ELSEVIER

Cancer Genetics 206 (2013) 103–115

Cancer Genetics

Analysis of genomic abnormalities in tumors: a review of available methods for Illumina two-color SNP genotyping and evaluation of performance

Kathleen Klein Oros [a], Suzanna L. Arcand [b], Jane Bayani [c], Jeremy A. Squire [d], Anne-Marie Mes-Masson [e,f], Patricia N. Tonin [b,g], Celia M.T. Greenwood [a,h,*]

[a] *Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada;* [b] *The Research Institute of the McGill University Health Centre, Montreal, QC, Canada;* [c] *Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada;* [d] *Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada;* [e] *Centre de recherche du Centre hospitalier de l'Université de Montréal, Institut du cancer de Montréal, Montreal, QC, Canada;* [f] *Department of Medicine, Université de Montréal, Montreal, QC, Canada;* [g] *Departments of Medicine and Human Genetics, McGill University, Montreal, QC, Canada;* [h] *Departments of Oncology and Epidemiology, Biostatistics and Occupational Health, McGill University, QC, Canada*

Several methods have recently been proposed for identifying copy number alterations (CNAs) in genomic DNA from tumors, using the signals arising from two-color genotyping technologies. Although copy number estimation in normal tissue has been well studied, methods developed for normal tissue tend to perform poorly when applied to tumors, due to normal cell contamination, varying levels of ploidy, and genetic heterogeneity within the tumor. Here we compare the performance of seven methods (DNA-Chip Analyzer software (dCHIP), GenoCNA software, allele-specific copy number analysis of tumors (ASCAT), OncoSNP software, genome alteration print (GAP) visualization, CNVpartition software plug-in for the Genome Studio software, and Partek Genomics Suite software) that have been established for two-color CNA analysis on the Illumina platform, using two ovarian cancer cell lines where spectral karyotyping analysis has also been performed, and two tissue samples, one from a highly malignant ovarian cancer and one from a benign ovarian tumor, all of which harbor significantly different genomic abnormalities. ASCAT shows very stable estimates of CNAs, as does OncoSNP when jointly analyzing paired normal DNA. We found the best performance, in general to be from ASCAT.

Keywords Tumor abnormalities, spectral karyotyping, copy number estimation, loss of heterozygosity, ploidy

Acquired somatic genomic alterations in cancer include alterations in the DNA sequence, copy number alterations (CNAs), and structural and numerical chromosomal rearrangements. Amplification and deletion of large segments of chromosomes have been found in nearly all major human tumor types (1–3), and the regions involved in copy number changes tend to be complex, with regions of CNA spanning between a few hundred to several million nucleotide bases. The gains of chromosomal material leading to an increase in oncogene activity and losses resulting in inactivation of tumor suppressor genes, have been associated with the development and progression of human malignancies (4). Therefore, methodologies enabling the identification of cancer-specific CNAs should facilitate the discovery of new genes with functional importance in cancer biology.

Single nucleotide polymorphism (SNP) genotyping microarrays provide a convenient high-throughput platform for the genome-wide probing of DNA CNAs and loss of heterozygosity (LOH) (5). Systems with allele-specific hybridization or allele-specific fluorescence allow detailed resolution of DNA abnormalities, since these genotyping platforms provide two intensity measurements for each probe locus, corresponding to the two SNP alleles. For CNA analysis, the two allele-specific intensities are usually transformed into two

Received September 20, 2012; received in revised form March 12, 2013; accepted March 13, 2013.

* Corresponding author.
E-mail address: celia.greenwood@mcgill.ca

2210-7762/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cancergen.2013.03.001

alternative measures: the logarithmic ratio of the total intensity (LRR), and the proportion of total intensity attributable to the B allele (BAF). In contrast, platforms using comparative genomic hybridization or LOH techniques (6) measure only one quantity per probe. Resequencing technologies also provide two data dimensions—the counts of each allele at each position—and therefore, resequencing data should prove to be extremely rich not only for estimating copy number, but also for accurately identifying breakpoints and for inferring how the genome has been rearranged. However, the analysis of resequencing data is challenging, and methods are still in development.

The estimation of copy number variations (CNVs) in normal tissue has been well studied (7,8). However, the methods designed for estimating CNVs in normal tissue tend to perform poorly when also used to estimate CNAs in tumor specimens. Although many CNVs occur in normal tissue, the regions affected tend to involve small, discrete segments of genomic DNA. The estimation of CNAs can be complicated by the presence of short CNVs with large variations in the number of repeat units. If such a CNV maps to a genomic region with a cancer-specific acquired somatic aneuploidy, then there may be discordance in apparent copy number between the CNV and the neighboring flanking genomic DNA. In addition, for tumor specimens, the estimation of CNAs is also complicated by the varying amounts of normal DNA contamination, intratumor genetic heterogeneity, complex chromosomal rearrangements including crossover events, and aneuploidy. Therefore, specific methods for estimating CNAs in tumor tissue have recently been developed.

Here we compare the performance of several recently established methods for inferring CNAs and LOH for samples evaluated on the Illumina two-color SNP platform. Specifically, we focus on methods that use both the LRR and BAF data; these methods include analysis via DNA-Chip Analyzer software (dCHIP) (9–11), GenoCNA software (12), allele-specific copy number analysis of tumors (ASCAT) (13), OncoSNP software (14), genome alteration print (GAP) visualization (15), CNVpartition software (Illumina, Inc. San Diego, CA) plug-in for the Genome Studio software (16), and Partek Genomics Suite software (Partek Incorporated, Saint Lois, MO) (17). We have restricted our analyses to methods that (a) estimate actual copy numbers, not simply gains or losses, (b) estimate LOH, and (c) do not require matched normal genotyping information as it is not always feasible to obtain matching normal tissue. Our comparative analyses were performed on genotyping data from the Illumina platform derived from two epithelial ovarian cancer cell lines and two epithelial ovarian tumor samples, which represented malignant and benign disease, respectively. Genotyping data from normal cells were also available for one of the tumor specimens. The ovarian cell lines used in the study, OV-90 and TOV-21G, were derived from spontaneously immortalized ovarian cancer tumor cells, which exhibit distinct genomic abnormalities and have been well characterized (18–20). For these two ovarian cancer cell lines, we also used spectral karyotyping (SKY) analysis to assess the anomalies estimated from genotyping data. Epithelial ovarian carcinomas tend to be abundant with tumor cells, often yielding high quality DNA suitable for large-scale genotyping studies. Also, malignant ovarian cancers often exhibit complex genomic anomalies, in contrast to benign disease, which often displays little or no evidence of gross genomic rearrangements (21,22). We demonstrate that although there are some algorithms, such as ASCAT, that exhibited better performance than others, most methods were not able to infer all types of genomic anomalies in either cancer samples or cell lines.

Materials and methods

Ovarian cancer cell lines

The ovarian cancer cell lines, OV-90 and TOV-21G, were derived from the long-term passages of malignant ascites from a stage IIIC patient with high grade undifferentiated ovarian adenocarcinoma and from a high grade clear cell ovarian carcinoma tumor sample from a stage III patient, respectively, as described previously (18). Both cancer cell lines were derived from patient samples obtained prior to chemotherapy (18) and were cultured in ovarian surface epithelium medium supplemented with 2.5 μg/mL amphotericin B, 50 μg/mL gentamicin, and 10% fetal bovine serum as described previously (18).

Ovarian tumor tissues

The ovarian tumor samples, TOV-490T and BOV-1207DT, were collected with informed written consent for research use from participants undergoing surgeries performed at the Centre hospitalier de l'Université de Montréal-Hôpital Notre Dame; these samples have been described previously (22). TOV-490T and S-490 were obtained from a high grade malignant papillary serous ovarian cyst adenocarcinoma tumor and peripheral blood leukocytes, respectively, from a stage IIIC, chemotherapy-naïve patient. Sample BOV-1207DT was obtained from a benign serous adenoma tumor (also from a chemotherapy-naïve patient).

Nucleic acid extraction

DNA was extracted from the ovarian cancer cell lines, fresh frozen tumor specimens, and peripheral blood leukocytes, as described previously (23).

High density genotyping

Genomic DNA (750 ng) from the ovarian cancer cell lines (OV-90 and TOV-21G), tumor specimens (TOV-490T and BOV-1207DT), and peripheral blood DNA (S-490) were genotyped using the Infinium technology with the Illumina Human 610-Quad BeadChip (Illumina, San Diego, CA). This BeadChip assays 620,901 markers, where over 560,000 markers are SNPs with an average spacing of 4.7 kb per marker (median spacing is 2.7 kb). Both genotyping and scanning using the BeadArray Reader were performed at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada). Genotyping results were viewed with the Genome Viewer module in the BeadStudio Data Analysis software v.2.2.22 (Illumina) using the LRR and BAF for each