Identification of novel deletion polymorphisms in breast cancer

AKIRA KOMATSU[1,3], KOICHI NAGASAKI[2], MINORU FUJIMORI[3], JUN AMANO[3] and YOSHIO MIKI[1,2,4]

[1]Department of Molecular Diagnosis, Cancer Institute, [2]Genome Center, Japanese Foundation for Cancer Research, Tokyo; [3]Department of Surgery, Shinshu University School of Medicine, Nagano; [4]Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

Received January 2, 2008; Accepted February 29, 2008

DOI: 10.3892/ijo_00000005

Abstract. Breast cancer is the most frequent cancer in females worldwide and it has long been known that multiple genetic rearrangements correlate with complex biology and clinical behavior. In addition, copy number variations (CNVs) of DNA sequences account for a significant proportion of normal phenotypic variation and may have an important role in human pathological variation. In this study, we carried out a high-density oligonucleotide array comparative genomic hybridization (CGH) analyses in a series of breast cancer cell lines to identify novel homozygous deletion loci. The results were confirmed by quantitative PCR (Q-PCR) and 4 genes, the REV1L, ZNF14, NPAS1 and APOBEC3B genes, were selected. Analyses of 30 microdissected human breast tumors and paired normal mammary tissue samples indicated that these homozygous deletions are small-scale deletion polymorphisms. The variation in copy number at the loci of the 4 genes in blood-derived DNA demonstrated the frequency of deletions including homozygous deletions and single copy variants to be higher in breast cancer patients than healthy females. Notably, the homozygous deletion of APOBEC3B involved part of exon 5 and seemed to be cancer-specific in some patients, indicating that this is a functionally important structural variant. These copy number changes may play an important role in breast cancer and array-CGH analyses can thus be expected to provide new insight into the genetic background of breast cancer.

Introduction

Breast cancer is the most common malignancy and the most deadly cancer among women worldwide. With advances in molecular biological analyses, a number of genes involved in the development and progression of breast cancer have been isolated and shown to have abnormalities in breast cancer patients. Breast cancer is a heterogeneous disease comprising tumors differing remarkably in clinical behavior (1) and numerous genes controlled by complex regulatory networks are thought to be involved in its development and progression.

CGH has been widely used to analyze the pattern of unbalanced genomic aberrations in several types of human cancer, including breast cancer (2-4). Recent studies have used higher resolution array-based CGH to show the enormous complexity of breast cancer cell genomes and have raised attention regarding genomic variation such as deletions, inversions and copy number variations (5). It is thought that these variations account for a significant proportion of normal phenotypic variation and have an important role in human pathological variation (6), but the structure and biological function of these variants remain largely unknown.

In the present study, we carried out a high-resolution array-based CGH analysis to scan for homozygous deletions in patients with breast cancer. We found the same chromosomal alterations described previously and further identified new small-scale homozygous deletions at 2q11.1, 19p13.3, 19q13.2 and 22q13.1 involving the REV1L, ZNF14, NPAS1 and APOBEC3B genes, respectively. We also compared the frequency of these deletions between breast cancer patients and healthy female volunteers, and determined the homozygous regions of these genes.

Materials and methods

Cell lines and clinical specimens. The following 25 human breast cancer cell lines were obtained from the American Type Culture Collection (Rockville, MD, USA): AU565, BT474, DU4475, HCC38, HCC70, HCC202, HCC1143, HCC1187, HCC1419, HCC1428, HCC1569, HCC1806, HCC1954, HCC2157, MCF7, MDA-MB-157, MDA-MB-231, MDA-MB-330, MDA-MB-361, MDA-MB-435S, MDA-MB-468, SK-BR-3, UACC812, UACC893, ZR-75-1, and were maintained under the conditions recommended by the supplier.

Samples of primary breast tumors and paired normal mammary tissues were obtained from a series of 30 patients, median age 56.9 (range 33-83) years who underwent surgery in 2003-2004 at the Japanese Foundation for Cancer

Correspondence to: Dr Yoshio Miki, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyoku, Tokyo 113-8510, Japan
E-mail: miki.mgen@mri.tmd.ac.jp

Abbreviations: CNVs, copy number variations; Q-PCR quantitative real-time PCR; CGH, comparative genomic hybridization

Key words: breast cancer, comparative genomic hybridization, copy number variation, deletion polymorphism, homozygous deletion

Table I. Forty-one homozygous deletion sites in 25 human breast cancer cell lines detected by array-CGH.

Probe no.	Chromosome position	Gene symbol	Description	Deleted cell line
A_14_P107780	chr1:193477721-193477780	CFHL3	Homo sapiens complement factor H-related 3 (CFHL3), mRNA [NM_021023]	HCC70, HCC2157
A_14_P116789	chr2:99562517-99562576	REV1L	Homo sapiens REV1-like (yeast) (REV1L), mRNA [NM_016316]	HCC202, HCC1419, HCC1428
A_14_P114658	chr2:219368746-219368805	STK36	Homo sapiens serine/threonine kinase 36 (fused homolog, *Drosophila*) (STK36), mRNA [NM_015690]	HCC1143, HCC2157
A_14_P126733	chr3:196270299-196270358	Unknown	Unknown	HCC1569, MDA-MB-157, MDA-MB-231, MDA-MB-468, UACC812
A_14_P100632	chr7:143586787-143586846	TPK1	Homo sapiens thiamin pyrophosphokinase 1 (TPK1), mRNA [NM_022445]	HCC1806, MDA-MB-330
A_14_P135989	chr8:55533892-55533937	SOX17	Homo sapiens SRY (sex determining region Y)-box 17 (SOX17), mRNA [NM_022454]	MCF7, MDA-MB-361
[a]A_14_P101143	chr9:21853204-21853263	[b]MTAP	Homo sapiens methylthioadenosine phosphorylase (MTAP), mRNA [NM_002451]	HCC38 MDA-MB-231, MCF7
[a]A_14_P121053	chr9:21957548-21957607	C9orf53	Homo sapiens susceptibility protein NSG-x mRNA, complete cds. [AF211119]	HCC38, MDA-MB-231, HCC1806, MCF7
[a]A_14_P130650	chr9:21968346-21968405	[b]CDKN2A	Homo sapiens cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4) (CDKN2A), transcript variant 4, mRNA [NM_058195]	HCC38, MDA-MB-231, HCC1806, MCF7
[a]A_14_P112983	chr9:21980522-21980581	[b]CDKN2A	The same as above	HCC38, MDA-MB-231, HCC1806
[a]A_14_P102897	chr9:21993651-21993710	[b]CDKN2B	Homo sapiens cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) (CDKN2B), transcript variant 2, mRNA [NM_078487]	HCC38, MDA-MB-231, HCC1806
[a]A_14_P103496	chr9:21998167-21998226	[b]CDKN2B	The same as above	HCC38, MDA-MB-231, HCC1806
[a]A_14_P127768	chr9:21998596-21998655	[b]CDKN2B	The same as above	HCC38, MDA-MB-231, HCC1806
[a]A_14_P118440	chr9:21998982-21999029	[b]CDKN2B	The same as above	HCC38, MDA-MB-231, HCC1806
[a]A_14_P106813	chr9:22136626-22136685	LOC401495	PREDICTED: Homo sapiens similar to ubiquitin A-52 residue ribosomal protein fusion product 1 (LOC401495), mRNA [XM_379625]	HCC38, MDA-MB-231
[a]A_14_P113515	chr9:22316165-22316224	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P129139	chr9:22690103-22690162	AX747623	Sequence 1148 from Patent EP1308459 [AX747623]	HCC38, MDA-MB-231
[a]A_14_P127990	chr9:22889584-22889643	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P116221	chr9:22992318-22992377	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P102923	chr9:23425976-23426035	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P128755	chr9:23649971-23650030	CR627240	Homo sapiens mRNA; cDNA DKFZp781D1719 (from clone DKFZp781D1719) [CR627240]	HCC38, MDA-MB-231
[a]A_14_P113744	chr9:23680273-23680332	ELAVL2	Homo sapiens ELAV (embryonic lethal, abnormal vision, *Drosophila*)-like 2 (Hu antigen B), mRNA (cDNA clone MGC:26319 IMAGE:4826082), complete cds. [BC030692]	HCC38, MDA-MB-231
[a]A_14_P106458	chr9:23779157-23779216	ELAVL2	The same as above	HCC38, MDA-MB-231
[a]A_14_P123655	chr9:24022718-24022777	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P124975	chr9:24136307-24136366	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P116740	chr9:24684752-24684811	Unknown	Unknown	HCC38, MDA-MB-231
[a]A_14_P107920	chr9:24873989-24874048	Unknown	Unknown	HCC38, MDA-MB-231
A_14_P111981	chr9:113502180-113502239	BU556905	AGENCOURT_10279007 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6591965 5', mRNA sequence [BU561092]	AU565, HCC1143, HCC1806, HCC1419, HCC1428, MDA-MB-157, SK-BR-3, UACC893

Table I. Continued.

Probe no.	Chromosome position	Gene symbol	Description	Deleted cell line
A_14_P130513	chr11:192958-193017	BET1L	Homo sapiens blocked early in transport 1 homolog (S. cerevisiae) like (BET1L), mRNA [NM_016526]	HCC1806, HCC1428
[a]A_14_P139049	chr13:47953947-47954006	[b]RB1	Homo sapiens retinoblastoma 1 (including osteosarcoma) (RB1), mRNA [NM_000321]	DU4475, HCC2157
[a]A_14_P108855	chr17:11985171-11985230	[b]MAP2K4	Homo sapiens mitogen-activated protein kinase kinase 4 (MAP2K4), mRNA [NM_003010]	DU4475, HCC1428
[a]A_14_P202132	chr17:11990074-11990133	Unknown	Unknown	DU4475, HCC1428
A_14_P138654	chr19:6407036-6407089	SLC25A23	Homo sapiens solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23 (SLC25A23), mRNA [NM_024103]	HCC1806, MDA-MB-330
A_14_P114245	chr19:19699485-19699544	ZNF14	Homo sapiens zinc finger protein 14 (KOX 6) (ZNF14), mRNA [NM_021030]	HCC1569, HCC1806, HCC70
A_14_P135824	chr19:40543992-40544045	Unknown	Unknown	MDA-MB-361, MDA-MB-435S
A_14_P125190	chr19:50174693-50174752	CLPTM1	Homo sapiens cleft lip and palate associated transmembrane protein 1 (CLPTM1), mRNA [NM_001294]	AU565, BT474, HCC1187, HCC1806, HCC1954, MDA-MB-231, SK-BR-3, UACC893
A_14_P135707	chr19:52220223-52220282	NPAS1	Homo sapiens neuronal PAS domain protein 1 (NPAS1), mRNA [NM_002517]	HCC1428, MCF7, UACC812
A_14_P125602	chr20:21635880-21635939	PAX1	Homo sapiens paired box gene 1 (PAX1), mRNA [NM_006192]	MDA-MB-361, UACC893
[a]A_14_P111340	chr22:22706765-22706820	[b]GSTT1	Homo sapiens glutathione S-transferase theta 1 (GSTT1), mRNA [NM_000853]	HCC1187, HCC1569. HCC1806, HCC1419, HCC202, HCC70, MDA-MB-231, MDA-MB-330, MDA-MB-468
A_14_P138833	chr22:37683612-37683671	APOBEC3A	Homo sapiens apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A (APOBEC3A), mRNA [NM_145699]	AU565, HCC38, MDA-MB-330, SK-BR-3
A_14_P120313	chr22:37709939-37709985	APOBEC3B	Homo sapiens apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B (APOBEC3B), mRNA [NM_004900]	AU565, HCC38, MDA-MB-330, SK-BR-3

Probe numbers are available from Agilent web resource (http://www.home.agilent.com). [a]Probes excluded from the present study; [b]known breast cancer-related gene.

Research Hospital without any preoperative therapy. Each tumor was diagnosed as invasive ductal carcinoma and was between 1.0 and 6.0 cm in maximal dimension. All samples were snap-frozen in liquid nitrogen within 1 h after surgery and stored at -80˚C. Before genomic DNA was prepared, Laser captured microdissection (LCM) using a Leica Microsystems AS LMD 600 (Leica, Wetzlar, Germany) was performed to ensure that only tumor cells were dissected. LCM was conducted in all the tumor samples.

Human blood genomic DNA samples were obtained from 50 healthy female volunteers, median age 35 (range 21-56) years and 50 pretreatment breast cancer patients, median age 51 (range 26-73) years who planned to receive neoadjuvant chemotherapy. All patients and volu-nteers were Japanese and provided written informed consent according to institutional regulations. The genomic DNA was extracted using a QIAamp™ DNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Human genomic DNA (normal female 46, XX) purchased from Promega (Madison, WI, USA) was used as a reference.

Array CGH analysis. Array-based CGH was performed with Agilent Human Genome CGH Microarray kit 44B (Agilent Technologies, CA, USA). This array is a 60-mer oligonucleotide-based microarray that contains 42,896 biological probes sourced from the National Center for Biotechnology Information. Direct labeling reactions were per-formed with 3 μg of genomic DNA and a BioPrime Array CGH Genomic Labeling Module (Invitrogen, CA, USA) according to the manufacturer's instructions in a volume of 50 μl with 60 μM

Table II. PCR primers for 16 homozygous deletion sites.

Probe Number	Chromosome position	Gene symbol	Forward primer for Q-PCR	Reverse primer for Q-PCR
A_14_P107780	chr1:193477721-193477780	CFHL3	AGAAGGAGTGCTTAGCGATGT	TATGCCTTGCGTGATCTTTC
A_14_P116789	chr2:99562517-99562576	REV1L	CCTGCAACCTAATTGCCTCT	CAACTGAAGCGTTAACATGGAC
A_14_P114658	chr2:219368746-219368805	STK36	AGTGGAAGCGGTGCACATAG	TGGGCCTCAAATAGGAACAG
A_14_P126733	chr3:196270299-196270358	Unknown	TTGGCTATCCTCACTCATGG	GTAGGCTGCAGACGTGAAAG
A_14_P100632	chr7:143586787-143586846	TPK1	CCGGTCAAGCAGAAAGTAATG	AGAGAATGAAAGGAGCAGTGG
A_14_P135989	chr8:55533892-55533937	SOX17	CCTCCGCTCGACGGCTAC	GCCAGCGTAGTCCGAGACCT
A_14_P111981	chr9:113502180-113502239	BU556905	AACCGACATGGGGTAACTGA	TCGTCTGCGTCTCTGTATTTTG
A_14_P130513	chr11:192958-193017	BET1L	GGTAACCCTGGCTCTCACAC	CAGCTCAGCCCTGTACACCT
A_14_P138654	chr19:6407036-6407089	SLC25A23	AGGCAGAGAACCCAGCAGAA	CGCAGCCCTTGAGACTTACA
A_14_P114245	chr19:19699485-19699544	ZNF14	GGCATGTATTCCCAAACACC	TGTGTGCTGTCCATTGGTTT
A_14_P135824	chr19:40543992-40544045	Unknown	TTCGGGGAGTTCGACACTAA	AAATCCACTTCCCCACACAC
A_14_P125190	chr19:50174693-50174752	CLPTM1	TGACTCTATGTGGGTCCCTAGAA	GGACATACCAACGCAGCAC
A_14_P135707	chr19:52220223-52220282	NPAS1	CGGCTTTTACAGCCTTCTTT	TGTGTCCCAGATGCTACTCC
A_14_P125602	chr20:21635880-21635939	PAX1	AGAGACCTCGGACATCTTGAAC	CAACAGGGGCTTGCATACTT
A_14_P138833	chr22:37683612-37683671	APOBEC3A	TCCTTGTGTAGCGGACCTGT	GGTCTGCATTTAGGTTCCTCTC
A_14_P120313	chr22:37709939-37709985	APOBEC3B	GGTTTCCCCTGTCTTTGTCC	GCGCTCCACCTCATAGCA

Probe numbers are available from Agilent web resource (http://www.home.agilent.com).

Cy5-dUTP (for the experimental sample) or Cy3-dUTP (for the reference, normal female 46, XX). The microarrays were hybridized in an Agilent microarray hybridization chamber, and hybridization was carried out for 40 h at 65˚C in a hybridization oven at 20 rpm. Following post-hybridization washes, the array was scanned with an Agilent Microarray G2565AA scanner. Fluorescence intensity was extracted using Agilent Feature Extraction 8.5 software.

CGH data analysis. The data obtained from the Agilent Feature Extraction software were analyzed using Agilent CGH analytics software version 3.4. For each spot on the array, $\log_2$-ratios of the Cy3-labeled experiment sample vs. the Cy5-labeled reference sample were calculated. Calculated ratios were plotted against chromosomal base-pair locations obtained from the University of California-Santa Cruz (UCSC) genome browser in May 2004 and displayed chromosomal deletions or amplifications at multiple zoom levels simultaneously.

PCR. Conventional PCR was performed with 10 ng of DNA template in a total volume of 25 μl using a Takara Ex Taq (Takara Biomedical, Japan). The reaction profile was one cycle of 94˚C for 5 min followed by 30 cycles of 94˚C for 30 sec, 60˚C for 30 sec, and 72˚C for 15 min. PCR products were assessed by ethidium bromide agarose gel electrophoresis. Primers were designed using Primer 3 to span a 100-300 bp non-repetitive region (Table I).

Q-PCR was performed with 5 ng of DNA template in a total volume of 25 μl using a POWER SYBR-Green PCR Master Mix (Applied Biosystems, CA, USA) and an ABI Prism 7900HT (Applied Biosystems). The reaction profile was one cycle of 95˚C for 10 min followed by 40 cycles of 95˚C for 15 sec and 60˚C for 1 min. At the end of the reaction, samples were subjected to a melting analysis to confirm the specificity of the amplicon. We designed the primer on including the CGH target probe or as near as possible using Primer 3 to span a 100-150 bp non-repetitive region. GAPDH was used as a reference. Primer sequences for the targets in this study are listed in Table II. Quantification is based on a standard curve made from serial dilutions of normal genomic DNA within the range 0.2-125 ng. Each experiment was performed in triplicate and means are reported. Copy number changes of target genes relative to GAPDH and the calibrator were determined by using the formula: $(S_{target}/S_{GAPDH})/(C_{target}/C_{GAPDH})$ where S_{target} and S_{GAPDH} are the mean quantity from sample DNA using the target and GAPDH, and C_{target} and C_{GAPDH} are the quantity from reference DNA using the target and GAPDH. Referring to previous studies (7-9), we separated the results into 4 groups as follows: 0 copy number group (mean calculation ratio was ≤0.2), 1 copy number group (0.3-0.7), 2 copy number group (0.8-1.2), and amplification copy number group (≥1.3). Hard to classify data were placed in the borderline group.

Results

DNA copy number variation in human breast cancer cell lines. First, we performed CGH on 25 human breast cancer cell lines and constructed array-CGH profiles. When the pattern of DNA copy number changes detected by array-CGH was visualized as a function of the nucleotide position of genes, gain or loss copy number changes were observed on all of the chromosomes. Noticeably, chromosomes 1q, 5p, 6p, 7p, 8q, 10p, 10q, 11q, 16p, 17q, 19p, 19q and 20q

contained more genes undergoing frequent copy number changes, whereas the high frequency gene deletions were observed on 8p, 9p, 17p, 18q, Xp and Xq. Several regions previously reported to have alterations in copy number in breast cancer were again identified in the current analysis, such as 1q, 5p, 7p, 8q, 11q, 17q and 20q (region of DNA copy number gains), and 9p and 17p (region of losses) (10-12).

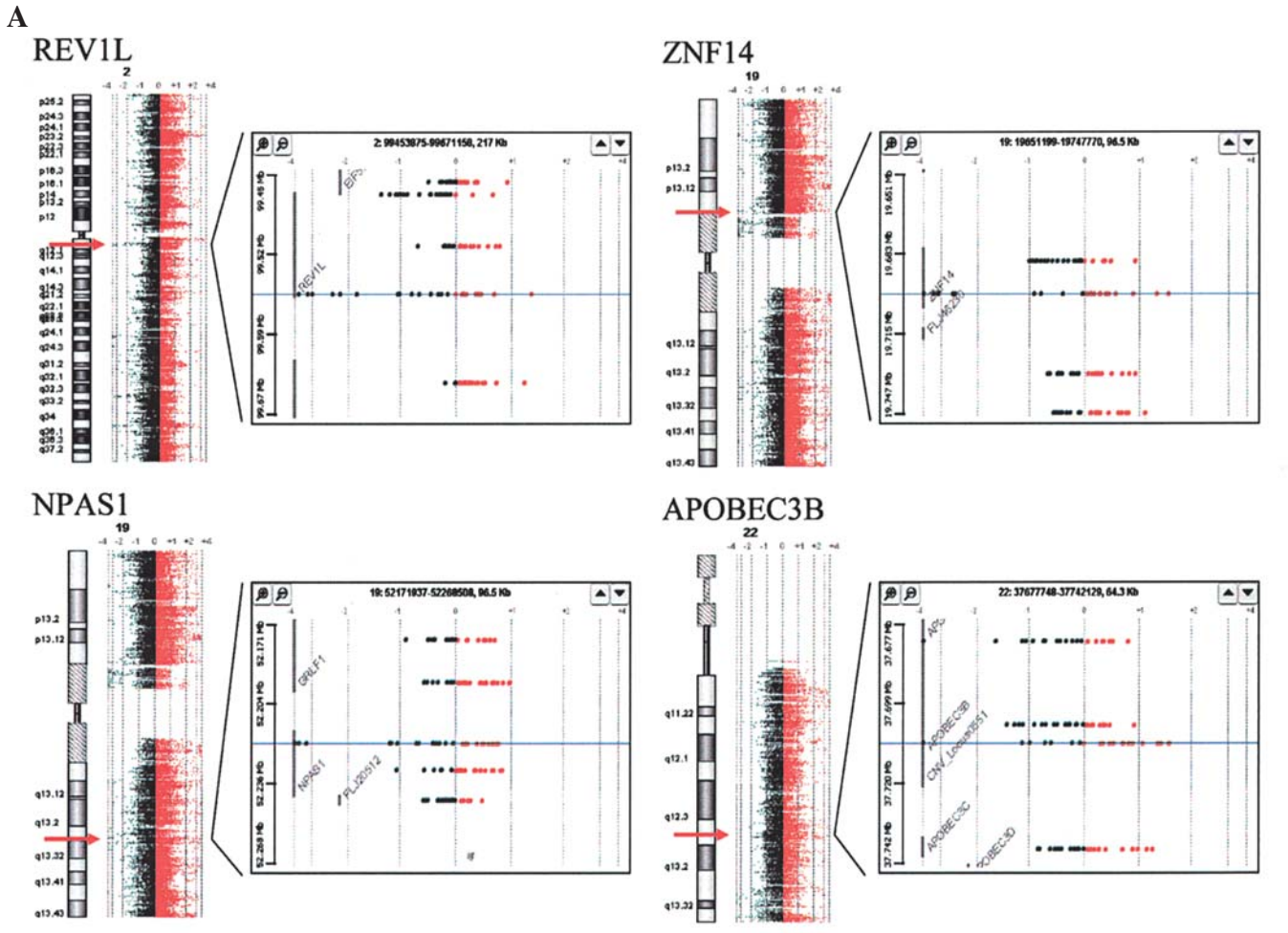
In the present study, we focused on sites of novel homozygous deletions in human breast cancer. We did not apply any filters to the CGH array data sets. A search was performed for genes that displayed a theoretical threshold of a 2.5-fold ($\log_2$ ratio = -2) decrease in the tumor sample (Cy5) compared to the reference (Cy3). However, the only loci with probes on these arrays whose measured $\log_2$ ratios were below -3 in at least 2 cell lines were selected as candidate homozygous deletion regions, because it was necessary to allow for a small margin of error and cell lines often develop artificial gene amplifications during long-term culturing. Forty-one probes ($\log_2$-ratio range -3.09 to -6.47) were selected as shown in Table I. Ten probes corresponded to known copy number changes of breast cancer-related genes, including CDKN2A (2 probes), CDKN2B (4 probes) (13), MTAP (1 probe) (14), RB1 (1 probe) (15), MAP2K4 (1 probe) (16) and GSTT1 (1 probe) (17). These 10 probes were then removed from our analyses. Furthermore, 23 probes including the 10 above were contiguous with at least 2 probes in a region of known copy number change at 9p.21.1-21.3 (3 Mb, 21 probes including CDKN2A, CDKN2B and MTAP), and 17p.11.2 (5 Kb, 2 probes including MAP2K4). Finally, 25 of 41 probes were excluded to identify sites of novel small-scale homozygous deletions. The remaining 16 probes were not contiguous and did not overlap known breast cancer-related genes (Table II). Next, Q-PCR was performed to confirm the array-CGH data for these 16 probes as candidate sites for homozygous deletions. Ten regions did not show a homozygous deletion in any of the 25 breast cancer cell lines whereas the remaining 6 regions showed homozygous deletions consistent with the CGH array data (data not shown). We selected these 6 regions for further analysis.

Homozygous deletion in primary breast cancer. To examine the prevalence of homozygous deletions in breast cancer, a Q-PCR analysis of 6 candidate sites was performed in 30 clinical samples prepared from microdissected primary breast cancer tissues. The 19q13.2 region including the NPAS1 gene was identified as the most frequent target of homozygous deletions (11/30, 36.7%). Five tumors (5/30, 16.7%) had homozygous deletions of 22q13.1 including the APOBEC3B gene. Similarly, 3 tumors (3/30, 10%) and 2 tumors (2/30, 6.7%) had homozygous deletions of 2q11.1 including the REV1L gene and 19p13.3 including the ZNF14 gene, respectively. One copy loss variants were also observed: REV1L in 9 tumors (30%), ZNF14 in 7 tumors (23.3%), NPAS1 in 2 tumors (6.7%), and APOBEC3B in 9 tumors (30%) (Fig. 1). However, no tumors showed a homozygous deletion of the BET1L gene (11p15.5) or chr19: 40543992-40544045 (19p13.1) and so these 2 regions were excluded (data not shown).

Subsequently, we performed Q-PCR with 30-paired normal mammary tissues to establish whether the 4 homozygous deletions described above were specific to tumors. Homozygous deletions of the REV1L, ZNF14, or NPAS1 gene were observed in normal mammary tissues obtained from the patients who had homozygous deletions in cancer tissues as shown in Fig. 1. Additionally, single copy loss variants were found at REV1L in 13 patients (43.3%), ZNF14 in 9 patients (30%), and NPAS1 in 9 patients (30%). The loss of one copy in normal tissue was observed in the same patients that had a single copy loss in the tumor. The results suggested that these 3 regions are deletion polymorphisms that are frequently observed in breast cancer patients. On the other hand, homozygous deletions of APOBEC3B in paired normal samples were observed in 3 of 5 tumor homozygous deletion cases, but 2 cases (sample number: 8824, 8962) had one copy loss in paired normal samples. In addition, one copy loss variant was also found at APOBEC3B in 9 tumors (30%), and 3 of 9 cases (sample number: 8720, 8734, 8743) had two copy of APOBEC3B in paired normal samples. Thus, the homozygous deletion or one copy loss of APOBEC3B seemed to be tumor-specific in some patients. These findings indicate that the APOBEC3B gene had both a deletion polymorphism and a specific deletion copy number change for breast cancers.

Frequency of deletion polymorphisms in breast cancer patients and healthy volunteers. To determine the incidence of homozygous deletions of the APOBEC3B, REV1L, ZNF14, and NPAS1 genes, we carried out Q-PCR with 50 pre-treatment breast cancer patients and 50 healthy female volunteers whose genomic DNA samples were prepared from peripheral blood. Homozygous deletions of the APOBEC3B gene were observed in 7 of the breast cancer patients (14%) and 2 (4%) of the volunteers (Fig. 2). Homozygous deletions of the other three genes, the REV1L, ZNF14, and NPAS1 genes, were detected in 4 (8%), 2 (4%) and 9 (18%) of the patients and in 4 (8%), 1 (2%) and 14 (28%) of the volunteers, respectively. Deletions, especially of the APOBEC3B gene, were more frequent among the breast cancer patients than the volunteers [odds ratio (breast cancer patient/healthy group) = 3.906, P=0.0806, 95% CI, 0.77-19.83]. The deletion polymorphisms (including one copy loss variants) of these 4 genes were slightly more frequent in breast cancer patients than healthy females except for NPAS1 (Fig. 2). In a series of 80 breast cancer samples (30 postoperative breast cancer patients and 50 preoperative breast cancer patients), no significant association between the influence of deletion polymorphisms and clinicopathological features including clinical stage, estrogen and progesterone receptor status, HER2 status, family history, and distant metastasis was found.

Detection of small-scale homozygous deletions. To determine the homozygous region of 4 genes, DNA isolated from blood samples was analyzed. The PCR analyses against genomic DNA using contiguous primers including the CGH target probe revealed that the homozygous deletions of REV1L, ZNF14, NPAS1 and APOBEC3B genes were >1.9, 7.5, 2.2 and 4 kb in size, respectively (Fig. 3). Although the homozygous deletions of all these genes except APOBEC3B did not include an exon domain, homozygous


A
REV1L
ZNF14
NPAS1
APOBEC3B
B
Breast cancer
Normal tissue
REV1L
ZNF14
NPAS1
APOBEC3B
Ratio (sample / reference)

C

REV1L

sample name	Breast cancer	Normal tissue
8880	0	0
8865	0	0
8758	0	0
8721	1	1
8734	1	1
8743	1	1
8794	1	1
8881	1	1
8883	1	1
8938	1	1
8962	1	1
8932	1	1
8710	1<border<2	1
8793	1<border<2	1
8824	1<border<2	1
8897	1<border<2	1
8704	2	2
8711	2	2
8720	2	2
8742	2	2
8771	2	2
8872	2	2
8877	2	2
8955	2	2
8843		2
8722		2
8964		2
8833		2
8894		2
8902		2<border<3
Reference	2	2

ZNF14

sample name	Breast cancer	Normal tissue
8932	0	0
8938	0	0
8734	0<border<1	1
8902	0<border<1	1
8720	1	1
8742	1	1
8794	1	1
8865	1	1<border<2
8883	1	1
8897	1	1
8955	1	1
8824	1<border<2	1
8877	1<border<2	2
8771	2	1<border<2
8704	2	2
8710	2	2
8711	2	2
8721	2	2
8722	2	2
8743	2	2
8758	2	2
8793	2	2
8833	2	2
8872	2	2
8880	2	2
8894	2	2
8962	2	2
8964	2	2
8843	2<border<3	2
8881		2
Reference	2	2

NPAS1

sample name	Breast cancer	Normal tissue
8720	0	0
8742	0	0
8743	0	0
8758	0	0
8794	0	0
8824	0	0
8877	0	0
8883	0	0
8894	0	0
8932	0	0
8964	0	0
8734	1	1
8938	1	1
8704	1<border<2	1<border<2
8771	1<border<2	1
8721	2	1
8793	2	1
8872	2	1
8881	2	1
8955	2	1
8962	2	1
8722	2	1<border<2
8865	2	2
8902	2<border<3	1<border<2
8710		2
8711		2
8833		2
8843		2
8880		2
8897		2
Reference	2	2

APOBEC3B

sample name	Breast cancer	Normal tissue
8894	0	0
8964	0	0
8711	0	0
8824	0	1
8962	0	1
8771	1	1
8742	1	1
8794	1	1
8704	1	1
8793	1	1
8938	1	1
8720	1	2
8734	1	2
8743	1	2
8880	1<border<2	2
8758	1<border<2	2
8710	2	2
8721	2	2
8722	2	2
8833	2	2
8865	2	1<border<2
8872	2	2
8877	2	2
8881	2	2
8897	2	2
8932	2	2
8955	2	2
8883	2<border<3	2
8843	2<border<3	2
8902		2
Reference	2	2

Figure 1. Detection of a small-scale homozygous deletion locus in breast cancer. (A) Detection of a small-scale homozygous deletion locus in breast cancer cell lines. The Log_2 ratios of 4 homozygous deletion sites (REV1L, NPAS1, ZNF14 and APOBEC3B) are plotted as a function of chromosomal position for hybridization of the 25 breast cancer cell lines to the CGH array. The locations of each probe are indicated by the red arrow on the chromosome and by the blue line in the enlarged view. (B) Validation of copy number changes in breast cancer patients by Q-PCR. Tumor and matched normal genomic DNA samples from 30 breast cancer patients were used to quantify the amplification of 4 candidate genes. Results are expressed as the fold-increase in the product relative to the reference DNA. Detailed criteria are described in Materials and methods. Each group is shown as follows: 0 copy number group (red or no bar), 1 copy number group (pink bar), 2 copy number group (yellow bar), amplification copy number group (orange bar) and border line group (gray bar). (C) Results are expressed as the relationship in copy number change between tumor and matched normal genomic DNA for 4 homozygous deletions. Boxes represent copy number. Each color in this figure corresponds to colors in B.

deletions of APOBEC3B were observed in part of exon 5. RT-PCR targeting exon 5 demonstrated that RNA of APOBEC3B was expressed in controls (healthy volunteers, wild-type), but was not detected in breast cancer patients that showed homozygous deletions (data not shown).

Discussion

CNVs account for a significant proportion of normal phenotypic variation and may have an important role in human pathological variation (6). CNVs can be simple in structure, such as tandem duplications, or may involve complex gains or losses of homologous sequences at multiple sites in the genome (18). According to recent studies using BAC-array and SNP-array technology, CNVs ranging in size from kilo bases to mega bases are more common in the human genome than previously thought (5,19), and several databases of deletions and amplifications are already available (20,21). On the other hand, it has long been known also that chromosomal deletions can lead to genomic disorders (22), though it was thought that chromosomal deletions causing severe disease were rare in populations. However, recent study has indicated that more benign deletions are widespread in the human genome (20), and several polymorphic deletions are related to carcinogenesis (23).

In the present study, 4 novel small-scale homozygous deletion loci were identified in breast cancer. We performed a genome-wide analysis of changes in DNA copy number, especially sites of homozygous deletions, in breast cancer, by way of oligo array-based CGH, on 44B microarrays. Scanning for homozygous deletions was undertaken in 25 breast cancer cell lines. The agilent 44B array interrogates over

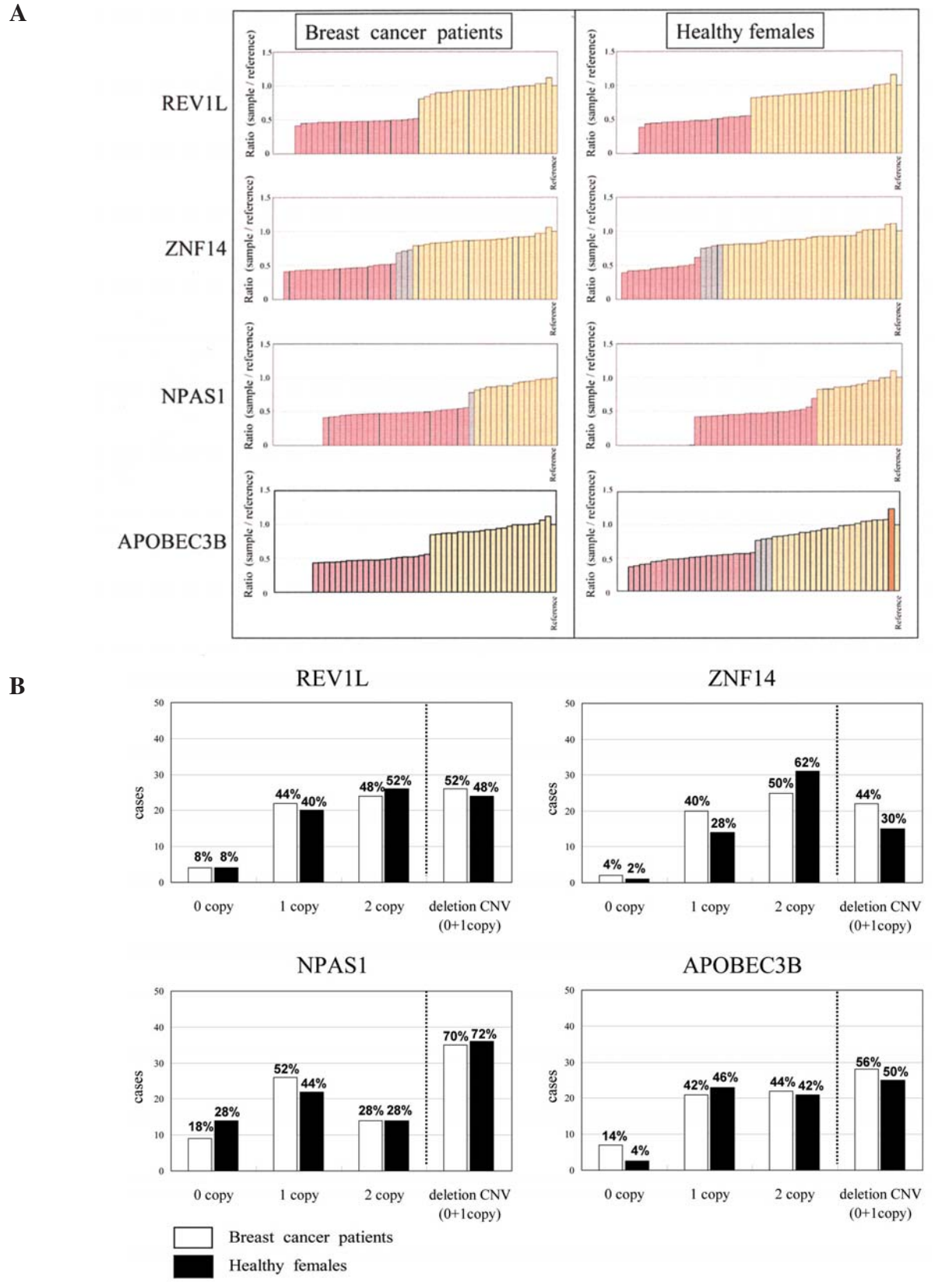


Figure 2. Frequency of deletion copy number changes in breast cancer patients and healthy volunteers. Genomic DNA samples of 50 pretreatment breast cancer patients and 50 healthy female volunteers were used to quantify the amplification of 4 genes. (A) Results are expressed as the fold-increase in the product relative to the reference DNA. Detailed criteria and colors are the same as in Fig. 1B. (B) Results are expressed as the relationship in copy number alteration between cancer patients and healthy females (except for border-line cases). The frequency of copy number alterations for each gene (frequency (%) = each cases/50) in breast cancer patients (indicated by white bars) and healthy females (black bars) is shown.

43,000 coding and non-coding human sequences. The average probe spatial resolution is approximately ~35 kb. The size of a microdeletion could be determined up to approximately 70 kb. We did not apply any filters to the CGH

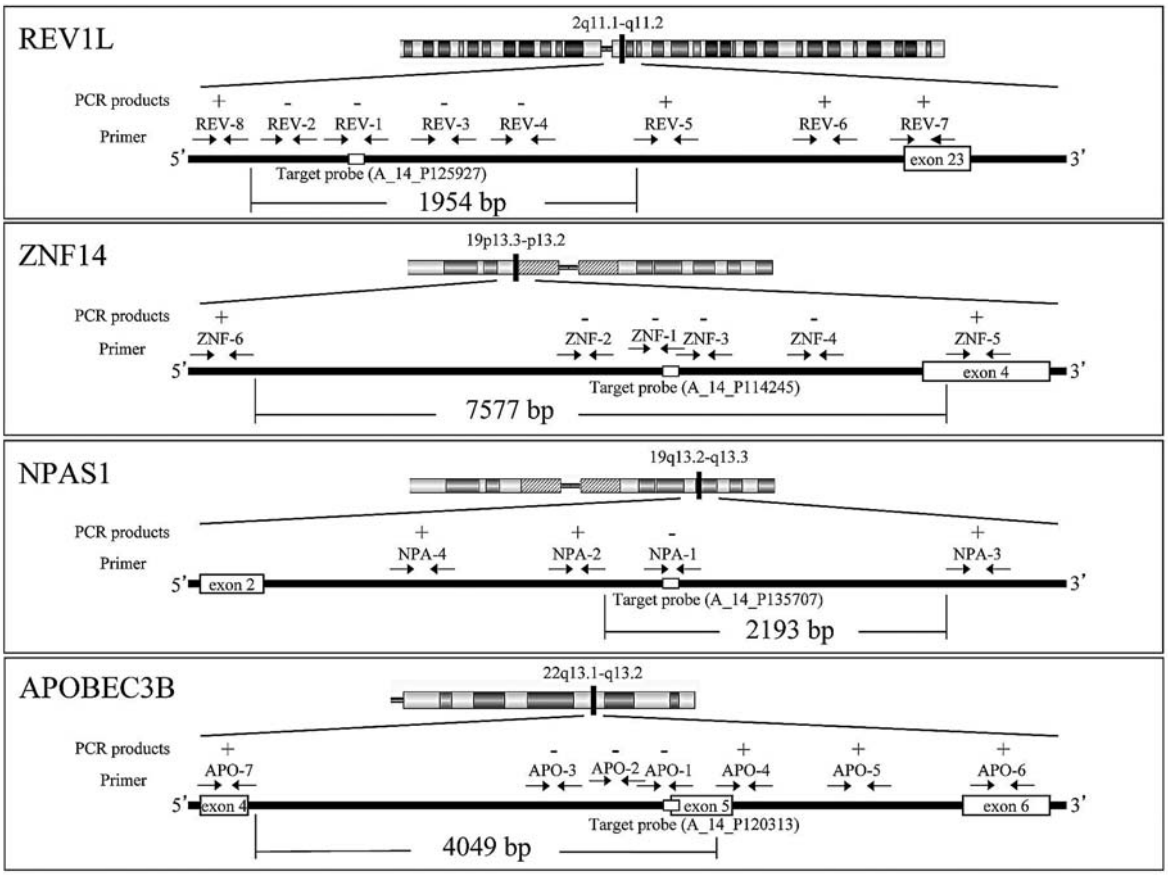


Figure 3. Schematic deletion map of the REV1L, NPAS1, ZNF14 and APOBEC3B loci. Locations of each probe for array CGH are indicated by black bars on each chromosome and by white boxes in the enlarged genomic sequence. PCR was designed to span each region (shown as a black arrow). Each primer name is shown above the arrow. Results of the PCR analysis for each locus are shown by '+' (retention) and '-' (homozygous deletion).

array data sets to identify candidate deletion loci, because the detection of microdeletions enables the identification of novel tumor suppressor genes, and biomarkers associated with cancer. Several genes previously shown to have homozygous deletions were also identified in the current study, including the CDKN2A, CDKN2B, MTAP, RB1, MAP2K4 and GSTT1 genes whose loci were detected as regions containing at least 2 consecutive probes. However, 4 homozygous deletion loci containing the REV1L, ZNF14, NPAS1 and APOBEC3B genes were recognized by one probe that was not contiguous. We have further demonstrated that the range of homozygous deletions of these 4 genes was >8 kb. This led us to consider the possibility that high-resolution CGH array technology allows the detection of novel small-scale deletion loci of breast cancer-related genes.

The analysis of breast cancer patients showed that homozygous deletions of REV1L, ZNF14, and NPAS1 were observed not only in microdissected tumor tissues but also in matched normal tissues. It was suggested that these homozygous deletions are polymorphisms. However, homozygous deletions of APOBEC3B showed tumor specificity in some cases. These findings suggested that the locus of the APOBEC3B gene had both deletion polymorphisms (copy number polymorphisms) and tumor-specific deletion copy number changes for breast cancer.

APOBEC3B is a member of the cytidine deaminase family. It is thought that APOBEC3B proteins may be RNA-editing enzymes and have roles in growth or cell cycle control (24,25). A large deletion polymorphism including the APOBEC3 gene cluster was recently identified. Kidd *et al* demonstrated a large deletion of 29.5 kb that removes the APOBEC3A and APOBEC3B genes (26). However, our results demonstrated that the homozygous deletion was >4 kb long, occurring within the APOBEC3B gene. Although this small-scale deletion was included in their large-scale polymorphism, our findings suggest that it is a novel small-scale deletion in the APOBEC3B gene.

It is worth noting that the homozygous deletion of APOBEC3B involved part of exon 5 and seemed to be tumor-specific in some patients. Furthermore, no transcription of APOBEC3B in breast cancer cells with a homozygous deletion at the APOBEC3B locus was detected (data not shown). The results suggest that this is a functionally important structural variant. There have been several reports that APOBEC3B played a role in innate cellular immunity inhibiting retroviral infections (such as HIV-1 and Mouse mammary tumor virus), the propagation of hepatitis B virus, and the retrotransposition of endogenous elements (27-30). Certainly, the role of APOBEC3B in breast cancer warrants further study.

The copy number variations at the loci of these 4 genes detected in blood-derived DNA by QPCR demonstrated that the frequency of deletions including homozygous deletions and single copy loss variants was slightly higher in breast

cancer patients than healthy females except for NPAS1. Although the association between deletions of these genes and breast cancer risk needs further analysis, the identification and investigation of such small-scale deletions using the present approach may help to elucidate their effect on disease susceptibility.

In conclusion, we subjected breast cancer samples to an array-CGH analysis and found 4 regions of small-scale deletion polymorphisms. Array-CGH can thus be expected to provide new insight into the genetic background of breast cancer.

Acknowledgements

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Simpson PT, Reis-Filho JS, Gale T and Lakhani SR: Molecular evolution of breast cancer. J Pathol 205: 248-254, 2005.
2. Roylance R, Gorman P, Harris W, *et al*: Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. Cancer Res 59: 1433-1436, 1999.
3. Albertson DG: Profiling breast cancer by array CGH. Breast Cancer Res Treat 78: 289-298, 2003.
4. Reis-Filho JS, Simpson PT, Gale T and Lakhani SR: The molecular genetics of breast cancer: the contribution of comparative genomic hybridization. Pathol Res Pract 201: 713-725, 2005.
5. Redon R, Ishikawa S, Fitch KR, *et al*: Global variation in copy number in the human genome. Nature 23: 444-454, 2006.
6. Tuzun E, Sharp AJ, Bailey JA, *et al*: Fine-scale structural variation of the human genome. Nat Genet 37: 727-732, 2005.
7. Berggren P, Kumar R, Sakano S, *et al*: Detecting homozygous deletions in the CDKN2A(p16(INK4a))/ARFp14(ARF)) gene in urinary bladder cancer using real-time quantitative PCR. Clin Cancer Res 9: 235-242, 2003.
8. Zhao X, Li C, Paez JG, *et al*: An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res 64: 3060-3071, 2004.
9. Aveyard JS and Knowles MA: Measurement of relative copy number of CDKN2A/ARF and CDKN2B in bladder cancer by real-time quantitative PCR and multiplex ligation-dependent probe amplification. J Mol Diagn 6: 356-365, 2004.
10. Orsetti B, Nugoli M, Cervera N, *et al*: Genetic profiling of chromosome 1 in breast cancer: mapping of regions of gains and losses and identification of candidate genes on 1q. Br J Cancer 95: 1439-1447, 2006.
11. Pinto AE, Roque L, Rodrigues R, Andre S and Soares J: Frequent 7q gains in flow cytometric multiploid/hypertetraploid breast carcinomas: a study of chromosome imbalances by comparative genomic hybridisation. J Clin Pathol 59: 367-372, 2006.
12. Nessling M, Richter K, Schwaenen C, *et al*: Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. Cancer Res 65: 439-447, 2005.
13. Musgrove EA, Lilischkis R, Cornish AL, *et al*: Expression of the cyclin-dependent kinase inhibitors p16INK4, p15INK4B and p21WAF1/CIP1 in human breast cancer. Int J Cancer 63: 584-591, 1995.
14. Christopher SA, Diegelman P, Porter CW and Kruger WD: Methylthioadenosine phosphorylase, a gene frequently codeleted with p16(cdkN2a/ARF), acts as a tumor suppressor in a breast cancer cell line. Cancer Res 62: 6639-6644, 2002.
15. Fung YK and T'Ang A: The role of the retinoblastoma gene in breast cancer development. Cancer Treat Res 61: 59-68, 1992.
16. Su GH, Song JJ, Repasky EA, Schutte M and Kern SE: Mutation rate of MAP2K4/MKK4 in breast carcinoma. Hum Mutat 19: 81, 2002.
17. Gudmundsdottir K, Tryggvadottir L and Eyfjord JE: GSTM1, GSTT1 and GSTP1 genotypes in relation to breast cancer risk and frequency of mutations in the p53 gene. Cancer Epidemiol Biomarkers Prev 10: 1169-1173, 2001.
18. Feuk L, Carson AR and Scherer SW: Structural variation in the human genome. Nat Rev Genet 7: 85-97, 2006.
19. Sebat J, Lakshmi B, Troge J, *et al*: Large-scale copy number polymorphism in the human genome. Science 305: 525-528, 2004.
20. Iafrate AJ, Feuk L, Rivera MN, *et al*: Detection of large-scale variation in the human genome. Nat Genet 36: 949-951, 2004.
21. Sharp AJ, Locke DP, McGrath SD, *et al*: Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77: 78-88, 2005.
22. Conrad DF, Andrews TD, Carter NP, Hurles ME and Pritchard JK: A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75-81, 2006.
23. Hicks J, Krasnitz A, Lakshmi B, *et al*: Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res 16: 1465-1479, 2006.
24. Jarmuz A, Chester A, Bayliss J, Gisbourne J and Dunham I: An anthropoid-specific locus of orphan C to U RNA-editing enzymes on Chromosome 22. Genomics 79: 285-296, 2002.
25. Conticello SG, Thomas CJ, Petersen-Mahrt SK and Neuberger MS: Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. Mol Biol Evol 22: 367-377, 2005.
26. Kidd JM, Newman TL, Tuzun E, Kaul R and Eichler EE: Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet 3: 583-592, 2007.
27. Doehle BP, Schafer A and Cullen BR: Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology 339: 281-288, 2005.
28. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK and Cullen BR: APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res 34: 89-95, 2006.
29. Stenglein MD and Harris RS: APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 281: 16837-16841, 2006.
30. Okeoma CM, Lovsin N, Peterlin BM and Ross SR: APOBEC3 inhibits mouse mammary tumour virus replication *in vivo*. Nature 445: 927-930, 2007.