

A survey of analysis software for array-comparative genomic hybridisation studies to detect copy number variation

Anis Karimpour-Fard,^{1*} Laura Dumas,² Tzulip Phang,³ James M. Sikela² and Lawrence E. Hunter¹

¹Center for Computational Pharmacology, University of Colorado School of Medicine, Aurora, CO 80045, USA

²Department of Pharmacology, Neuroscience and Human Medical Genetics Programs, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045, USA

³Department of Medicine, University of Colorado Denver School of Medicine, Denver, CO 80045, USA

*Correspondence to: Tel.: +1 (303)724 0274; Fax: +1 (303)724 3663; E-mail: anis.karimpour-fard@ucdenver.edu

Date received (in revised form): 27th August 2010

Abstract

Copy number variants (CNVs) create a major source of variation among individuals and populations. Array-based comparative genomic hybridisation (aCGH) is a powerful method used to detect and compare the copy numbers of DNA sequences at high resolution along the genome. In recent years, several informatics tools for accurate and efficient CNV detection and assessment have been developed. In this paper, most of the well known algorithms, analysis software and the limitations of that software will be briefly reviewed.

Keywords: copy number variants, CNV, deletion, insertion, duplication, aCGH

Background

Copy number variants (CNVs) are DNA sequences that are present in different amounts among individuals in a population. Copy number differences can confer a change in gene expression, phenotypic variation, disease susceptibility,^{1–5} and gene and genome evolution.^{6,7} Repetitive sequences that flank a specific genomic region can further facilitate a duplication or deletion of that region via the mechanism of non-allelic homologous recombination, which can occur when paralogous sequences in the genome mis-pair during meiosis.^{8–10} A key method used to study CNVs across individuals is that of array-based comparative genomic hybridisation (aCGH). The goal of aCGH experiments is to detect and compare the copy numbers of DNA sequences at high resolution along the genome. Several informatics tools currently exist for accurate and efficient CNV detection and assessment.

These tools assist in automated analysis of array CGH data and user-friendly copy number reporting for individual samples. The goal of the statistical algorithms used in these software programs is to call aberrations reliably, accurately and precisely.

The analysis of CNVs is broken down into several steps, including: (i) pre-processing and normalisation of the raw data; (ii) aligning data with its genome location, conducting segmentation analysis and providing statistical analysis to ensure the reliability of detection; and (iii) post-processing to assign biological meaning to the different states.

(i) Normalisation of the \log_2 ratios is typically conducted in an attempt to adjust for sources of systematic variation. Since these effects are often not known or measured, most aCGH methodologies incorporate global normalisation techniques, centring the data about the sample mean or median for a given hybridisation.¹¹ Normalisation remains imperfect, and an accurate estimation of

the copy number is unlikely. It is assumed, however, that changes in the observed, normalised \log_2 ratios correspond directly to changes in the true copy numbers.

(ii) From a statistical perspective, segmentation has received most attention, and many different schemes have been proposed. Three main methods include: a) segmenting chromosome-arrayed genotypes into discrete regions, with probes in each region presenting different signal intensity patterns to adjacent regions; and b) labelling particular segments that are inherently different in copy number from their expected value. Segmentation methods seek to identify the locations of \log_2 ratio mean change (ie change points or breakpoints) and to estimate the values of those means. All of these segmentation methods provide breakpoint locations but do not identify the associated genomic alterations as gains or losses. Because a primary objective of aCGH analysis is to identify regions of copy number gain and loss, follow-up methods have been proposed for this from segmentation results. Some of the studies have used a non-parametric estimate of the standard deviation to identify a global threshold for categorising segments.^{12,13} Another approach for identifying gains and losses based on segmentation results entails the combination of identified segments across chromosomes and a subsequent establishment of a no-change baseline.

The challenge with segmentation methods is that, in order to find the optimal segmentation, all possible change-points need to be evaluated, creating a combinatorial explosion. For example, if there are ten copy number segments positioned across 2,000 probe intensities, then there are 2000^{10} places in which these segments may lie (roughly a 1 followed by 33 zeros, or one decillion). Given that a chromosome may have as many as a hundred change-points, and that today's whole-genome arrays contain over 50,000 intensity values for a given chromosome, the search space can easily exceed $50,000^{100}$ possible change-points.¹⁴

This seemingly impossible series of calculations is what has led researchers to adopt various heuristics to make this process computationally viable, as was done with circular binary segmentation (CBS), or by avoiding segmentation altogether.

(iii) Finally, a post-processing step is necessary to assign biological meaning to the different states.

In this paper, a survey is presented of currently available analysis tools for aCGH data to detect copy number variation. Several of these tools provide user-friendly software, with visualisation tools and links to other databases (Table 1). Comparison of these methods is difficult because array platforms differ in probe type, size, varying resolutions and noise levels. A series of methods for performing this analysis are described below.

Hidden Markov model (HMM) and BioHMM

Fridlyand *et al.*¹¹ proposed an unsupervised HMM for identifying copy number changes on chromosomes. Marioni *et al.*¹⁵ described a new segmentation scheme, BioHMM, which extends the HMM approach of Fridlyand *et al.*,¹¹ to take account of the distance between adjacent clones or of clone quality that are likely to affect the segmentation.

CBS

Olshen *et al.*¹⁶ introduced another sophisticated method for aCGH analysis, CBS. This is a modification of the change-point approach, allowing for tertiary splits by connecting the two chromosomal ends. It splits the chromosomes into contiguous regions of equal copy number by modelling discrete copy number gains and losses. It then assesses the significance of the proposed splits by using a permutation reference distribution.

Several studies have shown that CBS is the most efficient method.^{16–19} They have also shown that optimal combination of the smoothing step and the segmentation step may result in improved performance.

Willenbrock and Fridlyand¹⁹ compared three publicly available methods for the analysis of aCGH data — DNACopy (CBS), the gain and loss analysis of DNA Gaussian-based approach (GLAD) and the 'cluster along chromosomes' (CLAC) approach — and they showed that segmentation by any of the

Table 1. Software for array CGH analysis

Software	Cost	Platform	Requirements	Method	Link
CGH Fusion	Cost	L, W	Java	Lowess normalisation, CBS, ratio thresholding	http://www.infoquant.com/index/cghfusion
Agilent Genomic Workbench Standard Edition 5.0.14	Cost	L,W	Java	UCSC Genome Browser Cluster multiple samples into groups based on their aberration profiles Hidden Markov model (HMM) for detecting probe-based aberrations Circular binary segmentation (CBS)	http://www.chem.agilent.com/en-us/products/instruments/dnamicroarrays/dnaanalyticssoftware/pages/default.aspx
infoQuant oneClickCGH	Cost	L,W		CBS-type algorithms, log-ratio thresholding etc	http://www.infoquant.com/index/oneclickcgh
CNAM	Cost	L,W,M		Univariate Multivariate	http://www.goldenhelix.com/SNP_Variation/CNAM/index.html
CytoSure	Cost	L,W,M		Direct thresholding, moving average thresholding, K-means clustering, HMM and CBS	http://www.ogt.co.uk/analysis_software.html
CNV workshop	Free	L,W,M	Java, R, Perl	CBS segmentation algorithm Log ratio for normalisation	http://sourceforge.net/projects/cnv
SnoopCGH	Free	L,W,M	Java	Haar wavelets Smith–Waterman algorithm permutations	http://snoopcgh.sourceforge.net/
PennCNV	Free	L, W		HMM	
QuantiSNP	Free	L,W	MATLAB Linux user (glibc2.6+)	Objective Bayes HMM (OB-HMM)	http://www.well.ox.ac.uk/QuantiSNP/
Ultrasome	Free	L,W	Stand-alone program		http://www.broadinstitute.org/ultrasome
CGHExplorer	Free	L,W,M	J2SEJRE	Thresholding, bootstrap-based method, analysis of copy errors (ACE), clustering along chromosomes (CALC) with false discovery rate (FDR)	http://www.ifi.uio.no/forskning/grupper/bioinf/Papers/CGH/
CGHPRO	Free	L,W	Java My SQL	Quantile smoothing Wavelet-based smoothing	http://www.mcu.edu.tw/department/biotech/en_page/CGcgh/
RJaCGH	Free	L,W,M	R	HMM	http://cran.r-project.org/web/packages/RJaCGH/index.html

Continued

Table 1. Continued

Software	Cost	Platform	Requirements	Method	Link
ADaCGH	Free	L,W,M	R Web-based	CBS, HMM, BioHMM, comparative genomic hybridisation (CGH) segmentation, gain and loss analysis of DNA (GLAD), wavelet-based smoothing, Smith–Waterman algorithm and analysis of copy errors	http://cran.r-project.org/web/packages/ADaCGH/
CLAC	Free	L,W,M	R	Clustering along chromosomes	http://cran.r-project.org/src/contrib/Archive/clac
CGHseg	Free	L,W,M	R	CGH segmentation	
ISACGH	Free	L,W,M	Web based	GLAD, CBS Fisher's exact test GO, KEGG pathway	http://www.gepas.org
aCGH	Free	L,W,M	R	HMM	
GLAD	Free	L,W,M	R	Gaussian-based likelihood Adaptive weights smoothing	http://www.bioconductor.org/packages/2.1/bioc/html/GLAD.html
CBS	Free	L,W,M	R	CBS	CBS algorithm has been implemented in the 'DNAcopy' package
DNAcopy	Free	L,W,M	R	CBS	http://www.bioconductor.org/packages/2.0/bioc/html/DNAcopy.html
snapCGH	Free	L,W,M	R	Limma, GLAD, DNAcopy, tilingArray and aCGH	http://www.bioconductor.org/packages/1.9/bioc/html/snapCGH.html
CGHcall	Free	L,W,M	R	Expectation maximisation (EM)	http://www.bioconductor.org/packages/release/bioc/html/CGHcall.html
cghMCR	Free	L,W,M	R	DNAcopy Minimal common regions (MCRs)	http://www.bioconductor.org/packages/bioc/html/cghMCR.html
BioHMM	Free	L,W,M	R	Heterogeneous HMM	http://www.bioconductor.org/packages/bioc/1.8/html/snapCGH.html
fused lasso	Free	L,W, M	R	Regression	http://cran.r-project.org/web/packages/cghFLasso/
aroma.affymetrix	Free	L,W,M	R	GLAD ChromosomeExplorer	http://www.aroma-project.org/

three methods aids downstream analyses of aCGH data. They also noted that DNACopy had the best operational characteristics in terms of its sensitivity and false discovery rate (FDR) for breakpoint detection, but it was not able to identify single clone aberrations. CBS has been implemented in DNACopy. Applying CBS to the same simulated data set, the authors were able to achieve a 0.06 median FDR with 0.88 sensitivity. Although effective for finding segments, and despite speed optimisations by Venkatraman and Olshen,²⁰ however, CBS is not computationally efficient for whole-genome analysis. For example, analysis on Affymetrix 500K data has been shown to take over 20 minutes per sample and roughly 45 minutes per sample on Illumina 550K data.¹⁴

Bottom-up agglomerative approach

In contrast to the top-down strategy employed by CBS, Wang *et al.* introduced a bottom-up agglomerative approach, CLAC,²¹ which enjoys better computational efficiency. CLAC builds hierarchical clustering-style trees along each chromosome arm (or chromosome), and then selects the ‘interesting’ clusters (genome regions with copy number gains/losses) by controlling the FDR at a certain level.

Multivariate method

The multivariate method segments all samples simultaneously, finding general copy number regions that may be similar across all samples. This method is preferable for finding very small copy number regions, and for finding conserved regions, possibly useful for association studies. The copy number analysis model (CNAM) is a commercial tool that uses two types of segmentation: univariate (on a per-sample basis) and multivariate (on a multi-sample basis).¹⁴

Other methods

While most segmentation methods employ parametric models for array CGH data, some non-parametric approaches that are free of

distribution assumptions have also shown success in calling gains and losses in array CGH data. Hsu *et al.*²² proposed to minimise noise from the array CGH data using wavelets before making inferences on the aberrations. Tibshirani *et al.*²³ developed a spatial smoothing approach using fused lasso regression for calling gains and losses. The regression framework of fused lasso brings great computational efficiency and can be easily generalised to other analyses involving CGH data. Jong *et al.*²⁴ used genetic local search algorithms and Willenbrock *et al.*¹⁹ used the adaptive weights smoothing method, GLADmerge (a modified version of GLAD²⁵), for combining segments obtained from GLAD, first within and then across chromosomes through hierarchical clustering in which clusters of segments are identified from the resultant dendrograms.

An ideal tool for the analysis of aCGH data should allow the user to choose among several of the algorithms. For the end-users, the web-based applications are the most suitable, since they do not require software installation and there are no concerns about the hardware. Some of the available tools are analysis of array-based comparative genomic hybridisation (ADaCGH)²⁶ and in silico array-CGH (ISACGH)²⁷ (see Table 1). Some of the tools are implemented in MATLAB or an executable file with a very simple interface which guides the user through the analysis (Table 1). A very helpful feature that exists in some of the tools is the ability to estimate the statistical significance of the detected copy number changes and then rank them accordingly. R (<http://www.r-project.org>) also has several packages for the analysis of aCGH data.

R packages

R is a powerful, yet flexible, statistical computing/programming environment. Its object-orientation programming scheme has made algorithm development easy and flexible and has attracted a huge developer community. R is platform independent, and works on all major computer operating systems. R has several packages for the analysis of aCGH data. These packages are freely available at the Comprehensive R Archive Network (CRAN) section of the website.²⁸

They include a CBS method (DNACopy)^{16,29} an unsupervised HMM approach¹¹ GLAD,²⁵ cghMCR,³⁰ the CLAC and method using the hierarchical clustering algorithm,²¹ a penalised least-squares regression³¹ and the wavelet approach.²² BioHMM is another integral part of the segmentation, normalisation and processing of aCGH data (snapCGH)³² R library.¹⁵ This library lets the user apply other segmentation schemes using common input and output data objects. Additionally, snapCGH works seamlessly with limma objects³³ and enables the use of pre-processing (and other) functions therein. RAN-aCGH is an R graphical user interface (GUI) for analysis and visualisation of aCGH data and includes several of the packages in R.³⁴ There are also a number of web-based applications, such as ADaCGH²⁶ and ISACGH,²⁷ for viewing and comparing outputs from multiple algorithms (Table 1).

Discussion and conclusion

Most of the methods do well in detecting the existence and the width of aberrations for large changes and high signal-to-noise ratio. None of the algorithms, however, reliably detected aberrations with small width and low signal-to-noise ratio.^{35–38} Several previous studies have compared the performance of these methods, as well as the segmentation schemes.^{17,19,39}

Lockwood *et al.*³⁹ reviewed 16 different tools that were used in visualisation or analysis of ACGH data.

Lai *et al.*¹⁷ compared ten different methods and found that HMM¹¹ performed poorly, with a high false-positive rate (~ 0.40 – 0.60) and low sensitivity (~ 50 – 80 per cent) with copy number segments.¹⁷ These authors showed that DNACopy²⁹ generally performed better than GLAD²⁵ and HMM with regard to detection of copy number alterations. Their results also indicated that HMM performed best for small aberrations (given a sufficient signal-to-noise ratio), and that GLAD did better than HMM for wider aberrations.¹⁷ They showed that simple smoothing algorithms such as lowess and wavelets are the fastest, and the HMM and CBS¹⁶ were the slowest. They also noted that only

CLAC²¹ and the array CGH expression integration tool (ACE)⁴⁰ incorporate the FDR. They also noted that some of the segmentation methods, such as CGHseg⁴¹ and CBS, consistently performed well.¹⁷

Wang compared several different segmentation methods and found that CGHseg appeared to be overly sensitive to outlier measurements, and thus would be more suitable for detecting single gene copy number changes.¹⁸ Her result showed that CLAC was conservative in handling outliers with opposite signs in the same alteration region and therefore tended to break large alteration segments into small blocks. CBS provided clean solutions for segmentation but had the limitation of detecting break points whose alteration signals were weak.¹⁸

The few early methods employed automatically to call gains and losses from aCGH data involved smoothing the \log_2 ratio vectors followed by applying certain thresholds.^{38,42,43} A common drawback of these methods was not taking into account the biological covariates, such as the distance between adjacent clones or clone quality, which are likely to affect the segmentation (ie some regions of the genome being densely covered, while others have larger gaps between probes).

aCGH analysis has come a long way, and the software packages have become more accurate and user friendly, but we are likely to see even more improvements in these software packages in the future.

Acknowledgments

This study was supported by NIH 5T15LM009451-03 for A.K.E, 5R01MH81203-2 and 2R01AA11853-11 for J.M.S. and L.D., P30CA46934 for T.P. and R01LM009254 and R01LM008111 for L.E.H.

References

1. Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J. *et al.* (2006), 'Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans', *Nature* Vol. 439, pp. 851–855.
2. Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R. *et al.* (2007), 'FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity', *Nat. Genet.* Vol. 39, pp. 721–723.
3. Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A. *et al.* (2005), 'The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility', *Science* Vol. 307, pp. 1434–1440.

4. McCarroll, M.E., Shi, Y., Harris, S., Puli, S. *et al.* (2006), 'Computational prediction and experimental evaluation of a photoinduced electron-transfer sensor', *J. Phys. Chem. B* Vol. 110, pp. 22991–22994.
5. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E. *et al.* (2007), 'Relative impact of nucleotide and copy number variation on gene expression phenotypes', *Science* Vol. 315, pp. 848–853.
6. Dumas, L., Kim, Y.H., Karimpour-Fard, A., Cox, M. *et al.* (2007), 'Gene copy number variation spanning 60 million years of human and primate evolution', *Genome Res.* Vol. 17, pp. 1266–1277.
7. Fortna, A., Kim, Y., MacLaren, E., Marshall, K. *et al.* (2004), 'Lineage-specific gene duplication and loss in human and great ape evolution', *PLoS Biol.* Vol. 2, p. E207.
8. Hurles, M. (2005), 'How homologous recombination generates a mutable genome', *Hum. Genomics* Vol. 2, pp. 179–186.
9. Sharp, A.J., Cheng, Z. and Eichler, E.E. (2006), 'Structural variation of the human genome', *Annu. Rev. Genomics Hum. Genet.* Vol. 7, pp. 407–442.
10. Shaw, C.J. and Lupski, J.R. (2004), 'Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease', *Hum. Mol. Genet.* Vol. 13 (Spec No 1), pp. R57–R64.
11. Fridlyand, J., Snijders, A.M., Pinkel, D., Albertson, D.G. *et al.* (2004), 'Hidden Markov models approach to the analysis of array CGH data', *J. Multivar. Anal.* Vol. 90, pp. 132–153.
12. Paris, P.L., Andaya, A., Fridlyand, J., Jain, A.N. *et al.* (2004), 'Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors', *Hum. Mol. Genet.* Vol. 13, pp. 1303–1313.
13. Rossi, M.R., Gaile, D., Laduca, J., Matsui, S. *et al.* (2005), 'Identification of consistent novel submegabase deletions in low-grade oligodendrogliomas using array-based comparative genomic hybridization', *Genes Chromosomes Cancer* Vol. 44, pp. 85–96.
14. CNAM. Available at http://www.helixtree.com/SNP_Variation/CNAM/index.html
15. Marioni, J.C., Thorne, N.P. and Tavare, S. (2006), 'BioHMM: A heterogeneous hidden Markov model for segmenting array CGH data', *Bioinformatics* Vol. 22, pp. 1144–1146.
16. Olshen, A.B., Venkatraman, E.S., Lucito, R. and Wigler, M. (2004), 'Circular binary segmentation for the analysis of array-based DNA copy number data', *Biostatistics* Vol. 5, pp. 557–572.
17. Lai, W.R., Johnson, M.D., Kucherlapati, R. and Park, P.J. (2005), 'Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data', *Bioinformatics* Vol. 21, pp. 3763–3770.
18. Wang, P. (2009), *Algorithms for Calling Gains and Losses in Array CGH Data*, Vol. 556. Humana Press, Hatfield, UK.
19. Willenbrock, H. and Fridlyand, J. (2005), 'A comparison study: Applying segmentation to array CGH data for downstream analyses', *Bioinformatics* Vol. 21, pp. 4084–4091.
20. Venkatraman, E.S. and Olshen, A.B. (2007), 'A faster circular binary segmentation algorithm for the analysis of array CGH data', *Bioinformatics* Vol. 23, pp. 657–663.
21. Wang, P., Kim, Y., Pollack, J., Narasimhan, B. *et al.* (2005), 'A method for calling gains and losses in array CGH data', *Biostatistics*, Vol. 6, pp. 45–58.
22. Hsu, L., Self, S.G., Grove, D., Randolph, T. *et al.* (2005), 'Denoising array-based comparative genomic hybridization data using wavelets', *Biostatistics* Vol. 6, pp. 211–226.
23. Tibshirani, R. and Wang, P. (2008), 'Spatial smoothing and hot spot detection for CGH data using the fused lasso', *Biostatistics* Vol. 9, pp. 18–29.
24. Jong, K., Marchiori, E., Meijer, G., Vaart, A.V. *et al.* (2004), 'Breakpoint identification and smoothing of array comparative genomic hybridization data', *Bioinformatics* Vol. 20, pp. 3636–3637.
25. Hupe, P., Stransky, N., Thiery, J.P., Radvanyi, F. *et al.* (2004), 'Analysis of array CGH data: From signal ratio to gain and loss of DNA regions', *Bioinformatics* Vol. 20, pp. 3413–3422.
26. Diaz-Uriarte, R. and Rueda, O.M. (2007), 'ADaCGH: A parallelized web-based application and R package for the analysis of aCGH data', *PLoS One* Vol. 2, p. e737.
27. Conde, L., Montaner, D., Burguet-Castell, J., Tarraga, J. *et al.* (2007), 'ISACGH: A web-based environment for the analysis of Array CGH and gene expression which includes functional profiling', *Nucleic Acids Res.* Vol. 35 (Web Server issue), pp. W81–W85.
28. The Comprehensive R Archive Network. Available at <http://www.bioconductor.org/CRAN/>
29. DNACopy. 'DNA copy number data analysis'. Available at <http://bioconductor.org/packages/19/bioc/html/DNACopy.html>
30. Aguirre, A.J., Brennan, C., Bailey, G., Sinha, R., Feng, B. *et al.* (2004), 'High resolution characterization of the pancreatic adenocarcinoma genome', *Proc. Natl. Acad. Sci. USA*, Vol. 101, No. 24, pp. 9067–9072.
31. Huang, T., Wu, B., Lizardi, P. and Zhao, H. (2005), 'Detection of DNA copy number alterations using penalized least squares regression', *Bioinformatics* Vol. 21, pp. 3811–3817.
32. Smith, M.L., Marioni, J.C., Hardcastle, T.J. and Thorne, N.P. (2006), 'snapCGH: Segmentation, normalization and processing of aCGH data', *Bioconductor Users Guide*.
33. Smyth, G.K., Michaud, J. and Scott, H.S. (2005), 'Use of within-array replicate spots for assessing differential expression in microarray experiments', *Bioinformatics* Vol. 21, pp. 2067–2075.
34. Kim, S. and Kim, B.-S. (2007), 'RAN-aCGH: R GUI tools for analysis and visualization of an array-CGH experiment', *Genomics Informatics* Vol. 5, pp. 137–139.
35. Ferreira, B.I., Alonso, J., Carrillo, J., Acquadro, F. *et al.* (2008), 'Array CGH and gene-expression profiling reveals distinct genomic instability patterns associated with DNA repair and cell-cycle checkpoint pathways in Ewing's sarcoma', *Oncogene* Vol. 27, pp. 2084–2090.
36. Hodgson, G., Hager, J.H., Volik, S., Hariono, S. *et al.* (2001), 'Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas', *Nat. Genet.* Vol. 29, pp. 459–464.
37. Lassmann, S., Weis, R., Makowicz, F., Roth, J. *et al.* (2007), 'Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal- and microsatellite-unstable sporadic colorectal carcinomas', *J. Mol. Med.* Vol. 85, pp. 293–304.
38. Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A. *et al.* (2002), 'Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors', *Proc. Natl. Acad. Sci. USA* Vol. 99, pp. 12963–12968.
39. Lockwood, W.W., Chari, R., Chi, B. and Lam, W.L. (2006), 'Recent advances in array comparative genomic hybridization technologies and their applications in human genetics', *Eur. J. Hum. Genet.* Vol. 14, pp. 139–148.
40. van Wieringen, W.N., Belien, J.A., Vosse, S.J., Achame, E.M. *et al.* (2006), 'ACE-it: A tool for genome-wide integration of gene dosage and RNA expression data', *Bioinformatics* Vol. 22, pp. 1919–1920.
41. Picard, F., Robin, S., Lavielle, M., Vaisse, C. *et al.* (2005), 'A statistical approach for array CGH data analysis', *BMC Bioinformatics* Vol. 6, p. 27.
42. Cheng, C., Kimmel, R., Neiman, P. and Zhao, L.P. (2003), 'Array rank order regression analysis for the detection of gene copy-number changes in human cancer', *Genomics* Vol. 82, pp. 122–129.
43. Lingjaerde, O.C., Baumbusch, L.O., Liestol, K., Glad, I.K. *et al.* (2005), 'CGH-Explorer: A program for analysis of array-CGH data', *Bioinformatics* Vol. 21, pp. 821–822.