Application Note

Genomics

Agilent

Trusted Answers

High-Performance Aberration Detection from Low DNA Input and Mosaic Samples with the Agilent SureTag Labeling Kit and CGH Microarrays

Author

Scott Happe, Xuan Shen, Nga Truong, and Valentina Maran, Agilent Technologies, Inc.

Abstract

Array comparative genomic hybridization (aCGH) is a powerful technology for studying copy number variations (CNVs) that are associated with disease. In this application note, we tested the limits of detection for low DNA input and mosaic analyses using the Agilent aCGH method.

Introduction

In prenatal diagnoses, aCGH is the preferred genetic tool to analyze cases where fetal ultrasound reveals signs of abnormalities in utero or when investigating the cause of stillbirth (Norton and Jackson, 2017; ACOG Practice Bulletin #162). While various non-invasive prenatal testing methods can use maternal blood to identify large fetal aneuploidy events, conclusive proof of chromosomal abnormalities and identification of small segmental gains and losses require the analysis of chorionic villi (CVS), amniotic fluid, or products of conception (Rose and Mercer, 2016; ACOG Practice Bulletin #163). Direct testing on these samples without culturing may be preferred to avoid growth bias and other artifacts (Carey, et al., 2014). However, with direct testing, samples are limited to as little as 50 to 100 ng of DNA for subsequent labeling and hybridization to CGH arrays (Rickman, et al., 2006).

In prenatal testing, minor clones with distinct CNVs may exist within the sample. For example, confined placental mosaicism may result in chromosomal abnormalities in extra-embryonic cells of the CVS sample that are not present in the fetus. Alternately, true fetal mosaicism may be detectable directly from the amniocentesis (Grati, 2014).

The interpretation of low-level aneuploidy results can be impacted by sample type, the use of primary or cultured cells, and sensitivity of mosaic detection. The Agilent SureTag Complete DNA Labeling kit (p/n 5190-4240) is designed to generate high-quality labeled DNA from a variety of sample types for use in aCGH experiments, even when availability of starting sample is limited.

To demonstrate the performance and sensitivity of the SureTag kit in combination with Agilent CGH arrays, we compared low-input (50 ng) results to those generated with the specified standard input (500 ng for 4x format; 200 ng for 8x format). In addition, we analyzed CNV detection under simulated mosaic conditions by mixing diploid DNA with DNA from cell lines containing known segmental chromosomal aneuploidies.

Experimental design

DNA derived from either de-identified blood or cell line samples was quantified and shown to meet standard input quality control (QC) requirements as specified in the Agilent CGH protocol (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Version 7.5, June 2016) (p/n G4410-90010).

Accurate quantification of starting sample is critical for array quality. This is even more important when working with a low starting amount. Both reference and sample need to be quantified using the same method and have starting material amounts precisely matched to minimize variation prior to starting the labeling process. For the experiments reported in this Application Note, DNA was measured with the Qubit 4 Fluorometer (p/n Q33226, Thermo Fisher Scientific) and the indicated input sample amount (50, 200, or 500 ng) was paired with the matching reference sample quantity.

The expected size of each CNV from three blood-derived specimens, based on prior combined array analyses from two different formats/platforms (Agilent GenetiSure Dx Postnatal Assay and Affymetrix CytoScan HD), is described in Table 1.

Table 1. ID and description of aberrations of samples used in these studies. The -F and -M suffixes represent female and male samples, respectively.

Sample ID*	Aberration Size	Gain or Loss
AG16-121-F	9.8 Mb	Loss
AG16-123-F	5.0 Mb	Gain
AG16-085-M	0.7 Mb	Gain

For testing on the ISCA array design (p/n G4827A-031746), mosaicism was simulated prior to labeling by mixing the three blood-derived samples that contained known aberrations at 100%, 40%, 20% and 10% ratios with respect to sex-matched Agilent Reference DNA. The resulting samples were tested at both 500 ng and 50 ng input (with matching quantities of reference DNA). Samples were processed using the standard Agilent CGH protocol referenced above with heat fragmentation; these samples were not subjected to a preamplification step and were labeled with Cy5 fluorescent dye.

Sex-matched Agilent Reference DNA was labeled in parallel with Cy3 fluorescent dye. After purification, sample-reference pairs were combined after matching specific activity as closely as possible. Sample pairs were hybridized for 24 hours to the ISCA CGH G3 8x60k arrays (AMADID 031746), washed, and scanned according to standard procedures. Agilent CytoGenomics Software (Version 5.0.1.6) was used to extract and analyze the data.

For mosaic samples, we developed a custom Long Low Algorithm to provide the best sensitivity while minimizing false-positive calls. Filters were set as follows: min size ≥ 300 kb, absolute mean Log Ratio filter ≥ 0.07 (minimal theoretical value to detect 10% mosaicism (Table 2) and Aberration Detection Method-2 ≥ 7. For these arrays and samples, we kept the same log ratio limit for both gains and losses. This could be modified in the normal analysis settings based on experimental design to allow a more consistent level of detection for gains/losses in the analyzed sample.

To test the new GenetiSure Cyto arrays, three cell line-derived samples (NA10636, NA10925, and NA12662 (Coriell Institute for Medical Research, Camden, NJ)) carrying known representative CNVs of various sizes were selected. Simulated mosaic samples were prepared as above. Ratios tested for these arrays were 100%, 50%, 20%, 15% and 10%. The mixed DNA samples were tested at the standard total input levels (500 ng for the 4x180k array format and 200 ng for the 8x60k array format) on three new GenetiSure Cyto arrays (8x60k CGH, p/n G5982B; 4x180k CGH, p/n G5983B; 4x180k CGH+SNP, p/n G5984B). Samples were subjected to heat fragmentation, labeling, hybridization, and slide scanning procedures as outlined in the standard protocol. Appropriate analysis methods with customized thresholds for mosaic CNV calling (minimum ADM-2 score down to 2.7; minimum absolute average log ratio down to 0.06 for gain and 0.07 for loss) were applied in the Agilent CytoGenomics software for CGH data analysis.

Results and Discussion

The first set of experiments focused on blood-derived samples applied to the 8x60k CGH ISCA array. The QC metrics produced by CytoGenomics were in the good to excellent range (as defined by the software) for both sample input levels and all mosaicism levels. A particularly important metric, DLRSD (Derivative of Log Ratio Standard Deviation), measures noise based on the variation in log ratios of adjacent probes across the array. In these experiments, it ranged from 0.097 to 0.15 (mean = 0.119) for 500 ng input and 0.13 to 0.19 (mean = 0.148) for 50 ng (Figure 1).

Other metrics (such as signal-to-noise ratio) passed according to standard CytoGenomics software thresholds (e.g. > 30 for signal-to-noise). While improved values for certain QC metrics are observed with 500 ng input compared to 50 ng, the ultimate detection of a given aberration at 50 ng input was not affected. Therefore, these data suggest that, in general, the relative magnitude of a given QC value does not reflect the assay performance at the aberration-detection level as long as these values are within the good-excellent ranges (see Figure 2).

As shown in the Genome Views in Figure 2, the expected aberration for each 100% sample was called at both the 500 ng and 50 ng input levels. These data demonstrate that, even at a tenfold lower input than the standard SureTag recommendation, aberrations from 700 kb to 10 Mb were able to be detected by the CytoGenomics software. Thus, SureTag provides high-resolution detection of gains and losses when input sample is limited, enabling the analysis of samples with limited starting material.

Table 2. Summary of quality control (QC) metrics for various mosaic levels across both DNA sample inputs.

DNA Input (ng)	Mosaic Level (%)	QC Metrics Summary (mean ± SD)		
		DLRSD	Green SignalToNoise	Red SignalToNoise
500	All	0.119 ± 0.018	88.6 ± 10.6	85.1 ± 14.3
	100	0.147 ± 0.008	84.2 ± 10.6	76.2 ± 15.9
	40	0.118 ± 0.004	91.2 ± 11.1	88.2 ± 14.8
	20	0.106 ± 0.004	92.3 ± 10.7	88.2 ± 12.8
	10	0.106 ± 0.010	86.8 ± 14.5	87.8 ±18.5
50	All	0.148 ± 0.017	32.6 ± 2.1	34.7 ± 2.4
	100	0.175 ± 0.011	31.4 ± 0.9	36.8 ± 1.2
	40	0.142 ± 0.003	34.2 ± 1.6	36.1 ±1.5
	20	0.138 ± 0.001	33.6 ± 3.1	34.0 ±2.5
	10	0.136 ± 0.002	31.0 ± 0.7	32.1 ± 1.2

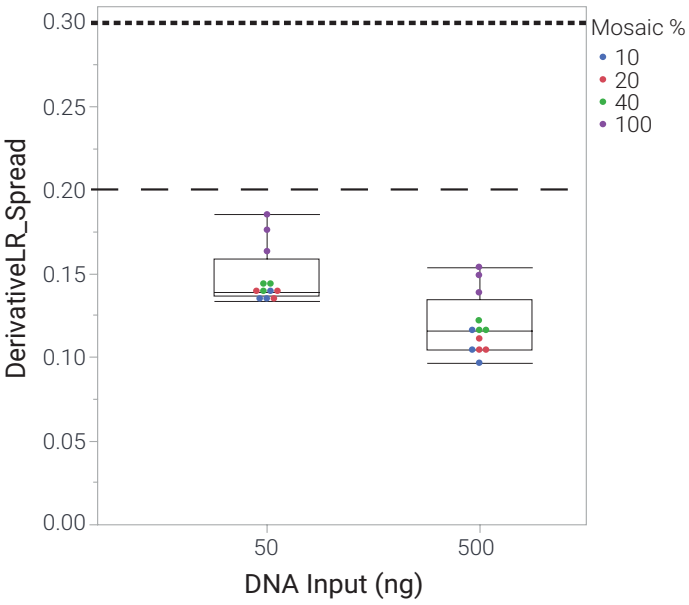

Figure 1. Box plots of DLRSD values for all samples in the study at either 500 ng or 50 ng. Note the reproducibility of replicate samples regardless of starting DNA amount, highlighting the robustness of the platform.

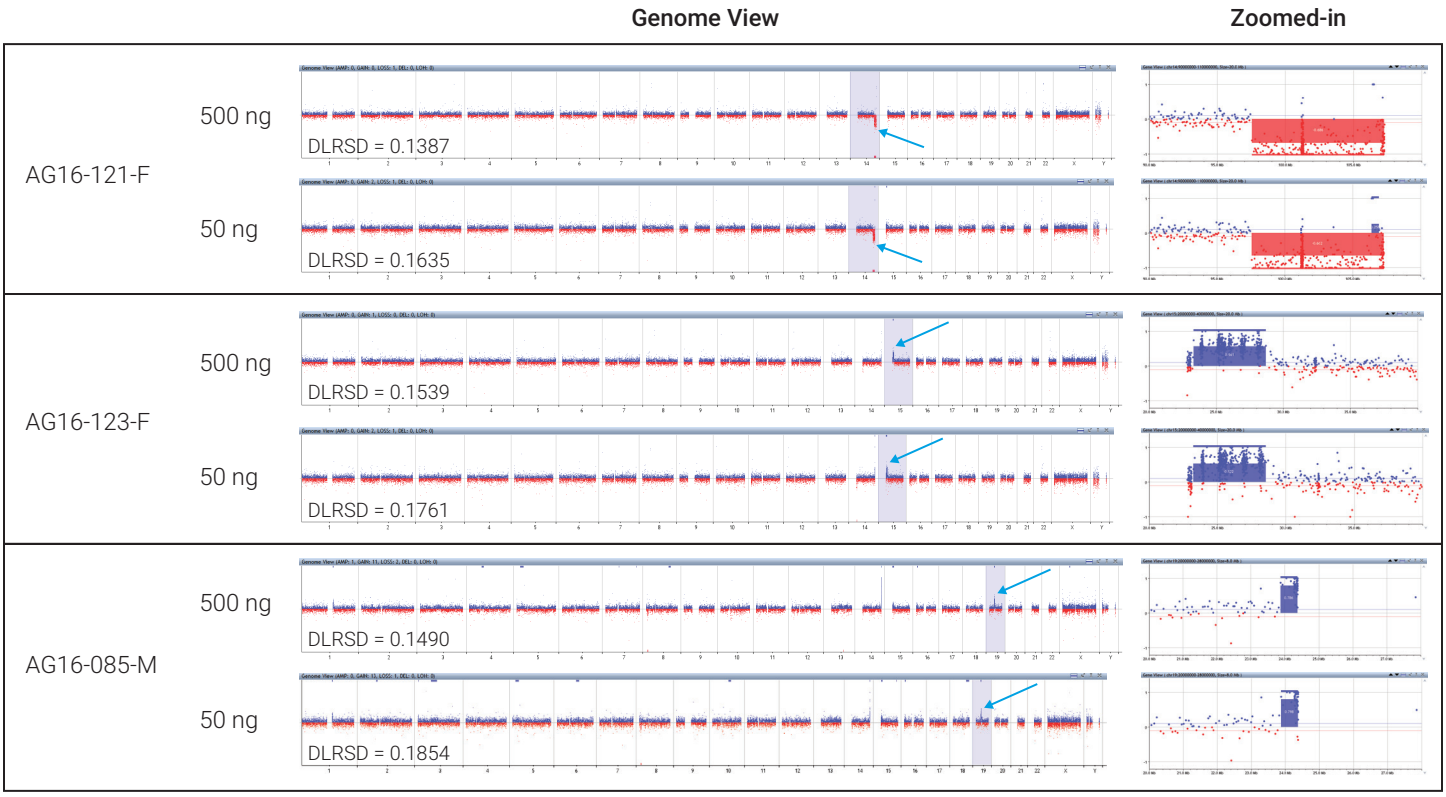

Figure 2. Genome views and DLRSD values of samples containing 100% aberration ratios. The left panel shows genome views of samples containing gains or losses of 9.8 Mb (AG16-121-F), 5 Mb (AG16-123-F), and 0.7 Mb (AG16-085-M) at 500 ng or 50 ng input. The right panel depicts a zoomed-in view of each aberration, demonstrating equivalent detection independent of sample input.

Next, the ability to detect mosaicism at both high and low input levels was evaluated. In practice, the standard CytoGenomics settings are used to analyze a sample. When mosaicism is suspected through observation of the data and log ratios of individual loci, a secondary analysis is performed to further characterize the prospective mosaic event. For these samples, the settings described in the experimental section were used for the secondary analysis (though settings may be modified depending on your sample in question).

In this study, aberrations from 700 kb to 10 Mb were detected at different mosaicism levels for both inputs (Figure 3). As an example, a 700 kb gain on chromosome 19 was readily observable down to the 20% level for both inputs (Figure 3A, B). The log ratio of each mosaicism level also correlated well with expected values (Figure 3C). Of note, the number of probes within each aberration showing significant log ratio deflection from zero remains similar for both low and standard inputs (data not shown).

In practice, the level of mosaicism that can be detected depends on a number of factors including the aberration's size, copy number, presence of log ratio compression, and number of probes within the evaluated region. In addition, when choosing algorithmic parameters for analysis, sensitivity must always be balanced with specificity.

We also tested simulated mosaic samples as part of verification studies for the new GenetiSure Cyto Arrays (p/n G5982B, G5983B, and G5984B). Overall, > 80% of targeted copy number gains and losses could be detected down to 15-20% mosaic levels, depending on the fragmentation method and array format. Examples of low-level mosaic detection are shown on the two CGH-only array formats (using heat fragmentation) with aberrations ranging from 2.9 to 19 Mb (Figures 4A and 4B). As expected, log ratio response is linear with respect to mosaicism level (Figure 4C).

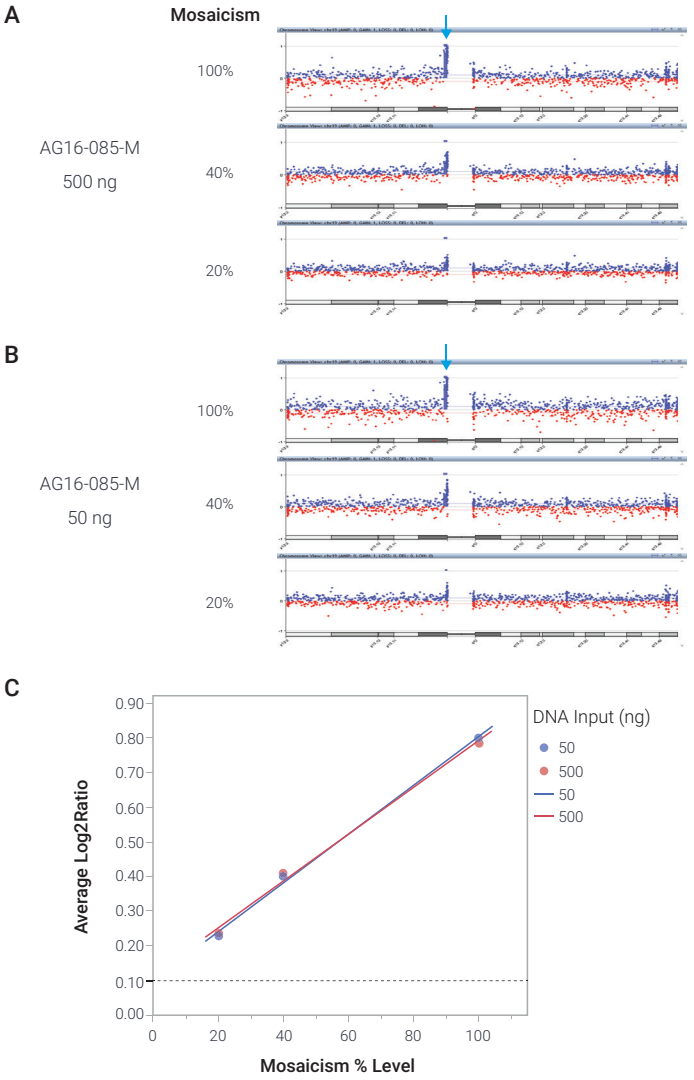

Figure 3. Detection of simulated mosaic samples down to the 20% level. A 700 kb gain was able to be detected down to the 20% level with both 500 ng DNA input (panel A) or 50 ng (panel B) with the SureTag labeling kit. The log ratio values of each sample correlate with expected mosaicism level (panel C).

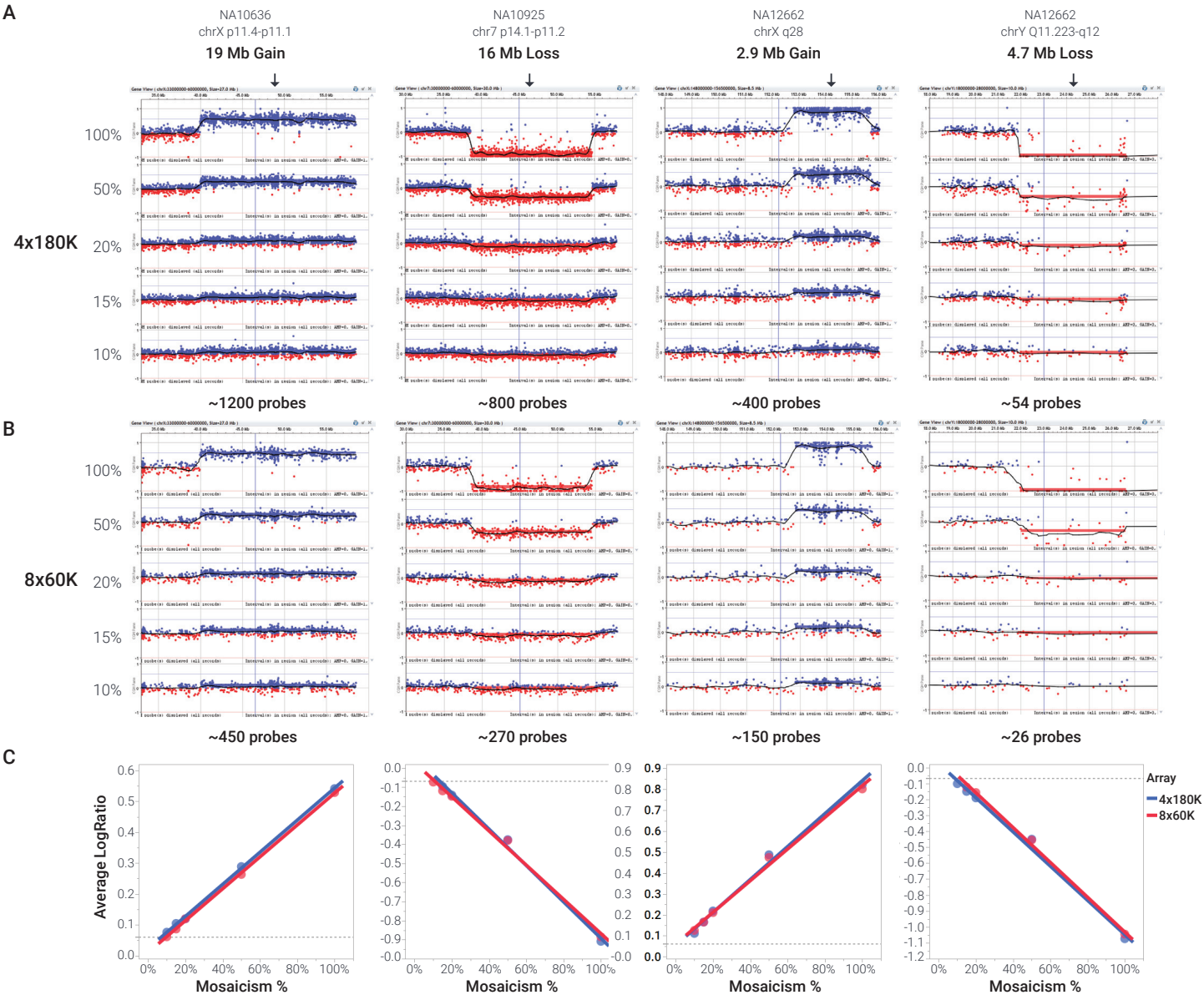

Figure 4. Detection of simulated mosaic CNVs with the new GenetiSure Cyto Arrays.

Conclusions

Taken together, these data demonstrate the ability of the SureTag labeling system and Agilent CGH arrays to detect low-level mosaicism even at low DNA inputs. High sensitivity and low DLRSD values down to 50 ng input was demonstrated for aberrations ranging in size from 700 kb to over 10 Mb. Mosaicism of 15 to 20% was observed using appropriate algorithms. SureTag and Agilent arrays are useful for aCGH analysis with limiting DNA for applications in prenatal and postnatal genetic analysis where analysis of mosaicism or heterogenous cell populations is important.

References

Carey, L.; Scott, F.; Murphy, K.; Mansfield, N.; Barahona, P.; Leigh, D.; Robertson, R.; McLennen, A. Prenatal diagnosis of chromosomal mosaicism in over 1600 cases using array comparative genomic hybridization as a first line test. *Prenat. Diagn*. **2014**, *34(5)*, 478-486.

Grati, F.R. Chromosomal Mosaicism in Human Feto-Placental Development: Implications for Prenatal Diagnosis. *J Clin Med*. **2014**, *3(3)*, 809-837.

Norton, M.E.; Jackson, M. and committee. Practice Bulletin No. 162: Prenatal Diagnostic Testing for Genetic Disorders. *Obstet Gynecol*. **2016**, *127(5)*, e108-122.

Rickman, L.; Fiegler, H.; Shaw-Smith, C.; Nash, R.; Cirigliano, V.; Voglino, G.; Ng, B.L.; Scott, C.; Whittaker, J.; Adinolfi, M.; Carter, N.P.; Bobrow, M. Prenatal detection of unbalanced chromosomal rearrangements by array CGH. *J Med Genet*. **2006**, *43(4)*, 353-361.

Rose, N.C.; Mercer, B.M. and committee. Practice Bulletin No. 163. Screening for Fetal Aneuploidy. *Obstet Gynecol*. **2016**, *127(5)*, 979-981.

www.agilent.com

PR7000-2625

Printed in the USA, October 2020
5994-2722EN

Agilent
Trusted Answers