

Minireview

# Identification of structural aberrations in cancer by SNP array analysis

Stefan Heinrichs and A Thomas Look

Address: Dana-Farber Cancer Institute, Department of Pediatric Oncology, Binney Street, Boston, MA 02215, USA.

Correspondence: A Thomas Look. Email: thomas\_look@dfci.harvard.edu

Published: 31 July 2007

*Genome Biology* 2007, **8**:219 (doi:10.1186/gb-2007-8-7-219)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/7/219>

© 2007 BioMed Central Ltd

## Abstract

Recent studies using single-nucleotide polymorphism arrays have pinpointed novel oncogenes and tumor suppressors involved in specific types of human cancers.

One of the most daunting, though rewarding, challenges in cancer medicine is to determine how specific genetic alterations in tumors may affect the prognosis and lead to targeted therapies for the individual cancer patient. Current methods of gene-expression profiling have revealed that tumor types previously thought to be homogenous from histological criteria alone often have different underlying molecular signatures [1-3]. Complex mutational events seem to have a major impact on the expression of specific genes that contribute to the induction and progression of cancer, and, therefore, on the aggressiveness of the tumor and the clinical outcome of therapy [3-5]. The precise assessment of tumor-cell heterogeneity has thus become a central focus of cancer investigations. The ultimate goal of these efforts is to identify disease subtypes that are driven by altered signaling pathways whose genetic defects correlate well with prognosis and that offer attractive targets for molecular intervention [6-11].

The longest-established method of diagnosing and differentiating tumor types is the detection of chromosomal aberrations by cytogenetic analysis. Molecular cytogenetic techniques, such as spectral karyotyping, fluorescence *in situ* hybridization and chromosome-based comparative genomic hybridization (CGH), substantially improved resolution and genome coverage compared with conventional cytogenetics. But these techniques still did not offer the resolution and genome coverage of microarray gene-expression profiling. This can provide clinically significant insights into the

heterogeneity of tumor cells and has been used to subclassify various human tumors [1-3,12], but it can sometimes be difficult to identify the truly relevant genes among the multiplicity of differences in gene expression recorded. Genomic methods that identify mutations directly and cover the whole genome at a similarly high resolution are required to help resolve such problems.

One attempt to improve the detection of structurally altered genomic regions combines classic CGH with the microarray platform, generating the array CGH technique, which relies on competitive hybridization of fragmented, labeled tumor DNA together with fragmented, but differentially labeled control DNA [13,14]. The microarray platform facilitates higher-resolution mapping of genomic regions that contain copy-number aberrations, such as amplifications and deletions, and the interpretation of data from array CGH studies is much more straightforward than that of conventional CGH. Another new microarray-based cytogenetic technique, high-resolution single-nucleotide polymorphism (SNP) array analysis, perhaps holds even greater promise for detailed structural examination of the cancer genome. SNP arrays allow the high resolution detection of loss of heterozygosity, a common event in tumorigenesis, in addition to the identification of DNA copy-number aberrations at a resolution similar to that of array CGH. A recent study of childhood acute lymphoblastic leukemia (ALL) by Mullighan *et al.* [15] illustrates the strength of SNP arrays for the identification of key genetic abnormalities in cancer.

### Advantages of SNP array analysis

A SNP is defined as a DNA sequence variation at one specific position in the genome that occurs in at least 1% of the human population. Almost all SNPs have only two alleles, and so the heterozygous genotype and the two types of homozygous genotypes can generally be unambiguously determined. On current microarray platforms, 300,000 to 500,000 SNPs can be genotyped simultaneously. Ideally, the tumor sample is analyzed in parallel with a normal - or 'germline' - sample from the same patient; if such a control sample is unavailable, algorithms can be used instead [16]. However, with this approach, the resolution will be lowered, and the data interpretation could be hampered due to the extensive somatic variation in copy number within human populations (so-called copy number variation, or CNV) [17]. As the signal obtained for each position on the array is quantitative, DNA copy number can be determined from it. At the same time, a discrete genotype designation is generated that can be used to detect regions of loss of heterozygosity by comparison with the patient's germline DNA. Loss of heterozygosity means the loss of one allele at a given position (or positions); it is classically associated with tumorigenesis when a 'good' copy of a tumor suppressor gene is physically lost as a result of the deletion of a chromosome or a chromosomal region, leaving the cancer cell with only one (usually defective) allele.

Copy-number analysis by comparison to a matched normal DNA control for each patient's tumor will rapidly detect gene amplification, low-copy gain and deletion with a high degree of confidence, even at the level of a single-copy gain or loss (Figure 1). To identify regions of loss of heterozygosity, one must infer genotype calls from a string of adjacent heterozygous SNPs, because homozygous germline genotypes are noninformative.

Most commonly, the loss of heterozygosity in tumor cells is a result of deletion of a region of a chromosome or of a whole chromosome, and SNP arrays identify these deleted regions as having loss of heterozygosity combined with a copy-number reduction. Loss of heterozygosity can, however, appear without a copy-number change - copy-neutral loss of heterozygosity. For example, a mutated tumor suppressor allele and its surrounding DNA can be copied and replace the other allele by somatic homologous recombination during the development of the neoplastic clone, resulting in a tumor cell that is homozygous for the mutated tumor suppressor allele and has a growth or survival advantage. This type of mutational event is known as uniparental disomy (UPD) and represents an important but largely overlooked mechanism for generating loss of heterozygosity. One of the advantages of SNP microarrays is that they are unique among genomic analysis methods in being able to identify UPD.

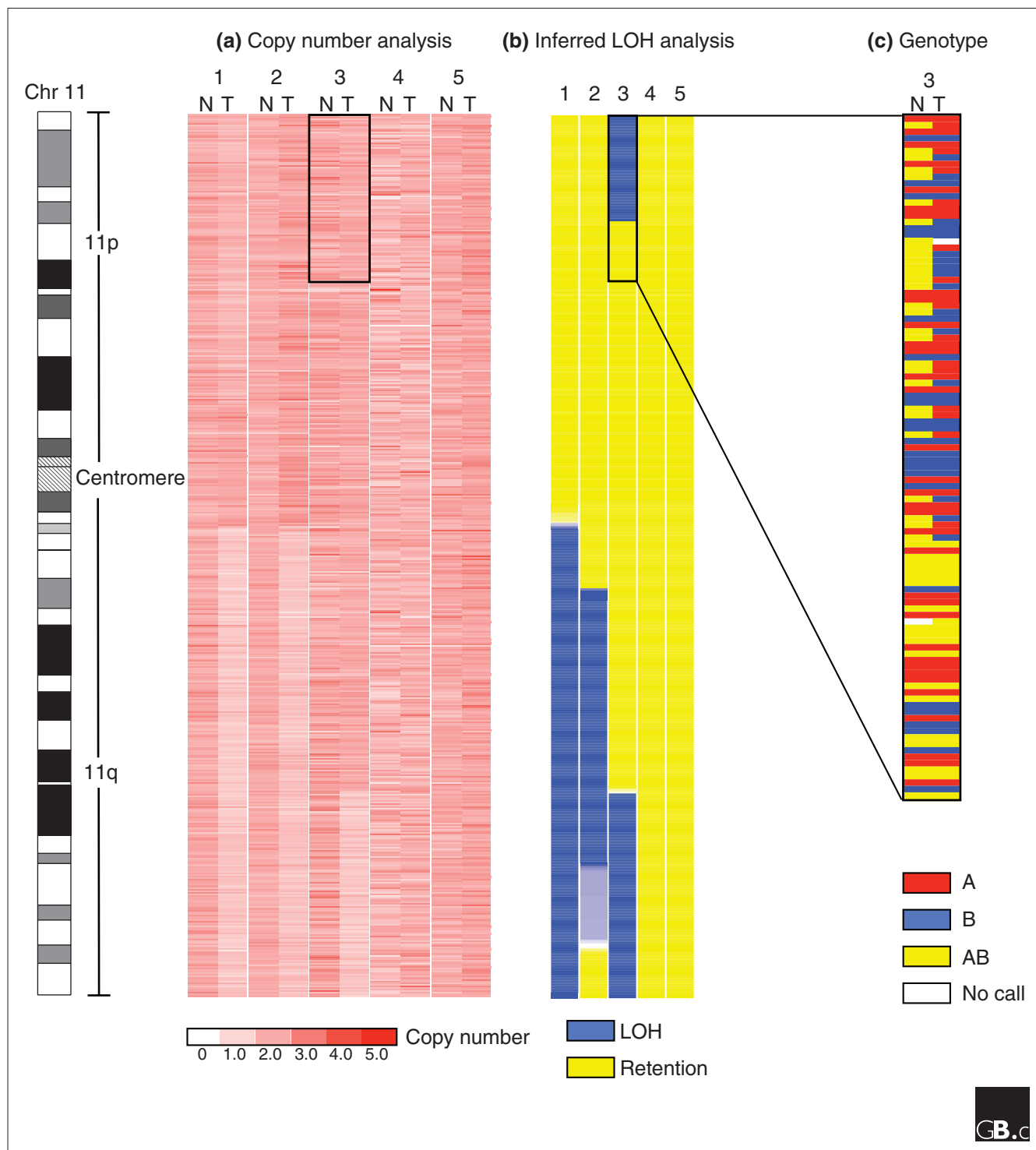
The study by Mullighan *et al.* [15] nicely illustrates the advantages of SNP arrays. The authors analyzed 192 cases of

pediatric B-cell-progenitor acute lymphoblastic leukemia (B-ALL), 94% of which had a matched control sample from a time when the patient's leukemia was in remission. Recurrent chromosomal abnormalities are a hallmark of early B-ALL and the karyotype is, therefore, used to classify subtypes of the disease [18]. Copy-number analysis of the B-ALL cases by Mullighan *et al.* [15] revealed an overall prevalence of deletions in all subgroups except the hyperdiploid cases (cases with more than 50 chromosomes in the leukemic clone), in which gains dominated.

The highest frequency of deletions was found in hypodiploid cases (cases with less than 45 chromosomes in the leukemic clone), and in cases in which the *ETV6* gene (on chromosome 12) and the *RUNX1* gene (on chromosome 21; both genes encode transcription factors) were fused as the result of a translocation. A deletion involving *ETV6* was detected in 33 of 46 cases also harboring this translocation between chromosomes 12 and 21. By contrast, cases with rearrangements affecting the *MLL* gene had a very low frequency of deletions and almost no amplifications. Altogether, the study identified more than 40 regions that were recurrently deleted in different patients, with three focal segments of chromosome 9 showing the highest overall frequency of deletions. At 9p21.3, a third of all cases had deletions in the tumor suppressor locus *CDKN2A* (encoding both p14-ARF and p16-INK4A), often occurring in the context of a region of UPD. A fifth of cases had a deleted *MLL* translocation partner gene *MLLT3* (*AF9*), located on 9p21. More than a quarter of the cases (56 of 192) showed a deletion at 9p13.2, a locus not previously identified as being involved in B-ALL.

Some informative cases had very focused deletions that pinpointed the *PAX5* gene as the likely target on chromosome band 9p13.2 [15]. Indeed, sequencing and functional studies by Mullighan *et al.* [15] led to the identification of *PAX5* as a highly tumor type-specific tumor suppressor gene in early B-cell lineage ALL. *PAX5* encodes a transcription factor that drives the differentiation of progenitor B cells by repressing self-renewal programs and activating genes specific for the B-cell lineage [19]. Mullighan *et al.* [15] found that haploinsufficiency rather than total loss of *PAX5* function predominated; the deletions were accompanied by mutation of the remaining allele in only a minority of cases and two cases were identified that had a heterozygous mutation without a deletion. Other genes involved in B-cell development were found to be deleted in some cases, including *EBF1*, a transcription factor obligatory for B-progenitor cell differentiation. Six of eight cases showed very focused deletions that affected only the *EBF1* locus and, therefore, were not detectable by conventional cytogenetic analysis.

The identification of *PAX5* and *EBF1* as new mutational targets in early B-lineage leukemogenesis shows the value of SNP array studies for selecting genes for detailed analysis. Like *PAX5*, the *EBF1* gene retained one wild-type



**Figure 1**  
 Illustration of SNP array analysis by example of matched neuroblastoma samples using the dChip software [25,26]. Normal (N) and tumor (T) DNA of five selected patients were hybridized to 10K Affymetrix SNP arrays (data kindly provided by R George [22]). **(a)** Copy numbers are shown as shades of red. Sample 1, 2 and 3 show a copy-number loss on 11q, whereas samples 4 and 5 are normal. **(b)** The inferred comparison of the genotype (loss of heterozygosity (LOH) analysis) results in a single lane per case, in which regions of LOH are depicted in blue and heterozygous regions are in yellow. Besides classical LOH with copy-number loss (11q region of samples 1-3), a region of UPD, identified by copy-neutral LOH, is identified in sample 3 on 11p. **(c)** The actual genotype calls for the UPD region and part of the adjacent region of sample 3 are shown in expanded form. The region of UPD shows only red (A) or blue (B) SNP calls, whereas other regions have the expected numbers of retained heterozygous alleles resulting in an AB call (yellow).

allele in the majority of the cases, supporting the idea that haploinsufficiency is an inherent property of some tumor suppressors [20,21]. In cases with defects in such genes, it may be possible to increase gene expression from the remaining allele.

Other work has also shown the power of SNP array analysis to identify the loss of functional tumor suppressors even in cases lacking chromosomal deletions, or gain of regions containing potential oncogenes. We have performed a matched control study by SNP array of 22 neuroblastoma patients [22] and identified chromosomal aberrations that had been previously implicated in neuroblastoma by more laborious analysis of loss of heterozygosity at individual loci. A subset of four cases showed loss of heterozygosity of 11p solely as a result of UPD, indicating that cells might not tolerate the haploinsufficiency generated by large deletions of some chromosomal regions. A matched control study of 14 basal cell carcinomas by Teh *et al.* [23] revealed that, in almost all cases, the region on chromosome 9q harboring the tumor suppressor gene *PTCH1* has undergone loss of heterozygosity. More than a third of these cases resulted from UPD, implying the duplication of a mutated allele.

Sellers and colleagues [24] have taken a different approach to exploiting the information provided by SNP arrays. To uncover novel signaling pathways in human cancers, they first examined the structural genomic aberrations of a cell line panel by SNP array copy-number analysis. Clustering of the cell lines according to their copy-number aberrations identified subgroups that showed amplifications and deletions in shared regions. One cluster, comprising six out of nine melanoma cell lines, showed a copy-number gain in a defined region of chromosome 3p. Comparison of the gene-expression profiles of the six melanoma cell lines with the other lines identified a small set of genes as highly expressed, only one of which, that encoding transcription factor MITF, was located within the chromosome 3p region. Additional studies established that MITF is a survival factor with oncogenic properties in melanoma.

Thus, SNP array technology can provide a global analysis of DNA copy-number alterations in human cancers while revealing important loss of heterozygosity due to UPD, which would be entirely missed by conventional cytogenetic analysis or array CGH. Identification of UPD in tumor cells allows genetically similar cases to be classified together for prognostic and therapeutic purposes in the absence of a cytogenetically apparent deletion. In addition, the finding of a UPD implies that a significant mutational or heritable epigenetic event has occurred within the duplicated region, thus providing a good reason for further detailed analysis at the DNA sequence level.

A cross comparison of all cases included in a SNP array study makes it possible to define shared regions of

copy-number change, loss of heterozygosity and UPD and to delineate both minimally deleted and minimally amplified regions. Thus, SNP array studies can pinpoint critical structurally altered regions within the genome of a particular type of cancer and contribute to the discovery of novel oncogenes or tumor suppressors, as shown by the study of Mullighan *et al.* [15]. The potential oncogenic function of genes located in amplified regions that are also overexpressed in the tumor cells can be tested functionally in animal models.

Ultimately, SNP array analysis should provide a way to reliably subclassify tumors on the basis of shared genetic abnormalities, so that patients can be assigned to the most appropriate therapies. This technology also seems especially promising as a way of implicating oncogenic pathways and initiating the search for targets that could be exploited in the development of molecular therapeutics. For a protein to be a useful therapeutic target within the cancer cell, it must have a driving role in a pathway controlling tumor initiation, the maintenance of the malignant phenotype or metastatic behaviors. Tumors acquire multiple critical genetic aberrations before they become clinically apparent, and, by the use of powerful technologies, such as SNP analysis and eventually whole genome resequencing, it should then be possible to target several of these defects to reverse tumor growth.

## Acknowledgements

We thank John R Gilbert and Rima V Kulkarni for editorial assistance.

## References

1. Ferrando AA, Neuberger DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, *et al.*: **Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia.** *Cancer Cell* 2002, **1**:75-87.
2. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, *et al.*: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
3. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltner JM, *et al.*: **The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.** *N Engl J Med* 2002, **346**:1937-1947.
4. Look AT: **Oncogenic transcription factors in the human acute leukemias.** *Science* 1997, **278**:1059-1064.
5. Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, Khazi D, McGrady PW, Seeger RC, Look AT, Shimada H, *et al.*: **Chromosome 1p and 11q deletions and outcome in neuroblastoma.** *N Engl J Med* 2005, **353**:2243-2253.
6. Kohn EC, Lu Y, Wang H, Yu Q, Yu S, Hall H, Smith DL, Meric-Bernstam F, Hortobagyi GN, Mills GB: **Molecular therapeutics: promise and challenges.** *Semin Oncol* 2004, **31**(1 Suppl 3):39-53.
7. Sawyers CL: **Making progress through molecular attacks on cancer.** *Cold Spring Harb Symp Quant Biol* 2005, **70**:479-482.
8. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, Deininger MW, Silver RT, Goldman JM, Stone RM, *et al.*: **Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia.** *N Engl J Med* 2006, **355**:2408-2417.
9. Greulich H, Chen TH, Feng W, Janne PA, Alvarez JV, Zappaterra M, Bulmer SE, Frank DA, Hahn WC, Sellers WR, *et al.*: **Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants.** *PLoS Med* 2005, **2**:e313.

10. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, et al.: **MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling.** *Science* 2007, **316**:1039-1043.
11. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, et al.: **Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib.** *N Engl J Med* 2004, **350**:2129-2139.
12. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al.: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**:531-537.
13. Maser RS, Choudhury B, Campbell PJ, Feng B, Wong KK, Protopopov A, O'Neil J, Gutierrez A, Ivanova E, Perna I, et al.: **Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers.** *Nature* 2007, **447**:966-971.
14. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al.: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207-211.
15. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, et al.: **Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia.** *Nature* 2007, **446**:758-764.
16. Beroukhi R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, Fox EA, Hochberg EP, Mellinghoff IK, Hofer MD, et al.: **Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays.** *PLoS Comput Biol* 2006, **2**:e41.
17. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al.: **Global variation in copy number in the human genome.** *Nature* 2006, **444**:444-454.
18. Armstrong SA, Look AT: **Molecular genetics of acute lymphoblastic leukemia.** *J Clin Oncol* 2005, **23**:6306-6315.
19. Cobaleda C, Schebesta A, Delogu A, Busslinger M: **Pax5: the guardian of B cell identity and function.** *Nat Immunol* 2007, **8**:463-470.
20. Fero ML, Randel E, Gurley KE, Roberts JM, Kemp CJ: **The murine gene p27Kip1 is haplo-insufficient for tumour suppression.** *Nature* 1998, **396**:177-180.
21. Ma L, Teruya-Feldstein J, Behrendt N, Chen Z, Noda T, Hino O, Cordon-Cardo C, Pandolfi PP: **Genetic analysis of Pten and Tsc2 functional interactions in the mouse reveals asymmetrical haploinsufficiency in tumor suppression.** *Genes Dev* 2005, **19**:1779-1786.
22. George RE, Attiyeh EF, Li S, Moreau LA, Neuberg D, Li C, Fox EA, Meyerson M, Diller L, Fortina P, et al.: **Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays.** *PLoS ONE* 2007, **2**:e255.
23. Teh MT, Blaydon D, Chaplin T, Foot NJ, Skoulakis S, Raghavan M, Harwood CA, Proby CM, Philpott MP, Young BD, et al.: **Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key somatic event.** *Cancer Res* 2005, **65**:8597-8603.
24. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhi R, Milner DA, Granter SR, Du J, et al.: **Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma.** *Nature* 2005, **436**:117-122.
25. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C: **dChip-SNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data.** *Bioinformatics* 2004, **20**:1233-1240.
26. **dChip Software: Analysis and Visualization of Gene Expression and SNP Microarrays** [<http://biosun1.harvard.edu/complab/dchip/>]