# Integrated analysis of copy number and loss of heterozygosity in primary breast carcinomas using high-density SNP array

HO CHING CHING[1], RAKESH NAIDU[2], MUN KEIN SEONG[1,4],
YIP CHENG HAR[3,4] and NUR AISHAH MOHD TAIB[3,4]

[1]Department of Pathology, Faculty of Medicine, University of Malaya,
50603 Kuala Lumpur; [2]School of Medicine and Health Sciences, Monash University
Sunway Campus, Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor Darul Ehsan;
[3]Department of Surgery, Faculty of Medicine; [4]University Malaya Medical Center,
University of Malaya, 50603 Kuala Lumpur, Malaysia

**Abstract.** Breast cancer is a heterogeneous disease, marked by extensive chromosomal aberrations. In this study, we aimed to explicate the underlying chromosomal copy number (CN) alterations and loss of heterozygosity (LOH) implicated in a cohort of Malaysian hospital-based primary breast carcinoma samples using a single nucleotide polymorphism (SNP) array platform. The analysis was conducted by hybridizing the extracted DNA of 70 primary breast carcinomas and 37 normal peripheral blood samples to the Affymetrix 250K *Sty* SNP arrays. Locus-specific CN aberrations and LOH were statistically summarized using the binary segmentation algorithm and hidden Markov model. Selected genes from the SNP array analysis were also validated using quantitative real-time PCR. The merging of CN and LOH data fabricated distinctive integrated alteration profiles, which were comprised of finely demarcated minimal sites of aberrations. The most prevalent gains (≥30%) were detected at the 8q arm: 8q23.1, 8q23.3, 8q24.11, 8q24.13, 8q24.21, 8q24.22, 8q24.23 and 8q24.3, whilst the most ubiquitous losses (≥20%) were noted at the 8p12, 8p21.1, 8p21.2, 8p21.1-p21.2, 8p21.3, 8p22, 8p23.1, 8p23.1-p23.2, 8p23.3, 17p11.2, 17p12, 17p11.2-p12, 17p13.1 and 17p13.2 regions. Copy-neutral LOH was characterized as the most prevailing LOH event, in which the most frequent distributions (≥30%) were revealed at 3p21.31, 5q33.2, 12q24.12, 12q24.12-q24.13 and 14q23.1. These findings offer comprehensive genome-wide views on breast cancer genomic changes, where the most recurrent gain, loss and copy-neutral LOH events were harboured within the 8q24.21, 8p21.1 and 14q23.1 loci, respectively. This will facilitate the uncovering of true driver genes pertinent to breast cancer biology and the development of prospective therapeutics.

## Introduction

Breast cancer is a heterogeneous disease characterized by a wide array of clinical behaviours. These are underpinned by widespread genomic instability, extending from dimunitive point mutations to gross chromosomal-based alterations. Accordingly, a chromosomal copy number (CN) gain can initiate the gain of oncogenic function, whilst the loss of heterozygosity (LOH) underlined by the deletion and mutational inactivation of the retained allele, can lead to the loss of function of tumour suppressor genes (TSGs) (1). Regions of recurrent chromosomal aberrations and LOH could therefore implicate the underlying of putative or known TSGs and oncogenes. Additionally, some of these aberrations have also been linked with biological parameters and states of disease progression (1,2). Of note, the harboured genes in regions of allelic loss could also pose potential predictive power and impart important information for therapeutic decisions amongst difficult cases of recurring and advanced breast carcinoma (3).

Conventionally, genome-wide detection of CN alterations and LOH were conducted via means of comparative genomic hybridization (CGH), array-CGH and widespread polymorphic microsatellite markers (4). LOH has been proposed to customarily serve as the instigation platform in searching for candidate markers relevant to the pathogenesis of breast cancer (3). While these methods permit the quantification of total genome dosage and identification of LOH events as displayed by the recurrent gains and losses within the 1q21, 8q23-24, 11q13, 17q12-21, 20q13 (2,5,6) and 1p36, 7q31, 8p21, 13q14, 16q24, 17p13 loci (7), respectively, the paucity of available markers and reduced spatial resolution of these approaches (4) has impeded the efforts in delineating the fine altered genomic boundaries and the revelation of true driver genes from bystander entities. The introduction of single nucleotide

*Correspondence to:* Dr Rakesh Naidu, School of Medicine and Health Sciences, Monash University Sunway Campus, Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor Darul Ehsan, Malaysia
E-mail: kdrakeshna@hotmail.com

polymorphism (SNP)-based arrays has offered superior investigative prospects for cancer studies that combine both the benefits of the CGH and LOH techniques (8). Most crucially, the advent of the SNP-based technology permits high-resolution analysis of all chromosomes with denser marker spacing (9), and enables the concurrent detection of genotypic alterations with quantification of DNA CN changes in individual patients (8). Accordingly, this allows for the discovery of a previously underappreciated copy-neutral LOH (CN-LOH) event which was manifested via mitotic nondisjunction, somatic recombination, or a deletion followed by a reduplication of the retained allele (10). Further elucidation of driver genes and mechanisms underlying the LOH course can also be attained as previously described (9).

In Malaysia, breast carcinoma constitutes for 31% of newly diagnosed female patients, making it the most prevalent female cancer across all age strata and major ethnic groups (11). There is an impartially higher incidence of early-onset in this population (<50 years old) and the overall age-standardized rate is lower (46.2 per 100,000) compared to the developed regions of the West (12). Thus, to identify the underlying genome-wide chromosomal aberrations implicated in primary breast carcinomas, we incorporated high-density Affymetrix 250K *Sty* SNP arrays to profile a cohort of primary breast tumours. By overlapping both the CN and LOH alteration data as acquired, we identified 5 distinctive integrated genomic aberration profiles. Many of the detected altered regions were smaller (smallest region detected, ~9 kb), and comprised of genes that could play pivotal roles in breast carcinogenesis. Despite a few previous studies of similar nature using SNP arrays (13-16), to the best of our knowledge, this is the first study that generates and describes the integrated alteration profiles in breast cancer in a hospital-based Malaysian population.

## Materials and methods

*Patients and samples collection*. This study was approved by the Ethics Committee of University Malaya Medical Centre (UMMC), Malaysia. Written informed consent was obtained from the involved patients prior to study commencement. A panel of 70 fresh primary breast carcinoma tissue samples (67 infiltrating ductal carcinomas, 2 ductal carcinomas *in situ* and 1 invasive papillary carcinoma) was acquired from patients who had undergone mastectomy procedures. Prior to sample collection, the tumour was grossly examined and the sample was taken from an area determined to be 'almost pure' tumour tissue by an experienced pathologist at the Department of Pathology, UMMC. Once the hematoxylin and eosin slides were prepared, the specimen was examined again and confirmed to be at least 85% epithelial tumour before being included in the study. All of the fresh tissues were collected immediately after surgery, snap-frozen, and stored at -80°C until further use. Peripheral blood samples were obtained from 37 normal volunteers with no history of any malignancies. Accordingly, the normal DNA samples were obtained from a similar population to the test samples. The peripheral blood samples were stored at 4°C immediately after collection. The clinicopathological details of the patients are summarized in Table I.

Table I. Summarized clinicopathological characteristics of clinical samples.

| Clinicopathological parameters | Number of cases (%) (n=70) |
|---|---|
| Age at diagnosis | |
| <50 years old | 17 (24) |
| ≥50 years old | 53 (76) |
| Histological grade | |
| Grade 1 | 1 (2) |
| Grade 2 | 35 (50) |
| Grade 3 | 31 (44) |
| Not available[a] | 3 (4) |
| Estrogen receptor status | |
| Positive | 38 (54) |
| Negative | 32 (46) |
| Progesterone receptor status | |
| Positive | 34 (49) |
| Negative | 36 (51) |
| Lymph node status | |
| Negative | 24 (35) |
| Positive | 43 (61) |
| Not available[a] | 3 (4) |

[a]Clinical data not available.

*DNA extraction for SNP array analysis*. Genomic DNA was extracted from the fresh-frozen tumour tissues and peripheral blood samples using QIAamp DNA Mini Kits and Gentra Puregene Blood Kits (Qiagen, Valencia, CA, USA), respectively. This was carried out as per the manufacturer's instructions with minor modifications. The acquired DNA yields were quantified and the purity of the extracted DNA was within the $A_{260}/A_{280}$ ratio range of 1.7 to 1.9. The eluted DNA was kept at -20°C until further use in SNP array analysis.

*Preparation of SNP array experiments*. The Affymetrix 250K *Sty* SNP array experiments were conducted according to the manufacturer's recommendations (Affymetrix Inc., Santa Clara, CA, USA). Briefly, 250 ng of the extracted genomic DNA of tumour and normal samples were digested with *Sty 1* and ligated to a universal adaptor sequence. The ligated DNA for each sample was PCR amplified under suggested conditions. The resultant products were validated by agarose gel electrophoresis, and were purified and concentrated using the QIAquick PCR Purification kit. The purified PCR products were then fragmented and end-labeled with biotinylated ddATP. The labeled DNA was hybridized overnight to arrays containing probes for *Sty 1*-cleaved DNA. Finally, the hybridized arrays were washed, incubated and stained with streptavidin R-phycoerythrin conjugates at the assigned fluidic station. To acquire the relevant raw data for subsequent analysis, the arrays were scanned with a GeneChip Scanner 3000 (Affymetrix Inc.) as per the manufacturer's instructions. Each SNP locus
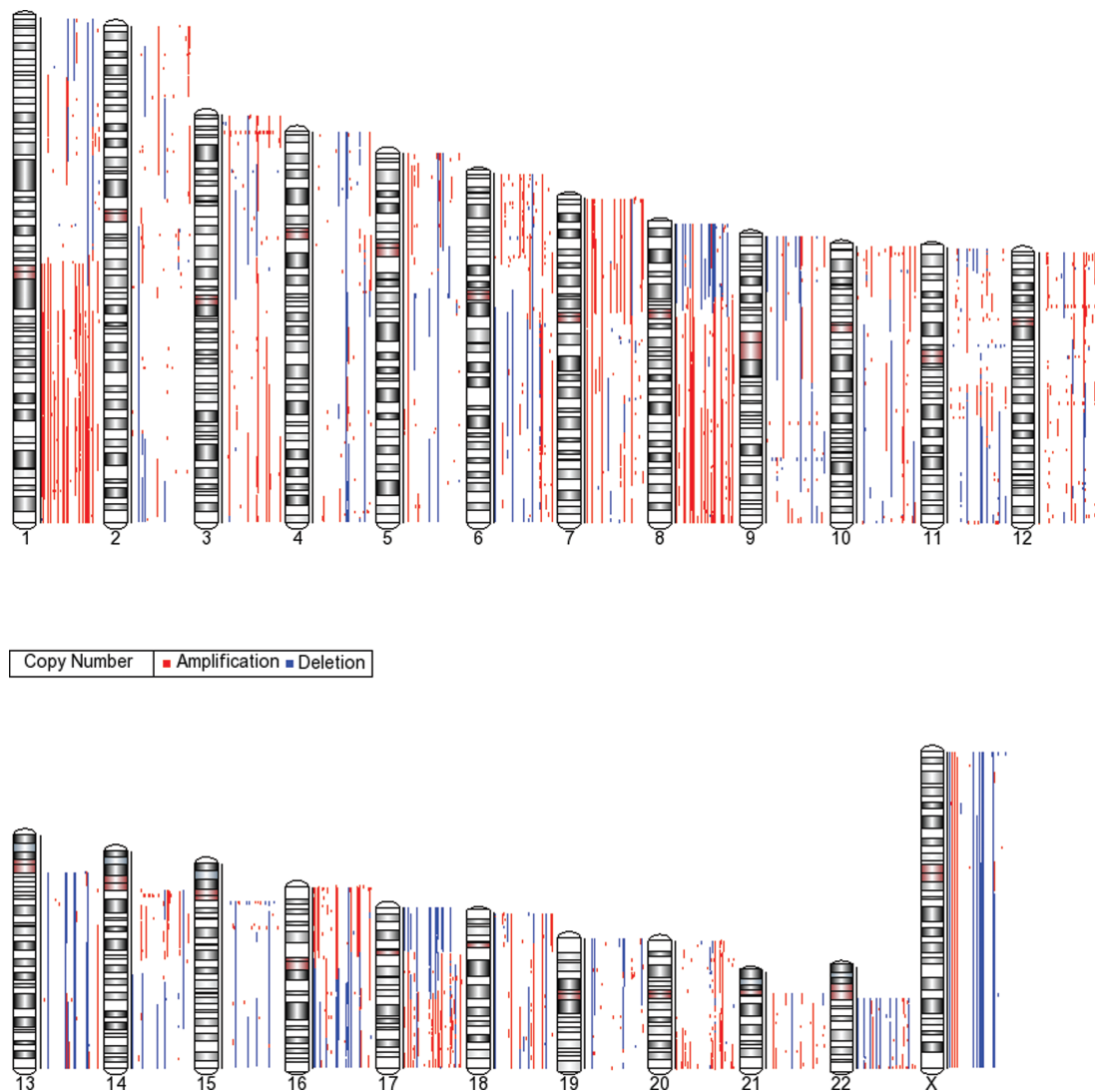
Figure 1. Genomic copy number alterations in primary breast carcinomas were generated using the Partek Genomic Suite binary segmentation algorithm. The karyoview represents the global copy number alterations which were manifested across the whole genome of the recruited 70 samples. Each columnized line denotes a single sample, in which all chromosomes were shown to display copy number aberrations. Gains are represented by red, whilst losses are represented by blue.

was interrogated by 10 probe quartets, in which each probe quartet consisted of a perfect match (PM) and mismatch (MM) probe for a given allele. On the whole, each SNP locus was denoted by 40 different 25-bp oligonucleotides. The intensities of the probe sets on the scanned images were extracted and saved in the allele intensity files (.CEL files) using GeneChip Operating Software (GCOS, Affymetrix). The genotype calls (.CHP files) were produced using the dynamic model algorithm by GeneChip Genotyping Analysis Software (GTYPE, Affymetrix) version 4.0. A stringent P-value cut-off threshold of 0.33 was applied and all the arrays were assessed for call rates based on the in-built QC metrics. Only samples with call rates of ≥93% were used for downstream analysis.

*Integration of CN change and LOH.* The Partek Genomics Suite version 6.5 (Partek Inc., St. Louis, MO, USA) was used for both CN and LOH analyses. Regions of CN abnormalities were detected using the binary segmentation algorithm. Prior to data analysis, the intensities of the individual probes of the

.CEL files were first adjusted for bias due to fragment length, and GC content using the Partek GC-wave correction which has been shown to improve CN calls (17). CN values were calculated from the summarized signal intensities of the probes, and were achieved by normalizing each sample to the set of pooled references (unpaired baseline) created in the unpaired CN workflow. The CN baseline was generated by loading in the 37 .CEL files of the normal samples and subjected to unpaired baseline creation. The .CEL files of the 70 tumour samples were loaded in a similar manner, and unpaired CN analysis with GC correction was carried out accordingly. The ensuing spreadsheet was analyzed using stringent parameters applied to the binary segmentation algorithm (deletion, P<0.0005, ≥15 markers, signal/noise ≥0.7; amplification, P<0.0002, ≥15 markers, signal/noise ≥0.3), to detect deleted and gained segments. Segments with a default mean of ≥2.3 copies were annotated as gained, and segments with a default mean of ≤1.7 copies were annotated as deleted. Amplifications were defined as segments assigned with ≥3.2
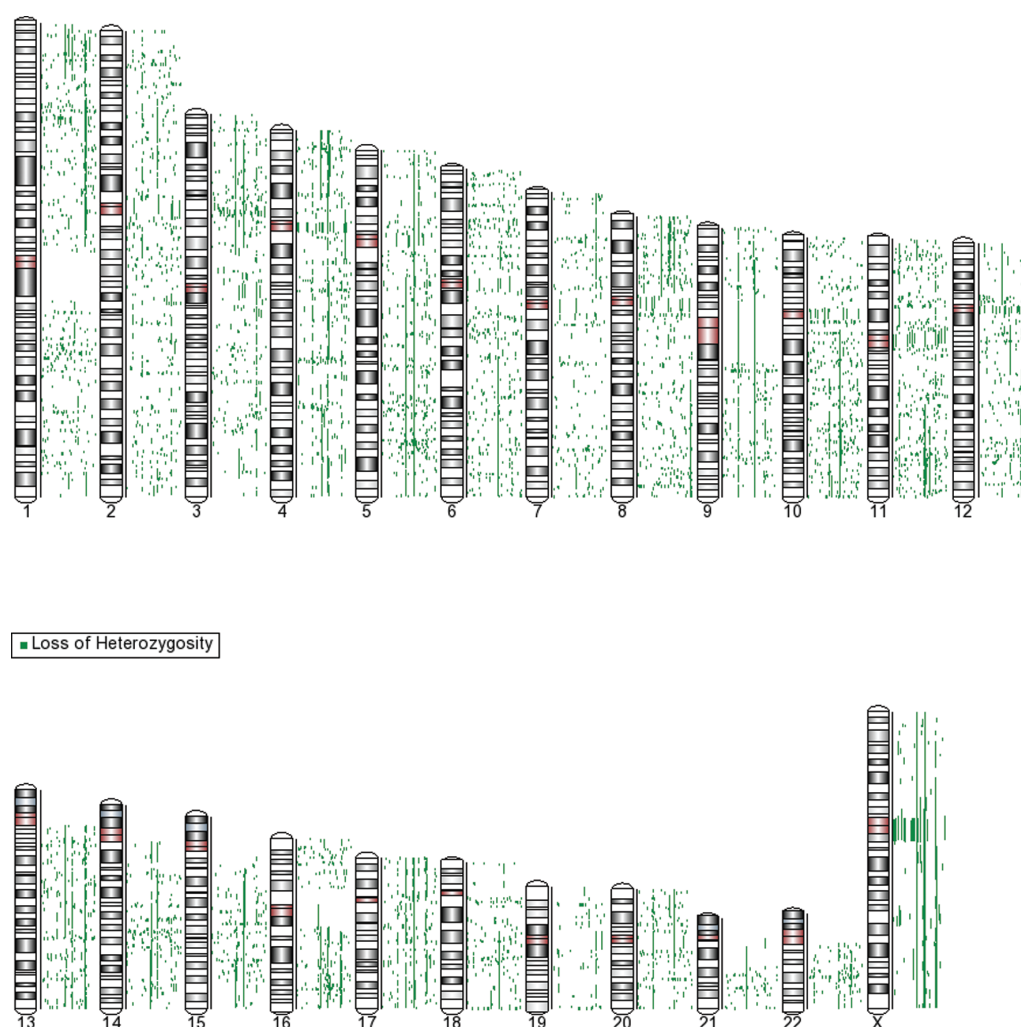
Figure 2. The plot illustrates the global LOH alterations of the 70 primary breast carcinoma samples across the whole genome. HMM was employed to detect the LOH events, in which the detected LOH states are indicated in green. In this karyoview, each columnized line signifies a single sample. Generally, all chromosomes displayed LOH aberrations.

copies, whilst high-grade amplifications were defined as segments assigned with ≥3.5 copies. Similarly, high-grade deletions were defined as segments set at a threshold value of ≤1.5 copies.

By utilizing the in-built default parameters, LOH regions were analyzed based on the hidden Markov model (HMM) within the Partek LOH workflow. The analysis was carried out using the genotyping results from the unpaired normal and tumour samples (.CHP files). The detected LOH regions were overlapped with the resultant CN regions generated to further obtain 5 integrated categories of aberrations. Subsequently, overall regions which are common over multiple samples were identified, followed by the subjection of individual filtering criteria upon the 5 integrated categories of aberrations for the acquisition of high-confidence regions of interest (amplification without LOH, n ≥11, mean copies ≥3.2; amplification with LOH, n ≥7, mean copies ≥2.7; deletion without LOH, n ≥10, mean copies ≤1.55; deletion with LOH, n ≥7, mean copies ≤1.50; CN-LOH, n ≥15, het rate ≤0.03, number of markers ≥10, where n represents the number of cases). This resulted in the identification of key altered cytobands (gained/ deleted) that were further characterized by a single or multiple

minimal common overlapping region/s (MCOR/s). Each key altered cytoband was assigned to the frequency rate of a single percentage value or a range of percentages according to the number of MCOR/s it was comprised of. Intermittently, MCORs of a cytoband could exhibit similar percentage values. Underlying genes as detected in these altered regions were annotated based on the RefFlat files from the National Centre for Biotechnology Information (NCBI) build 36.1, and the University of California, Santa Cruz (UCSC) hg18.

*Quantitative real-time PCR (qPCR).* CNs of selected genes obtained from the SNP array data analysis were validated by qPCR. The MYC, PVT1, ERRB2 and CCND1 genes were selected to represent the amplification CN, whilst the NRG1, TP53, PTPRJ and CASP1 genes were selected to represent the deletion CN. The qPCR experiments were carried out according to the manufacturer's recommendations, on a StepOnePlus™ System using TaqMan Genotyping Master Mix (Applied Biosystems, Foster City, CA, USA). The primers and TaqMan-based probes for the qPCR analysis were acquired from Applied Biosystems (TaqMan Copy Number Assays Database). Target gene CN was determined by relative quantification
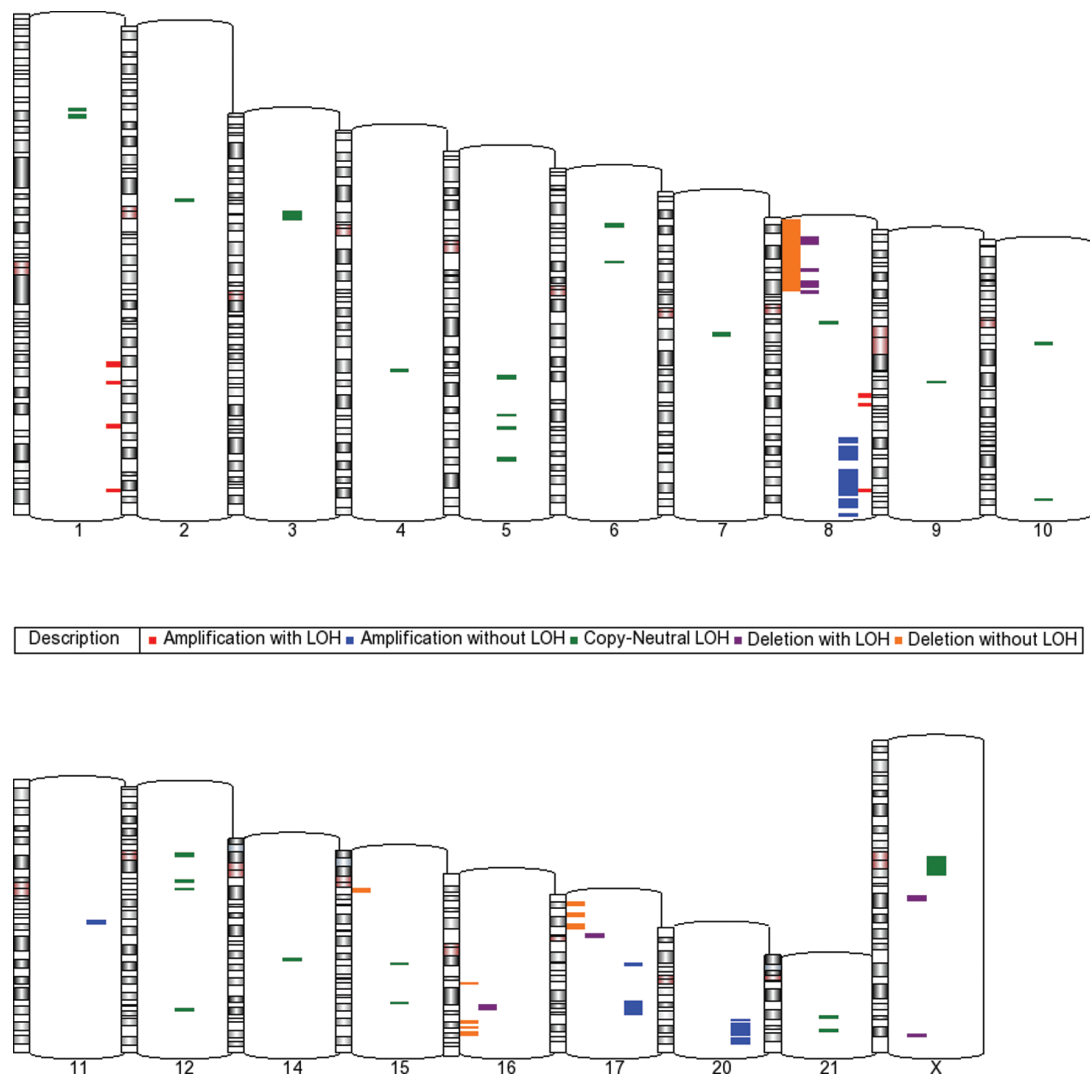
Figure 3. The integration analysis of copy number changes and LOH alterations was conducted via merging both the afore-mentioned data. This resulted in the generation of 5 integrated categories of aberrations, followed by subjection to individual filtering parameters. The plot shows the acquired high-confidence regions of interest, as distributed throughout the whole genome of the breast carcinoma samples. Accordingly, each of the integrated categories of aberration is represented as follows: Amplification without LOH (blue), amplification with LOH (red), deletion without LOH (orange), deletion with LOH (purple), and copy-neutral LOH (green).

using the comparative threshold cycle (ΔΔCt) method, with RNaseP selected as the reference gene assay. Calibrator samples from normal human genomic DNA of 5 individuals were also used as the basis for normalization with the relative target CN. A no-template control was included in all the qPCR experiments. All reactions were conducted in triplicate, under the following PCR conditions: Holding period of 10 min at 95˚C, followed by subsequent 40 thermal cycles, each of 15 sec at 95˚C and 1 min at 60˚C. Raw data obtained from the qPCR experiments were exported to CopyCaller™ Software for CN determination, and by setting RNaseP CN as 2. The analyzed CN calls were then reviewed for confidence level and z-score values.

**Results**

*Detection of overall CN and LOH aberrations.* From the CN analysis based on the stringent parameters as designated, 2,178 gained regions and 497 deleted regions were statistically identified. The composite CN profiles of the 70 study cases were generated based on the compilation of the acquired 2,675 somatic aberrations (Fig. 1). On the basis of the in-built default parameters, the Partek hidden Markov model detected 7,614 discrete LOH alterations (Fig. 2). High incidences of LOH aberrations (>20%) were distributed throughout all the autosomes and the X chromosome, except for chromosome 19. All the samples assayed showed CN changes and LOH alterations. Five integrated categories of aberrations were fabricated by overlapping both the total generated CN and LOH data, in which the overall attained and filtered high-confidence regions of interest were as portrayed in the global karyoview (Fig. 3).

*Amplification without LOH.* A total of 63 MCORs affecting the 8q (31-44%), 11q (19-23%), 17q (16-21%) and 20q (16-23%) arms were detected as regions of amplification without LOH alterations (Table II). Regions of amplification that were analyzed were defined as segments with a minimal estimated

Table II. Common chromosomal regions of amplification without LOH in primary breast carcinoma.

| Chr | Cytoband | Start | End | MCOR | Representative gene/s[a] | % of cases |
|---|---|---|---|---|---|---|
| 8 | 8q23.1 | 108344823 | 109621344 | 2 | ANGPT1, EIF3E | 31-36 |
| | 8q23.3 | 114054572 | 116645217 | 2 | CSMD3, TRPS1 | 34-41 |
| | 8q24.11 | 117993510 | 118313461 | 1 | C8orf85, SLC30A8 | 39 |
| | 8q24.13 | 123864370 | 126404360 | 7 | ZHX2, FBXO32, ANXA13, FAM91A1, TMEM65, MTSS1, NSMCE2, SQLE | 37-40 |
| | 8q24.21 | 128351093 | 129284858 | 3 | MYC, PVT1, MIR1208 | 43-44 |
| | 8q24.22 | 131797913 | 134561733 | 3 | ADCY8, LRRC6, ST3GAL1 | 40 |
| | 8q24.23 | 139371662 | 139741853 | 2 | FAM135B, COL22A1 | 33-36 |
| | 8q24.3 | 141052385 | 142253102 | 3 | EIF2C2, PTK2, DENND3 | 33-34 |
| 11 | 11q13.2-q13.3 | 69087157 | 69299379 | 1 | CCND1, FGF19, FGF4 | 21 |
| | 11q13.3 | 69299379 | 70001848 | 4 | FGF3, FADD, CTTN, PPFIA1, SHANK2 | 19-23 |
| 17 | 17q12 | 34860330 | 35233202 | 4 | MED1, PPP1R1B, STARD3, TCAP, ERBB2, GRB7, PNMT, IKZF3 | 16-17 |
| | 17q22 | 53346542 | 54812045 | 6 | SFRS1, VEZF1, EPX, LPO, MPO, MTMR4, PPM1E, TEX14, RAD51C, YPEL2 | 20-21 |
| | 17q23.1 | 54981042 | 55432943 | 3 | DHX40, CLTC, TMEM49, TUBD1, RPS6KB1,PTRH2 | 19-21 |
| | 17q23.1-q23.2 | 55541557 | 55756313 | 1 | CA4, USP32 | 21 |
| | 17q23.2 | 56653175 | 58137814 | 5 | BCAS3, BRIP1, INTS2, MED13, TLK2, MRC2 | 19-21 |
| | 17q23.3 | 58789143 | 59466014 | 3 | TANC2, MAP3K3, DDX42, FTSJ3, PSMC5, SCN4A | 19-21 |
| 20 | 20q13.13 | 48181225 | 48643301 | 2 | CEBPB, PTPN1 | 16-17 |
| | 20q13.2 | 50210482 | 52279149 | 4 | ZFP64, ZNF217, BCAS1, CYP24A1, PFDN4 | 17-21 |
| | 20q13.31 | 55077601 | 55608996 | 3 | BMP7, RBM38, PCK1 | 20-23 |
| | 20q13.32 | 56384360 | 57358466 | 3 | VAPB, GNAS, EDN3 | 21 |
| | 20q13.32 - q13.33 | 57846599 | 57932267 | 1 | SYCP2 | 23 |

MCOR, represents the number of minimal common overlapping region(s) contained within the specific cytoband region. [a]Representative genes are from the minimal sites encompassed in the region. Chr, chromosome.

CN of 3.2. Of note, the MCORs found in this category ranged from ~10 kb to ~0.8 Mb. Twenty-three regions exhibited high rates of recurrent amplifications (≥30%), as seen within the 8q arm: 8q23.1 (31-36%), 8q23.3 (34-41%), 8q24.11 (39%), 8q24.13 (37-40%), 8q24.21 (43-44%), 8q24.22 (40%), 8q24.23 (33-36%) and 8q24.3 (33-34%). Additionally, the highest number of MCORs (7 minimal sites) in this category was found to be encompassed within the 8q24.13 locus. The 3 MCORs which were denoted by MYC (8, 128351093-128853640), PVT1 (8, 129142521-129182398), and MIR1208 (8, 129182398-129284858) were encompassed within the most frequently amplified 8q24.21 amplicon. Moreover, high-grade amplification events (>3.5 copies) were seen in 16-23% of our samples, localized within the 11q13.2-q13.3 (21%), 11q13.3 (19-23%), 17q12 (16-17%), 17q23.1 (19-21%), 17q23.1-q23.2 (21%), 20q13.2 (17-21%) and 20q13.31 (20-23%) loci. The most frequent high-grade amplifications as perceived in each of these key chromosomes were 11q13.3, 23% (11, 69697090-69754235; FADD); 17q23.1, 21% (17, 54981042-55297505; DHX40, CLTC, TMEM49, TUBD1 and PTRH2); 17q23.1-q23.2, 21% (17, 55541557-55756313; CA4 and USP32); and 20q13.31, 23% (20, 55433083-55608996; PCK1).

*Deletion without LOH*. The integration analysis revealed a sum of 73 MCORs in this category, in which these regions of deletion without LOH alteration were found to be manifested within the 8p (16-23%), 15q (14%), 16q (14-19%) and 17p (20-21%) arms (Table III). The rate of recurrent losses was lower than that of amplifications, and the regions detected were defined by CN values of ≤1.55 copies. In particular, the MCORs as demarcated were noted to range from ~15 kb to ~1.2 Mb. The most frequently deleted regions were localized at the 8p21.1 locus (8, 28222208-29427118) occurring in 23% of the primary breast carcinoma samples. Other prevalent losses with rates of 16-21% were found to be linked with regions of the following cytobands: 8p12 (16-20%); 8p21.2, 8p21.1-p21.2, 8p21.3 and 8p21.3-p21.2 (16-21%); 8p22 (16-20%); 8p23.1, 8p23.2-p23.1 and 8p23.3 (16-21%); 16q22.3 (16%); 16q23.1 and 16q23.2 (17-19%); 17p11.2, 17p11.2-p12 and 17p12 (20-21%); 17p13.1 and 17p13.2 (20-21%) arms. A previously not well described region in breast cancer studies, 15q11.2 was also identified. This region spanned 2 MCORs of ~0.9 Mb (15, 19841337-19934211) and ~1.2 Mb (15, 18451755 -19646095), comprising of 11 genes that displayed high-grade deletions of <1.29 copies. Moreover, frequent high-grade

Table III. Common chromosomal regions of deletion without LOH in primary breast carcinoma.

| Chr | Cytoband | Start | End | MCOR | Representative gene/s[a] | % of cases |
|-----|----------|-------|-----|------|-------------------------|-----------|
| 8 | 8p12 | 29824801 | 33534036 | 5 | TMEM66, DCTN6, LEPROTL1, RBPMS, NRG1, C8orf41, RNF122 | 16-20 |
| | 8p21.1 | 27505194 | 29427118 | 5 | CLU, SCARA3, ELP3, FZD3, PNOC, DUSP4, EXTL3, INTS9 | 17-23 |
| | 8p21.2 | 23671390 | 27381108 | 10 | STC1, ADAM28, ADAM7, ADAMDEC1, NEFM, DOCK5, GNRH1, KCTD9, EBF2, BNIP3L, PPP2R2A, ADRA1A, PNMA2, PTK2B, CHRNA2 | 16-21 |
| | 8p21.2-p21.1 | 27381108 | 27443762 | 1 | EPHX2 | 20 |
| | 8p21.3 | 19588142 | 23365859 | 7 | LPL, LZTS1, GFRA2, DOK2, FGF17, NPM2, XPO7, BMP1, LGI3, SFTPC, BIN3, PPP3CC, EGR3, TNFRSF-genes | 17-20 |
| | 8p21.3-p21.2 | 23365859 | 23649933 | 1 | NKX3-1, SLC25A37 | 19 |
| | 8p22 | 12945618 | 18829463 | 9 | DLC1, TUSC3, MSR1, FGF20, CNOT7, ZDHHC2, PDGFRL, ASAH1, NAT1, NAT2, PSD3 | 16-20 |
| | 8p23.1 | 6711884 | 11772110 | 12 | DEFA4, DEFA6, DEFA1, DEFA3, DEFB-genes, SPAG11B, CLDN23, MSRA, SOX7, PINX1, MTMR9, BLK, GATA4, NEIL2, CTSB | 16-21 |
| | 8p23.2-p23.1 | 6069900 | 6442204 | 1 | ANGPT2, MCPH1 | 21 |
| | 8p23.3 | 438295 | 2168430 | 4 | C8orf42, ERICH1, ARHGEF10, MYOM2 | 19-21 |
| 15 | 15q11.2 | 18451755 | 19934211 | 2 | BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, NF1P1, POTEB, LOC727924, OR4M2, OR4N3P, OR4N4 | 14 |
| 16 | 16q12.2 | 54426169 | 54479842 | 1 | CES7 | 14 |
| | 16q22.3 | 73189489 | 73249954 | 1 | GLG1, RFWD3 | 16 |
| | 16q23.1 | 73343962 | 73495584 | 2 | FA2H, WDR59 | 17 |
| | 16q23.2 | 79067790 | 80061949 | 2 | CENPN, DYNLRB2, GAN | 19 |
| 17 | 17p11.2 | 16042253 | 16789645 | 1 | PIGL, TNFRSF13B, ZNF287 | 20 |
| | 17p12 | 14964753 | 15849960 | 2 | PMP22, CDRT1, TRIM16 | 21 |
| | 17p12-p11.2 | 15849960 | 16042253 | 1 | NCOR1, TTC19 | 20 |
| | 17p13.1 | 9696508 | 10783644 | 3 | GLP2R, GAS7, MYH1, MYH2, MYH3, MYH4, MYH8, SCO1 | 21 |
| | 17p13.2 | 3710323 | 5630436 | 3 | ATP2A3, P2RX1, UBE2G1, ALOX15, ARRB2, CHRNE, CXCL16, ENO3, GP1BA, KIF1C, MED11, MINK1, PELP1, PFN1, PSMB6, RNF167, SLC25A11, TM4SF5, USP6, DERL2, MIS12, NUP88 | 20-21 |

MCOR, represents the number of minimal common overlapping region(s) contained within the specific cytoband region. [a]Representative genes are from the minimal sites encompassed in the region. Chr, chromosome.

hemizygous deletions of <1.5 copies (19-21%) were also linked with the 8p arm, in which the most recurrent ones were noted in the regions: 8p23.1, 21% (8, 6711884-6755592; DEFB1); 8p23.2-p23.1, 21% (8, 6069900-6442204; ANGPT2 and MCPH1); and 8p23.3, 21% (8, 1707146-1789291; ARHGEF10). Of note, the 8p23.1 and 8p21.2 loci displayed highest number of MCORs in this category. Both regions accounted for 12 and 10 minimal sites, respectively.

*Amplification with LOH.* The analysis based on the delimited criteria resulted in 17 regions of LOH accompanied by CN

gain, notably localized within the 1q and 8q arms (Table IV). The 1q regions indicated included the 1q25.1 (10%), 1q25.3 (10%), 1q32.1 (11%) and 1q43 (10%) loci. High-grade amplifications (≥3.5 copies) were denoted by the genes implied within the 7 MCORs in all the 8q regions. The demarcated 8q sites encompassed the 8q21.2, 11-13% (REXO1L1 and ATP6V0D2); 8q21.3, 10-13% (CPNE3, WWP1, and TMEM64); and 8q24.22, 10% (SLA, TG and NDRG1) loci.

*Deletion with LOH.* Twenty-four regions of LOH resulting from CN loss were observed in the form of hemizygous CN

Table IV. Common chromosomal regions of amplification and deletion with LOH in primary breast carcinoma.

| Chr | Cytoband | Start | End | MCOR | Representative gene/s[a] | % of cases |
|---|---|---|---|---|---|---|
| **Amplification with LOH** | | | | | | |
| 1 | 1q25.1 | 171831910 | 173264373 | 6 | CENPL, DARS2, GAS5, ZBTB37, SERPINC1, RABGAP1L, GPR52, MRPS14 | 10 |
| | 1q25.3 | 181439268 | 181523841 | 1 | LAMC2, NMNAT2 | 10 |
| | 1q32.1 | 202899104 | 203127591 | 2 | LRRN2, NFASC | 11 |
| | 1q43 | 234761113 | 234801293 | 1 | HEATR1, LGALS8 | 10 |
| 8 | 8q21.2 | 86737728 | 87180823 | 2 | REXO1L1, ATP6V0D2 | 11-13 |
| | 8q21.3 | 87399989 | 91777414 | 3 | WWP1, CPNE3, TMEM64 | 10-13 |
| | 8q24.22 | 134041593 | 134384150 | 2 | SLA, TG, NDRG1 | 10 |
| **Deletion with LOH** | | | | | | |
| 8 | 8p12 | 30529807 | 35442637 | 6 | GTF2E2, RBPMS, GSR, PPP2CB, TEX15, WRN, NRG1, UNC5D | 10-14 |
| | 8p21.2 | 24200953 | 24446745 | 1 | ADAM28, ADAMDEC1 | 10 |
| | 8p23.1 | 8948883 | 11495681 | 3 | PPP1R3B, MTMR9, BLK | 10 |
| 16 | 16q21 | 65156974 | 65165991 | 1 | CKLF | 10 |
| | 16q21-q22.1 | 65165991 | 65310104 | 1 | CMTM3, CMTM4 | 10 |
| | 16q22.1 | 65310104 | 66920496 | 5 | CA7, NAE1, CDH16, CES2, PDP2, RRAD, CBFB, E2F4, HSF4, LRRC29, LRRC36, NOL3, TRADD, ACD, CENPT, CTCF, CTRL, DDX28, DPEP2, DPEP3, DUS2L, EDC4, LCAT, NUTF2, PSMB10, SLC12A4, THAP11, NFATC3, PRMT7 | 10 |
| 17 | 17p11.2 | 19750545 | 21028207 | 3 | AKAP10, USP22, DHRS7B | 10 |
| X | Xq21.1 | 77258931 | 78869429 | 3 | CYSLTR1, PGK1, TAF9B, P2RY10, ITM2A | 10 |
| | Xq27.3-q28 | 145061891 | 147263147 | 1 | FMR1, FMR1NB | 10 |

MCOR, represents the number of minimal common overlapping region(s) contained within the specific cytoband region. [a]Representative genes are from the minimal sites encompassed in the region. Chr, chromosome.

deletions (CN<1.5) across the genome of our test subjects (Table IV). Accordingly, the principal chromosomal sites being observed in this category include the 8p, 10-14% (8p12, 8p21.2 and 8p23.1); 16q, 10% (16q21, 16q21-q22.1 and 16q22.1); 17p, 10% (17p11.2); and Xq, 10% (Xq21.1 and Xq27.3-q28) arms. Fourteen percent of losses were linked with a MCOR of ~64 kb within the 8p12 region (8, 30529807-30593711) that was comprised of GTF2E2 and RBPMS genes. Other 13% of deletions originate from two distinct MCORs of ~0.1 Mb (8, 30682001-30810693) and ~88 kb (8, 30593711-30682001) of the 8p12 locus, encompassing the PPP2CB, TEX15 and GSR genes, respectively.

*CN-LOH*. A total of 42 regions affecting all the autosomes and the X chromosome were detected as CN-LOH, except for chromosomes 11, 13, 16, 17, 18, 19, 20 and 22 (Table V). The median size of these alterations was ~0.25 Mb (range, ~13 kb to ~9.7 Mb). Of the 42 detected regions, 34% of the samples displayed alterations that were linked with a MCOR of ~77 kb (14, 59775222-59851967) within the 14q23.1 locus, comprising of a single gene, PPM1A. Other prevailing MCORs of CN-LOH were manifested in a range of 30-33%, as dictated within the 3p21.31 (3, 49745036-49978636), 5q33.2 (5, 153073784-

153237829; 153315076-153410180), 12q24.12 (12, 110346958-110788418), 12q24.12-q24.13 (12, 110788418-111082704) and 14q23.1 (14, 65848833-66708208) loci. Generally, the regions of CN-LOH were small and interspersed throughout the whole genome. This suggests that the CN-LOH events obtained from the analysis arose from a specified mechanism.

*Quantitative validation of CN with real-time PCR (qPCR)*. For direct comparison analysis, representative DNA samples used for SNP array hybridizations were also used in the qPCR procedures. Generally, the CNs estimated by SNP array were comparable to the results obtained from qPCR for both amplification (PVT1, 3.69 for SNP array vs. 4.02 for qPCR; MYC, 3.69 vs. 4.56; CCND1, 3.71 vs. 5.15; ERBB2, 4.10 vs. 4.86) and deletion (CASP1, 1.53 vs. 1.53; NRG1, 1.52 vs. 1.51; PTPRJ, 1.33 vs. 1.72; TP53, 1.55 vs. 1.42) CN abnormality categories. Conversely, much higher values of copy-number amplification were observed by qPCR than by SNP array estimation for high-CN regions as seen in the CCND1 gene. This is in agreement with the notion that the hybridization intensities of SNP array analysis are often saturated at higher magnitude amplifications, leading to a generally lower estimated

Table V. Common chromosomal regions of copy-neutral LOH in primary breast carcinoma.

| Chr | Cytoband | Start | End | MCOR | Representative gene/s[a] | % of cases |
|---|---|---|---|---|---|---|
| 1 | 1p33 | 49209779 | 50467327 | 1 | ELAVL4 | 29 |
| | 1p34.1 | 45612230 | 46249174 | 2 | PRDX1, TMEM69, MAST2 | 21 |
| 2 | 2p11.2 | 86079885 | 86437981 | 2 | IMMT, PTCD3, REEP1 | 23-26 |
| 3 | 3p21.1 | 52239947 | 52801587 | 2 | BAP1, PHF7, STAB1, NEK4, PBRM1, SPCS1 | 21-23 |
| | 3p21.2-p21.1 | 51678104 | 51979357 | 1 | GRM2, PARP3, PCBP4 | 26 |
| | 3p21.31 | 48530086 | 51012298 | 5 | ARIH2, MST1R, RBM6, TRAIP, CACNA2D2, RASSF1, TUSC4, CISH, MAPKAPK3, DOCK3 | 23-30 |
| 4 | 4q26 | 119211846 | 119462492 | 1 | NDST3, PRSS12 | 21 |
| 5 | 5q22.2 | 112049571 | 112422434 | 2 | APC, SRP19, REEP5 | 23-26 |
| | 5q31.1-q31.2 | 130863839 | 131269677 | 1 | EGR1, ETF1, GFRA3 | 21 |
| | 5q33.2 | 153073784 | 153563613 | 3 | GRIA1, MFAP3, GALNT10 | 29-30 |
| 6 | 6p21.1-p12.3 | 45050258 | 45264421 | 1 | SUPT3H | 24 |
| | 6p22.1 | 26572768 | 27331446 | 2 | ABT1, BTN1A1, PRSS16 | 23-24 |
| 7 | 7q11.22 | 68921196 | 69505366 | 1 | AUTS2 | 29 |
| 8 | 8q11.22 | 50673450 | 51502012 | 1 | SNTG1 | 24 |
| 9 | 9q21.13 | 74130751 | 89373256 | 2 | TMC1, ZFAND5, DAPK1 | 21-29 |
| 10 | 10q11.22 | 49985086 | 50093612 | 1 | C10orf72 | 23 |
| | 10q26.2 | 127691801 | 127736971 | 1 | ADAM12 | 21 |
| 12 | 12q13.11 | 46691389 | 47150312 | 2 | ASB8, PFKM, H1FNT | 23-24 |
| | 12q13.13 | 50999013 | 51020106 | 1 | KRT83 | 23 |
| | 12q24.12 | 110346958 | 110788418 | 1 | ACAD10, ATXN2, SH2B3 | 31 |
| | 12q24.12-q24.13 | 110788418 | 111082704 | 1 | C12orf30, ERP29, TRAFD1 | 31 |
| 14 | 14q23.1 | 59775222 | 66708208 | 3 | PPM1A, SIX1, GPHN | 23-34 |
| 15 | 15q22.1 | 55995457 | 56132293 | 1 | ALDH1A2 | 26 |
| | 15q24.3 | 75293482 | 75628285 | 1 | HMG20A | 21 |
| 21 | 21q21.3 | 29154970 | 29406022 | 1 | CCT8, N6AMT1, USP16 | 23 |
| | 21q22.12 | 35987412 | 36107656 | 1 | MIR802 | 23 |
| X | Xp11.1-q12 | 57283339 | 67006471 | 1 | AR, EDA2R, HEPH, LAS1L, MSN, VSIG4 | 24 |

MCOR, represents the number of minimal common overlapping region(s) contained within the specific cytoband region. [a]Representative genes are from the minimal sites encompassed in the region. Chr, chromosome.

CN of gains (8). Nevertheless, the disparities could also be attributed to the different methodological basis between SNP array analysis and qPCR technique. All the analyzed CN calls were denoted by a z-score value of <1.75, at a confidence level of 95%.

**Discussion**

*Amplification without LOH.* The attained data in this analysis revealed a set of non-random recurrent genomic amplifications that were previously described and manifested within the 8q, 11q, 17q and 20q arms (2,14,18). Generally, the 8q amplicon which was denoted by 8 distinct regions, was gained in 31-44% of our samples. This is in agreement with previous reports that noted 27- 49% of gains (2,19,20) even though there were discordant cases accounted by the recruitment of high-grade breast cancer patients (21), or due to distinct

genetic profiles imprinted in diverse populations (22). Of note, the most prevalent amplifications as quantified in 31-41% and 33-44% of our cohort were localized within the 8q23 and 8q24 amplicons, respectively. In accordance, the rates exhibited are in agreement with ones from previous studies (14,15,18,23). Specifically, the most recurrent amplification overlaps were detected from the 3 MCORs within the 8q24.21 locus (~40 kb to ~0.5 Mb), present in 43-44% of our subjects. This is in agreement with previous studies (14,23). Each MCOR was designated by a single gene, including MYC and PVT1. The well established MYC oncogene was amplified in 44% of our samples, and has been linked with poor prognosis (24). Nonetheless, it has been proposed that PVT1 could be the driver of this locus as its amplification and overexpression were higher than that of MYC (14). A study of both these genes exhibited that only PVT1 silencing educed strong apoptotic response (25).

We were able to identify 2 distinct sites at 11q13 which were amplified to ≥3.5 copies in 19-23% of our cohort. This is in agreement with other studies that displayed a range of 8-24% of gains (5,14,18). Conversely, an earlier CGH-based study by Valladares *et al* (26) showed a lower rate (1/16 cases) that could be due to haplotype variations or differences in technical platforms used. This suggests that low resolution techniques would have missed a substantial fraction of genomic alterations, as the most recurrently affected MCOR in this amplicon was only 57 kb in length. Furthermore, the 4 demarcated MCORs in the 17q12 locus encompassed the ERBB2, STARD3 and GRB7 genes that demonstrated amplifications of ≥4 copies, detected in 16% of our samples. Similar rates were also formerly depicted in this region (18,23), in contrast to the study by Saito *et al* (27) which displayed a gain of 71%. This disparity could be due to the incorporation of breast cancer cell lines in their analyses, which have been linked with higher rates of aberration compared to primary tumours. Of note, the amplification rates of 19-21% as denoted by the 17q23 region in this study, comprised of the candidate RPS6KB1 gene. This is in agreement with a previous study by Haverty *et al* (14). Our analysis also defined 5 sites in the 20q13 amplicon that were gained in 16-23% of our subjects, corresponding with the 3-26% of gains reported by others (5,23). Candidate genes localized in this region include ZNF217, PTPN1 and SYCP2 (14,28).

*Deletion without LOH*. Our data are in agreement with previous studies, which exhibited widespread CN losses on 8p, 16q and 17p in panels of breast cancer specimens. Generally, the 8p, 16q and 17p loci which were demarcated into 10, 4 and 5 distinct sites, respectively, were lost in 16-23%, 14-19% and 20-21% of our samples, accordingly. The results from our study are in agreement with ones from earlier studies, in which 8p, 16q and 17p losses were documented in scales of 12-29% (2,20,23), 13-22% (2,6,29) and 12-21% (2,5), respectively. Nevertheless, intermittent variations have also been shown, as exemplified in the higher loss rates at 8p (21), 16q (30) and 17p (31) or lower rates at the 8p locus (22). The disparities could be attributable to factors, such as the ones mentioned above. The limitations of CGH to further segment CN data to deletion with/without LOH could also ascribe to the higher rates as demonstrated in earlier studies.

Genes found in this category can act as TSGs under the haploinsufficiency model, in which the loss of one allele via deletion/mutation results in gene dosage diminution that may lead to ample phenotypic changes contributing to tumourigenesis (32). The haploid deletion of the NKX3-1 gene which was mapped to the 8p21.2-p21.3 locus in 19% of our samples, has been linked with haploinsufficiency in the development of prostate cancer and prostate intraepithelial neoplasia (PIN) (33). The NKX3-1 gene has also formerly been suggested to function as an ER repressor (34). Its loss could thus be vital in hormone-responsive breast carcinogenesis. One of the most prevalent regions of deletion without LOH overlap was a 0.98 Mb region on 8p21.1 (8, 28448462-29427118) present in 23% of our samples. Localized to this region is DUSP4, a dual-specificity phosphatase that inactivates the MEK/ERK pathway, and that is down-regulated in clinical breast samples (35). Notably, the 4 MCORs (~23 kb to ~1.1 Mb) of 8p23.1

comprised of many genes from the defensin family, in which the 8p23.1-p23.2 locus detected also consists of the candidate gene, ANGPT2 (15). These genes displayed hemizygous deletions of ≤1.5 copies, common in 19-21% of our study subjects. Conversely, Naylor *et al* (31) showed loss rates of 40% within the noted regions. The discrepancies could be due to the smaller sample size they analyzed, resulting in possibly skewed frequency representations.

As noted, the 17p arm which was lost in 20-21% of our study set was segregated into 5 distinct sites which included the 17p13.2 locus, concurring with an earlier study by Stange *et al* (5). We also identified a deleted 15q11.2 site in 14% of our cases which has been poorly described in previous breast cancer studies. This minimal or broader region as a whole has only been scarcely noted, sometimes being associated with the basal-like subtype (21,36) and higher grade tumours (37). This region comprised of the BCL8, CXADRP2, POTEB, GOLGA6L6, GOLGA8C, LOC646214, NF1P1, LOC727924, OR4M2, OR4N3P and OR4N4 genes that displayed high-grade deletions of ≤1.29 copies. No report has yet to indicate the putative carcinogenesis-related function of these indicated genes. Accordingly, further functional analyses are required to discover the possible implications of these genes in the pathogenesis of breast cancer.

*Integration of LOH accompanied by CN alterations*. By profiling both tumour and constitutional DNA on a series of polymorphic markers, some of the most striking LOH events in breast cancer were manifested on the arms of 1p, 1q, 3p, 6q, 7q, 8p, 11p, 13q, 16q, 17p, 17q, 18q, 19p, 21q and 22q (7,16). As most of the LOH events noted in earlier studies have not been discerned into complementing CN profiles, the rates displayed were higher in comparison to those observed in each of our integrated LOH with CN profile. We were able to detect LOH events accompanied by CN gains as opposed to Hawthorn *et al* (15), who used a similar platform and software in their integration study. As the rate of these alterations was considerably low (10-13%), the discrepancies were most possibly due to the differences in sample size analyzed. In our current study, we incorporated a larger cohort that enabled the accurate detection of less prevalent aberrations. The effects of this alteration could be attributed to the elimination of the wild-type function of an activated oncogene or due to erroneous LOH callings caused by allelic imbalances (9). Nonetheless, an earlier study on neuroblastoma (9) has also further substantiated the existence of such an alteration in our analysis. In particular, LOH events without accompanying CN description were also noted by Loo *et al* (16) on the 8q21.3 region. At this site, 2 candidate genes which were amplified in 10% of our samples include WWP1 and CPNE3. WWP1 is a prospective oncogene in breast cancer, which encodes for an E3 ubiquitin ligase that negatively regulates the TGF-β tumour suppressor pathway (38). CPNE3 is a member of the $Ca^{2+}$-dependent phospholipid-binding proteins family, which was overexpressed upon ERBB2 amplification in cancer cells (39).

We exclusively identified regions of LOH accompanied by CN loss at the 8p, 16q, 17p and Xq regions, in agreement with earlier reports (16,40). The detected low incidences (10-14%) are also in accordance with similar recent studies (15,16), which noted losses that ranged from 11-17% (16). We did not

detect any homozygous deletion and it has been addressed as a considerably rare event in primary breast cancer (31). In this category, the genes on the retained alleles are thought to be inactivated via intra-chromosomal deletions, haplo-insufficiency, mutations, or epigenetic silencing. Additionally, several genes which were deleted without LOH were also detected at the noted similar loci. Some were postulated to encounter their '2nd hit' here, or this may merely indicate the existence of multiple clones in the tumour samples that can be altered via diverse mechanisms.

LOH events on the 8p12 site were detected in 10-14% of our samples, as opposed to previous studies which exhibited a higher range of 45-≥50% (16,41). The discrepancies could be ascribed to the variations in recruited clinical samples and methodologies, or it could portray the total LOH representations at a specific locus due to the scarcity of downstream integration analysis. Accordingly, some of the functionally appealing genes harboured in this site include GTF2E2, PPP2CB, WRN and NRG1. Even though NRG1 has been noted as being mitogenic, it was found to be deleted in 10% of our subjects. This is in agreement with a previously suggested TSG role for this gene, in which its depletion was shown to increase breast cancer cell proliferation (42). Additionally, the reported 10% of LOH events on the detected 8p23.1 locus in this study have also been indicated in an earlier analysis which recruited Indian population-based clinical samples (43).

LOH events on the 16q arm have been documented at a range of 28-90% (40). The 3 discrete sites which were deleted in 10% of our samples include the 16q21, 16q21-q22.1 and 16q22.1 loci. The 16q21-q22.1 locus harboured the recently described CMTM3, that was notably downregulated in breast tumour samples (44). The preferential loss of the 17p arm has also been linked with breast cancer (7) in a range of 45-≥50% (16,45). The only distinct chromosome 17 site that was found in 10% of our samples is the previously not well-described 17p11.2 region, which has been associated with ovarian and prostate cancers (46,47). We also detected poorly characterized deletion with LOH regions in breast cancer on the Xq21.1 and Xq27.3-q28 loci. These losses have been reported, but are specific to breast cancer cell lines due to the low incidences in primary tumours (23). Nonetheless, LOH events on the Xq27-q28 locus in breast cancer and hereditary prostate cancer (48,49) have also been noted. Albeit no putative gene has been indicated, the 10% losses acquired may still suggest the harbouring of potential X-linked TSGs, in which the loss of the active X allele occurs in a single step upon X-inactivation (50). Alternatively, these genes may have also escaped the X-inactivation mechanism, where they are still principally able to conform to Knudson's 'two-hit' TSG inactivation (50).

Concurrent analysis of the genotypic and CN data by SNP array has revealed the rise of CN-LOH/acquired uniparental disomy (aUPD) events, which are characterized by the loss of one allele, followed by the reduplication of the remaining allele. As such, this mechanism is vital in the inactivation of TSGs or activation of oncogenes, via the homozygous manifestation of pre-existing mutations or methylations (51). The segmental and widespread aUPD regions detected are thought to arise from multiple mitotic recombination events. Most of the LOH regions documented in this study have been associated with CN-LOH events, according to earlier reports (15,16). To

the best of our knowledge, many of the identified aUPD regions including those that were prevalent (≥30%) in our study, such as 14q23.1, 3p21.31, 5q33.2 and 12q24.12-q24.13, have not been formerly reported. This indicates putative mutational targets that have yet to be discovered.

CN-LOH events at 12q24.12 have been linked with 31% of our cohort, and noted by Hawthorn *et al* (15). Other identified sites such as 3p21.31, 4q26, 5q22.2, 5q31.1, 5q31.2, 5q33.2, and 14q23.1 have been associated with LOH events, even though no correlation with CN states has been shown (16). Additionally, Loo *et al* (16) in their study also noted high rates of LOH events accompanied by diploid CN. Some of these regions may thus be linked with CN-LOH, as indicated in the present study. The most prevalent region of CN-LOH overlap was a ~77 kb site on 14q23.1, present in 34% of our study subjects. This minimal region consists of a single gene, PPM1A, a SMAD phosphatase which inhibits TGF-β mediated growth arrest upon ectopic expression (52).

In our study, the 3p21.31 locus was further demarcated into 5 distinct MCORs that were altered in 23-30% of our samples. Localized within a ~0.11 Mb overlap (3, 50352133-50461061) and altered in 29% of our cohort, is the putative RASSF1 TSG that produces two main splice variants. *RASSF1A* is very frequently silenced via CpG island hypermethylation in breast cancer cases (53). It has also been suggested to regulate the G1-S transition via the modulation of the CCND1 level (54). Of note, we uncovered an interesting ~10 Mb region (X, 57283339-67006471) of isodisomy which spanned from Xp11.1 to Xq12. This isodisomy was detected in 24% of our samples, and has been formerly linked with sporadic basal-like breast cancers resulting in the loss of normal Xi, duplication of the active X allele, and the lack of CpG island methylation on the X-linked genes (55). This could subsequently perturb the normal cellular network and contribute to the basal-like breast carcinogenesis. Despite limitation in direct comparison, these effects could also explicate part of the underlying mechanisms in the breast tumourigenesis of our sample set.

In summary, our study provides the first comprehensive integrated profiles of CN changes and LOH alterations in a cohort of Malaysian hospital-based primary breast carcinoma samples. From the 5 integrated groups of genomic alterations, the most prevalent gains and losses as perceived were harboured within the 8q24.21 and 8p21.1 loci, respectively. Additionally, we also characterized CN-LOH events as the most ubiquitous aberrations within the LOH expanse, in which many of the detected regions have not been previously indicated. This includes regions that were prevalent (≥30%) in our study, such as 3p21.31, 5q33.2 and 12q24.12-q24.13, with the most frequent being at the 14q23.1 locus. By defining different categories of aberrations as exemplified by the current study, putative underlying mechanisms which target different groups of genes in breast carcinogenesis could be discerned. This in turn will facilitate further complementation studies that are required in order to delineate the true driver genes.

### Acknowledgements

## References

1. Wang ZGC, Lin M, Wei LJ, *et al*: Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res 64: 64-71, 2004.
2. Rennstam K, Ahlstedt-Soini M, Baldetorp B, *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 63: 8861-8868, 2003.
3. Miyakis S and Spandidos DA: Allelic loss in breast cancer. Cancer Detect Prev 26: 426-434, 2002.
4. Zhou XF, Rao NP, Cole SW, Mok SC, Chen ZG and Wong DT: Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizing high density single nucleotide polymorphism arrays. Cancer Genet Cytogenet 159: 53-57, 2005.
5. Stange DE, Radlwimmer B, Schubert F, *et al*: High-resolution genomic profiling reveals association of chromosomal aberrations on 1q and 16p with histologic and genetic subgroups of invasive breast cancer. Clin Cancer Res 12: 345-352, 2006.
6. Zudaire I, Odero MD, Caballero C, *et al*: Genomic imbalances detected by comparative genomic hybridization are prognostic markers in invasive ductal breast carcinomas. Histopathology 40: 547-555, 2002.
7. Miller BJ, Wang D, Krahe R and Wright FA: Pooled analysis of loss of heterozygosity in breast cancer: a genome scan provides comparative evidence for multiple tumor suppressors and identifies novel candidate regions. Am J Hum Genet 73: 748-767, 2003.
8. Zhao XJ, 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. George RE, Attiyeh EF, Li SL, *et al*: Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. PLoS One 2: e255, 2007.
10. O'Keefe C, McDevitt MA and Maciejewski JP: Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. Blood 115: 2731-2739, 2010.
11. Lim GC and Halimah Y: Second Report of the National Cancer Registry. Cancer Incidence in Malaysia 2003. National Cancer Registry. Kuala Lumpur, 2004.
12. Yip CH, Taib NA and Mohamed I: Epidemiology of breast cancer in Malaysia. Asian Pac J Cancer Prev 7: 369-374, 2006.
13. Argos M, Kibriya MG, Jasmine F, *et al*: Genomewide scan for loss of heterozygosity and chromosomal amplification in breast carcinoma using single-nucleotide polymorphism arrays. Cancer Genet Cytogenet 182: 69-74, 2008.
14. Haverty PM, Fridlyand J, Li L, *et al*: High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer 47: 530-542, 2008.
15. Hawthorn L, Luce J, Stein L and Rothschild J: Integration of transcript expression, copy number and LOH analysis of infiltrating ductal carcinoma of the breast. BMC Cancer 10: 460, 2010.
16. Loo LWM, Ton C, Wang YW, *et al*: Differential patterns of allelic loss in estrogen receptor-positive infiltrating lobular and ductal breast cancer. Genes Chromosomes Cancer 47: 1049-1066, 2008.
17. Diskin SJ, Li MY, Hou CP, *et al*: Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res 36: e126, 2008.
18. Chin SF, Teschendorff AE, Marioni JC, *et al*: High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 8: R215, 2007.
19. Kallioniemi A, Kallioniemi OP, Piper J, *et al*: Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. Proc Natl Acad Sci USA 91: 2156-2160, 1994.
20. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J and Kallioniemi OP: Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosomes Cancer 21: 177-184, 1998.
21. Natrajan R, Lambros MB, Rodriguez-Pinilla SM, *et al*: Tiling path genomic profiling of grade 3 invasive ductal breast cancers. Clin Cancer Res 15: 2711-2722, 2009.
22. Ghaffari SR, Sabokbar T, Pour PN, *et al*: Comparative genomic hybridization (CGH) analysis of chromosomal aberrations in Iranian patients with invasive ductal carcinoma breast cancer. Asian Pac J Cancer Prev 9: 66-70, 2008.

23. Tsuji K, Kawauchi S, Saito S, *et al*: Breast cancer cell lines carry cell line-specific genomic alterations that are distinct from aberrations in breast cancer tissues: comparison of the CGH profiles between cancer cell lines and primary cancer tissues. BMC Cancer 10: 15, 2010.
24. Deming SL, Nass SJ, Dickson RB and Trock BJ: C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. Br J Cancer 83: 1688-1695, 2000.
25. Guan YH, Kuo WL, Stilwell JL, *et al*: Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clin Cancer Res 13: 5745-5755, 2007.
26. Valladares A, Salamanca F, Madrigal-Bujaidar E and Arenas D: Identification of chromosomal changes with comparative genomic hybridization in sporadic breast cancer in Mexican women. Cancer Genet Cytogenet 152: 163-166, 2004.
27. Saito S, Morita K and Hirano T: High frequency of common DNA copy number abnormalities detected by bacterial artificial chromosome array comparative genomic hybridization in 24 breast cancer cell lines. Hum Cell 22: 1-10, 2009.
28. Kao J, Salari K, Bocanegra M, *et al*: Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 4: e6146, 2009.
29. Bae JS, Choi JS, Baik SH, *et al*: Genomic alterations of primary tumor and blood in invasive ductal carcinoma of breast. World J Surg Oncol 8: 32, 2010.
30. Li J, Gromov P, Gromova I, *et al*: Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. Proteomics 8: 5038-5052, 2008.
31. Naylor TL, Greshock J, Wang Y, *et al*: High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast Cancer Res 7: R1186-R1198, 2005.
32. Santarosa M and Ashworth A: Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 1654: 105-122, 2004.
33. Magee JA, Abdulkadir SA and Milbrandt J: Haploinsufficiency at the Nkx3.1 locus: a paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation. Cancer Cell 3: 273-283, 2003.
34. Holmes KA, Song JS, Liu XLS, Brown M and Carroll JS: Nkx3-1 and LEF-1 function as transcriptional inhibitors of estrogen receptor activity. Cancer Res 68: 7380-7385, 2008.
35. Armes JE, Hammet F, de Silva M, *et al*: Candidate tumor-suppressor genes on chromosome arm 8p in early-onset and high-grade breast cancers. Oncogene 23: 5697-5702, 2004.
36. Chin K, DeVries S, Fridlyand J, *et al*: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10: 529-541, 2006.
37. Richard F, Pacyna-Gengelbach M, Schlüns K, *et al*: Patterns of chromosomal imbalances in invasive breast cancer. Int J Cancer 89: 305-310, 2000.
38. Chen CS, Zhou ZM, Ross JS, Zhou W and Dong JT: The amplified WWP1 gene is a potential molecular target in breast cancer. Int J Cancer 121: 80-87, 2007.
39. Heinrich C, Keller C, Boulay A, *et al*: Copine-III interacts with ErbB2 and promotes tumor cell migration. Oncogene 29: 1598-1610, 2010.
40. Rakha EA, Green AR, Powe DG, Roylance R and Ellis IO: Chromosome 16 tumor-suppressor genes in breast cancer. Genes Chromosomes Cancer 45: 527-535, 2006.
41. Dahiya R, Perinchery G, Deng GR and Lee C: Multiple sites of loss of heterozygosity on chromosome 8 in human breast cancer has differential correlation with clinical parameters. Int J Oncol 12: 811-816, 1998.
42. Chua YL, Ito Y, Pole JCM, *et al*: The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. Oncogene 28: 4041-4052, 2009.
43. Bhattacharya N, Chunder N, Basu D, *et al*: Three discrete areas within the chromosomal 8p21.3-23 region are associated with the development of breast carcinoma of Indian patients. Exp Mol Pathol 76: 264-271, 2004.
44. Wang Y, Li JS, Cui Y, *et al*: CMTM3, located at the critical tumor suppressor locus 16q22.1, is silenced by CpG methylation in carcinomas and inhibits tumor cell growth through inducing apoptosis. Cancer Res 69: 5194-5201, 2009.
45. Phelan CM, Borg A, Cuny M, *et al*: Consortium study on 1280 breast carcinomas: allelic loss on chromosome 17 targets subregions associated with family history and clinical parameters. Cancer Res 58: 1004-1012, 1998.

46. Phillips N, Ziegler M, Saha B and Xynos F: Allelic loss on chromosome 17 in human ovarian cancer. Int J Cancer 54: 85-91, 1993.

47. Simard J, Dumont M, Labuda D, *et al*: Prostate cancer susceptibility genes: lessons learned and challenges posed. Endocr Relat Cancer 10: 225-259, 2003.

48. Roncuzzi L, Brognara I, Cocchi S, Zoli W and Gasperi-Campani A: Loss of heterozygosity at pseudoautosomal regions in human breast cancer and association with negative hormonal phenotype. Cancer Genet Cytogenet 135: 173-176, 2002.

49. Xu JF, Meyers D, Freije D, *et al*: Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20: 175-179, 1998.

50. Loupart ML, Adams S, Armour JAL, Walker R, Brammar W and Varley J: Loss of heterozygosity on the X-chromosome in human breast cancer. Genes Chromosomes Cancer 13: 229-238, 1995.

51. Tuna M, Knuutila S and Mills GB: Uniparental disomy in cancer. Trends Mol Med 15: 120-128, 2009.

52. Lin X, Duan XY, Liang YY, *et al*: PPM1A functions as a Smad phosphatase to terminate TGF beta signaling. Cell 125: 915-928, 2006.

53. Feng W, Orlandi R, Zhao N, *et al*: Tumor suppressor genes are frequently methylated in lymph node metastases of breast cancers. BMC Cancer 10: 378, 2010.

54. Shivakumar L, Minna J, Sakamaki T, Pestell R and White MA: The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol Cell Biol 22: 4309-4318, 2002.

55. Richardson AL, Wang ZGC, De Nicolo A, *et al*: X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell 9: 121-132, 2006.