Malaysian J Pathol 2013; 35(1) : 33 – 43

REVIEW

SNP array technology: an array of hope in breast cancer research

Ching Ching HO* *MMedSc*, Kein-Seong MUN* *MBBS, MPath* and Rakesh NAIDU** *MSc, PhD*

**Department of Pathology, Faculty of Medicine, University of Malaya and **School of Medicine and Health Sciences, Monash University Sunway Campus*

Abstract

Breast cancer is the most common malignancy in women worldwide. The incidence of breast cancer in Malaysia is lower compared to international statistics, with peak occurrence in the age group between 50 to 59 years of age and mortality rates of 18.6%. Despite current diagnostic and prognostic methods, the outcome for individual subjects remain poor. This is in part due to breast cancers' wide genetic heterogeneity. Various platforms for genetics studies are now employed to determine the identity of these genetic abnormalities, including microarray methods like high-density single-nucleotide-polymorphism (SNP) oligonucleotide arrays which combine the power of chromosomal comparative genomic hybridization (cCGH) and loss of heterozygosity (LOH) in the offering of higher-resolution mappings. These platforms and their applications in highlighting the genomic alteration frameworks manifested in breast carcinoma will be discussed.

Keywords: breast carcinoma, microarray, single-nucleotide-polymorphism, comparative genomic hybridization, loss of heterozygosity

INTRODUCTION

Worldwide, breast cancer remains the most common form of malignancy in women with approximately 1.38 billion cases reported in the year 2008.[1] In the United States of America, breast carcinoma was the most prevalent non-cutaneous form of malignancy in 2010, accounting for about 207,090 new cases,[2] and contributing to a total of 40,230 cases of projected mortality amongst female patients. In Malaysia, 31.3% of the total number of new cancer cases in female patients was breast cancer, and the highest age-standardized rate (ASR) (59.9 per 100 000) was seen in the Chinese community, followed by Indians (54.2 per 100 000) and Malays (34.9 per 100 000).[3] Although Malaysia's breast cancer incidence was lower when compared to Western statistics,[4] the incidence was high when compared with other prominent Asian countries such as China (Beijing), Japan (Hiroshima), Korea (Seoul), and India (Chennai).[3] The peak incidence in Malaysia was in the age group of between 50 to 59 years old,[5] except for Indian patients who peak after the age of 60 years old.[3] The mortality of breast carcinoma in Malaysia have been documented at a disconcerting number of 1,716, which is approximately 18.6% of the total cases.[1]

In current practice, clinical factors (anatomy, morphology and pathology) are utilized to categorize breast carcinomas for diagnostic, prognostic, and therapeutic purposes. Despite the well established criteria in the assessment of these clinical factors, outcome for individual subjects remain difficult to predict. Even with well-worked out diagnoses, current existing therapeutic plans could not optimally treat all breast carcinoma cases.[6] This is because, like most human cancer, breast carcinoma's heterogeneous nature is underpinned by a wide array of genetic alterations. These encompass those that extend from gross and structural-based chromosomal aberrations to diminutive point mutations. The alterations involved manifest at the molecular level to affect established equilibrium on the controlling and functional aspects of individual genes, as well as impinge upon the appropriate cellular network which ultimately leads to chaotic genomic behaviour that favours carcinogenesis. Uncovering these

Address for correspondence and reprints: Associate Professor Dr. Rakesh Naidu, School of Medicine and Health Sciences, Monash University Sunway Campus, Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor Darul Ehsan, Malaysia. Telephone: + 603 55146345 Email: kdrakeshna@hotmail.com

genetic aberrations is paramount for a more comprehensive aetiologic understanding of breast cancer's natural and progressive life story. Genetic studies employing genome-wide platforms are vital in discovering important genetic events such as chromosomal copy number aberrations or loss of heterozygosity that may lead to the activation of oncogenes or the inactivation of tumour suppressor genes (TSGs). This in turn could facilitate efforts in the detection of genetic markers that are involved in breast cancer risk evaluation, prognosis and diagnosis. In the future, the full mapping of such genetic aberrations may be incorporated into the development of individualised treatment regimens.

Recent emergence of new breast carcinoma subtypes, e.g. basal-like subtype (triple-negative breast tumours), has further emphasized the need for development of novel molecular-based therapeutic options that allow for more specific and discreet targeted treatment. The understanding of the underpinning mechanisms of genomic aberrations that underline these tumours is even more critical now than before.

In the recent years, many microarray approaches or genome-wide detection tools have been developed to answer the needs of comprehensive screenings for genomic alterations across the entire genome. These platforms permit cross-assessment of individual subjects on various factors ranging from genomics to proteomics, and allow correlation with clinical endpoints. In this review, we will provide an outline of current microarray methods, especially the newly developed high-density single-nucleotide-polymorphism (SNP) oligonucleotide arrays which combines the power of CGH and LOH in offering higher-resolution mappings. Their applications in highlighting the genomic alteration frameworks manifested in breast carcinoma will also be discussed.

1. Microarray platforms

1.1 Pre-single-nucleotide polymorphism (SNP) array

Since the 1970s, karyotyping has played a vital role in our understanding of human cancer via study of global genetic alterations at chromosomal level. Although this technique has brought forth identification of TSGs and tumorigenic fusion genes, its general application remains limited due to the requirement for metaphase cells analyses that are difficult to obtain in many solid tumours, and need for secondary verification of detected chromosome breakpoints using methods such as fluorescent *in situ* hybridization (FISH). The idea of DNA microarray methodology was first proposed by Fodor and colleagues,[7] where multiple arrays with pre-designated spots of chemical compounds were constructed using photolithography technology. By manipulating the DNA characteristics in complementary strands binding, study samples can be evaluated on multiple genomic and gene expression features in a single experiment. In the recent years, numerous microarray platforms with specific analytical functions have been developed. These encompass those which cater for gene expression, genotyping, epigenetic, as well as array comparative genomic hybridization analyses.[8]

In 1992, the development of chromosomal comparative genomic hybridization (cCGH) technique revolutionized genome-wide study of copy number aberrations (deletions and amplifications) by doing away with the necessity for cell cultures.[9] Genomic DNA from disease (test) and constitutional/reference samples were differentially labelled by fluorochromes and hybridized together to the normal human metaphase slides, in the presence of human Cot-1 DNA (un-labelled blocking DNA) to mask repetitive chromosomal elements.[9,10] The relative fluorescence ratio intensities were then detected along the target chromosomes and interpreted, resulting in cytogenetic copy number profiles that depicted gains and losses of genetic materials.[9] Although this method had a vast impact on the insights of chromosomal rearrangements in cancer biology,[11] nonetheless, this technique was generally cumbersome. One of its major drawbacks was the limited copy number mapping resolutions, typically ranging from 10 to 20 Mb for deletions, ≥ 250 kb for amplifications, and 2 Mb for gains.[12]

In 1997 and 1998, the second generation CGH platforms with higher-resolution were introduced in the forms of matrix-based CGH[13] or array-CGH,[14] where the second name is now more widely used. These improvements were largely driven by the Human Genome Project, in which the metaphase slides employed in cCGH were replaced by arrayed genomic elements of known locations, accurately spotted on glass slides or other supporting materials using glass capillaries or metal pins. These arrays conveyed locus-by-locus measurement of gene dosage changes (copy number), based on the

fluorescence ratio obtained from the arrayed elements, plotted relative to their locations in the genome.[14] Genomic DNA extracted from paraffin-embedded source have been utilized in studies employing these platforms,[15] as only fairly small quantity of starting material was required in each experiment.[16] Similar to cCGH, the test and reference DNA were labelled with different fluorochromes and co-hybridized to the arrays. Subsequent to washing the aCGH platforms from unbound by-products, they were laser-scanned to generate signal ratios representing copy number profiles via apposite extraction software *in silico*.[10] Based on the nature of the DNA sequences utilized to construct these arrays, there were two principal types of platforms used in the aCGH approach: Clone and PCR-product arrays [bacterial artificial chromosome (BAC), complementary DNA (cDNA), and genomic PCR products] and Oligonucleotide-based microarrays.[17] Resolutions of the various platforms were dependent on the genomic distribution and size of the arrayed probes.[18] The ability to design specific oligonucleotide probes and the availability of high probe density platforms via industrial manufacturing has made oligonucleotide arrays the highest-resolution platforms to date.[17] Together with the continuous progress in analytical software development, oligonucleotide arrays are now documented as the most regularly used microarray platforms.

While these platforms allowed the enumeration of a genome segment, there were some shortcomings regarding conceptual issues and methodological drawbacks. Some of the concerns relevant to BAC clones-based aCGH platforms were reduced spatial sensitivity and specificity due to incorporation of large genomic clones, as well as redundant sequence copy repeats such as segmental duplications, common repeat elements such as *Alu*, and segments with extensive sequence similarity such as pseudogenes.[19] Likewise, despite cDNA-based aCGH imparting an improved mapping resolution down to single-gene level,[20] this approach also suffered similar drawbacks as previously described, predominantly in the detection of low copy number aberrations.[19] This may be ascribed to the imperfect hybridization biases created between the labeled genomic DNA to the arrayed cDNA probes as designated.[17,21] Moreover, aCGH techniques could not perceive genomic aberrations unrelated to relative changes in copy number, such as inversions or balanced translocations, and polyploidy. Nevertheless, the key shortfall of these methodological approaches was the inability of parallel detection of chromosomal loss of heterozygosity (LOH) events.

Conventionally, chromosomal alterations which included allelic imbalance was identifiable by LOH events at polymorphic loci.[21] These events were uncovered via allelotyping analyses of the tumour tissues against corresponding constitutional samples. Studies involving this concept usually utilized restriction fragment length polymorphisms (RFLPs) or polymorphic microsatellite markers distributed throughout the whole human genome.[22-24] In addition, LOH screenings were further employed to detect allelic losses that were ascribed to mechanisms such as mitotic recombination, mitotic non-disjunction, or gene conversion.[21] In spite of these, the resolutions offered by these techniques were still inevitably low due to the paucity of available markers.[17,21] This consequently hampered the efforts in high-density genotyping analyses that were essential for genome-wide screening purposes.

1.2 In current focus: high density SNP microarray analysis

Genomic polymorphisms which were found to be widely dispersed in the human genome such as microsatellites, RFLPs, and single-nucleotide polymorphisms (SNPs) have had a long history of being employed as the basis for LOH analyses. According to Wang and colleagues[25] single nucleotide DNA sequence aberrations or single nucleotide polymorphisms (SNPs) were the most prevalent form of human genetic variations. To date, more than ten million SNPs have been identified, with approximately three million of them being genotyped. High-density SNP array platforms encompassed oligonucleotide probes tiled systematically to detect the two alleles of a specific SNP locus, in which both the homozygous and heterozygous genotypes could be discerned accordingly. Due to their even-spacing, stability, and abundance across the human genome, SNPs provided exceptional foundation for high-resolution and high-density genome-wide screenings for genetic aberrations. In that sense, SNPs imparted significant investigative prospect for disease analyses, when compared to other forms of polymorphic or non-polymorphic markers. This platform was initially incorporated in large scale genotyping studies, particularly those emphasizing familial disease

association, or linkage-based analyses.[17,21] These arrays were quickly adapted into the field of cancer genomic studies, due to their ideal features as demonstrated.

As mentioned, this platform was originally designed in the year of 1998 for large-scale genotyping analyses.[25] In this version of 558 loci-embedded platforms, SNPs present in the samples were amplified via multiplex polymerase chain reaction (PCR), where primer pairs from these various loci were integrated in one single reaction and hybridized onto the SNP arrays.[25] However, limitations such as primer dimers formation and sample preparation would still be cumbersome for high-density SNP array analyses. Higher confidence in the array was then achieved via interrogation of supplementary offset probes for each of the involved SNP locus.[26] Further improvements were administered, in which specific assays such as whole genome sampling assay (WGSA) were developed to counter issues pertinent to cross-hybridizations and non-specific signals generation.[27,28] In these protocols, the complexity of the human genome was reduced to approximately 2%, via selective restriction digestion and selective PCR amplification on only the relevant sequence fragments with designated sizes.[17,27,28] At that time, this imparted the highest resolution amongst other commonly used platforms (BAC and cDNA-based arrays), in which the widespread allele determination across the genome resulted in an excellent foundation for large-scale LOH analyses. Since then, SNP array platform has been promptly integrated into many analyses involving human malignancies, such as breast,[29] lung,[30] and bladder cancer.[31]

An important advancement in SNP array was to allow concurrent extraction of both the genotype and hybridization intensity (utilized to produce copy number data) of each probe set, thereby allowing it to act as a single platform for cross-validation and complementation functions.[21] Recently, there were also studies which have shown that LOH events could be uncovered without the need for matching constitutional samples, i.e. via the utilization of contiguous extension of homozygous markers.[32-34] The advent of this approach overrode the conventions and limitations implicated by previous techniques, in providing high-resolution and high-throughput whole-genome analysis with denser marker spacing.[35] It also permited parallel detection of genotypic aberrations with discrete DNA copy number changes in individual subjects.[36-41] Subsequently, this allowed for the revelation of previously underappreciated copy neutral LOH (CN-LOH) events, which were manifested by means of mitotic non-disjunction, somatic recombination, or deletion ensued by reduplication of the retained allele.[42,43] Further explication of mechanisms underlined in the LOH incidence was also attainable, whereby they could be exhibited in the forms of recombination, non-disjunction, hemizygous deletion, or preferential amplification of one parental allele with or without the loss of the second one.[21] Thus, SNP array-based technology offered an excellent basis for genome-wide study of diseases marked by intricate array of genomic aberrations, such as breast cancer.

2. Application of array-based approach in breast cancer research

2.1 CGH and conventional LOH approaches in determining breast cancer genomics profiles

Array-based analysis was a very demanding methodological approach in the field of breast cancer research, due to the complexity of genomic alterations manifested in breast cancer cells. These platforms have been widely used to uncover the genomic profiles of breast cancer, in the effort to reveal potential underlying driver genes that might play significant diagnostic, prognostic, and therapeutic roles, tailored for personalized regimens.

Array-based studies have been employed in breast cancer research for the detection of genome-wide copy number or gene expression aberrations, molecular subtype classifications, as well as correlation with clinical outcomes. Genome-wide gene expression profiling have recently brought forth more refined breast cancer classification models,[44-46] where the established sets of 'intrinsic genes' have delineated six distinct molecular classes: ERBB2, Luminal A, Luminal B, Basal-like, Normal-like, and Claudin-low subtypes.[44-47.] These discoveries have propelled a striking paradigm shift in breast carcinoma prediction schemes, in which a high number of recent studies were found to focus on relating patterns of genetic alterations with clinical endpoints.[48,49] Apart from gene expression studies, many genomic-based analyses have also been conducted and correlated with patients' outcomes,[50,51] where the precendents of these genomic aberrations were shown to complement well with the formerly characterized gene expression-based subtypes.[51,52] Based on recent

discoveries of distinct copy number aberrations within breast tumours,[39,40,53] three consensus genomic-based profiles emerged,[54] namely 'simplex' (gains or losses typically manifested on whole chromosomal arms of 1p and/or 16p, with 16q losses, respectively; associated with ER positive or Luminal-A transcriptional patterns), 'amplifier' (clustered focal DNA amplifications commonly localized in 8p12, 8q24, 11q13, 12q15, 17q12, and 20q13 loci; associated with Luminal B and ERBB2-subtypes), and 'saw-tooth' (segmented low-level copy-number gains were preferentially found at the 10p loci, whilst preferential losses were localized within the 3p, 4p, 4q, 5q, 14q, 15q, and 17q regions; generally associated with basal-like or triple negative tumours). These studies showed that pervasive copy number aberrations in breast cancer are closely associated with deregulation of their respective gene-expression levels.[55]

Hence, there is a need to further characterize the genomic aberrations involved in breast carcinogenesis, particularly that related to the stability of DNA and the fact that copy number alterations frequently characterized the principal genetic events that lead to tumorigenesis. This is crucial, especially in the search of relevant driver genes from passenger entities. Chromosomal copy number aberrations in the form of gains/amplifications might initiate oncogenic gains of functions, while LOH events as underscored by deletion and mutation of the remaining allele, might bring forth the loss of function of TSGs.[51] In the light of these events, regions of recurring chromosomal alterations and LOH were often set as the basis for the search of underlying oncogenes and TSGs such as HER-2, MYC, CCND1, BRCA1, BRCA2, and TP53.

Progress in the expansion of array-based CGH technology has outlined some of the most significant regions of chromosomal aberrations across the breast cancer genome[48,50,56,57] in various populations.[58-60] To complement these findings, genome-wide detection of LOH incidents was performed via widespread polymorphic microsatellite markers.[61] By employing these methods, significant regions of recurrent gains and losses in breast cancer were shown to localize amongst the 1q21, 8q23-24, 11q13, 17q12-21, 20q13,[49,50,56,62-64] and 1p36, 7q31, 8p21, 13q14, 16q24, 17p13 loci,[61] correspondingly. In addition, the uncovered genomic aberrations posed potential predictive power, in that they were associated with various biological parameters and different states of progression.[50,51] These could be denoted by the prominent correlations of 16q losses with good prognosis,[49] amplifications at 8p11, 11q13, 17q12, and/or 20q13 with markedly lower survival rates, as well as LOH events at 1p, 13q, 17p and 17q with worse clinical outcomes.[65] Subsequently, the implicated regions could also harbour putative marker genes that might play important roles in prediction strategies. Identification of these genes would therefore expand our understanding of breast carcinogenesis. It would also impart new targets for potential therapeutics development, such as the targeting of HER-2 amplification in metastatic patients by a humanized monoclonal antibody, trastuzumab.[65,66]

While the aforementioned techniques channelled openings for the quantification of copy number changes and discovery of LOH events, the scarcity of available markers and limited spatial resolution has hampered the processing of more specific demarcation of the altered genomic boundaries, as well as distinguishing putative driver genes from the wide network of passenger genes. The recent emergence of SNP-array platforms has extended the analytical prospect for breast cancer research by overriding the shortcomings implicated by earlier CGH and LOH methods, whilst retaining their benefits.[36,39,40,67,68] These analyses has revealed distinct sites of chromosomal alterations which encompassed prospective genes with consistent driving roles in tumour biology. The integrated analyses of both the copy number and LOH data has led to the detection of CN-LOH events and further elucidation of potential mechanisms underlying the LOH incidents.[39-41]

2.2 Single-nucleotide polymorphism (SNP)-based platforms in determining breast cancer genomics profiles

Originally, SNP array platform was designed to genotype multiple SNPs concurrently. Hence, some of the earliest cancer analyses utilized SNP arrays to detect LOH and allelic imbalance.[30,69,70] In breast cancer, LOH events were also detected using SNP array and were shown to be associated with discrete tumour sub-groups via hierarchical clustering, therefore agreeing with prior gene-expression analyses.[51] From a study by Wang and colleagues,[71] LOH incidents were found to be more prominent in phyllodes tumour, when compared with fibroadenomas. With the expansion of high-density SNP arrays, the genomic resolution offered was higher than most of the other array platforms. Of note, since

the release of the Affymetrix 10K SNP array system, genomic copy number analyses using this platform have been explored in various malignancies.[72-75] In a comparison study based on breast cancer cell lines, SNP array has succeeding in producing the highest resolution amongst the cDNA and BAC arrays.[67] Similar outcomes were also shown in the context of LOH studies, where comparable patterns with microsatellite allelotype[76] and SSLP[69] analyses were obtained, but with additional information, as well as with better accuracy and resolution.

In general, the patterns acquired by SNP array analyses on genomic copy number changes in breast cancer were consistent to those attained in earlier CGH-based protocols .[36,38-41,53,67,68] Nevertheless, smaller and better defined minimal sites of aberrations were localized, in which a recent study based on Affymetrix 250K *Sty* SNP system using Malaysian clinical-based primary breast cancer samples has narrowed the boundaries of altered genomic region to as low as ~ 9 kb.[41] Some of these minimal common overlapping regions included fewer prospective genes per locus,[39,40] with some even down to a single gene.[41] This in turn will assist in the complex efforts of discerning putative driver genes that may play critical roles in breast carcinogenesis from the wide network of bystander entities.

One of the most unique features of the latest SNP-array application is the integration analysis of both the copy number states and LOH events. The integration of the data has revealed an exclusive sub-group of deletion which was not accompanied by LOH incidence.[41] Based on this study, the genes harboured within these regions could act as TSGs under the haploinsufficiency model.[41] The loss of one allele in these cases via deletion or mutation has been proposed to result in ample phenotypic changes, which could contribute to tumorigenesis.[77, 78] By convention, LOH regions have been set as the instigation platform for the search of TSGs, in accordance to the Knudson's 'Two-Hit' paradigm.[79] Nonetheless, the discovery of the 'deletion without LOH' sub-group[43] concurred with recent allusions that have challenged the universality of the 'Two-hit' model, by channelling a new opening for the pursuit of TSGs.

The outcomes of integration analyses could also be incorporated into plausible mechanisms underlining LOH alterations.[35] LOH may not always be accompanied by deletion or copy number loss in some cancer cases, where many proposed alternate mechanisms have been described. New discrete integrated categories of aberrations have been established, including 'amplification with LOH', and 'copy-neutral LOH' profiles. The effects of 'amplification with LOH' alterations may be ascribed to the elimination of the wild-type function of an activated oncogene, close localization between an oncogene and TSG, or due to invalid LOH callings initiated by allelic imbalances.[35] The manifestation of this form of alterations have been observed in breast cancer,[41] and other malignancies such as neuroblastoma,[35] ovarian cancer,[80] and glioblastom[37] – further substantiating its impact in carcinogenesis. In a recent breast carcinoma study, these aberrations were found to localize conspicuously within the 1q and 8q loci, where the 8q21.3 site encompassed the candidate gene, WWP1.[41] WWP1 has been proposed to be a prospective oncogene in breast cancer. The siRNA knockdown of this gene was found to express strong apoptotic response in ER-positive breast cancer cell lines.[81]

In cancer research, one of the most rewarding applications of the SNP array was in its capacity to uncover acquired uniparental disomy or CN-LOH events. These events signify an example of genomic abnormality that occurs with no net change in copy number, yet the aberration can contribute to carcinogenesis. The term uniparental disomy or UPD was first coined in the year 1980, where Engel[82] described that atypical diploid zygotes may occur consequently from the fusion of a nullisomic gamete with a disomic gamete. UPD events can be in isodisomy or heterodisomy forms: in which heterodisomy takes place when the sequences from both the transmitting parents' homologues are present, whilst isodisomy occurs when two of its homologues originate from only one parental source.[83] Hence, isodisomy permits the homozygous manifestation of prior existing mutations carried by a heterozygous parent.[82] Albeit UPD events were initially noted in cases of developmental disorders, similar mechanisms have been shown to be pivotal in the activation of oncogenes or inactivation of TSGs amongst cancer cases.[43] Acquired UPD (aUPD) or CN-LOH events were only found in a portion of cells, and were characterized by the reduplication of a chromosomal segment or allele, alongside with the loss of its respective homologous region. In addition, CN-LOH/aUPD also plays a crucial role in affecting the patterns of epigenetic behaviours, as TSGs were frequently inactivated through

aberrant methylation.[84] Hence, it is of paramount importance to further characterize implicated CN-LOH sites in breast cancer as they often revealed aberrantly expressed or imprinted genes which were mutated or atypically methylated in the presence of diploid copy number states. Some of the proposed underlying mechanisms of CN-LOH included mitotic recombination, non-disjunction, or gene conversion.[42,43,83] Similar to deletion, earlier studies into various forms of tumour, including breast carcinoma, have suggested that mitotic recombination may act as the 'second hit' in the course of LOH, resulting in the loss of the wild-type allele of putative TSGs.[35,37,41,42,85,86] The number of breakpoints implicated in the emergence of CN-LOH is indicative of the possible pathways underpinning the events. Generally, segmental and small CN-LOH regions interspersed across the whole genome tend to originate from multiple mitotic recombination incidences. Conversely, large CN-LOH regions with no or less breakpoints were thought to rise from non-disjunction mechanism or single mitotic recombination.[42]

Some of the earliest documentations of CN-LOH events in breast cancer were noted from as early as the 1990s.[9,85,87] In 1992, a study using fluorescent *in situ* hybridization (FISH) and RB1 genomic probe exhibited discordance between the FISH-based RB1 copy number outcomes from those as expected of LOH analysis.[9] These authors proposed that duplication of the chromosome 13 region comprising the mutant RB1, or somatic recombination could be the reason such results were observed. Another group of researchers also postulated similar mechanisms for the inactivation of TSGs in specific chromosomal locations in breast cancer, such as those localized in the 3p locus.[87] Despite the fact that CN-LOH incidence has been uncovered in these studies, labour-intensive methods such as *in situ* hybridization and microsatellite-based system were still employed. Consequently, the progress and verification of such studies were time-consuming and inevitably protracted.

With the advent of SNP-based technology, such limitations were overridden, i.e. integration analysis of both copy number and genotyping data was achieved on a single platform. The majority of the LOH regions reported in recent breast cancer studies were demonstrated to be CN-LOH events.[38,40,41,68] A study using the Illumina 109K SNP array system and a unique bioinformatics approach, ASCAT (allele-specific copy number analysis of tumors), has found that this form of aberration was mostly found in chromosomes 2, 3, 4, 6, and 12.[68] These authors also implied that the rate of actual loss (loss of one allele, possibly combined with gain of the other allele) may be higher than previously detected, especially when individual allelic content was not taken into account.[68] According to Ching et al,[41] the most frequently implicated CN-LOH regions documented in the cohort of Malaysian-based clinical breast cancer samples were found in the chromosomal arms of 14q23.1, 3p21.31, 5q33.2, and 12q24.12-q24.13. The most ubiquitous region of CN-LOH overlap was localized on the 14q23.1 locus, found in 34% of the study subjects. This small site of aberration comprised of a single gene PPM1A,[41] which was shown to diminish the TGF- mediated growth arrest.[88] Thus, CN-LOH aberrations on the PPM1A gene may lead to functions effects..[41] UPD events spanning throughout the X-chromosome have also been uncovered in breast cancer studies.[41,89] A study has associated such aberrations with sporadic basal-like breast cancers,[89] whilst others have narrowed down the boundaries of these X-chromosomal alterations to Xp11.1 to Xq12.[41] Isodisomy of the X-chromosome was linked with the loss of the normal inactive X chromosome (Xi). It was also shown to initiate the up-regulation of a group of X-chromosomal genes, leading to possible disruption of normal cellular pathways in breast cancer.[89,90]

Another group of researchers who had integrated transcript expression, copy number, and LOH analysis in breast carcinoma (infiltrating ductal carcinoma subtype) research noted that some of the genes harboured within the CN-LOH regions, such as HSP1A1 in the 6p21.32 locus were up-regulated. They have suggested that the missing allele of the involved genes may be acting in a suppressive capacity.[40] In contrast, a colorectal cancer-based study demonstrated that the genes within the CN-LOH regions did not experience any changes in their expression levels.[86] This may indicate that copy-neutral LOH incidence can behave very differently in different tumour types, either in its underlying mechanisms or route of manifestation.

3. Advantages, limitations, and an array of prospects for SNP-based technology

The advantages of the SNP array technology in high-density and high-throughput genetic analysis across the whole genome have been

clearly validated. In contrast to CGH and microsatellite-based techniques, SNP array technology's unique features and spatial resolution are clearly advantageous in terms of parallel processing capacity and magnitude of genomic interrogation therein. However, the genomic coverage of the oligonucleotide probes in this platform was found to be inconsistent in some cases, partly due to challenging regions such as those of segmental duplications.

The design of robust genotyping SNP assay can be complicated, depending largely on the underlying state of the genome at a particular locus. One of the methods that can be utilised to alleviate this issue is to increase the redundancy at each query locus.[10] Next-generation arrays have included supplementary non-polymorphic probes, such as those in the Affymetrix SNP Array 5.0 system, in order to facilitate the evaluation of total genome dosage within areas that were out of the ideal SNPs coverage range.[17]

At present, the ongoing development of the whole genome SNP array technology has the prospect to include nearly all the existing SNPs into a single platform for experimentation. In this light, the resolution of the genomic aberrations detected by high density SNP array analysis will suffice as complementation for other studies as described below.

One the areas which warrant extensive study is the incorporation of a larger sample cohort and transcriptional profiling to the genomic frameworks as determined via SNP arrays. By identifying parallel changes in both of these analyses, areas of significant interest for gene discovery can be revealed. This in turn will facilitate further denomination of key driver genes from a large network of downstream passenger entities that are driven by genomic abnormalities in the presence of gene expression changes. These genes could act more precisely as molecular targets, in particular those associated with pre-metastasis breast carcinogenesis. Studies of such nature have been conducted in other tumour types, seeking candidate driver genes, such as those in ovarian cancer.[39,80] In that sense, SNP array analysis, in a larger perspective, principally serves as the primary screening tool for the identification of potential genetic players that may dictate further experimental verification. Nonetheless, parallel progress in advanced bioinformatics algorithms and data mining have strengthened the plausibility of hypotheses produced in this manner. In the near future, information generated in this manner could be integrated with specific online database platform in order to identify related protein-protein interactions and specific key pathways that may play crucial role in the pathogenesis of breast cancer.

It is also worthwhile for future studies to incorporate tissue microarray (TMAs) techniques for further verification of generated genomic profiles. While DNA-based microarrays such as the SNP-based platforms are constructive in assessing genomic alterations across the whole genome in a smaller group environment, TMAs are excellent channels to validate the 'putative markers' from the SNP array analysis in large cohort of patients. This reflects the present line of research which is focused in integrating various methodological approaches in the effort of characterizing ideal molecular targets with significant influence in the diagnosis and prognosis of breast cancer patients.

It is thus becoming clearer that the future of breast cancer treatment choices will rely on the discovery and effective unraveling of new key markers via high-density genome-wide technologies such as SNP arrays. Ultimately, this will lead to the expansion of optimal diagnostic and therapeutic avenues tailored for individual breast cancer patients in the face of future.

ACKNOWLEDGEMENTS

This study was financially supported by E-ScienceFund grant (02-02-10-SF0016) from Ministry of Science, Technology and Innovation, Malaysia.

Conflict of interest statement:
The authors declare that they have no conflict of interest related to the publication of this manuscript.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Breast cancer incidence and mortality worldwide in 2008 summary. GLOBOCAN 2008, Cancer incidence and mortality worldwide: IARC Cancer Base No 10. International Agency for Research, Lyon, France. 2010.
2. American Cancer Society. Cancer Facts & Figures 2010. Atlanta: American Cancer Society; 2010.
3. Lim GCC, Sanjay R, Halimah Y, editors. Third Report of the National Cancer Registry. Cancer Incidence in Malaysia 2003 - 2005. National Cancer Registry, Ministry of Health, Malaysia; 2008
4. Yip CH, Taib NA, Mohamed I. Epidemiology of breast cancer in Malaysia. Asian Pac J Cancer Prev. 2006; 7 (3): 369-74.

5. Lim GCC, Halimah Y, editors. Second Report of the National Cancer Registry. Cancer Incidence in Malaysia 2003. National Cancer Registry, Ministry of Health, Malaysia; 2004
6. van der Vegt B, de Bock GH, Hollema H, Wesseling J. Microarray methods to identify factors determining breast cancer progression: potentials, limitations, and challenges. Crit Rev Oncol Hematol. 2009; 70 (1): 1-11.
7. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science. 1991; 251 (4995): 767-73.
8. Hoheisel JD. Microarray technology: Beyond transcript profiling and genotype analysis. Nat Rev Genet. 2006; 7 (3): 200-10.
9. Kallionemi A, Kallionemi OP, Sudar D, *et al.* Comparative Genomic Hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992; 258 (5083): 818-821.
10. Garraway LA, Sellers WR. Array-based approaches to cancer genome analysis. Drug Discov Today: Disease Mechanisms. 2005; 2 (2): 171-7.
11. Kallioniemi A, Visakorpi T, Karhu R, Pinkel D, Kallioniemi OP. Gene copy number analysis by fluorescence in situ hybridization and comparative genomic hybridization. Methods. 1996; 9: 113-21
12. Kallioniemi A, Kallionemi OP, Piper J, *et al.* Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. Proc Natl Acad Sci U S A. 1994; 91 (6): 2156-60.
13. Solinas-Toldo S, Lampel S, Stilgenbauer S, *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer. 1997; 20 (4): 399-407.
14. Pinkel D, Segraves R, Sudar D, *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet. 1998; 20: 207-11.
15. Johnson NA, Hamoudi RA, Ichimura K, *et al.* Application of array CGH on archival formalin-fixed paraffin-embedded tissues including small numbers of microdissected cells. Lab Invest. 2006; 86 (9): 968-78.
16. Guillaud-Bataille M, Valent A, Soularue P, *et al.* Detecting single DNA copy number variations in complex genomes using one nanogram of starting DNA and BAC-array CGH. Nucleic Acids Res. 2004; 32 (13): e112.
17. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet. 2007; 39: S16-S21.
18. Coe BP, Yistra B, Carvalho B, Meijer GA, MacAulay C, Lam WL. Resolving the resolution of array CGH. Genomics. 2007; 89 (5): 647-53.
19. Mantripragada KK, Buckley PG, Diaz de Stahl T, Dumanski JP. Genomic microarrays in the spotlight. Trends Genet. 2004: 20; 87-94.
20. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet. 1999; 23 (1): 41-6.
21. Zhou XF, Rao NP, Cole SW, Mok SC, Chen ZG, Wong DT. Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizing high density single nucleotide polymorphism arrays. Cancer Genet Cytogenet. 2005; 159 (1): 53-7.
22. Driouch K, DorionBonnet F, Brifford M, Champeme MH, Longy M, Lidereau R. Loss of heterozygosity on chromosome arm 16q in breast cancer metastases. Genes Chromosomes Cancer. 1997; 19 (3): 185-91.
23. Lasko D, Cavenee W, Nordenskjold M. Loss of constitutional heterozygosity in human cancer. Ann Rev Genet.1991; 25: 281-314.
24. Thiagalingam S, Laken S, Willson JKV, *et al.* Mechanisms underlying losses of heterozygosity in human colorectal cancers. Proc Natl Acad Sci U S A. 2001; 98 (5): 2698-702.
25. Wang DG, Fan JB, Siao CJ, *et al.* Large scale identification, mapping, and genotyping of single-nucleotide polymorphisms in human genome. Science. 1998; 280 (5366): 1077-82.
26. Mei R, Galipeau PC, Prass C, *et al.* Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res. 2000; 10: 1126-37.
27. Kennedy GC, Matsuzaki H, Dong SL, *et al.* Large-scale genotyping of complex DNA. Nat Biotechnol. 2003; 21 (10): 1233-7.
28. Matsuzaki H, Loi H, Dong S, *et al.* Parallel genotyping of over 10,000 SNPs using one-primer assay on a high-density oligonucleotide array. Genome Res. 2004; 14: 414-25
29. 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. 2004; 64 (1): 64-71.
30. Janne PA, Li C, Zhao X, *et al.* High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. Oncogene. 2004; 23: 2716-26
31. Koed K, Wiuf C, Christensen LL, *et al.* High-density single nucleotide polymorphism array defines novel stage and location-dependent allelic imbalances in human bladder tumors. Cancer Res. 2005; 65: 34-45.
32. Huang J, Wei W, Zhang J, *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum Genomics. 2004; 1 (4): 287-99.
33. Beroukhim R, Lin M, Park Y, *et al.* Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. PLoS Comput Biol. 2006; 2 (5): e41.
34. Calhoun ES, Hucl T, Gallmeier E, *et al.* Identifying allelic loss and homozygous deletions in pancreatic cancer without matched normals using high-density single-nucleotide polymorphism arrays. Cancer Res. 206; 66 (16): 7920-8.
35. George RE, Attiyeh EF, Li SL, *et al.* Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. Plos One. 2007; 2 (2): e255.
36. 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. 2008; 182 (2): 69-74.
37. Kuga D, Mizoguchi M, Guan Y, *et al.* Prevalence of copy-number neutral LOH in glioblastoma revealed by genomewide analysis of laser-microdissected tissues. Neuro Oncol. 2008; 10 (6): 995-1003.
38. 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. 2008; 47 (12): 1049-66.
39. Haverty PM, Fridlyand J, Li L, *et al.* High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer. 2008; 47 (6): 530-42.
40. Hawthorn L, Luce J, Stein L, Rothschild J. Integration of transcript expression, copy number and LOH analysis of infiltrating ductal carcinoma of the breast. BMC Cancer. 2010; 10: 460.
41. Ho CC, Naidu R, Mun KS, Yip CH, Taib NA. Integrated analysis of copy number and loss of heterozygosity in primary breast carcinomas using high-density SNP array. International Journal of Oncology. 2011; 39(3): 621-633.
42. O'Keefe C, McDevitt MA, Maciejewski JP. Copy neutral loss of heterozygosity: A novel chromosomal lesion in myeloid malignancies. Blood. 2010; 115 (14): 2731-9.
43. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. Trends Mol Med. 2009; 15 (3): 120-8.
44. Perou CM, Sorlie T, Eisen MB, *et al.* Molecular portraits of human breast tumours. Nature. 2000; 406 (6797): 747-52.
45. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011; 5 (1): 5-23.
46. Sorlie T, Perou CM, Tibshirani R, *et al.* Gene expression patterns of breast carcinomas distingiush tumor subclasses with clinical implication. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98 (19): 10869-10874.
47. Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, *et al.* Characterization of a naturally occuring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Research. 2009; 69 (10): 4116-4124.
48. Chin K, DeVries S, Fridlyand J, *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006; 10 (6): 529-541.
49. Zudaire I, Odero MD, Caballero C, *et al.* Genomic imbalances detected by comparative genomic hybridization are prognostic markers in invasive ductal breast carcinomas. Histopathology. 2002; 40 (6): 547-55.
50. 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. 2003; 63 (24): 8861-8.
51. 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. 2004; 64 (1): 64-71.
52. Chin K, De Vries S, Fridlyand J, *et al.* Genomic and trasncriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006; 10 (6): 529-41
53. Fang M, Toher J, Morgan M, Davidson J, Tannenbaum S, Claffey K. Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. Cancer. 2011; 117 (10): 2024-34.
54. Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR. Genomic instability in breast cancer: Pathogenesis and clinical implications. Mol Oncol. 2010; 4 (3): 255-66.
55. Pollack JR, Sorlie T, Perou CM, *et al.* Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A. 2002; 99 (20): 12963-8.
56. 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 Bio. 2007; 8 (10): R215
57. 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. 2005; 7 (6): R1186-R1198.
58. 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. 2008; 9 (1): 66-70.
59. Valladares A, Salamanca F, Madrigal-Bujaidar E, Arenas D. Identification of chromosomal changes with comparative genomic hybridization in sporadic breast cancer in Mexican women. Cancer Genet Cytogenet. 2004; 152 (2): 163-6.
60. Packeisen J, Nakachi K, Boecker W, Brandt B, Buerger H. Cytogenetic differences in breast cancer samples between German and Japanese patients. J Clin Pathol. 2005; 58 (10): 1101-3.
61. Miller BJ, Wang D, Krahe R, 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. 2003; 73 (4): 748-67.
62. 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. 2005; 65 (2): 439-47.
63. 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. 2006; 12 (2): 345-52.
64. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosomes Cancer. 1998; 21 (3): 177-84.

65. Emi M, Yoshimoto M, Sato T, *et al.* Allelic loss at 1p34, 13q12, 17p13.3, and 17q21.1 correlates with poor postoperative prognosis in breast cancer. Genes Chromosomes Cancer. 1999; 26 (2): 134-41.
66. Goldenberg MM. Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. Clin Ther. 1999; 21 (2): 309-18.
67. 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. 2004; 64 (9): 3060-71.
68. van Loo P, Nordgard SH, Lingjaerde OC, *et al.* Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010; 107 (39): 16910-5
69. Lindblad-Toh K, Tanenbaum DM, Daly MJ, *et al.* Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat Biotechnol. 2000; 18: 1001-5.
70. Primdahl H, Wikman FP, von der Maase H, Zhou XG, Wolf H, Orntoft TF. Allelic imbalances in human bladder cancer: genome-wide detection with high-density single-nucleotide polymorphism arrays. J Natl Cancer Inst. 2002; 94 (3): 216-23
71. Wang ZC, Buraimah A, Iglehart JD, Richardson AL. Genome-wide analysis for loss of heterozygosity in primary and recurrent phyllodes tumor and fibroadenoma of breast using single nucleotide polymorphism arrays. Breast Cancer Res Treat. 2006; 97 (3): 301-9.
72. Rauch A, Ruschendorf F, Huang J, *et al.* Molecular karyotyping using an SNP array for genomewide genotyping. J Med Genet. 2004; 41 (12): 916-22.
73. Bignell GR, Huang J, Greshock J, *et al.* High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res. 2004; 14 (2): 287-95.
74. Rubin MA, Varambally S, Beroukhim R, *et al.* Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. Cancer Res. 2004; 64 (11): 3814-22.
75. Zhao X, Weir BA, La Framboise T, *et al.* Homozygous deletions and chromosome amplifications in human lung carcinoma revealed by single nucleotide polymorphism array analysis. Cancer Res. 2004; 65 (13): 5561-70.
76. Hoque MO, Lee CC, Cairns P, Schoenberg M, Sidransky D. Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. Cancer Res. 2003; 63 (9): 2216-22.
77. Santarosa M, Ashworth A. Haploinsufficiency for tumor suppressor genes: When you don't need to go all the way. Biochim et Biophys Acta 2004; 1654 (2): 105-22.
78. Payne SR, Kemp CJ. Tumor suppressor genetics. Carcinogenesis. 2005; 26 (12): 2031-45.
79. Knudson AG. Mutation and cancer: Statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971; 68 (4): 820-3.
80. Gorringe KL, Ramakrishna M, Williams LH, *et al.* Are there anymore ovarian tumor suppressor genes? A new perspective using ultra high-resolution copy number and loss of heterozygosity analysis. Genes Chromosomes Cancer. 2009; 48 (10): 931-42.
81. Chen CS, Zhou ZM, Ross JS, Zhou W, Dong JT. The amplified WWP1 gene is a potential molecular target in breast cancer. Int J Oncol. 2007; 121 (1): 80-7.
82. Engel E. A new genetic concept: Uniparental disomy and its potential effect, isodisomy. Am J Med Genet. 1980; 6 (2): 137-43.
83. Robinson WP. Mechanisms leading to uniparental disomy and their clinical consequences. Bioessays. 2000; 22 (5): 452-9.
84. Stebbing J, Bower M, Syed N, Smith P, Yu V, Crook T. Epigenetics: An emerging technology in the diagnosis and treatment of cancer. Pharmacogenomics. 2006; 7 (5): 747-57.
85. Murthy SK, DiFrancesco LM, Ogilvie T, Demetrick DJ. Loss of heterozygosity associated with uniparental disomy in breast carcinoma. Mod Pathol. 2002; 15 (12): 1241-50.
86. Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. Carcinogenesis. 2007; 28 (1): 38-48.
87. Chen LC, Matsumura K, Deng G, *et al.* Deletion of two separate regions on chromosome 3p in breast cancers. Cancer Res. 1994; 54: 3021-4.
88. Lin X, Duan XY, Liang YY, *et al.* PPM1A functions as a Smad phosphatase to terminate TGF beta signalling. Cell. 2006; 125 (5): 915-28.
89. Richardson AL, Wang ZGC, De Nicolo A, *et al.* X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell. 2006; 9 (2): 121-32.
90. Liao DZJ, Du QQ, Yu BW, Grignon D, Sarkar FH. Novel perspective: Focusing on the X chromosome in reproductive cancers. Cancer Invest. 2003; 21 (4): 641-58.