Detecting copy-neutral LOH in cancer using Agilent SurePrint G3 Cancer CGH+SNP Microarrays

Application Note

Authors

Paula Costa
Anniek De Witte
Jayati Ghosh

Agilent Technologies, Inc.
Santa Clara, CA USA

Abstract

The Agilent SurePrint G3 CGH+SNP microarray platform enables high-resolution CGH analysis, making it a powerful tool for detecting genome-wide copy number and copy-neutral genetic aberrations. Now, the platform has been extended to support hematologic cancer research. Algorithms developed for the concurrent analysis of CGH and SNP probes now support the analysis of samples with the genomic complexity often observed in cancers as a consequence of copy number changes, tumor heterogeneity, and admixtures of tumor and normal cells. In parallel with the algorithm improvements, a cancer-specific 4x180K CGH+SNP microarray was designed by the Cancer Cytogenomics Microarray Consortium (CCMC). Here, we used the cancer microarrays to study copy number (CN) and copy-neutral loss of heterozygosity (cnLOH) in aneuploid chronic lymphocytic leukemia (CLL) and constructed a mosaic model to calculate clonal fractions in aberrant samples. We conclude that the new computational methods developed for cancer sample analysis accurately determined the genotypes, total copy numbers, and clonal fractions in aneuploid samples highly mixed with normal cell populations.

Agilent Technologies

Introduction

Cancer research has greatly benefited from the use of array CGH for high-resolution, genome-wide CN measurements. The simultaneous analysis of single nucleotide polymorphisms (SNPs) can provide insight into regions of acquired homozygosity that may harbor tumorigenesis-related loci. The high resolution and sensitivity of the Agilent CGH+SNP microarrays are suitable for the discovery of aberrations exhibiting abnormal CN and/or cnLOH events and allelic imbalances arising from genomic instability during cancer development and progression. Nonetheless, aneuploidy, polyclonality resulting from multiple clonal cell populations with related but distinct aberrations, and mosaicism due to heterogeneous admixtures of tumor and normal cells create significant research challenges when analyzing cancer genomes. In the quest to decipher tumor complexity, Agilent has expanded the power and sensitivity provided by our CGH+SNP combined assay with novel computational methods capable of determining clonal fraction, total CN, and allele-specific CN in aneuploid tumor samples.

Methods

Microarray design

The CCMC designed a CGH+SNP microarray containing both CGH and SNP probes targeting genomic regions known to be associated with cancer. This SurePrint G3 Cancer CGH+SNP 4x180K microarray is available as a catalog kit of 3 slides (P/N G4869A). The array contains ~20,000 cancer-associated CGH probes, ~60,000 SNP probes, and backbone probes relatively evenly distributed across the genome, enabling simultaneous CN and cnLOH detection (Tables 1 and 2). CGH probes densely cover exons and flanking regions for more than 500 cancer-associated genes. They also target 130 newly discovered cancer-associated and subtelomeric regions.

Researchers can also design custom CGH+SNP microarray formats focused on regions of interest (Table 2) using eArray, Agilent's free web-based oligo design tool, from over 28 million CGH probes and probes for approximately 60,000 SNPs.

Table 1. SurePrint G3 Cancer CGH+SNP 4x180K microarray specifications

Format	4x180K
Design ID	030587
Total features	180,880
Cancer-associated CGH probes	20,000 ▪ 1 probe/0.5–1 Kb (min. 1 probe/exon, max. 200 probes/gene) ▪ 15 Kb flanking regions of each gene
Cancer-associated genes and regions	
- *Sanger cancer genes*	427
- *Other cancer-associated genes*	~100
- *Known cancer-associated genomic regions and subtelomeric regions*	~130
Distinct SNPs	59,647 (1 probe/SNP)
Genome build	hg19

Table 2. Catalog SurePrint G3 Cancer and Custom SurePrint G3 CGH+SNP Microarrays

	Catalog Kit	Custom Microarray			
	4x180K	1x1M	2x400K	4x180K	8x60K
Number of Arrays/Slide	4	1	2	4	8
Number of Slides	3	1	1	1	1
P/N	G4869A	G4882A	G4883A	G4884A	G4885A

Sample preparation, hybridization, and imaging

Genomic DNA (gDNA) from hematology-oncology samples and multiple cell lines, obtained from Coriell Cell Repositories (http://ccr.coriell.org/), were processed following the standard enzymatic labeling workflow described in the Agilent *Oligonucleotide Array-Based CGH for Genomic DNA Analysis – Enzymatic Labeling for Blood, Cells or Tissues (with a High Throughput option)* manual, version 6.4 (P/N G4410-90010). The gDNA was digested with *AluI* and *RsaI* restriction endonucleases, enabling the identification of SNPs located within the enzymes' recognition sites. The experimental samples were labeled with Cy5 dye and hybridized to the SurePrint G3 Cancer CGH+SNP 4x180K microarrays (P/N G4869A). A reference sample with known genotype was labeled with Cy3 and hybridized to the same microarrays. The microarray slides were scanned at 3-micron resolution on an Agilent microarray scanner and the images were extracted using Agilent Feature Extraction software v10.10 integrated in Agilent CytoGenomics 2.0 (P/N G1662AA–G1667AA).

Algorithms for total CN and cnLOH profiling

CGH+SNP algorithms, developed to analyze and genotype diploid genomes with constitutional aberrations, have been previously described[1]. Here we present an extension of these algorithms, implemented in Agilent CytoGenomics 2.0 software, which support the determination of the clonal fraction of aneuploid cells, and the total and allele-specific CN of aberrant clones in cancer genomes.

The CGH probes measure the log ratio of the comparative probe signal intensity of the experimental versus the diploid reference sample. Based on these log ratios, the genome is divided into diploid and aneuploid intervals. For the aberrant segments, the total CN is determined from the average CGH log ratio of each region. Figure 1.1 exemplifies the distribution of CGH log ratios for regions of diploidy and aneuploidy. The algorithm shifts the log ratios so that the peak corresponding to the sample's diploid genome is centered at zero. Negative log ratios are indicative of copy losses and positive log ratios result from copy gains in the test sample. The peaks of the log ratio distribution are assigned distinct copy numbers: the largest peak is considered diploid and all the other peaks are assigned copy numbers that will produce a linear relationship between CN and observed signal ratios. A good line fit indicates that the CN of the peaks has been correctly assigned and that those aberrations most likely occurred in a single aberrant clone. In mosaic samples, non-diploid regions can have a fractional total CN; for example, in a genome where 50% of the contributing cells have a hemizygous deletion in a specific region and 50% of the cells are normal diploid, the total CN assigned to the aberrant interval is 1.5. Because of the accuracy and low noise of the Agilent microarray platform, the slope of the ratio versus CN fit can be used to determine the fraction of aneuploid cells in a mosaic sample[2]. The maximum theoretical aneuploid fraction is 1.0, with slopes in the range of 0.89–0.96 commonly observed for monoclonal samples due to slight log ratio compression in the assay.

The SNP probes on the CGH+SNP microarrays are designed to include *Alu/Rsa* sites and the SNPs are genotyped by measuring the number of uncut alleles using the restriction digestion step in the protocol. For SNP CN profiling, the SNP probe raw log ratios are first adjusted for the known genotype of the reference to report absolute allele-specific CN of the uncut variant allele[1]. The allele-specific CN of the uncut allele is inferred from the total CN, as measured by the CGH probes neighboring the SNP site. Each SNP probe present in a diploid region is genotyped as AA (0 uncut copies) if the alleles are homozygously cut, AB (1 uncut copy) if the variant alleles are heterozygous, or BB (2 uncut copies) if they are homozygous and not digested by the enzymes (Table 3, Figure 1.2). Segments of cnLOH, where no heterozygous alleles are observed, are reported as either AA or BB. In aberrant genomic locations harboring single-copy losses, SNP probes are either cut or uncut and the resulting genotype is simply A or B. In amplified regions, additional A or B alleles are detected by the SNP probes and the genotypes become increasingly more complex (Table 3). For mosaic non-diploid samples, both the normal and aberrant clones contribute to the log ratios in any aberrant regions. Hence, it is necessary to separate out the aberrant clone algorithmically from the data to determine the exact allele-specific CN of a SNP in an aberrant region.

In Agilent CytoGenomics 2.0 software, only the allele-specific CN for the aberrant clone is displayed in the SNP panel (Figure 1.2B). If more than one *aberrant* clone is present in the sample, the algorithm identifies the allele-specific CN of the major aberrant clone; for minor clones the allele-specific CN will not be precisely determined.

After SNP CN has been assigned to all genomic locations, segments of cnLOH are located by identifying genomic regions with a statistically significant decreased density of heterozygous SNP calls in the experimental sample.

Table 3. Genotype corresponding to the number of uncut alleles

Genomic status	Genotype	No. uncut allele
Normal diploid genome	AA, AB, BB	0, 1, 2
Diploid genome with cnLOH	AA, BB	0, 2
Hemizygous LOH	A, B	0,1
Trisomy	AAA, AAB, ABB, BBB	0, 1, 2, 3
Tetrasomy	AAAA, AAAB, AABB, ABBB, BBBB	0, 1, 2, 3, 4

Results

CN and cnLOH analysis in CLL samples

CN alterations and cnLOH events were commonly found in the cancer genomes studied using the enhanced algorithms as implemented in Agilent CytoGenomics 2.0 software. In CLL samples, significant heterogeneity was observed among tumors of the same class. For example, in the case shown in Figure 1.1A, only one population is contributing to the sample genome, as indicated by the histogram of the CGH probe distribution. In this sample, three major aberrations were identified in 100% of the cells: a cnLOH on the long arm of chromosome 9 (data not shown), a cnLOH encompassing most of chromosome 13, and a 2.8 Mb homozygous deletion on chromosome 13 within the cnLOH (Figure 1.2A). This region is commonly deleted in CLL samples[3,4]. In a second CLL sample, the tumor population is admixed with normal diploid cells at a clonal fraction of ~66% (Figure 1.1B). A copy gain of the distal half of the long arm of chromosome 5 and a hemizygous deletion of most of the long arm of chromosome 6 were detected in the aberrant clone (Figure 1.2B).

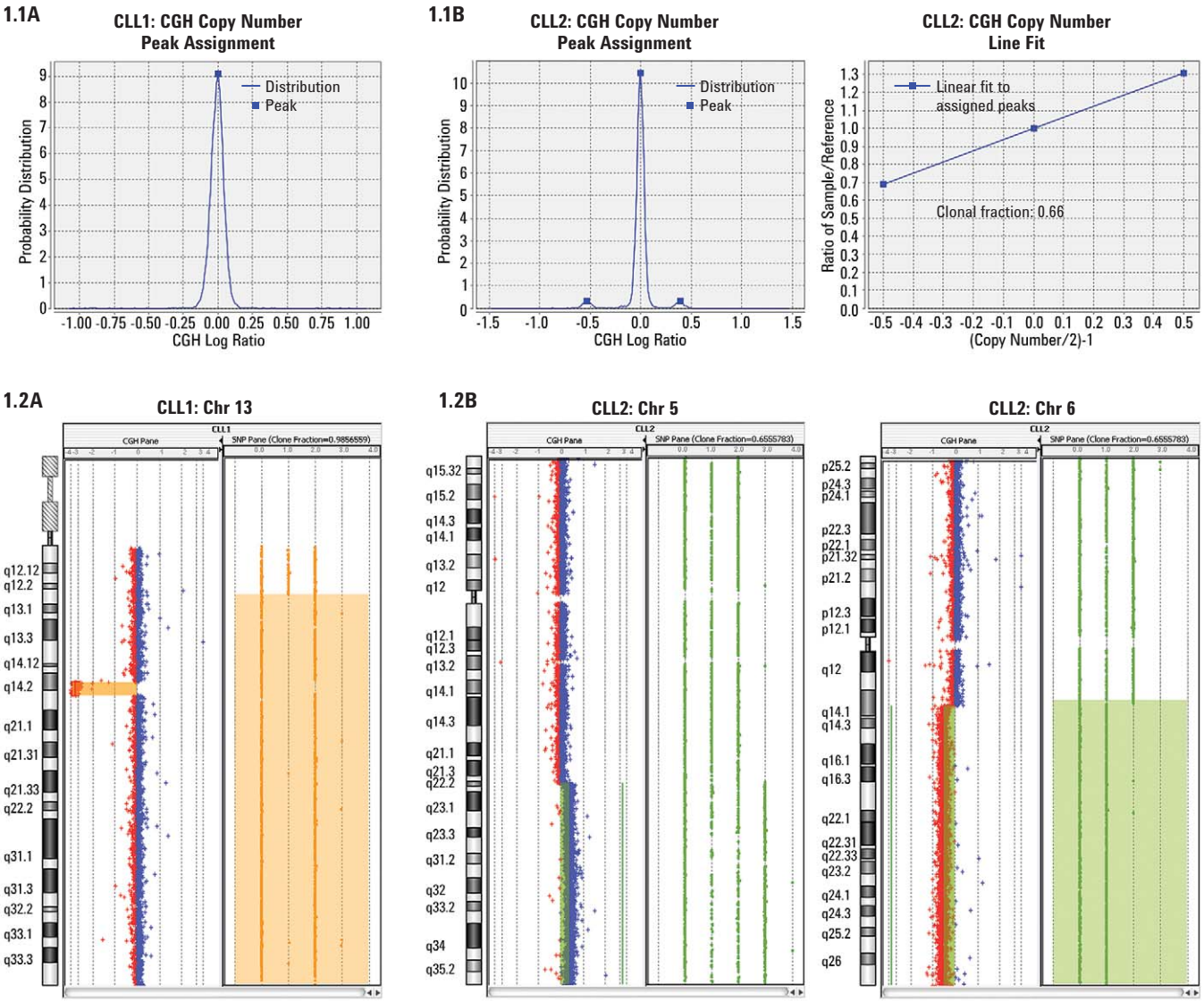


Figure 1. *CN and cnLOH analysis in CLL samples.* **1.1**. *Histogram of the CGH probes log ratio distributions and the peaks detected for the samples CLL1* (**1.1A**) *and CLL2* (**1.1B**). *In sample CLL1 only one peak was detected centered at 0; the majority of the genome from all contributing cells is diploid. For sample CLL2 three peaks were detected and used to calculate the clonal fraction of 0.66.* **1.2**. *Chromosome view in Agilent CytoGenomics 2.0 software showing the CGH log ratios and SNP CN for affected chromosomes in samples CLL1* (**1.2A**) *and CLL2* (**1.2B**): **A**. *Homozygous deletion within a larger region of cnLOH detected in chromosome 13.* **B**. *A copy gain on chromosome 5 and a hemizygous copy loss on chromosome 6 were identified in the aberrant clone. Settings for CGH aberration calling: ADM-2, threshold 6, minimum of 5 probes, ≥ 0.15 log ratio.*

The Mosaic Model system

To assess the ability of the algorithm to detect low-level CN mosaicism, samples from monozygotic twins were used. One twin has an amplification on chromosome 22 that is not present in the other twin. The sample from the affected twin was diluted with the sample from the unaffected twin at ratios ranging from 0 to 100%, in increments of 10%, for a total of 11 different sample mixtures. Total and allele-specific CN and clonal fractions as low as 20% were accurately determined (Figure 2).

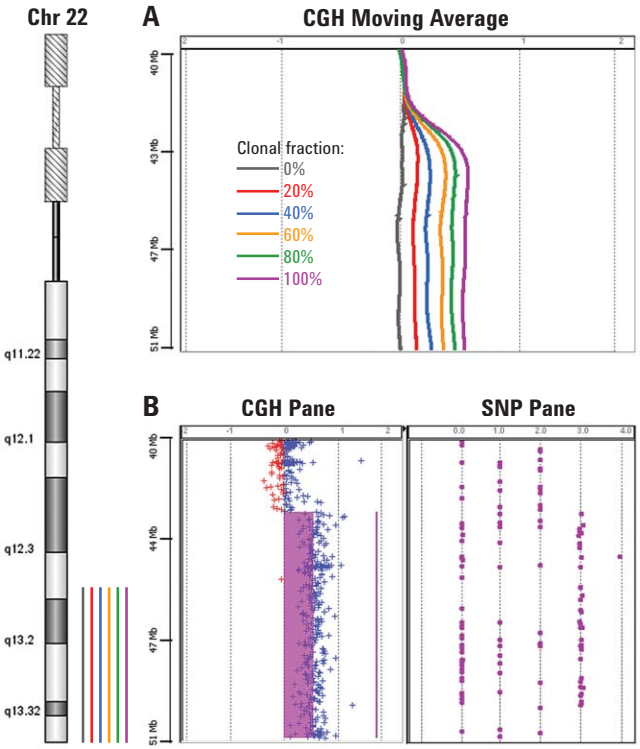


Figure 2. *Serial dilution of the affected and unaffected monozygotic twins.* **A**. *Moving average lines of the CGH probes log ratios for the distal part of the long arm of chromosome 22 with a copy gain in the aberrant sample and diploid for the matched normal, and the intermediate log ratios resulting from the serial fractional dilutions 20%, 40%, 60% and 80%.* **B**. *CGH and SNP CN profile for the 100% aberrant sample. Settings for CGH aberration calling: ADM-2, threshold 6, minimum of 5 probes, ≥ 0.1 log ratio.*

For clonal fraction determination, serial dilutions of normal and aberrant genomes were performed for two additional sample pairs, each with one sample having a trisomy of chromosome 21. The observed clonal fractions matched the expected values from the serial dilution as low as 20% (Figure 3). When the copy gain constituted only 10% of the sample, the shift in log ratios corresponding to the CN change was below the detection threshold. However, since monosomies have larger log ratio shifts than trisomies (-1.0 versus +0.58), hemizygous deletions present in only 10% of the cells would be measurable with the same detection thresholds. Valli *et al.* have reported the ability to detect levels of abnormal cells as low as 8% in acquired mosaicism[5].

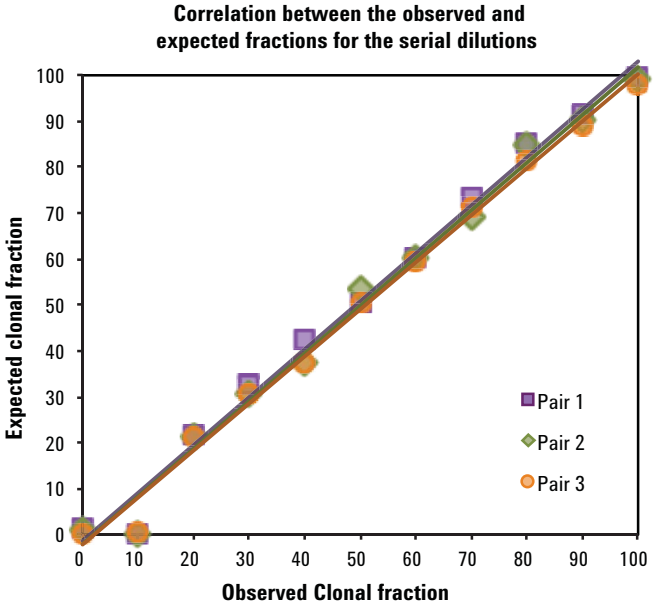


Figure 3. *Comparison between the expected and the observed clonal fractions calculated for the mixed aberrant and matched normal samples. The Y axis represents the actual fraction of the aberrant clone in the corresponding dilution series and the X axis represents the fraction of aberrant clone as calculated in Agilent CytoGenomics 2.0. A linear trendline represents the strong correlation between the expected and the observed clonal fractions.*

Conclusion

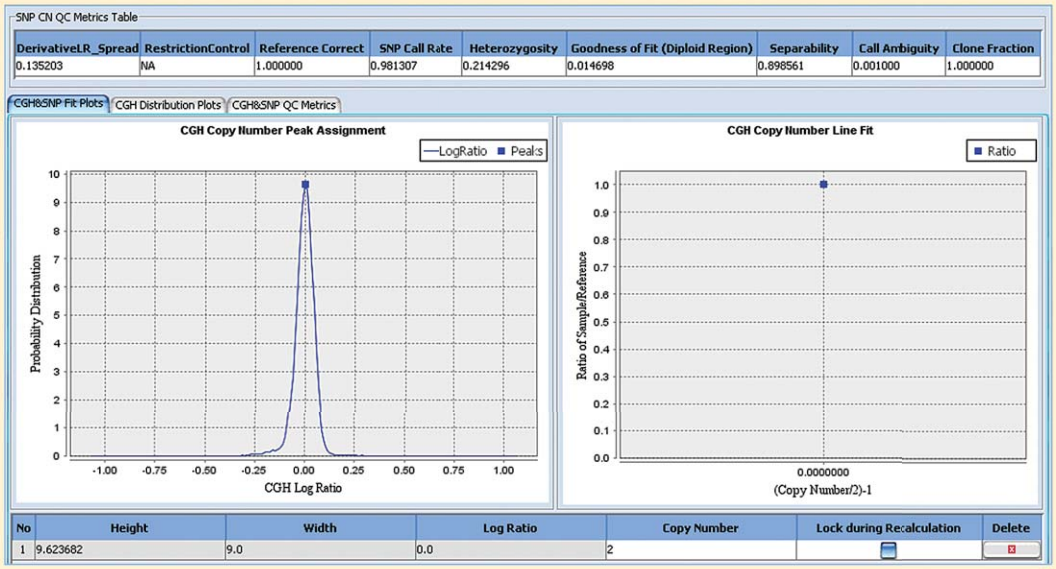
The CGH+SNP analysis workflow has been extended in Agilent CytoGenomics 2.0 software with enhanced analysis algorithms, providing a powerful tool in cancer cytogenetic research studies. For the aneuploid cancer samples analyzed, the SNP genotypes and total CN were successfully determined, even when contaminated with normal cells. We were able to determine the precise fraction of aneuploid cells in cancer samples and in the mosaic model with as little as 20% aberrant cell content for single copy amplifications. In conclusion, Agilent CGH+SNP microarrays, combined with algorithms with extended capability to study mosaic cancer samples, enable unprecedented power to detect CN changes and cnLOH in hematologic cancer research.

Appendix: Manual peak assignment

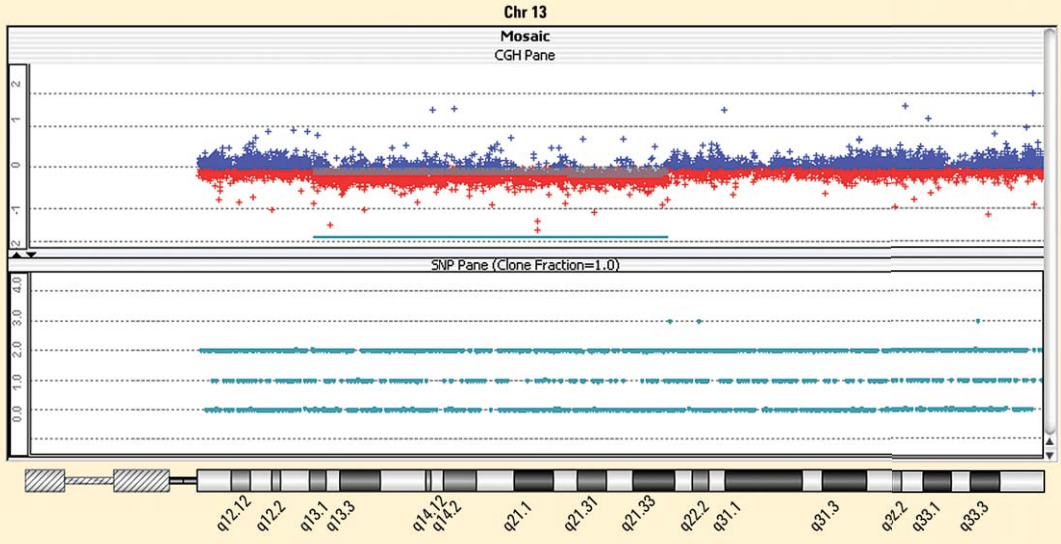
For certain samples that only contain small aberrations or for which the tumor fraction is very low (i.e., the aberrant clone is significantly diluted with a diploid genome), the default workflow in Agilent CytoGenomics 2.0 might not be able to detect a peak from the CGH log ratio probes distribution. In these cases manual intervention may be required for proper CN assignment and clonal fraction computation. The section below discusses the tool available to manually reassign, delete, and add peaks to the CGH probe distribution plot.

1. For a cancer sample analyzed using default settings, even though aberrant peaks were not detected in the CGH probes distribution due to low tumor content, a shoulder was observed at an approximate log ratio of -0.2 (A) and an aberration was identified in the chromosome 13 CGH data at low negative log ratio (B).

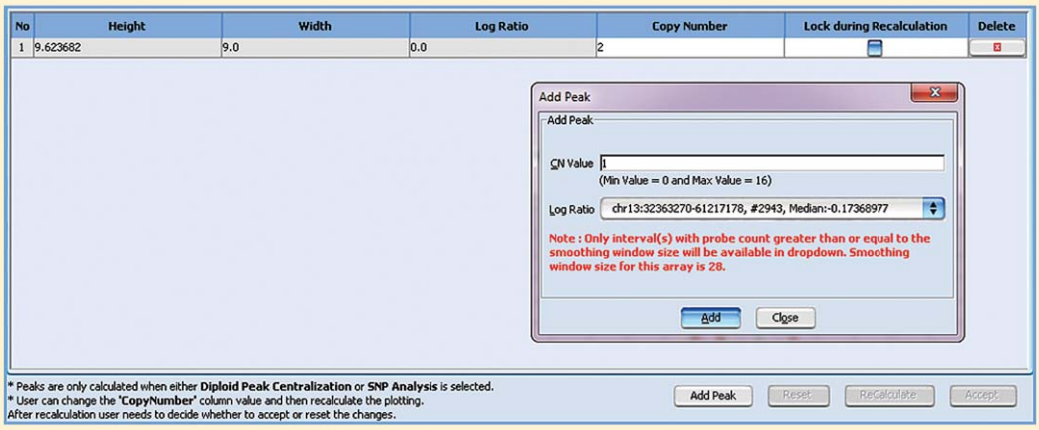
A



B

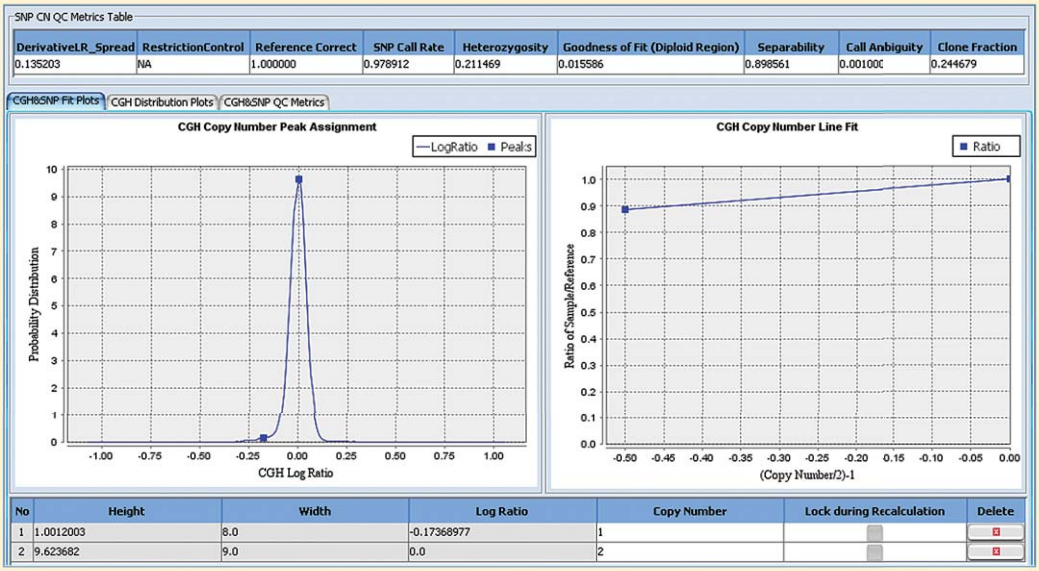


2. In the CGH&SNP Fit, peaks can be added to or deleted from the CGH probes distribution histogram. For a specific CN to be added to the plot, the median log ratio of an aberrant interval is selected from the drop-down menu.

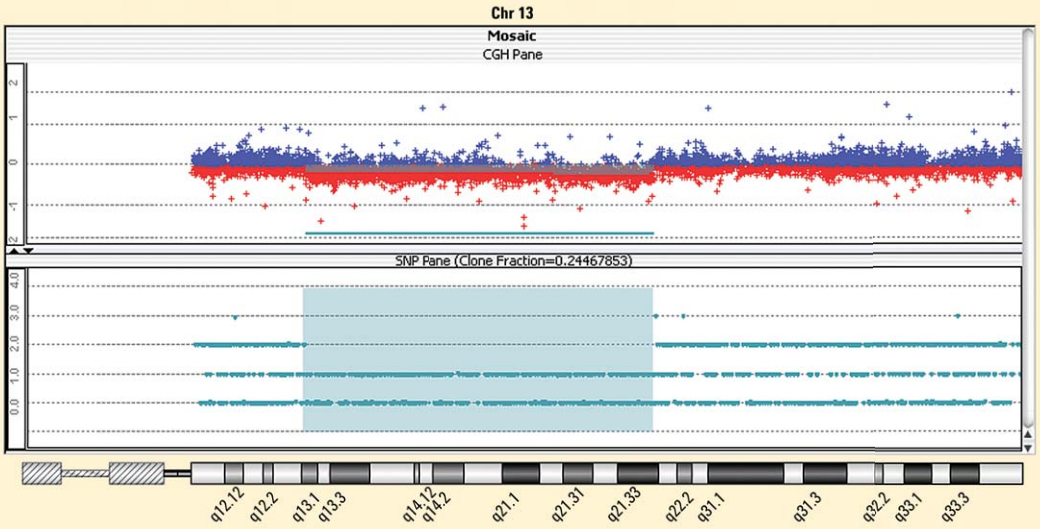


3. Following manual assignment of a new peak, clone fraction (A) and SNP CN (B) are recalculated for the aberrant clone to reflect the single copy loss in ~24% of the cells.

A



B



References

1. Application Note: Simultaneous detection of copy number and copy-neutral LOH using a single microarray. (5990-6274EN)
2. Curry *et al. CRC Press.* 2008;233–44.
3. Fitchett *et al. Cancer Genet Cytogenet.* 1987;24(1):143–50.
4. Juliusson *et al. N Engl J Med.* 1990;323:720–4.
5. Valli *et al. Mol Cytogenet.* 2011;4:13.

Required Agilent CGH Processing Components

Description	Part Number
SureTag DNA Labeling Kit	5190-3400
Human Cot-1 DNA	5190-3393
Agilent Oligo aCGH Hybridization Kit (25) or (100)	5188-5220 or 5188-5380
Agilent Oligo aCGH Wash Buffer 1 and 2 Set	5188-5226
Hybridization Chamber, stainless	G2534A
Hybridization Chamber Gasket Slides	Varies by array format and quantity
Hybridization Oven	G2545A
Hybridization Oven Rotator Rack	G2530-60029
SureScan Microarray Scanner	G4900DA
Agilent CytoGenomics	G1662AA–G1667AA

For more information and to learn more:
www.agilent.com/cgh_snp

Find an Agilent customer center in your country:
www.agilent.com/genomics/contact

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Asia Pacific
inquiry_lsca@agilent.com

Europe
info_agilent@agilent.com

Printed in the USA, September 29, 2011
5990-9183EN

Agilent Technologies