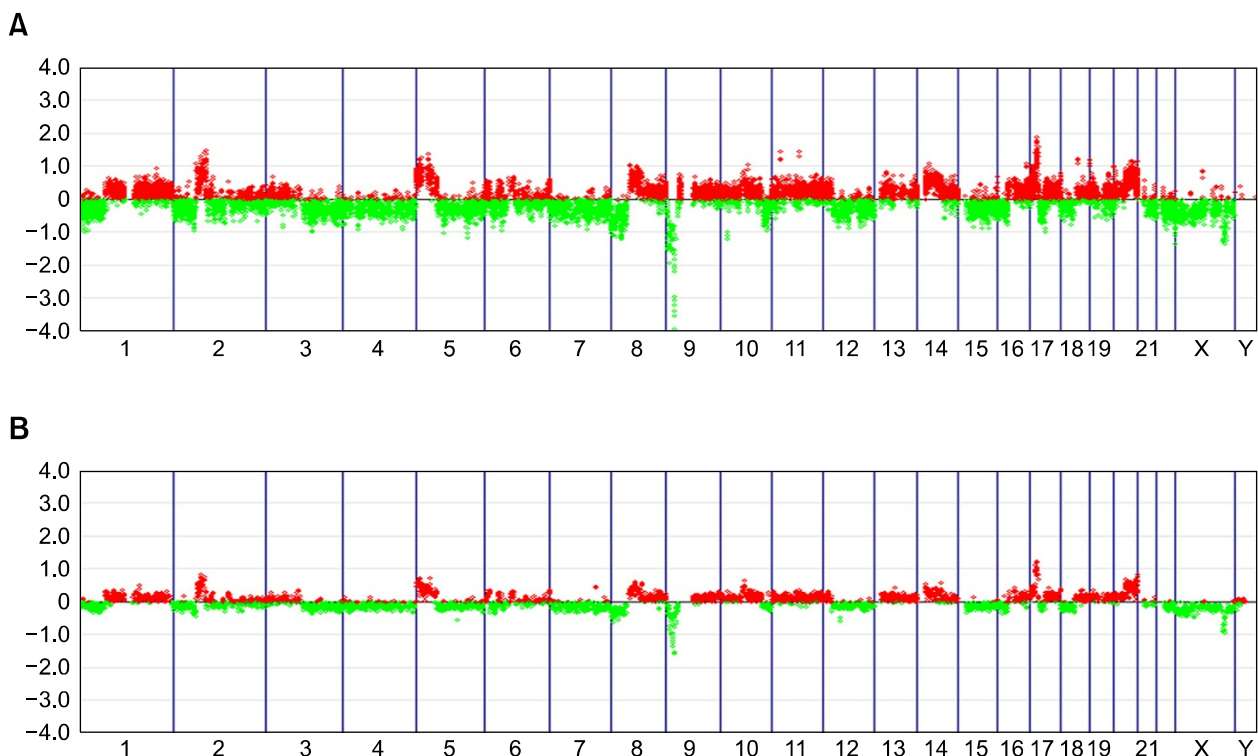
To achieve high-resolution copy number analysis integrated with analysis of expression on the same platform, we established human 30k oligoarray-based genome-wide copy number-expression analysis system using the platform originally designed

for expression analysis. In this platform, 60mer-sized oligo probes covering over 17,500 genes are spotted across the whole chromosome, which is basically the same to long oligoarrays for whole genome copy number analysis (Brennan *et al.*, 2004; Carvalho *et al.*, 2004). In this study, we explored the applicability of this system for integrated genome and transcriptome analysis using cancer cell line.

## Results

### Comparison of CNA detection ability between oligoarray and BAC array

Firstly, to see how accurately single copy difference would be detected by two different platforms, we hybridized normal male and female genomic DNA onto the two platforms, oligoarray (30k, 100 kb resolution on average) and BAC array (3k, 1mb resolution on average). Although both platforms successfully identified the single copy difference of X chromosome between male and female, oligoarray detected the difference more accurately than BAC array. In oligoarray-CGH results, mean  $\log_2$



**Figure 1.** Whole genome array-CGH profiles of the MDA-MB-231 using 2 different arrays. (A) Profile by 30k oligoarray and (B) Profile by 3k BAC array. X axis represents individual chromosomes and Y axis represents the signal intensity ratio (tumor/normal) in  $\log_2$  ratio. Red dots represent the probes above ratio zero and green dots represent below zero.

**Table 1.** Genomic alterations identified by both BAC and oligoarrays in MDA-MB-231.

CNA ID	Clone	Chr	Map position	Size (mb)	Cytoband	Cancer-related genes
G1	PH_hs_0040486- PH_hs_0010160	2	61,203,137- 66,652,876	5.4	2p15-2p14	
G2	PH_hs_0017517- PH_hs_0027696	2	68,736,091- 85,662,429	16.9	2p14-2p11.2	<i>ANTXR1, TIA1, TGFA, MOBKL1B, DOK1, TMSB10, KCMF1, CAPG</i>
G3	PH_hs_0015309- PH_hs_0027712	5	258,178- 39,418,807	39.1	5p15.33-5p13.1	<i>AHRR, TERT, SRD5A1, AMACR, SKP2, GDNF</i>
G4	PH_hs_0030540- PH_hs_0031517	6	71,660,593- 76,019,546	4.3	6q13-6q14.1	<i>DDX43, EEF1A1</i>
G5	PH_hs_0013046- PH_hs_0035433	8	50,571,933- 81,077,958	30.5	8q11.22-8q21.13	<i>RB1CC1, SDCBP, ASPH, CRH, COPS5, SULF1, NCOA2</i>
G6	PH_hs_0031183- PH_hs_0003766	10	57,787,294- 63,623,495	5.8	10q21.1-10q21.2	<i>CDC2</i>
G7	PH_hs_0039119- PH_hs_0028475	14	19,987,030- 61,092,886	41.1	14q11.2-14q23.1	<i>APEX1, NDRG2, MMP14, PSME1, PSME2, PRKD1, NFKBIA, RPS29, PTGER2, LGALS3</i>
G8	PH_hs_0038941- PH_hs_0019422	16	45,061,443- 46,478,325	1.4	16q11.2-16q12.1	
G9	PH_hs_0020121- PH_hs_0027026	17	1,368,228- 19,584,383	18.2	17p13.3-17p11.2	<i>SERPINF1, HIC1, ALOX15, PELP1, PLD2, GP1BA, PFN1, XAF1, ALOX12, GABARAP, CLDN7, TNFSF12, TNFSF13, TP53, ALOX15B, PER1, AURKB, NTN1, MAP2K4, FLCN, PEMT, LLGL1, MAPK7, ALDH3A1</i>
G10	PH_hs_0032483- PH_hs_0039471	20	29,538,943- 62,139,615	32.6	20q11.21-20q13.33	<i>ID1, BCL2L1, DNMT3B, E2F1, ASIP, RBL1, SRC, BLCAP, TGM2, TOP1, PLCG1, HNF4A, WISP2, STK4, SLPI, WFDC2, UBE2C, MMP9, CD40, EYA2, SNAI1, CEBPB, ZNF217, BCAS1, AURKA, TFAP2C, BMP7, LAMA5, GATA5, BIRC7, PTK6, TNFRSF6B</i>
L1	PH_hs_0035564- PH_hs_0025172	1	48,232,776- 52,907,016	4.6	1p33-1p32.3	<i>CDKN2C</i>
L2	PH_hs_0002020- PH_hs_0032953	2	50,001,041- 59,330,250	9.3	2p16.3-2p16.1	<i>RTN4</i>
L3	PH_hs_0030124- PH_hs_0000100	8	398,441- 39,904,628	39.5	8p23.3-8p11.21	<i>MCPH1, ANGPT2, DEFB4, CLDN23, PINX1, GATA4, CTSB, DLC1, TUSC3, FGF20, CNOT7, PDGFRL, MTUS1, NAT1, NAT2, LPL, LZTS1, KIAA1967, TNFRSF10B, TNFRSF10C, TNFRSF10D, TNFRSF10A, NKX3-1, STC1, PTK2B, CLU, SCARA3, PBK, FZD3, WRN, EIF4EBP1, BAG4, FGFR1, INDO</i>
L4	PH_hs_0024727- PH_hs_0032012	9	3,237,130- 26,831,705	23.5	9p24.2-9p21.2	<i>JAK2, RLN2, CD274, IFNB1, IFNA1, MTAP, CDKN2A, CDKN2B, TUSC1</i>
L5	PH_hs_0016939- PH_hs_0026494	23	1,465,314- 16,685,297	15.2	Xp22.33-Xp22.2	<i>CD99, TMSB4X, FIGF, BMX, GRPR</i>
L6	PH_hs_0004181- PH_hs_0015032	23	22,175,974- 30,487,677	8.3	Xp22.11-Xp21.2	<i>SAT1</i>
L7	PH_hs_0009133- PH_hs_0024909	23	123,334,353- 133,461,868	10.1	Xq25-Xq26.2	<i>GPC3</i>

signal intensity ratio value of the chromosome X probes in the female-versus-male hybridization was 0.78 (Standard Deviation (SD)=0.29) and that of autosomes was 0.0027 (SD=0.21). However, in BAC array-CGH results, mean  $\log_2$  ratio value of the chromosome X probes was 0.51 (SD=0.16)

and that of autosomes was -0.0020 (SD=0.06). Female versus male array-CGH plots from both platforms are available in the Supplemental Data Figure S1.

To examine the sensitivity of two arrays for detecting regional CNAs, we analyzed the same

**Table 2.** High copy alterations identified by oligoarray in MDA-MB-231.

CNA ID	Chr	Map position (bp) <sup>a</sup>	Size (mb)	Cytoband	Cancer-related genes
AMP1	2	61203137-61259674	0.06	2p15	
AMP2	2	71059975-71153730	0.09	2p13.3	
AMP3	2	73848768-74066608	0.22	2p13.2-2p13.1	
AMP4	2	79216764-80728801	1.51	2p12	
AMP5	2	84530324-85422904	0.89	2p11.2	<i>TMSB10, KCMF1</i>
AMP6	10	70123063-70221062	0.10	10q21.3	
AMP7	11	70938048-71392241	0.45	11q13.4	
AMP8	16	78191149-78329932	0.14	16q23.1-16q23.2	
AMP9*	17	12480926-18861433	6.38	17p12-17p11.2	<i>FLCN, PEMT, LLGL1</i>
AMP10	18	44899146-45272187	0.37	18q21.1	
AMP11	19	1728486-1803459	0.07	19p13.3	
AMP12	19	63631888-63638588	0.01	19q13.43	
AMP13	20	39238969-39241018	0.002	20q12	
AMP14	20	43315080-43430446	0.12	20q13.12	<i>SLPI</i>
AMP15	20	47422122-48178121	0.76	20q13.13	<i>SNAI1</i>
AMP16	20	49011077-49013562	0.002	20q13.13	
AMP17	20	56169424-56327399	0.16	20q13.32	
AMP18	20	60943121-61007784	0.06	20q13.33	
HD1	5	75006020-75064124	0.06	5q13.3	
HD2	5	137251112-137513312	0.26	5q31.2	
HD3	8	10383238-10659979	0.28	8p23.1	
HD4	8	12658287-12985408	0.33	8p23.1-8p22	<i>DLC1</i>
HD5	8	25756053-25945580	0.19	8p21.2	
HD6	8	27783832-28265239	0.48	8p21.1	
HD7*	9	3441038-6320859	2.88	9p24.2-9p24.1	<i>JAK2, RLN2, CD274</i>
HD8	9	12699155-14078289	1.38	9p23	
HD9	9	17493556-18899521	1.41	9p22.2-9p22.1	
HD10	9	19222270-26831705	7.61	9p22.1-9p21.2	<i>IFNB1, IFNA1, MTAP, CDKN2A, CDKN2B, TUSC1</i>
HD11	10	20410067-20608787	0.20	10p12.32-10p12.31	
HD12	11	36443684-36456476	0.01	11p12	
HD13	22	49312510-49316289	0.004	22q13.33	
HD14	23	123334353-125125459	1.79	Xq25	

<sup>a</sup>bp, base pair. \*High copy alterations identified in both BAC and oligoarrays.

cancer cell line using two platforms (Figure 1). All the chromosomal alterations identified by BAC array-CGH were consistently detected by oligoarray (Table 1). Mean size of the CNAs identified by both arrays was 18 mb. However, many of the relatively smaller sized-CNAs were detected only by oligoarray. In total, 122 alterations which were not detected in BAC array were identified by oligoarray, of which mean size was 1.79 mb. Detailed information of the 122 alterations is available in the Supplemental Data Table S1.

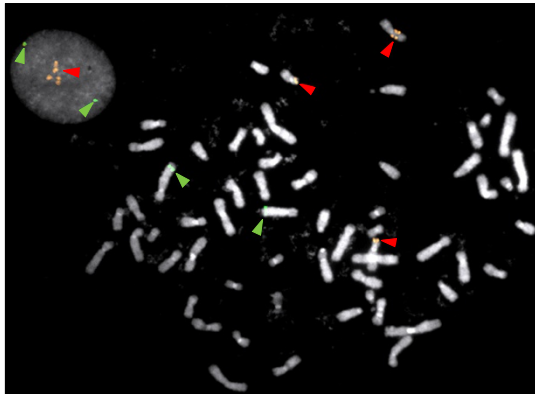
Most of the high copy changes, *i.e.* amplifications (AMP) and homozygous deletions (HD), were also identified only by oligoarray. In total, 18 AMPs and 14 HDs (mean size = 0.88 mb) were detected by oligoarray (Table 2). Among them, only 1 AMP (17p12) and 1 HD (9p24) were detected by both BAC- and oligoarrays. All the others were detected only by oligoarray under our criteria.

#### Validation of CNAs identified by oligoarray-CGH

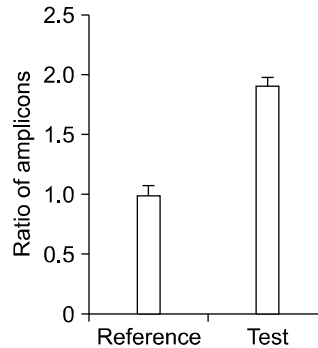
We firstly selected a copy number gained region (17p12) consistently identified by both oligo- and BAC arrays. Both FISH and genomic qPCR demonstrated the existence of the copy number gain in this region, which agrees with array-CGH results (Figure 2). In FISH analysis, median number of red signals targeting 17p12 (3 per each cell,  $n = 107$ ) was significantly higher than number of green signals targeting 2q35, diploid control region (2 per each cell,  $n = 107$ ) ( $P < 0.0001$ ). In genomic qPCR, the signal intensity ratio of 17p12 was approximately two times higher than that of the diploid control region.

Next, to further validate the reliability of oligoarray-CGH, we performed genomic qPCR targeting 7 CNA regions, gains on 1q23.3, 16q23.1, 18q21.1 and 19q13.43, and losses on 6p12.3, 10q26.13 and 22q13.31 which were detected by oligoarray

A



B



**Figure 2.** Validation of the 17p12 copy number gain identified by oligoarray-CGH. (A) FISH analysis of the MDA-MB-231. The signal number of 17p12 (red arrow) is higher than the signal number of 2q35 where no CNA was identified by oligoarray-CGH (green arrow). (B) Genomic qPCR analysis of MDA-MB-231. The signal intensity ratio of the test DNA (MDA-MB-231) is 1.9 (SD = 0.09).

only, and also targeting one non-CNA region (2q35) (Figure 3). All the genomic qPCR profiles of these 7 CNAs and one non-CNA region were consistent with the oligoarray-CGH results under our criteria.

**Integrated analysis of copy number and expression profiles**

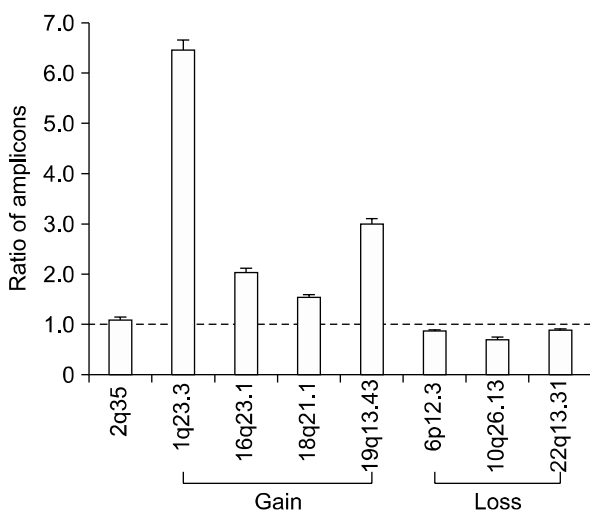
Combined interpretation of both DNA copy number and RNA expression profiles can provide new insights into biological effect of copy number alterations. To explore this possibility, we performed global gene expression analysis for the same cell line using the same oligoarray platform. As a whole, mean DNA copy number levels were significantly correlated with the mean RNA expression levels

( $R^2 = 0.92$ ) (Figure 4A). Figure 4B and C show examples of the correlations at amplified region (17p12-p11.2) and deleted region (9p22.1-9p21.2). In the peak of the amplified region (17p12), a putative oncogene *PMP22* is located, of which RNA expression was 7.6 times higher than normal breast tissue. In the deletion region on 9p22.1-9p21.2, *CDKN2A&B* tumor suppressor genes are located and their RNA expression was 6.8 times lower than normal breast tissue.

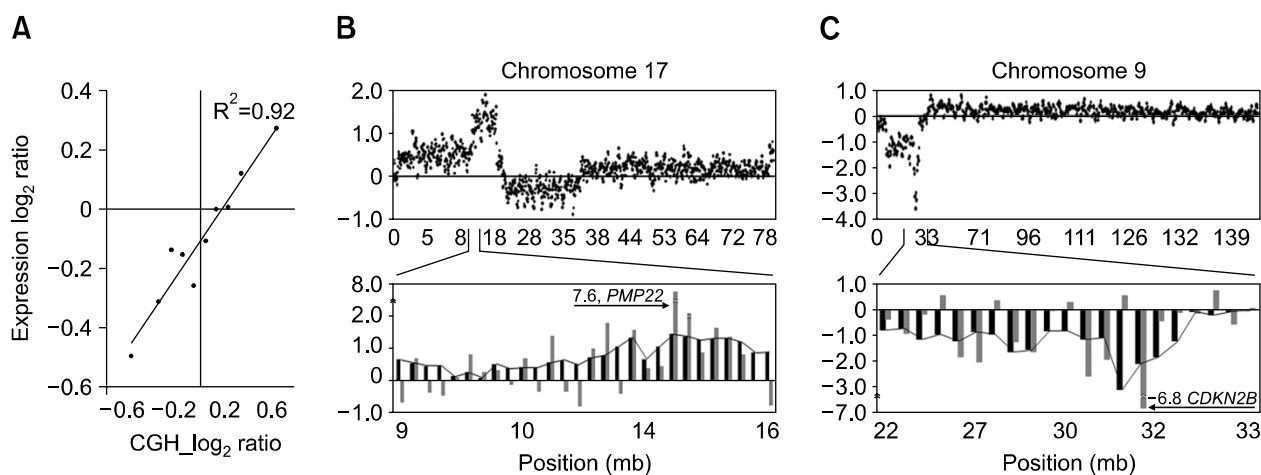
**Discussion**

In this study, we established the oligoarray-based, whole-genome copy number-gene expression analysis system and evaluated its accuracy and reliability. As a first step, we compared the CNAs detected by this 30k oligoarray and the 3k BAC array, because the validity of the BAC array has been well evaluated in identifying CNAs in various tumors (Kim *et al.*, 2005, 2006, 2008). This oligoarray identified the single copy difference more accurately than the BAC array and all CNAs detected by the BAC array were detected by the oligoarray, which suggests the reliability of the oligoarray we used.

Seventeen CNAs detected by both platforms are the major genomic aberrations in MDA-MB-231. Many of these CNAs such as gains of 5p15.33-13.1, 8q11.22-8q21.13, 17p11.2, and losses of 1p32.3, 8p23.3-8p11.21, and 9p21 were consistently identified in previous studies on breast cancer (Han *et al.*, 2006; Jones *et al.*, 2008). Interestingly, other 122 CNAs detected by the oligoarray only were approximately 10 times smaller than those detected by BAC-array. Among the 32 AMPs or HDs detected by the oligoarray, only 2 were detected by the BAC array, which reflects the higher resolution



**Figure 3.** Genomic qPCR results of 7 CNA regions identified by the oligoarray only (gains on 1q23.3, 16q23.1, 18q21.1 and 19q13.43, and losses on 6p12.3, 10q26.13 and 22q13.31), and a non-CNA region (2q35).



**Figure 4.** Integrated analysis of copy number and expression profiles. (A) Correlation analysis of copy number status and expression levels. X axis represents the array-CGH signal intensity ratio (tumor/normal) in  $\log_2$  scale and Y axis represents the expression signal intensity ratio (tumor/normal) in  $\log_2$  scale. Tumor, MDA-MB-231; Normal, normal female genomic DNA (B) Example of the correlation at an amplified region (17p12-p11.2). The arrow indicates the highest value of expression signal intensity ratio (7.6 in  $\log_2$  scale, *PMP22* gene). (C) Example of the correlation at a deleted region (9p22.1-p21.2). The arrow indicates the lowest value of expression signal intensity ratio (-6.8 in  $\log_2$  scale, *CDKN2AB* gene). In B and C, upper box represents whole chromosome arm plot and lower box represents copy number-expression signal intensity ratios in the region selected from the upper box. Both DNA copy number signal intensity ratio (solid bar) and RNA expression signal intensity ratio (gray bar) of each probe are represented in  $\log_2$  scale. X axis represents the chromosomal position in mb scale and Y axis represents the signal intensity ratio in  $\log_2$  scale.

of the platform as reported previously (Carvalho *et al.*, 2004; Brennan *et al.*, 2004; Ylstra *et al.*, 2006).

We further validated the CNAs identified by oligoarray using genomic qPCR and FISH. We selected 9 regions (8 CNA regions and 1 region without any CNAs) for validation. Of the 8 CNAs, 1 was identified in both platforms (6 mb-sized, 17p12) and the other 7 were detected only by the oligoarray. In case of the CNA detected by both arrays, both genomic qPCR and FISH analysis demonstrated the consistent copy number gain in this region. Genomic qPCR targeting the 7 oligoarray-only CNAs and 1 non-CNA region also showed the results consistent with the array-CGH result. It suggests that higher resolution oligoarray can detect smaller CNAs accurately which would be missed by the BAC-array.

A key advantage of this integrated system is the feasibility of integrated interpretation of both DNA copy number and RNA expression profiles. If genomic DNA and cDNA from the same sample is used for copy number analysis and transcriptome analysis, respectively, using the same platform, systemic errors could be minimized when exploring the relationship between copy number status and gene expression. When we assessed this possibility using MDA-MB-231, mean DNA copy number and RNA expression levels showed highly significant correlation. This result agrees with the previous reports which have demonstrated the correlation between DNA copy number alterations and gene expression in diverse cancers, although all

previous studies assessed copy number alterations and expression patterns separately using different arrays (Heidenblad *et al.*, 2005; Chin *et al.*, 2006; Dehan *et al.*, 2007; Vincent-Salomon *et al.*, 2008). In this study, we used normal Korean breast tissue RNA as reference of MDA-MB-231 RNA expression, because the matching RNA to Promega DNA was not available, which means that the reference DNA for array-CGH and the reference RNA for expression profiling of MDA-MB-231 were from different sources. To validate whether the oligoarray-CGH profiles of MDA-MB-231 using Promega genomic DNA and the DNA from the normal breast tissue are consistent with each other, we performed oligoarray-CGH of MDA-MB-231 using the normal breast DNA as reference in addition to oligoarray-CGH of MDA-MB-231 versus Promega DNA. The genome-wide CNA profiles obtained from MDA-MB-231 versus Promega DNA and from MDA-MB-231 versus normal breast DNA were largely consistent and the CNA-RNA expression correlation was highly significant too (Supplemental Data Figure S2 and S3).

There are several limitations in this study. First, we did not perform qPCR validation for every CNA detected by the oligoarray only. Although all the qPCR results from the regions we selected were consistent with oligoarray-CGH calls, we cannot assume that all the other CNAs which were not validated are true. Further validation will be required to calculate the exact validity of this oligoarray. Second, since we did not compare the

performance of this oligoarray system with currently available higher resolution oligoarray chips such as Agilent 244k or NimbleGen 1M array, we may miss even smaller CNAs. However, since main purpose of this study is to establish a reasonable tool for screening the chromosome alterations in cancer and combined interpretation of copy number and expression profiles using the same array, the resolution of 100 kb probe interval of this oligoarray is suitable enough to detect most of the CNAs precisely, given that most chromosomal alterations in cancer are mb-sized. In addition, the extremely higher resolution arrays mentioned above cannot be directly applied for gene expression analysis and, due to the cost, they are not affordable for researchers with large cancer sets analysis. Third, although we improved the array-CGH conditions to get higher signal to noise ratios and lower SDs in this study, mean SDs in oligoarrays was still higher than those in BAC arrays. Therefore, more conservative CNA defining approach such as increasing the detection threshold might be preferable.

In conclusion, this 30k oligoarray-CGH system can be a reasonable approach for analyzing whole genome CNAs and RNA expression profiles at an affordable cost.

## Methods

### Cell culture and DNA extraction

We used the MDA-MB-231 breast cancer cell line obtained from American Type Culture Collection (ATCC, Manassas, VA). The cell line was maintained in RPMI1640 medium containing 10% FBS (Hyclone, Logan, UT) under 5% CO<sub>2</sub>. Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) and quantified using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE). Female genomic DNA (Promega, Madison, WI) was also prepared as a normal control for array-CGH analysis.

### Oligo Array-CGH

We used 30k whole-genome human oligoarrays with approximately 100 kb resolution (Human OneArray™, Phalanx Biotech, Palo Alto, CA). In brief, 2 µg of genomic DNA from the MDA-MB-231 cell line was labeled with Cy3-dCTP and female control DNA (Promega, Madison, WI) was labeled with Cy5-dCTP using BioPrime Labeling Kit (Invitrogen, CA) according to manufacturer's instructions. Dye labeled DNA was purified by BioPrime spin columns (Invitrogen, Carlsbad, CA) and precipitated with 100 µg of human Cot-1 DNA (Roche, Mannheim, Germany). The labeled DNA pellet was dissolved in 50 µl of DIG hybridization buffer (Roche, Mannheim, Germany), where 600 µg of yeast t-RNA (Invitrogen, Carlsbad, CA) was

added. DNA was denatured for 10 min at 70°C and incubated for 1 h at 37°C before being applied on the oligoarray slide, which was pre-hybridized for 2 h at 37°C with 50 µl of DIG hybridization buffer containing 540 µg of herring sperm DNA. The labeled DNA solution applied on the array was incubated in MAUI hybridization machine (BioMicro, Utah) for 48 h at 37°C. The slides were washed serially in solution I (2× SSC, 0.1% SDS) for 4 min (1 time) at room temperature, in solution II (0.1× SSC, 0.1% SDS) for 3 min (2 times) and in solution III (0.1× SSC) for 30 s (3 times) followed by a rinse in DW for 10 s. Finally, the slides were spin-dried for 3 min at 900 rpm.

### BAC Array-CGH

To validate the copy number analysis results, an independent array-CGH was performed using a large insert clone array covering the entire human genome at 1 mb resolution. All the Array-CGH procedures including DNA labeling, pre-hybridization and hybridization using MAUI hybridization station (BioMicro Systems, Salt Lake city, UT) were performed as described elsewhere (Kim *et al.*, 2006).

### Scanning and CNA data analysis

Arrays were scanned using GenePix 4000B scanner (Axon Instruments, Sunnyvale, CA) and feature extraction was processed using GenePix Pro 6.0. Data processing, normalization, and re-aligning of raw array-CGH data were performed using CGHscape software (<http://www.irgcp.com/software/CGHscape>) (Jeong *et al.*, 2008). Mapping of each oligo-probe and BAC clone was performed according to the genomic location in the Ensembl Homo\_sapiens 43\_36e build. Information on the whole oligo-probe set is available on the Phalanx biotech homepage (<http://www.phalanxbiotech.com/Support/Files.html>). We performed the standard deviation (SD)-based CNAs identification as described previously with some modifications (Kim *et al.*, 2006, 2008). In brief, chromosomal gain or loss was assigned when the normalized log<sub>2</sub> intensity ratio of each data point exceeded or fell below ± 5SD derived from normal control hybridizations. Regional copy number change was defined as DNA copy number alteration which stretches over 2 or more consecutive large insert clones, but not across an entire chromosomal arm. High-level amplification of clones was defined when their intensity ratios were higher than 1.0 in log<sub>2</sub> scale, and vice versa for homozygous deletion.

### Expression microarray analysis

Total RNA of the MDA-MB-231 breast cancer cell line and normal breast tissue was isolated using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instruction. The quantity and quality of extracted RNA was assessed using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). First-strand cDNA was synthesized from total RNA using SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA). We labeled 5 µg of both cDNAs using BioPrime Labeling Kit (Invitrogen, Carlsbad, CA) and co-hybridized onto the same oligoarray for array-CGH. Hybridization, slide washing and data processing were basically same as array-CGH analysis. Gene expression

analysis was performed using CGHscape software (Jeong *et al.*, 2008).

### Genomic quantitative PCR analysis

For the quantitative PCR (qPCR) validation, we chose four CNA regions and one region without CNA. A genomic region which showed no genomic alteration on array-CGH data (13q32.1) was used as internal control. Primer sequences of the target regions and internal control are shown in Supplemental Data Table S1. Genomic qPCR was performed using Mx3000P qPCR system and MxPro Version 3.00 software (Stratagene, La Jolla, CA). Twenty  $\mu$ l of real-time qPCR mixture contains 20 ng of genomic DNA, SYBR Premix Ex Taq TM II (TaKaRa Bio, Japan),  $1 \times$  ROX, and 10 pmole of primers. H6ST3 was used as internal control in each procedure. Thermal cycling conditions consisted of one cycle of 30 s at 95°C, followed by 45 cycles of 5 s at 95°C, 10 s at 55-60°C and 30 s at 72°C. After amplifying PCR reaction, melting curve analysis was performed to confirm specific amplification. Relative quantification was performed by the  $\Delta\Delta$ CT method (Livak and Schmittgen, 2001). When the mean genomic dosage ratio of the region between the MDA-MB-231 and female control DNA ( $\Delta\Delta$ CT of target and internal control) was  $> 1.2$  or  $< -0.8$ , we defined the region as copy number gain or loss, respectively.

### FISH

For FISH analysis, target-specific probes (17p12) and control-specific diploid probes (2q35) were labeled with the Dig nick translation mix kit and Biotin nick translation mix kit (Roche, Mannheim, Germany). The labeled probes were mixed with salmon sperm DNA, human Cot-1 DNA in hybridization mixture (50% formamide, 10% dextran sulfate,  $2 \times$  SSC). After denaturing at 75°C for 10 min, the labeled probes were hybridized onto the denatured chromosomes and incubated overnight at 37°C. The slides were washed in 50% formamide in  $2 \times$  SSC at 45°C for 30 min and in  $2 \times$  SSC for 5 min. After a rinse in  $4 \times$  SSC/0.1% Tween20, they were incubated under cover slips with fluorescein avidin DCS (Vector Laboratories Inc., Burlingame) and Anti-Digoxigenin-Rhodamin (Roche, Mannheim, Germany) at 37°C for 1 h. After a rinse in  $4 \times$  SSC/0.1% Tween20 for 15 min, the slides were washed sequentially and counterstained with 4',6-diamidino-2-phenylindole (Vector Laboratories Inc.). FISH images were observed with a DMRXA2 (Leica Microsystems, Wetzlar, Germany).

### Statistical analysis

Difference in FISH signal numbers between target-specific probes and control-specific probes were examined using Mann-Whitney test because the copy number is a discrete variable. For CNA-expression correlation analysis, we used an average CGH and expression level for each ten-percentile of probes. Stata version 10 (Stata, College Station, TX) was used for all statistical analyses and *P* value lower than 0.05 was considered significant.

### Supplemental data

Supplemental Data include three figures and a table and

can be found with this article online at [http://e-emm.or.kr/article/article\\_files/SP-41-7-02.pdf](http://e-emm.or.kr/article/article_files/SP-41-7-02.pdf).

### Acknowledgements

This study was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (0405-BC02-0604-0004).

### Reference

- Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* 2003;12: R145-52
- Brennan C, Zhang Y, Leo C, Feng B, Cauwels C, Aguirre AJ, Kim M, Protopopov A, Chin L. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res* 2004;64:4744-8
- Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* 2004; 57:644-6
- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006;10:529-41
- Croce CM. Oncogenes and cancer. *N Engl J Med* 2008;358: 502-11
- Dehan E, Ben-Dor A, Liao W, Lipson D, Frimer H, Rienstein S, Simansky D, Krupsky M, Yaron P, Friedman E, Rechavi G, Perlman M, Aviram-Goldring A, Izraeli S, Bittner M, Yakhini Z, Kaminski N. Chromosomal aberrations and gene expression profiles in non-small cell lung cancer. *Lung Cancer* 2007;56:175-84
- Han W, Han MR, Kang JJ, Bae JY, Lee JH, Bae YJ, Lee JE, Shin HJ, Hwang KT, Hwang SE, Kim SW, Noh DY. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* 2006;6:92
- Heidenblad M, Lindgren D, Veltman JA, Jonson T, Mahlamäki EH, Gorunova L, van Kessel AG, Schoenmakers EF, Höglund M. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. *Oncogene* 2005; 24:1794-801
- Jeong YB, Kim TM, Chung YJ. CGHscape: A software framework for the detection and visualization of copy number alterations. *Genomics & Informatics* 2008;6:126-9
- Jones AM, Mitter R, Springall R, Graham T, Winter E, Gillett C, Hanby AM, Tomlinson IP, Sawyer EJ; Phyllodes Tumour Consortium. A comprehensive genetic profile of phyllodes tumours of the breast detects important mutations, intra-



tumoral genetic heterogeneity and new genetic changes on recurrence. *J Pathol* 2008;214:533-44

Kim MY, Yim SH, Kwon MS, Kim TM, Shin SH, Kang HM, Lee C, Chung YJ. Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. *Gastroenterology* 2006; 131:1913-24

Kim TM, Yim SH, Shin SH, Xu HD, Jung YC, Park CK, Choi JY, Park WS, Kwon MS, Fiegler H, Carter NP, Rhyu MG, Chung YJ. Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q. *Int J Cancer* 2008;123:2808-15

Kim TM, Yim SH, Lee JS, Kwon MS, Ryu JW, Kang HM, Fiegler H, Carter NP, Chung YJ. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res* 2005; 11:8235-42

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* 2001;25:402-8

Lynch M. Gene duplication and evolution. *Science* 2002; 297:945-7

Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998; 20:207-11

Vincent-Salomon A, Lucchesi C, Gruel N, Raynal V, Pierron G, Goudefroye R, Reyal F, Radvanyi F, Salmon R, Thiery JP, Sastre-Garau X, Sigal-Zafrani B, Fourquet A, Delattre O; breast cancer study group of the Institut Curie. Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. *Clin Cancer Res* 2008;14:1956-65

Yim SH, Chung YJ. Current status and future clinical applications of Array based Comparative Genomic Hybridization. *Genomics & Informatics* 2004;2:113-20

Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res* 2006;34:445-50