

Integrated Biology Solutions

# 60-mer Oligo-Based Comparative Genomic Hybridization

## Application Note

### Authors

Anniek De Witte, Manoj Nair,  
 Kshama Mehta, Jayati Ghosh,  
 Alicia Scheffer, Francisco Cifuentes  
 Agilent Technologies  
 Santa Clara, CA, USA

Javier Suela  
 CNIO, Centro Nacional de  
 Investigaciones Oncologicas  
 Madrid, Spain

Maria Garcia  
 Cancer Genomics Program  
 University of Cambridge  
 Cambridge, UK

### Introduction

Chromosome rearrangements are commonly associated with multiple disease states such as cancer and many developmental syndromes, including Down's syndrome and autism spectrum disorders. The identification and analysis of these genomic rearrangements has been instrumental to the advancement of research in these areas as well as the diagnosis of these diseases. Researchers and clinicians can employ several techniques to detect chromosome changes, which include karyotyping, FISH (Fluorescence In-situ Hybridization), SKY (Spectral Karyotyping) and conventional CGH (Comparative Genomic Hybridization). However, each of these techniques is extremely labor intensive and often require specially trained staff to perform these tests in a consistent manner.

Another major technical disadvantage is the low genomic mapping resolution along the chromosome, where only large chromosomal aberrations can be easily identified (Table 1a). Recently, microarray-based CGH (aCGH) techniques have revolutionized the field of chromosomal aberration detection. However, aCGH techniques relying on cDNA probes or BAC (Bacterial Artificial Chromosomes) clones still require significant human and technical efforts to manufacture the microarrays as well as to maintain and validate the cDNA and BAC libraries. Moreover, cDNA and BAC based probes are too long to ensure specific hybridization and the high resolution needed to detect focal chromosome alterations that researchers are looking for (Table 1b).

Agilent Technologies, Inc. has developed an oligo aCGH platform that enables

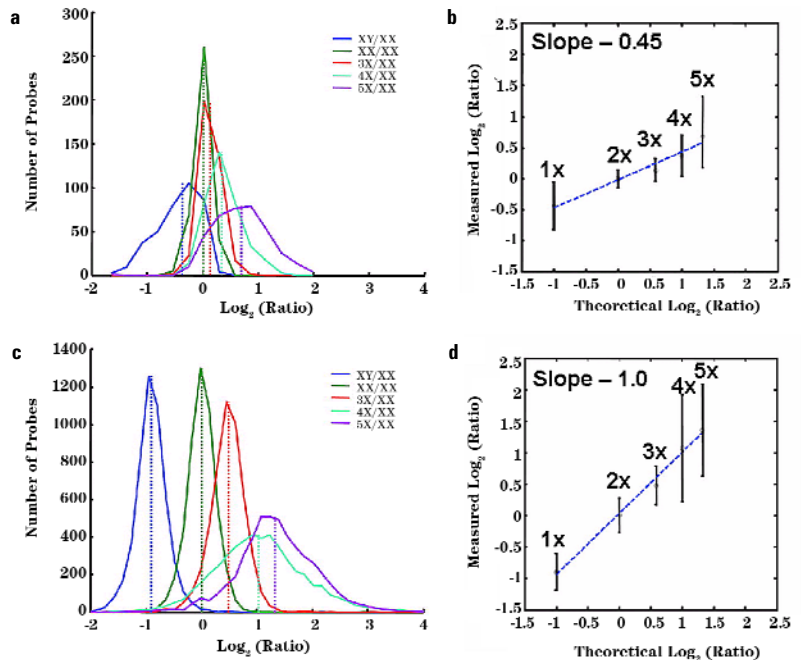
		Resolution	Coverage
<b>a Cytogenetics</b>	Karyotyping	> 10 Mb	Complete
	SKY	> 2 Mb	Complete
	Traditional CGH	> 2 Mb (cytoband)	Complete
	FISH (interphase)	≥ 20 Kb	Probe Specific
	FISH (metaphase)	≥ 100 Kb	Probe Specific
<b>b aCGH</b>	BAC	100 Kb (Spectral Genomics - 2 Mb)	Complete
	cDNA	2 Kb	Genes Only
	Oligo (60-mer)	0.06 Kb	Complete

Table 1. Coverage and resolution of different cytogenetic and microarray-based CGH techniques.

researchers to study genome-wide DNA copy number variations at a high resolution. The probes on Agilent aCGH microarrays are 60-mer oligonucleotides synthesized *in situ* using Agilent's inkjet SurePrint™ technology, which not only ensures consistent and high quality oligonucleotide probes, but also provides the greatest flexibility for customized microarray designs. The probes represented on the Agilent Human Genome aCGH microarray have been selected using algorithms developed specifically for CGH application, assuring optimal performance of these probes in detecting DNA copy number changes in non-reduced complexity samples. Protocols, reagents and general workflow have been optimized to maximize performance and reproducibility while ensuring ease-of-use. Our results demonstrate the effectiveness of the optimization in the probe design and microarray protocols, and more importantly, the ability of our oligo aCGH platform to detect chromosomal alterations throughout the human genome with high sensitivity and precision.

### Materials and Methods

Agilent offers two different protocols for use with CGH microarrays allowing users to directly label the genomic DNA (direct) or first phi-29 amplify the genomic DNA (amplified) before labeling. Genomic DNA quality assessment, amplification (if applicable), digestion, labeling, hybridization and wash procedures are the same as specified in the Agilent standard aCGH manual (Version 2.0) (Pub # G4410-90010). Protocols for the X-chromosome series experiments can be found in Reference 1. Male versus female experiments were performed following the amplified protocol with changes in hybridization times and rotation speeds according to the specific experimental conditions.



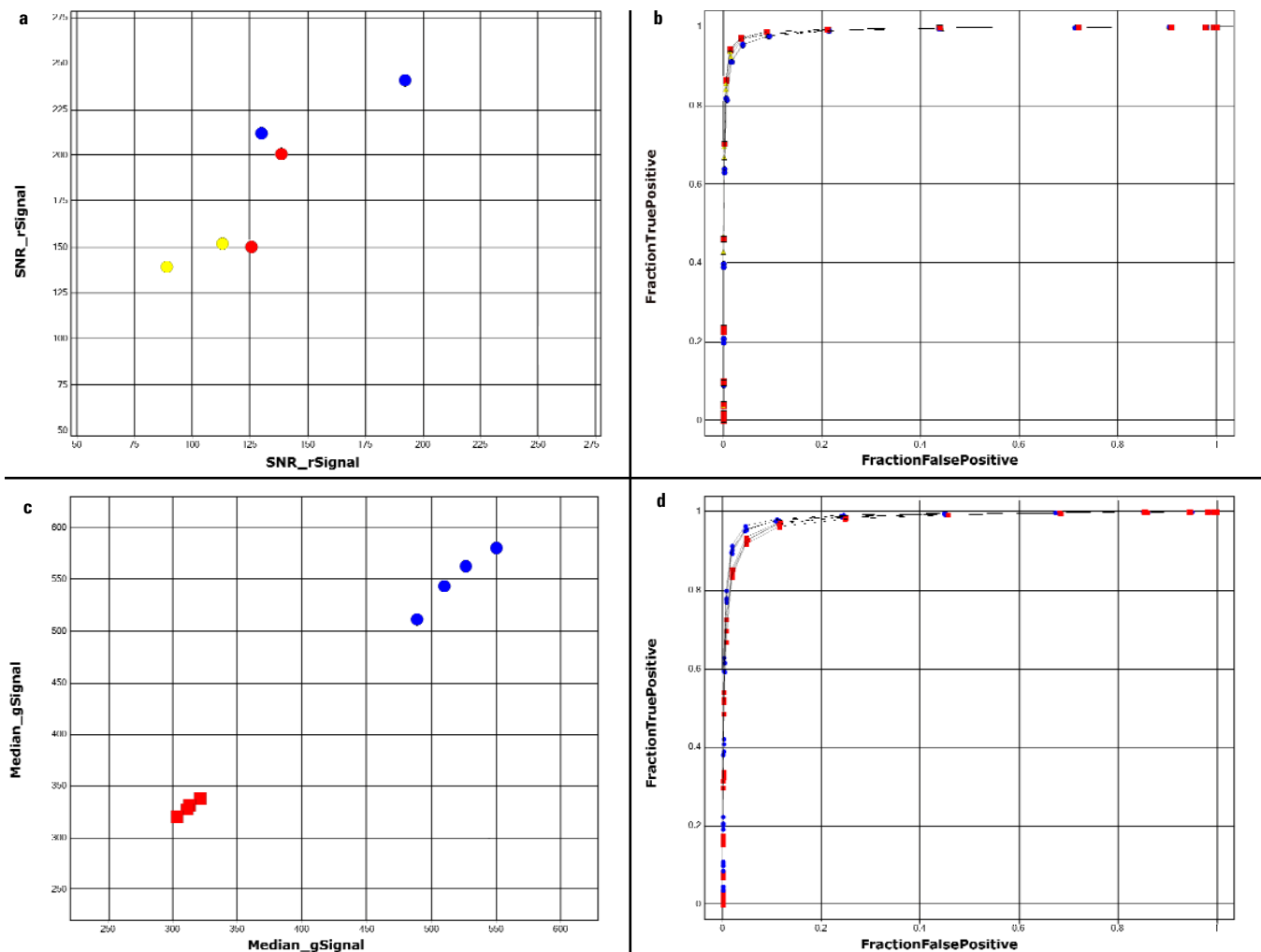
**Figure 1. Performance comparison between microarrays containing 60-mer probes optimized for gene expression and for CGH application. Distribution of Log Ratios and median values (dashed lines) of X chromosome probes from hybridizations with cell lines carrying different numbers of X chromosome (see color codes in the figure) on the gene expression (a) and aCGH (c) microarrays. Measured mean Log Ratios of the X-chromosome probes are plotted versus the theoretical Log Ratios on gene expression (b) and aCGH (d) microarrays from the hybridizations with cell lines containing different numbers of X chromosome.**

### Development of the Human Genome CGH Microarray

To confirm the capability of detecting DNA copy number changes with 60-mer oligonucleotide probes, the Agilent Human 1A Gene Expression microarray was used in the initial "proof-of-principle" study (Ref. 1). Several cancer cell lines, as well as a series of cell lines carrying different numbers of chromosome X (up to five copies), were used for this study. Probes on this microarray were optimized for gene expression, not for CGH experiments. Therefore, probes from the gene expression microarray that did not represent unique genomic sequences were filtered out and not included in the data analysis. Using Agilent gene expression microarrays, it was possible to detect the different copy numbers of the X chromosome in the X-series cell lines tested (Fig. 1a,b).

In addition, results obtained using genomic DNA from cancer cell lines were comparable to published data previously profiled on BAC and/or cDNA microarrays (not shown). These results demonstrated the feasibility of using *in situ* synthesized 60-mer oligonucleotide probes for CGH experiments. Nonetheless, the absolute numbers that were measured and the separation among the cell lines with various X chromosome numbers were clearly not optimal with the gene expression microarrays.

A set of new probe design algorithms and in-silico probe selection criteria were developed specifically for CGH application. Unlike gene expression microarrays, probes for CGH are not confined to the 3' end of the expressed sequences, and can also be located in intronic or intergenic regions. Special care was taken to select probes with a very narrow range of  $T_m$  (melting



**Figure 2. Optimization of hybridization conditions. (a)** Comparison of signal to noise ratios (SNR) at various rotation speeds: 5 RPM (yellow), 10 RPM (red) and 20 RPM (blue). SNR is calculated as the median signal intensity divided by the standard deviation of the negative controls. SNR in the green channel is plotted vs. the SNR in the red channel. Two data points per condition are shown. **(b)** Comparison of measurement accuracy in detecting X chromosome DNA copy number difference in male/female model system at different rotation speeds. The fraction of false positives is plotted vs. the fraction of true positives, i.e. the ROC curve. Two ROC curves per condition are shown. **(c)** Comparison of median signal intensities between 17-hour (red) vs. 40-hour (blue) hybridization. The median signal intensity in the green channel is plotted vs. that of the red channel. Four data points per condition are shown. **(d)** Comparison of measurement accuracy in detecting X chromosome DNA copy number difference in male/female model system at two different hybridization times, i.e. 17 hour (red) and 40 hour (blue). The fraction of false positives is plotted vs. the fraction of true positives, i.e. the ROC curve. Four ROC curves per condition are shown.

temperature) to ensure optimal performance and specificity of all probes at the given hybridization and wash conditions. A prototype aCGH microarray was designed with coverage on chromosome X, 18, 17 and 16. The microarrays were again tested with the X-series cell lines. The results showed a significant improvement in the accuracy of these CGH specific probes in detecting copy number changes in comparison with the probes optimized for gene expression microarrays (Fig. 1c,d). Genomic DNA

from cell line GM50122, derived from a patient with 18q deletion syndrome was also hybridized to the prototype microarray. This cell line contains a single copy loss at the distal region of the q arm of chromosome 18. The breakpoint has been mapped and sequenced to the exact nucleotide position. The probes corresponding to chromosome 18 detected the single copy loss with the breakpoint localized to the genomic position previously described (data not shown, see Ref. 1). The

empirical validation results obtained with the X chromosome cell line series and the GM50122 deletion cell line were used to further improve the aCGH probe design algorithms. A set of ~120,000 semi-final probes that had gone through this in-silico selection process was further tested in the male/female model system to remove any additional dye-biased or poor-performing probes. The best set of ~43,000 probes were finally selected and represented on the final Agilent Human Genome aCGH microarray.

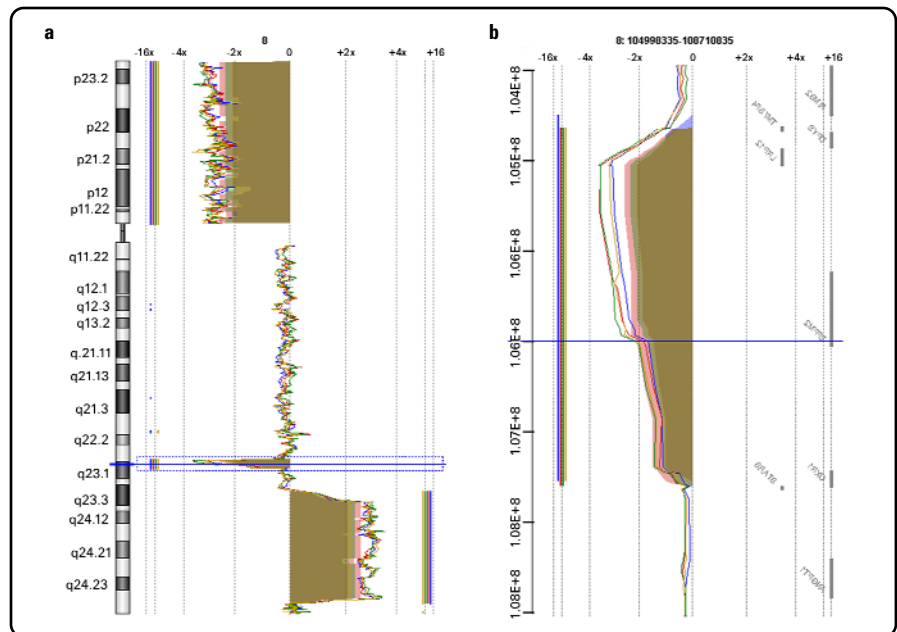
The Agilent Human Genome aCGH microarray (part number G4410B) consists of approximately 43,000 60-mer oligonucleotide probes on an industry standard 1" x 3" glass slide. It provides genome-wide coverage with an average spatial resolution of ~30-35kb (resolution is calculated by using the total size of the non-repeated sequences of the genome divided by the total number of features). UCSC Genome Browser hg17 (NCBI Build 35) was used as the source of the content. The probe selection is biased toward genes, with 84% of the probes residing in intragenic regions (approximately 50% in introns and 50% in exons) and 16% in intergenic regions. Well-characterized genes are each represented by at least one probe while ~ 1100 cancer-relevant genes are each represented by two or more probes (selected based on literature and feedback from oncology specialists).

### Optimization of Hybridization Conditions using Male/Female Model System

To assess the performance of the Human Genome aCGH microarray and optimize the assay conditions, a series of hybridizations were performed using female versus male genomic DNA. This model system is ideal to evaluate several performance parameters because the exact DNA copy numbers are known. Therefore, the number of true positives at a given rate of false positives can be readily calculated and compared. With this information a conventional ROC curve can be built to measure the results accuracy (Ref. 2). In addition, a set of other common metrics such as background noise levels, average signal intensities and reproducibility were generated from the same experiment.

#### Hybridization mixing

Mixing during hybridization greatly impacts data quality; good mixing results



**Figure 3. Inter-array reproducibility. Representation of the moving average for Chromosome 8 (window size = 10 probes)[(a) Chromosome View and (b) Gene View of the q23.1 region] from four replicate hybridizations using genomic DNA from HT29 cell line.**

in a greater probability of target molecules finding their corresponding probes, which translates into higher specific signal intensities on the microarray, thus a higher signal to noise ratio. In general, signal intensities on aCGH microarrays are much lower than that of gene expression. Therefore, hybridization using the Agilent standard gene expression conditions (i.e. 17 hours at 4 RPM) is not optimal for CGH experiments. On Agilent microarrays, hybridization cocktail mixing is facilitated by a bubble inside the hybridization chamber that rotates freely around the microarray surface. The bubble is limited to a fixed rate of rotation around the hybridization chamber, so simply increasing the rotation speed does not necessarily translate into better mixing. A modification to the hybridization cocktail composition was introduced to allow faster bubble rotation over the microarray surface. Comparison of the signal intensities obtained using the new aCGH hybridization buffer at 5, 10 and 20 RPM demonstrated a linear increase in signal intensities in both channels without affecting the noise levels, resulting in an

overall increase in signal to noise ratios (Fig. 2a). The ROC curves at the higher rotation speeds were comparable to that of the original curve. In all cases, the area under the curve was greater than 95% (Fig. 2b). Even higher rotation speeds (i.e. >20 RPM) were tested. Uneven back-ground noise and poor reproducibility resulted, therefore higher rotation speeds are not recommended (data not shown).

#### Hybridization time

Length of hybridization time is another key parameter that has a significant impact on the data quality. When the hybridization time is too short, reproducibility is poor because the probes do not reach the equilibrium needed for generating reliable results. When the hybridization time is too long, the background noise levels tend to increase and it is also not desirable from a practical stand point. Both situations result in poor/low signal to noise ratios. Performance of the Agilent aCGH microarrays was compared at two different hybridization

times, 17 hours and 40 hours. Average signal intensities of the autosomal probes ranged between 250 to 350 counts with 17-hour hybridization. The signal intensities were nearly doubled (between 500-550 counts) with 40-hour hybridization (Fig. 2c). The area under the ROC curve also improved from 93% to 95% when the hybridization time was increased from 17 hours to 40 hours (Fig. 2d). Longer hybridization (i.e. >40 hours) only marginally improved the performance both in terms of signal intensity and the ROC area (data not shown). Based on these results, the optimal hybridization conditions for Agilent aCGH microarrays with the new aCGH hybridization cocktail is standardized to 20 RPM rotation speed with a 40-hour hybridization.

## Detection of Chromosomal Aberrations in Cancer Cell Lines and Tumor Samples

Two well known cancer cell lines, HT29 and HCT116 (both established from colorectal cancer patients) were used to evaluate the detection of copy number changes with Agilent Human aCGH microarrays. Both cell lines have been previously characterized using various cytogenetic techniques (karyotyping, FISH and SKY) as well as BAC CGH microarrays. Four hybridizations of each of the two cell lines versus a commercial reference genomic DNA (2 per polarity) were performed according to standard Agilent CGH protocols. The results were compared with previously published data using other techniques.

All Quality Metrics obtained from this experiment confirmed that there were no major problems with sample handling and manipulation. Fig. 3 shows how well the moving average of all four replicate microarrays overlapped for the HT29 cell line. To further improve confidence in the results, the combined data from the replicate dye-swap experiments further

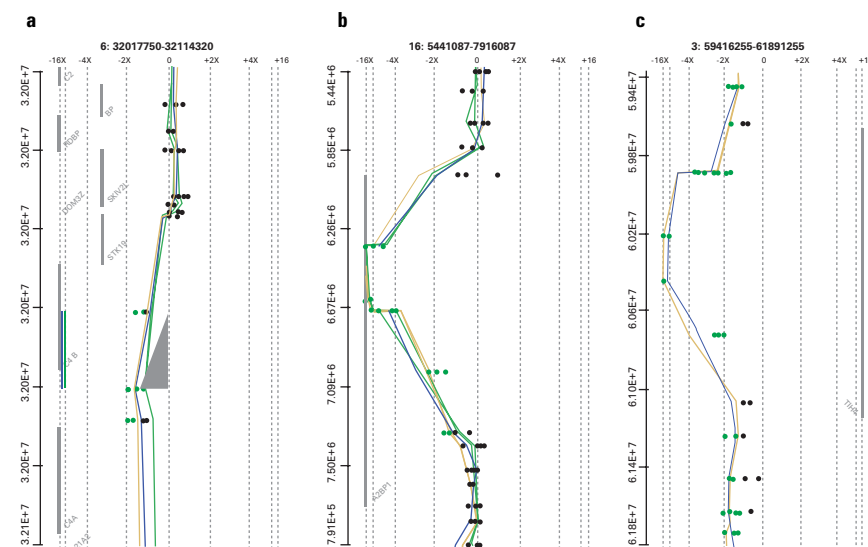
decreased the noise levels and the DLRSpread (see CGH Analytics insert for explanation) significantly dropped from 0.17-0.20 for the HT29 hybridizations down to 0.13.

The Z-score algorithm (see CGH Analytics insert for explanation) was used to analyze the above experiments. The selected window size was 1Mb, with a Z-score threshold set to 2.5. The threshold can be varied and defined by the users to accommodate for the quality of the microarray data. Using the Agilent Human Genome aCGH microarrays, all aberrations previously described for the HT29 and HCT116 cell lines were detected, including single copy deletions, homozygous deletions (Fig. 4), as well as chromosome gains and high copy number amplifications (data not shown). In addition, several small regions of gains and losses that were attributed by a single probe were detected. They were present in all four replicates, suggesting that it is unlikely due to random noise and can be confidently attributed to the presence of a micro-deletion or a micro-amplification. For further validation and refinement of the aberration boundaries, Agilent provides high-definition custom

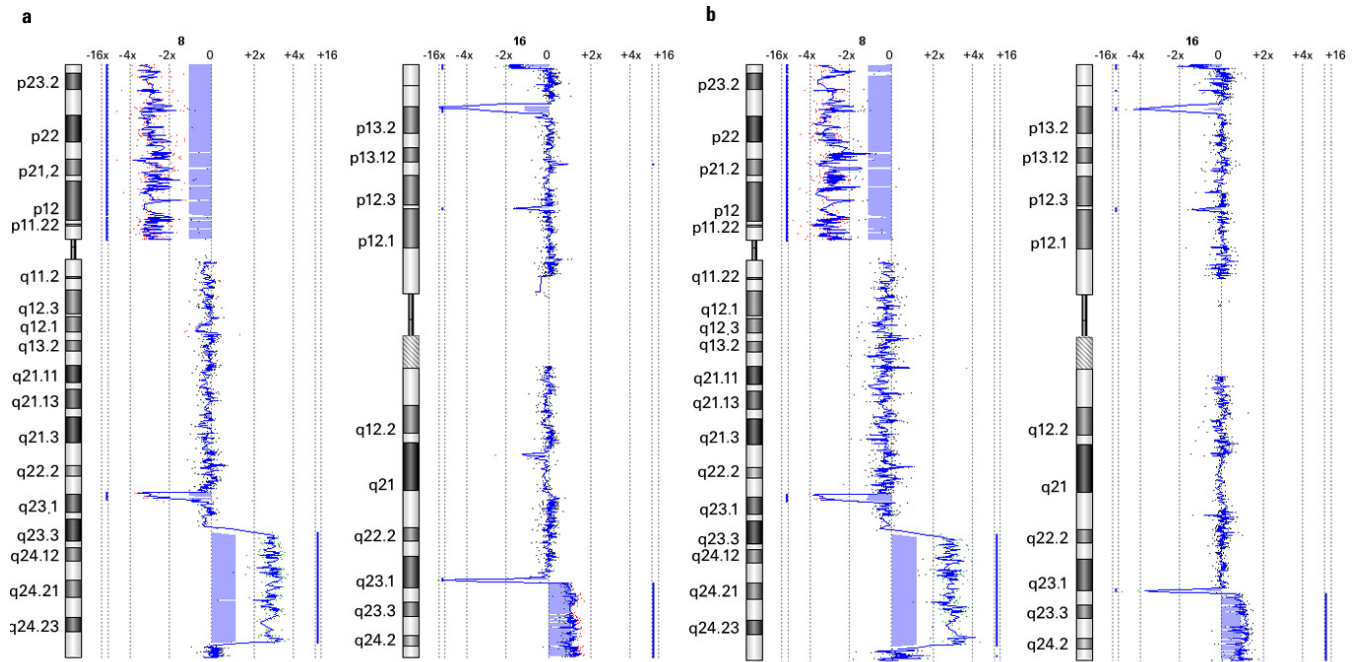
microarrays that can be easily designed from a master database consisting of ~4 million computationally validated CGH probes (average probe spacing ~400bp) using an online web tool called eArray (5989-2483EN).

## Amplified Samples

In cases where genomic DNA is very limited, such as samples from needle biopsy, LCM and sorted cell samples, a Phi-29 DNA polymerase based whole genome amplification protocol is provided. Although Phi-29 DNA polymerase enzyme introduces minimum bias from the original sample DNA, the enzyme requires genomic DNA template with high integrity (>50kb fragment size) in order to have a successful amplification. Heavily degraded DNA may not serve as a good template for whole genome amplification. The results obtained from HT29 and HCT116 cell lines after going through one round of amplification were compared to results from non-amplified starting DNA materials from the same cell lines. Although the Derivative Log Ratio Spread (DLRSread) for amplified



**Figure 4. Examples of various chromosomal alterations detected using Human aCGH microarray. (a) Single copy loss of a region of less than 50 Kb in size on Chromosome 6 of HCT116 cell line. (b) Homozygous deletion mapped to a single gene on Chromosome 16 of HCT116 cell line. (c) Single gene amplification on Chromosome 3 of HT29 cell line.**



**Figure 5. Comparison of direct vs. phi29-amplified protocols. Chromosome 8 from HT29 cell line and Chromosome 16 from HCT116 cell line using direct (a) and phi29-amplified protocol (b). Results are visualized in CGH Analytics.**

samples is slightly higher than that for non-amplified samples, the aberration calls obtained were comparable between the two protocols (Fig. 5).

### Diluted and Mosaic Samples

Tumor specimens are usually heterogeneous, containing both cancerous and non-cancerous cells. Even cancer cells from the same specimen can carry different chromosome alterations. These factors can result in an overall dilution of the chromosome aberration calls. To assess the ability of the Agilent aCGH platform to accurately measure the DNA copy number changes in a heterogeneous population, genomic DNA from the HT29 cell line was mixed with normal female genomic DNA at increasing percentages. The results show that the same aberration calls can be made with as much as 70% contamination of normal female DNA (i.e. 30% of HT29 DNA) (Fig. 6). However, for many small aberrations and single copy gains and losses, detection is more likely at levels less than 50% contamination of normal DNA (data not shown).

### Tumor Samples

There is a great variety in tumor isolation and handling techniques, however the key to successful CGH experiments with tumor samples, in spite of the high degree of variations, is to obtain high quality genomic DNA. To illustrate the performance of Agilent's aCGH platform on real biological samples, genomic DNA obtained from a patient diagnosed with Acute Myeloid Leukemia (liquid tumor) was used, according to Agilent's standard aCGH protocol. The Quality Metrics indicated that the quality of the patients' genomic DNA was good and that the labelin-g and hybridization procedures were performed successfully. The profiles showed a variety of chromosome alterations such as homozygous and heterozygous deletions, different copy number amplifications and in general different chromosome alterations of variable size across the whole human genome (Fig. 7 a,b,c). A solid tumor sample taken from a patient with ovarian cancer was also tested. Tumor samples were biopsied from the patient,

first pre-treatment (Fig. 7d), and then as a resection following completion of an appropriate chemotherapy treatment course (Fig. 7e). Differences were observed in chromosomal composition between the two samples, which established the basis for understanding the underlying biological processes associated with treatment response.

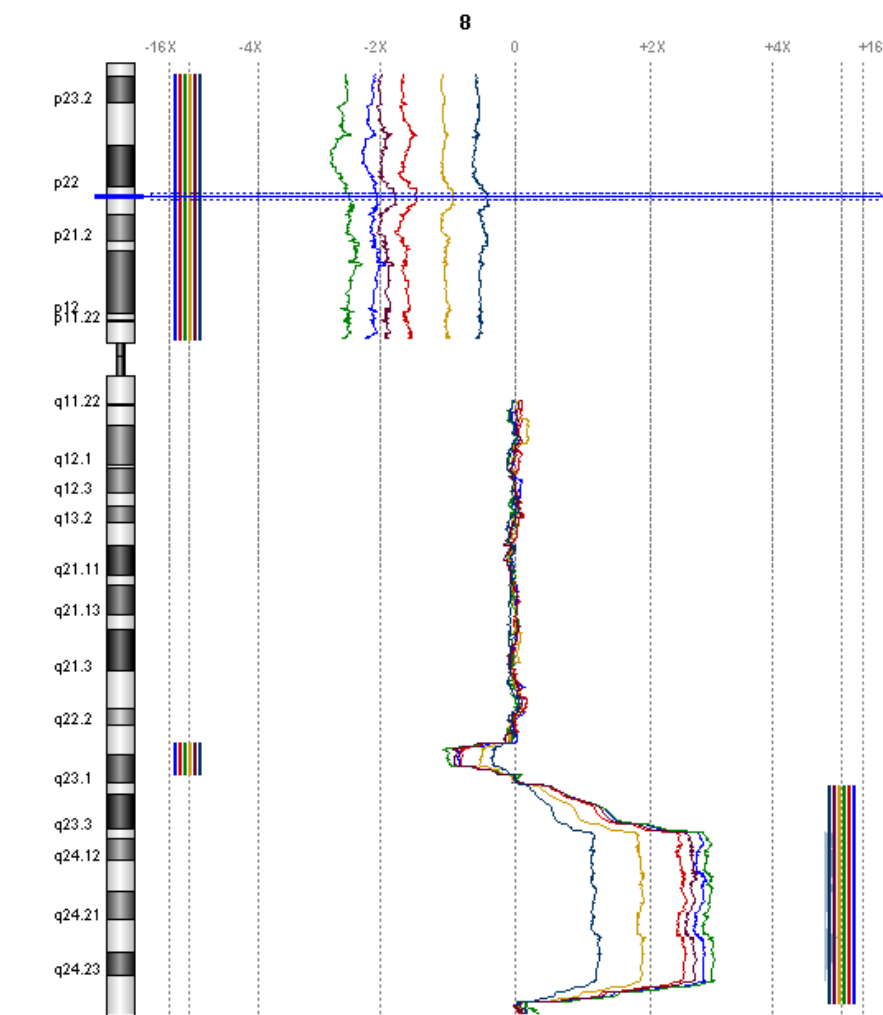
### Discussion

Agilent Technologies has developed a new 60-mer oligonucleotide based microarray to detect genomic DNA copy number changes, a technique that is commonly called Comparative Genomic Hybridization (CGH). The development of this microarray started with the "proof of concept" study demonstrating that microarray-based CGH could be performed using oligonucleotides (60-mers). The initial results obtained using Agilent 60-mer oligo gene expression microarrays were encouraging, but far from optimal. The complexity of the genome is around 100 times higher than that of the transcriptome. Many assumptions made for probes designed

to measure gene expression are therefore not valid for probes intended for CGH. 60-mer probes can be designed to hybridize with unique genomic DNA sequences. However, this is mathematically impossible with shorter oligonucleotides. Examining all possible combinations of 25-mer or 30-mer probes for the whole genome sequence, it was impossible to obtain a collection of short oligonucleotides that would individually hybridize with a unique genomic sequence. This illustrates the importance of the probe design and optimization for a particular application, in this case CGH. Probes that have been designed for other applications, such as single nucleotide polymorphism (SNP) detection or gene expression profiling are not likely to give the same level of robust performance, thus compromising the quality of the results obtained for CGH application.

The Agilent Human Genome aCGH microarray contains approximately 43,000 probes that have gone through thorough in-silico selection and an empirical validation process. The narrow  $T_m$  range for all the probes on the microarray enables common optimal processing conditions for virtually all the probes, ensuring high sensitivity and specificity.

Performance was optimized using model systems, such as cell lines carrying different numbers of Chromosome X, that allowed comparison of the results with those predicted so that the conditions could be improved. A software program was also developed, CGH Analytics, that enables the user to visualize and analyze data obtained from aCGH microarrays. Two different algorithms are implemented in the software to help users make correct and reliable aberration calls. The software allows the user to filter data, export it and import it in different formats, and visualize it in many different ways, offering a great degree of flexibility needed for CGH data analysis. The



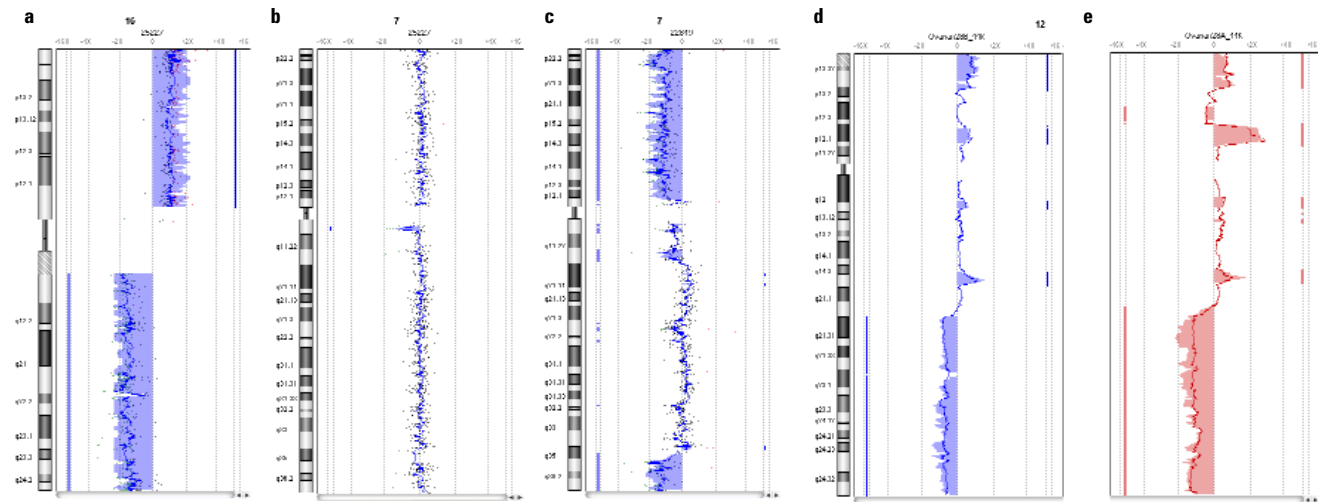
**Figure 6. Chromosomal aberrations detected in heterogeneous HT29 samples. Chromosome 8 view with HT29 cell line at increasing dilutions with normal genomic DNA (from 90%, 80%, 70%, 50% to 30% purity respectively).**

software also provides a very useful Quality Metrics panel that researchers can use to determine the success of particular experiments.

CGH Analytics enables the user to visualize and analyze data obtained from aCGH microarrays. Two different algorithms are implemented in the software to help users make reliable and statistically-relevant aberration calls. The software also allows the user to filter data, export and import with different file formats, and visualize results at multiple zoom-in levels, offering a great degree of flexibility needed for CGH data analysis. Additionally, the software provides a

useful Quality Metrics panel that researchers can use to assess the success of a particular microarray experiment.

Results from the two different cell lines derived from cancer patients demonstrate how the Agilent aCGH platform can be used to obtain high quality data pertaining to genomic DNA copy number changes. In addition to the detection of all aberrations previously described with high accuracy, the high resolution of Agilent microarrays allowed us to detect aberrations so small that they were never described before with any other microarray CGH technique. Agilent also offers High Definition microarrays that can be used



**Figure 7. Examples of various chromosomal aberrations in different tumor samples. (a) Gain of p-arm and loss of q-arm on Chromosome 16; (b) single focal deletion on Chromosome 7 and (c) a mix of alterations on the same chromosome from two patients diagnosed with Acute Myeloid Leukemia. Aberrations on Chromosome 12 of ovarian cancer samples obtained from the same patient before (d) and after (e) chemotherapy treatment.**

to "zoom-in" at any particular region of interest of the genome. This is possible thanks to a database of 4 million probes that have been specially designed to detect DNA copy number changes. The wide range of aberrations that Agilent aCGH microarrays can detect proves the great versatility and accuracy that the Agilent aCGH platform offers. Moreover, Agilent aCGH microarrays produced accurate results using DNA obtained from a patient diagnosed with Acute Myeloid Leukemia and a patient diagnosed with ovarian cancer. This demonstrates that the Agilent aCGH platform can be successfully used with real patient samples from very different origins, and therefore obtained using different techniques, to achieve accurate results. The genomic DNA used on the Agilent platform is always non-complexity reduced, allowing the broadest use of CGH by simplifying

the preparation of targets and the hybridization of the sample to any microarray design of interest.

In summary, Agilent has developed a complete CGH solution, consisting of microarrays, reagents, hardware and software, for studying chromosomal abnormalities associated with cancer and other genetic diseases. It fully complements our well-established gene expression product portfolio. Agilent has also created a database with 4 million computationally validated CGH probes for designing custom High Definition microarrays. The use of this complete Agilent aCGH solution will lead to new discoveries in the CGH field.

To learn more about Agilent's oligo aCGH solution, visit us at [www.agilent.com/chem/goCGH](http://www.agilent.com/chem/goCGH).

## REFERENCES

1. Michael T. Barrett, Alicia Scheffer, Amir Ben-Dor, Nick Sampas, Doron Lipson, Robert Kincaid, Peter Tsang, Bo Curry, Kristin Baird, Paul S. Meltzer, Zohar Yakhini, Laurakay Bruhn, and Stephen Laderman. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *P.N.A.S.* 2004 Dec; 101(51): 17764-17770.
2. J.A. Hanley, B.J. McNeil. The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.

## CGH Analytics Insert

Agilent's CGH Analytics software (v. 3.2) was used to visualize the chromosomal aberrations presented in this publication. This software uses two methods to make aberration calls: Z-score and ADM-1 (Aberration Detection Method 1). Both methods require a reliable measurement of the log ratio noise, which is defined as the Derivative of the Log Ratio Spread.

### Derivative Log Ratio Spread (DLRSpread)

To make aberration calls, CGH Analytics software needs the input of a measure of log ratio noise. In earlier versions of the software, a set of calibration microarrays consisting of a normal sample vs. normal sample was used. These microarrays have extensive stretches of chromosomes along which the genomic copy number should remain constant. The shape and the spread of the log ratios are calculated from the calibration microarrays and those statistics values are compared to sample microarrays with various chromosomal abnormalities to make amplification or deletion calls. Experience shows that this method, although quite useful, often underestimates the noise of the log ratios for the sample microarray. The noise estimate is more accurately calculated from the selected microarray itself.

Observations have shown that instead of using the standard deviation of the log ratio, a more robust estimate of noise is attained by calculating the spread of the log ratio differences between consecutive probes (DLRSpread) along all chromosomes (refer to CGH Analytics manual for more details). This eliminates the need for the calibration microarrays. The DLRSpread is also one of the most important metrics included in the CGH Analytics Quality Metrics panel. The ability of this metric to measure the noise of the log ratios independently of

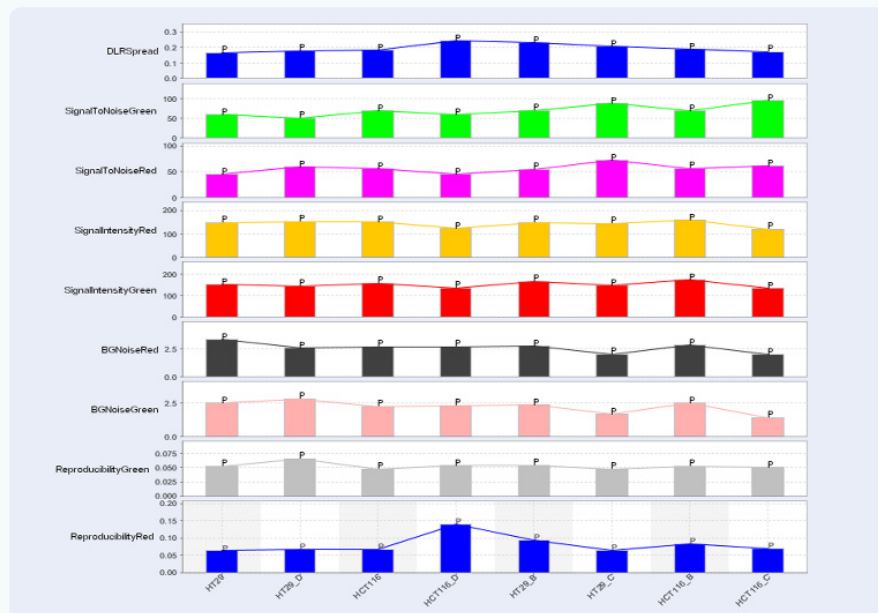


Figure A. Quality Metrics panel in CGH Analytics.

the number and severity of aberrations found, makes it instrumental in assessing the overall quality of each independent microarray experiment.

### Z-Score

CGH Analytics provides a fast method for the statistical analysis of aberrant regions based on hypergeometric Z-scores. This can be used for a quick determination of statistically interesting regions. However, once an interesting chromosome or region is found, it may be useful to follow this analysis with a more detailed search for aberrant regions (see alternative algorithm "ADM-1").

The "Z-score" aberration detection algorithm searches for chromosomal intervals containing an exceptionally high number of probes with log ratios significantly different from zero. The length of the intervals to be searched is set by the user in the Moving Average/Window box. The window should include at least three probes, or no aberrations can be found. A probe has a log ratio "significantly" different from zero if it is more than a certain distance

away from zero. That distance is calculated by multiplying the DLRSpread by a number called "Threshold". The "Threshold" is also user specified in the main user interface (UI). The number of non-zero probes in an interval is "exceptionally" high if there is a very low probability that at least that many non-zero probes would be observed by chance in an interval specified by the user. This probability is described by a hypergeometric distribution. The algorithm is described in more detail in the CGH Analytics manual.

### ADM-1

CGH Analytics provides Aberration Detection Method 1 (ADM-1), an alternative algorithm for finding the aberrant interval in a genome. The "ADM-1" aberration detection algorithm searches for chromosomal intervals of any length, for which the mean interval log ratio is significantly different from zero. An interval has a log ratio "significantly" different from zero if it is more than the "Threshold" standard deviations from zero. All possible intervals along a chromosomal arm are ranked by

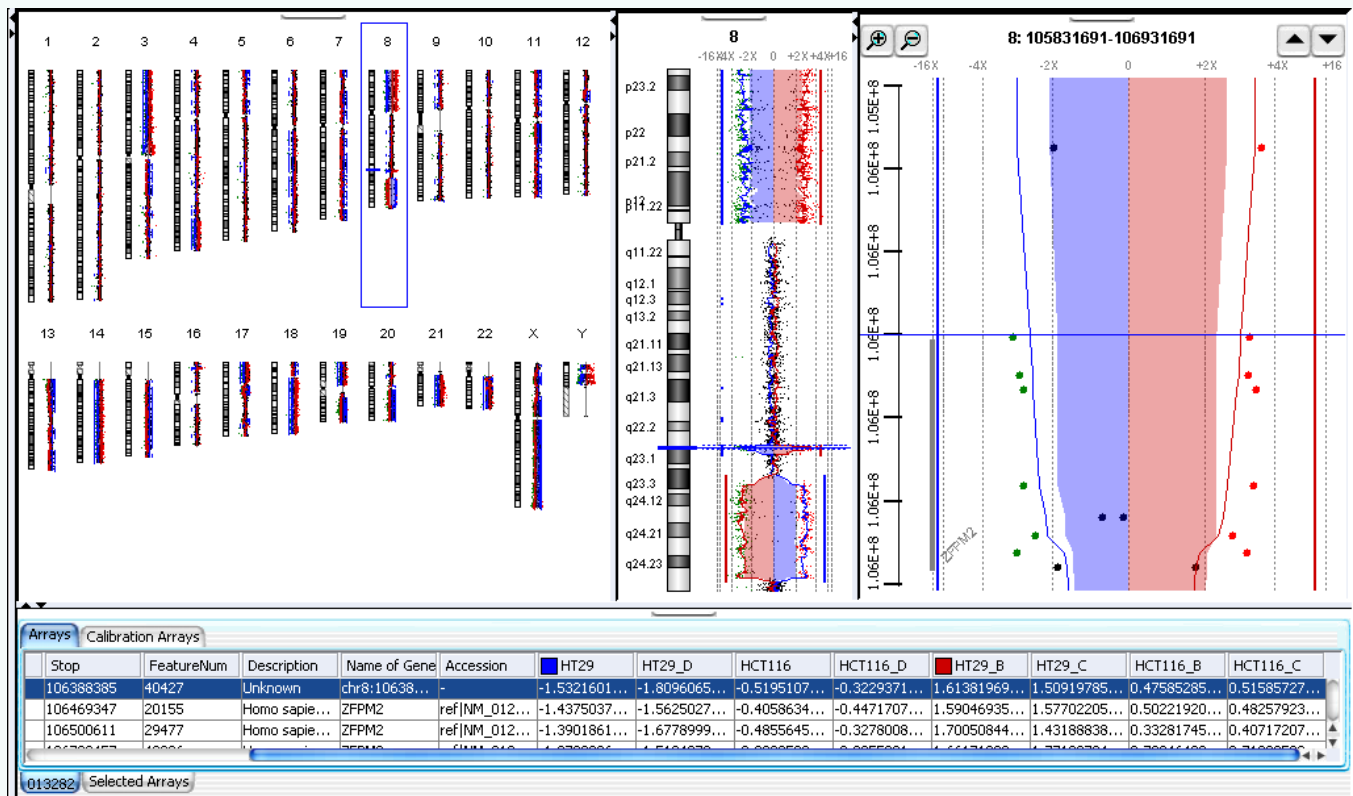


Figure B. CGH Analytics main visualization window.

the significance of their deviation from zero, and the most significant, above the user-specified threshold, are reported. This principle assigns statistical scores to intervals (I) of signals as described in the CGH Analytics manual. The scores are designed to reflect the statistical significance of the observed consistency of the high or low signals, and are useful in several levels of the DNA copy number (DCN) data analysis.

Here, 8 microarray profiles corresponding to HT-29 and HCT-116 cancer cell line experiments were imported into CGH Analytics, half of them with polarity 1 (cancer cell line DNA labeled with Cy-5 and Female or Male DNA labeled with Cy-3) and the other half of the opposite polarity, i.e.-1. CGH Analytics has an importer that allows users to name the individual profiles as well as to assign their polarities. Once the data is loaded into

the software, users can group the profiles that they want to visualize at once by creating an experiment and assigning the selected profiles to it.

The data quality of each microarray can be assessed using the Quality Metrics report provided in CGH Analytics (Fig. A). The DLRSread measures the probe to probe Log Ratio noise. High values of this metric (i.e. >0.3) usually indicate problems with sample DNA quality or amplification failure. Low Signal Intensity (i.e. <50) is likely due to too little input DNA, loss of DNA during the purification step or a faulty labeling reaction. High Background Noise (i.e. >10) can result from poor slide handling, problems with the wash procedures or contamination of any of the reagents used to process the microarrays. Signal-To-Noise Ratio (SNR) is the direct result of dividing Signal Intensity over Background noise. A SNR value greater

than 30 is highly desirable in order to significantly differentiate the true biological differences from the random or systematic noise. The Reproducibility value is calculated using the Log ratio measurements of the 200 replicated probes (x3) across the microarray. Values higher than 0.2 are indicative of poor hybridization mixing, large bubbles, leakage or other catastrophic failures. The Quality Metrics guideline provided in the CGH Analytics manual reflects what Agilent has typically observed for reference cell lines that are analyzed using the standard Agilent aCGH protocol on Human Genome aCGH microarrays. The Quality Metrics values are likely to be worse when deviations from this occur.

The Quality Metrics values for all 8 microarrays are within the standard guideline, indicating that there was no major problem with the quality of

the samples, and the microarray experiments were performed successfully. Reproducibility measurements of all non-control probes across the 4 replicate experiments revealed an average Standard Deviation of the Log Ratios of 0.052 for the HCT116, and 0.046 for HT29.

CGH Analytics provides four different views of the data (Fig. B). The Genome view gives a high level overview of all the chromosomes within a genome. The chromosome of interest can be selected for further investigation at a higher resolution in the Chromosome view, while the Gene View displays the raw data on a probe-by-probe level for the selected chromosome. The individual probes can be visualized side-by-side with the ReqSeq genes that are pre-loaded into the software, allowing a fast analysis of the genes and regions involved in particular chromosome abnormalities. The Table View contains the probe location, annotation and raw Log Ratio value for each of the probes on all the microarrays in the selected experiment.

CGH Analytics allows simultaneous visualization of the raw ratio value (via scatter plot) and the Log Ratio Error envelope for each probe, the moving average values and the aberration calls (using Z-score or ADM-1 algorithms). The users can self-define the moving average window size, using the number of probes or molecular distances. In addition, there are three filter categories that can be used to eliminate outliers prior to downstream data analysis, including aberration filters, microarray level filters and feature level filters.

Furthermore, CGH Analytics can combine data from microarrays that share a given attribute. The program will combine the Log Ratios from the same probes creating a new "combined microarrays" data set. When combining the same probes within a single microarray (intra-array) or from replicated microarrays of same polarity, the algorithm assigns different weights to each of the data points based on the Log Ratio Error value for each of the individual data points. However, when

combining dye-swap pairs, the algorithm takes the arithmetic average of log ratios from the same probes from opposite polarity in order to minimize any dye bias. The final compiled results for a given experiment can be presented using any of the three reporting summaries provided within the software. The text aberration summary includes the aberration scores and log ratios of the probes included in the aberrant regions across the genome. The graphical aberration summary shows a heat map of the gains and losses on a selected chromosome for each of the profile in a selected experiment. The graphical penetrance summary indicates the percentage of the selected microarrays that share the same gains and losses for each probe on the selected chromosome.

Go to [www.agilent.com/chem/goCGH](http://www.agilent.com/chem/goCGH) to learn more details or download a trial license of the CGH Analytics software.

### Agilent Online Resources At Your Fingertips

- Ask the Experts - Learn how to get the most out of your Agilent products.
- Special offers - Benefit from savings on Agilent products and services.
- Events calendar - See Agilent at leading industry conferences and trade shows.
- Web links - Go directly to detailed information on the Web.

### About Agilent's Integrated Biology Solutions

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms and speed drug discovery. Engineered for sensitivity, reproducibility and workflow productivity, Agilent's integrated biology solutions include instrumentation, microfluidics, software, microarrays, consumables and services for genomics, proteomics and metabolomics applications.

### For More Information

Learn more:

[www.agilent.com/chem/goCGH](http://www.agilent.com/chem/goCGH)

Buy online:

[www.agilent.com/chem/store](http://www.agilent.com/chem/store)

Find an Agilent customer center in your country:

[www.agilent.com/chem/contactus](http://www.agilent.com/chem/contactus)

#### u.s. and canada

1-800-227-9770

[agilent\\_inquiries@agilent.com](mailto:agilent_inquiries@agilent.com)

#### europa

[info\\_agilent@agilent.com](mailto:info_agilent@agilent.com)

#### asia pacific

[adinquiry\\_aplsca@agilent.com](mailto:adinquiry_aplsca@agilent.com)

Research use only. Information, descriptions and specifications in this publication are subject to change without notice.

Agilent Technologies shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance or use of this material.

SurePrint is a U.S. registered trademark of Agilent Technologies, Inc.

© Agilent Technologies, Inc. 2006  
Printed in the U.S.A. February 21, 2006  
5989-4848EN

