Abstract

Screening of the entire human genome using high-density single nucleotide polymorphism array (SNPA) has become a powerful technique used in cancer genetics and population genetics studies. The GeneChip® Mapping Array, introduced by Affymetrix, is one SNPA platform utilised for genotyping studies. This GeneChip system allows researchers to gain a comprehensive view of cancer biology on a single platform for the quantification of chromosomal amplifications, deletions, and loss of heterozygosity or for allelic imbalance studies. Importantly, this array analysis has the potential to reveal novel genetic findings involved in the multistep development of cancer. Given the importance of genetic factors in leukaemogenesis and the usefulness of screening the whole genome, SNPA analysis has been utilised in many studies to characterise genetic aberrations in childhood acute lymphoblastic leukaemia.

Keywords: B- and T-cell acute lymphoblastic leukaemia, child, cytogenetic aberrations, single nucleotide polymorphisms, oligonucleotide array sequence analysis, medical sciences

Introduction

Leukaemia is a cancer of the haemopoietic system that affects white blood cells, which are normally responsible for combating infections. All blood cells are produced in the bone marrow (although this is not the case in utero) and develop from stem cells. Stem cells possess the potential to develop into 2 different types of white blood cell, lymphocytic or myelocytic. Mature lymphoid cells are classified as B-cells or T-cells and arise from a common lymphoid stem cell in the bone marrow (1). Normally, white blood cells grow, divide, and reproduce in an orderly and controlled manner. However, this process becomes uncontrollable in leukaemia where the cells continue to divide but not mature. Leukaemia is characterised by a diffuse replacement of the bone marrow by neoplastic cells, which leads to suppression of normal haematopoiesis. Leukaemia can be classified as acute or chronic, and lymphoid or myeloid, resulting in 4 main types of disease: acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), and chronic myeloid leukaemia (CML). Childhood leukaemias are primarily acute and progress rapidly.

Genetic factors have a significant impact on the development of childhood ALL. Recent advances in cytogenetic and molecular analyses have led to a better understanding of important genes that are involved in leukaemogenesis. Conventional cytogenetic analysis has been the standard method of identifying genetic aberrations, including changes in chromosome structure and number. Cytogenetic evidence indicating that chromosomal translocations may have important prognostic implications in ALL has been used for more than a decade to assign patients to risk-directed therapy (2). It has been reported that acquired chromosomal abnormalities in the leukemic blast cells from patients with ALL strongly associate with the biology of the disease, and designate the genes involved in leukaemogenesis (3). According to Greaves, biological subsets of leukaemia that have prognostic importance can be defined by different chromosomal and gene abnormalities (4). Recurrent chromosomal translocations may provide valuable clues to the identification of fusion genes that can lead to the characterisation of known or new oncogenes or tumour suppressor genes critical for leukaemogenesis. Gene fusion events can result in altered mechanisms to activate abnormal cell growth. These mechanisms include enhanced gene expression, fusion of transcription factors, altered protein kinase activity, and deletion of tumour suppressor genes.
Genes (2). Leukaemia cells may differ from patient to patient in terms of molecular pathology, and a number of distinct genetic subtypes of ALL have been identified according to differences in cell surface markers (at the time of presentation), clinical behaviour, and response to individual therapeutic agents (5).

Genetic alterations in childhood ALL

Numerical and structural chromosomal changes
Most of the karyotypes in ALL samples have both numerical and structural chromosomal changes. Structural abnormalities include chromosomal translocations, deletions, inversions, and other rearrangements involving genes with oncogenic potential. Numerical changes can be classified into groups according to the chromosome number such as hypodiploidy, hyperdiploidy, triploidy, and tetraploidy (3). Hyperdiploidy and hypodiploidy occur in 25% to 30%, and 1% of childhood ALL, respectively (6,7).

Molecular genetic alterations in childhood ALL
Determining the cytogenetic and molecular characteristics of childhood ALL provides a valuable tool for monitoring the disease during treatment, for the prediction of prognostic outcome, and for the management of future therapy. Molecular identification that provides detailed knowledge of important genes, such as oncogenes and tumour suppressor genes, may provide valuable clues to mechanisms of leukaemogenesis. Complete loss of an allele or an increase in copy number (CN) of one allele relative to the other can lead to the formation of allelic imbalance, commonly found in malignancies. A previous study has reported 2 forms of allelic imbalance: gene CN amplification, which reveals the presence of oncogenes; and loss of heterozygosity (LOH), which is usually associated with the presence of tumour suppressor genes (8).

Gene mutations and amplifications
Abnormalities in the coding region of genes may affect the normal function of its protein and, in turn, impact cell division, cell growth, cell death, and the DNA damage response. Various mutations in tumour suppressor genes, proto-oncogenes/oncogenes, and DNA repair genes may contribute to tumour progression. Mutations in some tumour suppressor genes, including \( p16^{INK4a} \) and \( p53 \), have been reported in childhood ALL (9–11). A number of oncogenes, such as \( FLT3 \) and members of the ras family, are found to play a role in the formation of ALL. FMS-related tyrosine kinase-3 (\( FLT3 \)) is a receptor tyrosine kinase expressed in early haematopoietic progenitors that plays an important role in haematopoietic development (12). Activating mutations of \( FLT3 \) were discovered initially in AML and are found commonly in 2 subtypes of childhood ALL: infant ALL with MLL translocations and hyperdiploid ALL (13,14). Previous studies reported that the incidence of N-ras mutations is 10% in childhood ALL, but the role of this mutation is not fully understood (15,16). The Notch signalling pathway is essential in T-cell development and NOTCH1 mutations are found frequently in adult and paediatric T-cell ALL (17).

Gene amplification is a common event in many types of solid tumours. However, the occurrence of this aberration in acute leukaemia is quite rare (18). Multiple copies of \( AML1 \) (\( RUNX1 \)), on duplicated chromosome 21, were identified in 20 patients with ALL (childhood and adult), suggesting that chromosome 21 and \( AML1 \) amplification were important in leukaemogenesis (19). A further study from this group found the same \( AML1 \) amplification on 8 additional cases of childhood ALL, emphasising this as a cytogenetic subgroup in ALL and suggesting its role as an indicator of poor prognosis in ALL (20). This abnormality is known as intrachromosomal amplification of chromosome 21 with amplification of \( AML1 \) (\( iAMP21 \)) (21,22).

High level amplification of \( AML1 \) has also been reported in 2 cases of childhood ALL detected by fluorescence in situ hybridisation (FISH) and comparative genomic hybridisation (CGH) analysis (23). In ALL, \( AML1 \) is commonly involved in translocation t(12;21)(p13;q22), which leads to the \( TEL/AML1 \) (\( ETV6/RUNX1 \)) fusion gene. A previous study also found that \( AML1 \) amplification was present in childhood ALL but not in adult cases, and it was not associated with \( AML1 \) mutation (24). A few studies have reported on \( ABL1 \) amplification in T-cell ALL cases. The first study discovered multiple copies of \( ABL1 \) in 5 of 210 paediatric T-cell ALL cases (25). Two other studies also detected \( ABL1 \) amplification in T-cell ALL patients and T-ALL cell lines (26,27). The study showed that \( ABL1 \) amplification was the amplification of the \( NUP214-ABL1 \) fusion detected in 5 of 85 T-ALL patients (27). This gene is the target of many recurrent translocations seen in different leukaemia subtypes, mostly in t(9;22) (q34;q11.2); this particular translocation results in the formation of the \( BCR-ABL1 \) fusion gene and is one of the cytogenetic hallmarks of CML (28).
Loss of heterozygosity

Loss of heterozygosity (LOH) of human chromosomal regions is one of the most frequent genetic events found in many types of malignancies. Investigation of LOH and its effect on allelic imbalance (29) in childhood ALL may provide important information about the genetic basis of the disease because frequent allelic deletions in tumour cells are usually indicative of the inactivation of tumour suppressor genes. Previous studies have suggested that the loss of tumour suppressor gene activity is an important event in the development of cancer. Takeuchi et al. (30) reported that inactivation of tumour suppressor genes by mutation of one allele and loss of the second allele is a crucial pathway of leukaemogenesis in childhood ALL. Informative microsatellite markers are used as an indirect method to confirm LOH and to search for inactivated tumour suppressor genes (30). Several different mechanisms at the molecular or cytogenetic level have been considered to account for LOH: deletion, gene conversion, single or double homologous and non-homologous mitotic recombination, translocation, chromosome breakage and loss, chromosomal fusion or telomeric end-to-end fusion, or whole chromosome loss with or without accompanying duplication of the retained chromosome (31).

Previously, a large number of LOH studies using microsatellite markers in childhood ALL have been performed by a Japanese group and collaborators; these studies have found that LOH of chromosomes 6q, 9p, 11q, and 12p are frequent in childhood ALL (32–36). Baccichet et al. investigated LOH using 49 highly polymorphic markers distributed over 13 chromosomal arms and found that the highest rates of allelic losses were observed in 9p and 12p regions, which were deleted in 29% and 32% of childhood ALL patients, respectively (37). They found no LOH on chromosomes 3p, 5q, 11p, 11q, 13q, or 18q (37). Cavé et al. indicated that 12p12-13 alterations at the molecular level are present in about 27% of children with B-lineage ALL, which is a higher percentage than had previously been reported by standard chromosome analysis (38). LOH on chromosome 12p12-13 was detected in 26 to 47% of childhood ALL samples analysed (34,39,40), suggesting that inactivation of a tumour suppressor gene on this region, possibly the ETV6 and CDKN1B, may play a role in leukaemogenesis (40,41). Baccichet and Sinnet investigated a number of childhood ALL samples and found that 31% (17 of 55) of samples had LOH on 12p12-13 (42). According to Takeuchi et al., in the relapse of childhood ALL, chromosome 9p was the most frequent site for LOH (occurring in 15 of 38 informative cases); they also report that the 4q and 17q regions appear to have an important role in ALL relapse (43). In 2004, Heerema et al. performed cytogenetic analyses for chromosome 7 and reported that the critical region of loss of chromosome 7 in paediatric ALL at presentation may be on the p-arm (44).

Epigenetic changes by methylation

DNA methylation is a natural modification of human DNA and is catalysed by a DNA methyltransferase enzyme that adds a methyl group to the carbon 5 position of the cytosine ring in CpG nucleotides (45). Methylation, the major epigenetic modification of human genomic DNA, takes place only at cytosine bases that are located 5’ to the guanosine residue in a CpG dinucleotide within the mammalian genome (46,47). CpG islands are GC-rich DNA regions, typically 0.5 to 4.0 kb in length, which remain unmethylated in normal tissues when found in the 5’ region of genes (46,48). These CpG islands are located frequently within the promoter regions of human genes. Methylation within CpG islands has been associated with transcriptional inactivation and functional silencing of the corresponding gene due to chromatin compaction (49,50). Robertson and Wolffe reported that CpG islands can also be found in non-coding intergenic areas rich in highly repetitive DNA elements (51).

Both DNA hypermethylation and hypomethylation might play important roles in the tumourigenic process, but increased methylation at CpG islands has been, by far, the most studied process that has a clear role in carcinogenesis (50). It has been documented that aberrant methylation may occur during the early stages of carcinogenesis, and that distinct types of cancer exhibit specific patterns of methylation changes (52). It has been proposed that methylation of the promoter region should be included in Knudson’s two-hit hypothesis for inactivation of tumour suppressor genes (53,54). The first hit may be a mutation in the DNA sequence or promoter methylation, as was suggested. The second inactivating hit may be either LOH or a further mutational or methylation event in the second allele. Hypermethylation of CpG islands in a human tumour was first reported in 1986 (55). Multiple genes, such as tumour suppressor genes and DNA repair genes, have been shown to be inactivated by hypermethylation of CpG islands in many types of cancer, including haematological malignancies and solid tumours (50).
Application of high density single nucleotide polymorphism array to determine allelic imbalance in childhood ALL

Over the past few years, the application of advanced technologies, based on the large collections of mapped single nucleotide polymorphism (SNP), to genetic studies has yielded important findings to help elucidate the molecular mechanisms responsible for the pathogenesis of human cancers. A SNP is the most common type of DNA sequence variation that is found at one specific position in the human genome and occurs in at least 1% of the human population (56). In 1998, Wang et al. illustrated large-scale identification, mapping, and genotyping of SNPs in the human genome, which led to the development of the genotyping chips that allow for the simultaneous genotyping of 500 SNPs (57).

Investigation of allelic imbalance can also be performed using a variety of techniques such as conventional cytogenetic analysis (karyotyping), FISH, Southern blotting, microsatellite analysis using highly polymorphic markers, and CGH. These techniques are either of low resolution or laborious and require the use of relatively large amount of DNA if the entire genome is to be examined. It has been suggested that the most reliable method for characterising allelic imbalances should have the ability not only to provide locus-specific genotype but also to quantify accurately the copy number of each allele (8). Recently, oligonucleotide microarrays designed for whole genome, SNP analysis have been developed, which are proving to be a powerful methodology. The SNP mapping array (SNPA), Affymetrix HuSNP GeneChip containing 1494 SNPs, was initially applied for the detection of sequence polymorphisms (58). Several studies have subsequently utilised the Affymetrix HuSNP GeneChip to identify LOH or allelic imbalance in tumour tissues from breast, bladder, prostate and small-cell lung cancer (59–62). Array-based CGH is available and provides a high resolution for CN analysis, but it cannot readily detect chromosomal LOH without CN change (63).

Development of the high-density SNPA (GeneChip Human Mapping Array by Affymetrix) started with the introduction, in 2004, of a chip that could genotype more than 10,000 SNPs. Currently, a version of a chip containing up to 900,000 SNPs is available. The GeneChip Human Mapping Array is a high-resolution method for genome-wide screening and a powerful technique to detect genetic aberrations, such as LOH and CN alterations related to deletion or amplification in primary solid tumours, cancer cell lines, and leukaemias (8,63–65). Importantly, this array can characterise LOH due to acquired isodisomy (AID), which reveals the presence of LOH with no alteration of CN. This technology has been described as a fast, cost-effective, and reliable approach for whole genome AID screening (66). A previous study by the Leukaemia Research Group at Newcastle University used SNPA analysis to detect LOH using lymphoblasts from a limited cohort of children with ALL and compared samples obtained at presentation to those obtained at relapse (64). This study was extended to observe a larger group of 78 patients at presentation and 12 at relapse, and it revealed aberrations on chromosome 9p in 26% of these patients (67). Also, they found that genomic deletion analysis showed a high degree of concordance between the 3 techniques (SNPA, array CGH, and FISH) but some deletions detected by array CGH were below the resolution of FISH (67). Genome-wide SNP analysis provides a good approach for obtaining more genetic information about the cancer cells and targeting the involvement of candidate genes in leukaemogenesis. A study of genome-wide analysis on childhood ALL by Kuiper et al. reported that loss of the 9p21.3 region was the most common lesion and revealed both CDKN2A and CDKN2B as functional candidates (68). They also found loss of the 9p13.2 region that affected only PAX5 in 20% (8 of 40) of childhood ALL cases. Furthermore, this array analysis has the potential to reveal novel genetic findings involved in the multistep development of cancer. A study by Mullighan et al. revealed new, recurrent genetic alterations in childhood ALL, which indicates the power of SNP analysis in the study of cancer (69). Recently, a comprehensive analysis utilising 5 separate assays (mutation, methylation, SNP, array CGH and FISH) has been reported to assess the inactivation of the target gene CDKN2A in childhood ALL (67). They discovered that CDKN2A deletion is a significant secondary abnormality in childhood ALL, which strongly correlated with the phenotype and genotype.

SNPA technology has an additional advantage over cytogenetics, array CGH, and FISH in that it can identify LOH associated with no CN alterations, which indicate the presence of copy number-neutral (CNN) LOH (66). This phenomenon has revealed the involvement of AID. A recent report using high-density SNP analysis in a cohort of 98 patients with childhood ALL revealed CNN LOH (also known as AID
and uniparental disomy, UPD) in 8% of patients (67). AID is a frequent alternative mechanism for allelic imbalance in a variety of cancers and occurs when part or all of both chromosomes of an individual pair are derived from only one parent. However, AID is a rare occurrence in the general population (70). Identification of large regions of homozygosity, which did not correspond to deletions by SNPA analysis, were found in approximately 20% of AML cases but not in the DNA of remission cases, suggesting that it was an acquired abnormality (65). This same study also confirmed the presence of a normal CN by FISH within the homozygous regions, which then suggested that these areas consisted of AID occurring as a result of mitotic recombination. One of the mechanisms that can cause phenotypic abnormalities due to AID is homozygosity for an autosomal recessive gene. This event, as an acquired abnormality, can be correlated with homozygosity of a particular gene, suggesting that mitotic recombination acts to remove the wild type allele (71). Previous studies have shown that this mechanism of AID is associated with the presence of homozygous mutations of JAK2 in myeloproliferative disease (72) and CEBPA, FLT3, and RUNX1 in AML (73), suggesting that it is an important, novel mechanism for oncogene activation and/or tumour suppressor gene inactivation.

Comparing CN data obtained from SNPA with data obtained from conventional cytogenetics and FISH proved that SNPA is an efficient technique for CN estimation of individual chromosomes in individual patient samples. This SNPA is of great potential value in cases where cells fail to divide for the performance of standard cytogenetics and for centres where access to cytogenetic services is unavailable. It also validates DNA analysis for the majority of tumour types that do not yield suitable material for karyotyping. Within the past year, new computer software tools have been developed by Affymetrix for data analysis of SNPA; these tools provide better and easier interpretation of results when compared with the manipulation of SNPA data in Excel spreadsheets. These application tools, available free from the Affymetrix website (http://www.affymetrix.com), include CNAT version 4.0, Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) Analysis Tool (BAT), and Integrated Genome Browser (IGB). CNAT 4.0 can be used to perform analysis on data from 10 K, 100 K, and 500 K arrays. It also allows combination of data from different sized arrays (virtual array). BRLMM is an improved genotype calling algorithm that allows more accurate SNP calls and can be used with 100 K and 500 K array sets. IGB is a powerful tool linked to the Ensemble database to help researchers visualise their results against the entire human genome. The IGB software can compare multiple genomic annotation sets from various data sources and, thus, is very useful for targeting areas of interest. SNPA technology has been described as a promising approach to link oncogenic pathways and to initiate the search for targets that could be exploited in the development of molecular therapeutics (56).

Conclusion

SNPA technology is a useful method to determine LOH or allelic imbalance in childhood ALL. Frequent occurrences of AID in childhood ALL may implicate this event as contributing to the presence of allelic imbalance, but the mechanisms involved in this pathogenesis remain elusive. This genome-wide screening approach is believed to develop into a powerful method for the identification of molecular targets which may be important in tumourgenesis in childhood ALL. SNPA technology has the strength to provide global analysis of CN alterations in human cancers and is capable of revealing the occurrence of LOH due to AID, which would be missed by conventional cytogenetics and CGH. SNPA technology has been shown to have accuracy, reproducibility, and improved calling algorithms. Screening the entire genome using this array requires only 250 ng of genomic DNA. This technology may provide crucial data to target any genes associated with leukaemogenesis as well as provide information to gain a better understanding for the mechanisms by which CN changes occur. These possible genes of interest may be important for future molecular therapy.

Using genome-wide SNPA, a large number of candidate genes have been found; however, it is difficult to select those genes for further investigation. Thus, priority has been given to genes which are known to be associated with tumourgenesis and located in common regions of deletion, amplification, or copy-neutral LOH (AID). These common aberrant regions revealed by SNPA must be validated by other methods, such as microsatellite analysis to confirm a LOH event, and quantitative real-time PCR to confirm the presence of gene amplification. This approach could facilitate pinpointing critical regions within the genome as well as candidate genes in those regions.
Acknowledgements

The author would like to thank her supervisory team, Professor Andrew Hall, Dr Julie Irving, and Ms Marian Case from the Northern Institute for Cancer Research, Newcastle University, United Kingdom, for reviewing the contents of this paper. A part of this paper was written during the author’s PhD studies at Newcastle University, United Kingdom.

Correspondence

Dr Sarina Sulong
PhD (Newcastle University)
Human Genome Centre
School of Medical Sciences
Universiti Sains Malaysia Health Campus
16150 Kubang Kerian, Kelantan, Malaysia
Tel: +609-7657922
Fax: +609-7658914
Email: ssarina@kck.usm.my

References


