Agilent

Trusted Answers

In silico Probe Coverage Comparison Between Agilent GenetiSure Cyto vs Legacy Microarrays

Authors

Xuan Shen, Didier Goidin, Tri Doan, Valentina Maran, and Scott Happe, Agilent Technologies, Santa Clara, USA

Abstract

Copy number variations (CNVs) are a hallmark feature of constitutional disorders. Microarray-based CGH has been the gold-standard technique for identification of CNVs in the clinical setting. In this technical note, we performed an *in silico* analysis of content in our latest CGH and CGH+SNP microarrays, GenetiSure Cyto, versus three legacy microarrays (CytoScan HD and CytoScan 750K from "Competitor A", and CytoSNP 850K from "Competitor B"). We demonstrate GenetiSure Cyto arrays have comprehensive coverage of clinically-important targets involved in CNVs, with minimal coverage bias vs legacy arrays. Furthermore, the design flexibility of the Agilent microarray platform allows for focused exon- level coverage, allowing detection of smaller aberrations that may be missed by legacy arrays as well as reducing confirmatory testing burdens for labs. We envision these new arrays will likely add value to any lab conducting routine testing of constitutional samples.

Introduction

The past decade has seen tremendous progress of genomic technologies and their utility in clinical applications. For example, chromosomal microarrays (CMAs) have been widely used for cytogenetics applications in genetic laboratories around the world. Over the past 10 years, CMAs have been established as the first approach for evaluating pre- and postnatal samples and detecting copy number variations (CNVs) that may be associated with genetic disorders.[1][2][3]

To assess the significance of these findings, cytogeneticists often refer to reputable public databases (such as ClinGen, DECIPHER/DDG2P, and OMIM) to determine the clinical significance of their findings. These databases (and others) that aim to catalog, curate, and characterize these genomic findings—as well as their relationship to disease—have also been expanding in both size and scope as new data is generated. This constantly changing landscape makes it necessary to periodically reassess and update the contents of existing CMA designs.

The previous generation of Agilent CGH and CGH+SNP array designs used by cytogeneticists includes the SurePrint G3 ISCA arrays. These arrays were designed over ten years ago as an outcome of the collaborative efforts in the International Standards for Cytogenomic Arrays (ISCA) consortium, a predecessor of what is now the ClinGen database. Given how much the field has learned in the decade since their inception, a content update to these arrays was warranted.

Experimental

In pursuit of this, Agilent has recently designed a new generation of human cytogenetic microarrays—the GenetiSure Cyto CGH and CGH + SNP microarrays. They are available in three formats commonly used in analysis of constitutional DNA samples: 4x180K CGH (AMADID 085589), 8x60K CGH (AMADID 085590), and 4x180K CGH+SNP (AMADID 085591). Compared to the existing ISCA arrays, these new GenetiSure Cyto arrays take advantage of the vastly expanded information now available from relevant databases and are able to cover a much broader range of targets. The total number of targeted genes and regions has been increased from about 500 for the ISCA arrays to more than 3,600 for the new GenetiSure Cyto arrays.

As with all solid-phase array technologies, only a limited number of probes can be placed within a given array format. By optimizing probe selection and placement strategies we can achieve a balanced, high-performance probe distribution across targeted regions and the genomic backbone. Probes on each array format have been selected from Agilent's proprietary HD probe database for optimized coverage across the expanded range of targets. In addition to enforcing a minimum probe coverage (five probes per target) across most targets, additional probe enrichments have been applied in selected regions of higher clinical interest. Separately curated backbone probe sets are incorporated to maintain an essential probe coverage level across the whole genome.

The updated probe distribution profiles are expected to provide an enhanced CNV detection performance for cytogenetics applications. With the old ISCA designs, target probes were placed over a comparatively smaller number of targets, leading to overly high probe densities in some targeted regions. As illustrated in Figure 1, probes are redistributed in the new GenetiSure Cyto designs to not only cover an increased number of targets, but also avoid the overcrowding issue observed on certain loci in the old ISCA designs. However, the genomic backbone (e.g., regions between targets) is still adequately covered, with increased probe densities towards autosomal sub-telomeric regions and allosomal pseudo-autosomal regions (PAR).

Each of the three new array designs consists of 50 unique spike-in probes for sample tracking purpose. Adding the spike-ins to samples ensures the identity of the sample throughout the process and gives the user confidence that there was no mix-up of samples during wet lab or analysis processing[7]. In addition, on each of the new arrays, extra empty space (equivalent to about 500 or about 1500 unused feature positions on the 8x60k and 4x180k formats, respectively) has been reserved for optional custom content. Researchers can add content as needed using our user-friendly SureDesign web portal.

Each of the three GenetiSure Cyto arrays is suited for cytogeneticists with different needs. While ensuring enhanced coverage on all >3600 targets, the 8x60k CGH array provides the best sample testing throughput within the group. The 4x180k CGH+SNP array incorporates an expanded SNP probe set from the GenetiSure Postnatal Research Array 2X400K (p/n G5974A), enabling absence of heterozygosity (AOH) calling down to 2.5 Mb in size. Lastly, the 4x180K CGH array includes exon-focused probes for a selected set of 103 multiplex ligation-dependent probe amplification (MLPA) gene targets. MLPA assays are routinely used in clinical diagnostics to detect copy number variation related to genetic diseases, particularly at the exon level. Exon-focused probe enrichment on this array allows a minimum coverage of three probes per exon (including adjacent ±200 bp flanking regions) and mitigates the need to perform MLPA. Alternately, in labs that routinely perform MLPA, this allows for an easier comparison between CMA and MLPA results.

ISCA Design
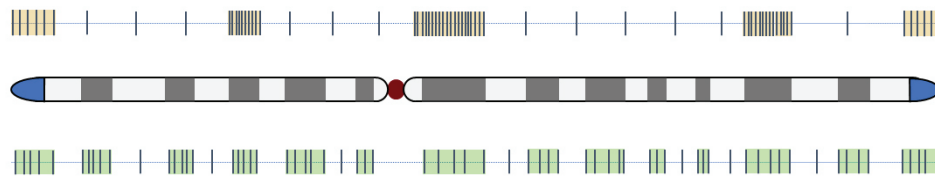
GenetiSure Cyto Design

Figure 1. Illustration of Broader Target Coverage by the New GenetiSure Cyto Array Designs. A schematic comparison of different probe distribution profiles between the old ISCA designs and the new GenetiSure Cyto designs. Greyed-out regions on the chromosome indicate the targets intended to be covered based on updated knowledge. Probe positions are indicated by thin vertical bars for the ISCA (top) and the GenetiSure Cyto designs (bottom), respectively. By optimizing probe placement, the GenetiSure Cyto design provides more balanced coverage within the same constraints of total probe numbers.

Results and Discussion

The aforementioned exon-focused strategy was expected to be a major advantage of the GenetiSure Cyto 4x180k CGH array. Exon-specific probes can provide superior resolution and sensitivity for detection of small exonic CNVs with potential clinical consequences. To demonstrate this advantage, a representative high-density microarray (CytoScan HD) from Competitor A and another medium-high-density microarray (CytoSNP 850K) from Competitor B were selected for comparison with the GenetiSure Cyto 4x180k CGH array for exon-level probe coverage on MLPA gene targets. the two comparator arrays contain roughly 2.7 million and 850,000 unique probes (markers), respectively, far exceeding that available on the Agilent 4x180k array. However, more probes do not equate to better performance. Agilent GenetiSure Cyto arrays require only ≥5 consecutive probes to make a CNV call versus ≥25 probes for competitor A and ≥10 probes for competitor B. The corresponding probe annotation files were obtained and analyzed along with the probe profile for the GenetiSure Cyto 4x180k CGH array. *In silico* comparison of probe distribution was performed across the three arrays based on probe coordinates (hg38) and visualized in the Integrative Genomics Viewer (IGV) software v2.11.9.[4]

As illustrated in Figure 2, the new Agilent GenetiSure Cyto 4x180k CGH array applies exon-focused coverage on two selected MLPA gene examples: FANCA and MID1. Probes are precisely placed in or near the exons, enabling an efficient coverage focused on coding region CNVs. In contrast, the two comparator arrays have probes placed less optimally for exon coverage. The SNP-focused probes on the CytoSNP 850K array, albeit with a relatively even distribution across the full range of the targeted genes, are clearly insufficient to enable aberration detection at the exon level.
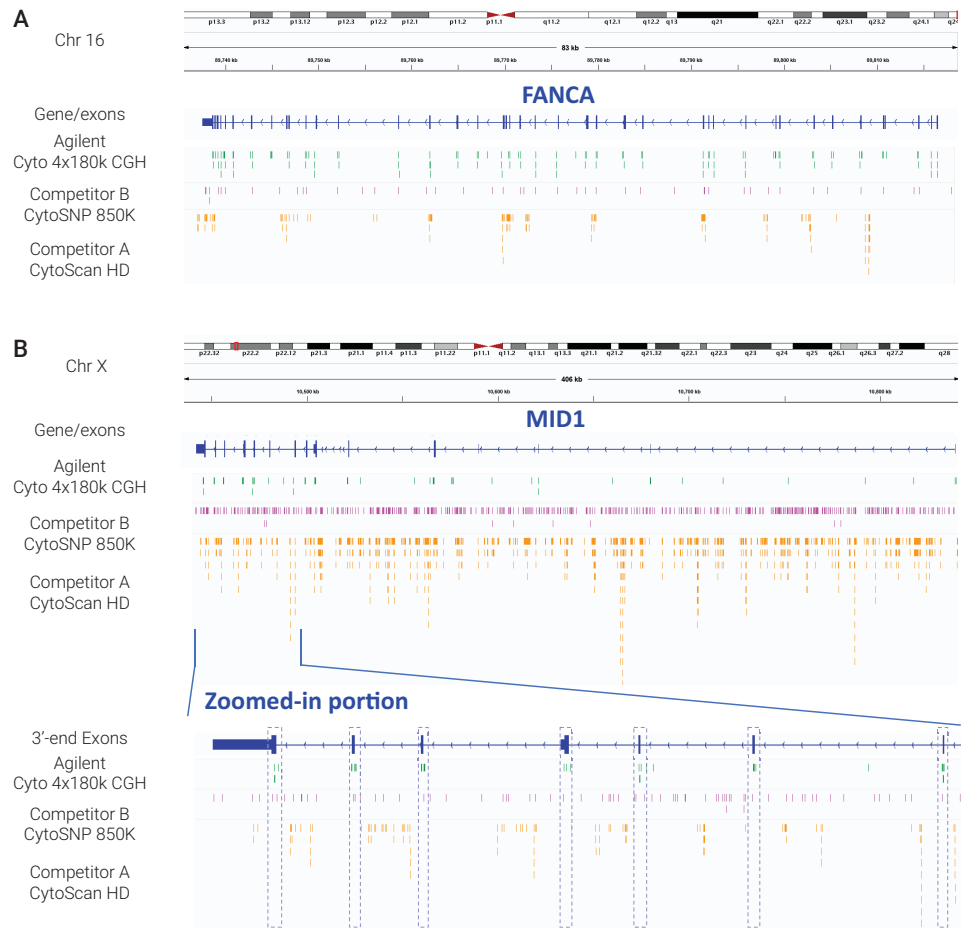

Figure 2. Examples of Exon-focused Coverage by GenetiSure Cyto 4x180K CGH Array. Two examples of exon-focused probe placement on MLPA gene targets by the GenetiSure Cyto 4x180K CGH array (AMADID: 085589), compared with a high-density array from Competitor A (CytoScan HD) and a medium-high-density array from Competitor B (CytoSNP 850K). In each example, the entire gene range and all exon positions are indicated, along with probe positions from each array design. Panel A: the FANCA gene on chromosome 16, approximately 80 kb. Panel B: the MID1 gene on chromosome X, approximately 400 kb. A smaller 3'-end portion of MID1 containing 7 exons is zoomed in to show further details.

The inconsistent or non-existing coverage on exons is more evident on the CytoScan HD array. Despite an overall greater number of probes placed within each gene target, many of the CytoScan HD probes fall into the intronic regions. This discrepancy is particularly evident in the case of the MID1 gene (Figure 2B), which spans a large range (approximately 400kb) with uneven exon distribution (mostly toward the 3' end of the gene). A significant portion of the probes available on the CytoScan HD array are located in the 5' exon-less regions. Even within the exon-rich portion of the gene (left side), the CytoScan HD probes are almost all located in the intronic regions.

As with our genomic surveying tools, a remaining challenge is how to best utilize the limited assets available on an established array platform to cover the growing spectrum of genomic targets of interest. Some competitors have opted to use a simple probe distribution approach that assigns consistent probe spacing either across the entire genome or throughout the full collection of regions of interest. We believe this strategy, although generally adequate in capturing large aberrations, is prone to picking up variants of unkown significance (VOUS) and incidental findings; additionally, it may miss the opportunity to detect smaller aberrations in certain focused regions of high clinical interest. To demonstrate the latter, a probe coverage analysis in stratified target groups (Figure 3, panels A and B) of clinical interest was conducted using two new Agilent GenetiSure Cyto arrays (4x180k CGH and 8x60k CGH) and three representative arrays with similar utilities (CytoScan HD and CytoScan 750K from Competitor A, and CytoSNP 850K from Competitor B). Although not included in this direct comparison, the Agilent GenetiSure Cyto 4x180k CGH+NSP array carries a similar 60k CGH probe set used in the 8x60k CGH array format and is thus expected to perform similarly for CNV detection.
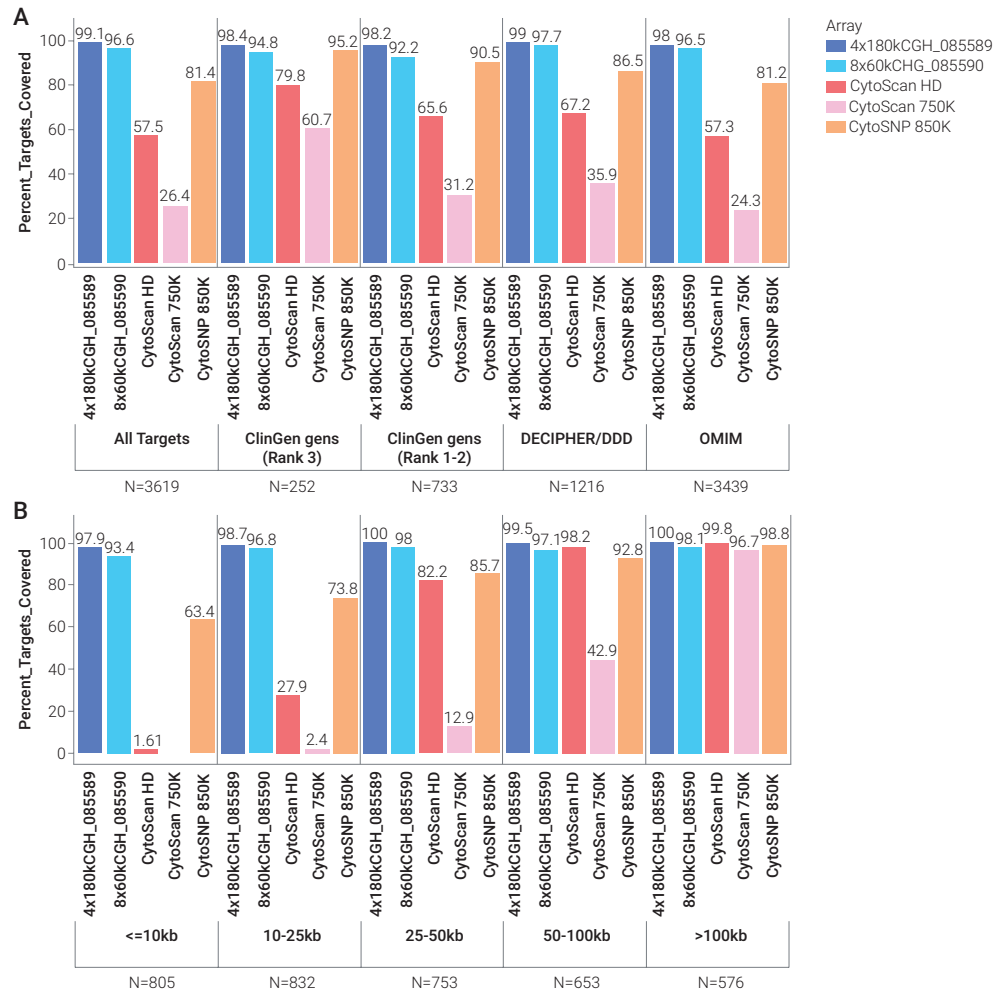

Figure 3. Comparison of Clinical Target Coverage between GenetiSure Cyto Arrays and CytoScan HD and 750K Arrays from Competitor A and CytoSNP 850K Array from Competitor B. Probe coverage efficiencies on clinically relevant targets were assessed for two representative GenetiSure Cyto arrays (4x180k CGH AMADID: 085589; 8x60k CGH AMADID: 085590) and three comparable arrays from Competitor A (CytoScan HD and CytoScan 750K) and Competitor B (CytoSNP 850K). For each category of targets assessed, the percentage of targets exhibiting sufficient coverage in each array design is plotted for comparison. The criteria for sufficient coverage are defined as presence of at least 5 probes per target for each GenetiSure Cyto array, at least 25 probes for each CytoScan array, or at least 10 probes for the CytoSNP 850K array. Panel A: coverage comparison for all clinically relevant targets combined, as well as for individual target groups from each primary source database. The number of targets encompassed in each group is labeled under each subgraph (N). Note that there are common targets that overlap multiple databases. Gene targets from ClinGen are further divided based on ranks of supporting evidence (Rank 3: sufficient evidence; Ranks 1 and 2: limited evidence). Panel B: coverage comparison stratified by size of the target (kb).

Probe profiles were obtained similarly as described above. Additional probe coverage statistics in the targeted regions were computed using the R statistical programming language. CNV detection sensitivity was assessed based on the minimum number of probes considered "sufficient" for CNV calling per target, which was defined according to manufacturers' recommendations for probe cutoffs: ≥25 probes for the two arrays from Competitor A,[5] ≥10 probes for the array from Competitor B,[6] versus ≥5 probes for the two Agilent GenetiSure Cyto arrays. A total of 3619 gene targets from a composite of clinically relevant targets from various databases were assessed for sufficient probe coverage on each target across the five array designs. The results are summarized in Figure 3 with different stratifications. We note that, due to the use of longer DNA oligonucleotide probes (approximately 60-mers versus 25-mers for Competitor A or 50-mers for Competitor B), Agilent GenetiSure Cyto arrays offer superior target binding specificity when hybridized under recommended conditions. This provides significantly lower cross-hybridization noise, reduces the number of datapoints required for signal smoothing/deconvolution, and enables effective CNV detection at comparable confidence levels with fewer probes.

For all the three competitor arrays (CytoScan HD, CytoScan 750K, and CytoSNP 850K), the total number of unique probes (approximately 2.7 million, 750,000, and 850,000, respectively) is numerically greater than that of Agilent GenetiSure Cyto arrays. However, only 10 to 14% of probes are located in the 3619 targeted regions (roughly 275,000 for CytoScan HD, 81,000 for CytoScan 750K, and 116,500 for CytoSNP 850K), resulting in 57.5%, 26.4%, and 81.4% of targets with sufficient coverage, respectively. For the two Agilent GenetiSure Cyto arrays, a higher percentage (26 to 35%) of probes are located in the targeted regions: approximately 44,000 for the 4x180k array and 20,000 for the 8x60k array. This contributes to an overall sufficient target coverage of 99.1% and 96.6%, respectively (Figure 3A). When these clinical targets are further stratified into separate categories based on the source database, the two Agilent GenetiSure Cyto arrays also manifest clear advantages in target coverage across all categories. High coverage efficiencies are consistently observed(98 to 99% for the 4x180k CGH design; 92 to 98% for the 8x60k CGH design).

By contrast, the three arrays from competitors trail behind by variable margins in these categories (57 to 80% for CytoScan HD; 24 to 61% for CytoScan 750K; and 81 to 95% for CytoSNP 850K), despite each design possessing a far greater number of total probes. This clearly demonstrates the advantage of our clinical target-oriented design strategy. The differences in target coverage are also further manifested in the size-dependent comparison. The relationship between coverage efficiency and target size is compared in Figure 3B. Although the three arrays from competitors demonstrate similarly high coverage for the largest targets (>100 kb) as the two Agilent GenetiSure Cyto arrays, only Agilent GenetiSure Cyto arrays maintain consistent and sufficiently high levels of coverage for smaller targets. The coverage efficiencies of the competitor arrays show a clear decreasing trend related to the target size. Particularly, for both CytoScan arrays from Competitor A, extremely poor performance is noted on the targets with smallest sizes (≤10 kb).

Conclusion

The *In silico* analyses we conducted on the new Agilent GenetiSure Cyto arrays and similar array offerings from competitors suggest superior performance by the Agilent arrays, thanks to their current content and optimized probe coverage across a wide range of clinically relevant gene targets with varying sizes. As a renewed tool set for cytogeneticists, the Agilent GenetiSure Cyto arrays allow flexibility in targeting exon-, gene-, or backbone-level CNVs and are expected to further strengthen the use of array comparative genomic hybridization technology in cytogenetics research and applications.

References

1. Miller, D. T.; Adam, M. P.; Aradhya, S.; Biesecker, L. G.; Brothman, A. R.; Carter, N. P.; Church, D. M.; Crolla, J. A.; Eichler, E. E.; Epstein, C. J.; Faucett, W. A.; Feuk, L.; Friedman, J. M.; Hamosh, A.; Jackson, L.; Kaminsky, E. B.; Kok, K.; Krantz, I. D.; Kuhn, R. M.; Lee, C.; Ostell, J. M.; Rosenberg, C.; Scherer, S. W.; Spinner, N. B.; Stavropoulos, D. J.; Tepperberg, J. H.; Thorland, E. C.; Vermeesch, J. R.; Waggoner, D. J.; Watson, M. S.; Martin, C. L.; Ledbetter, D. H. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *Am. J. Hum. Genet*. **2010**, *86* (5), 749–764

2. Wapner, R. J.; Martin, C. L.; Levy, B.; Ballif, B. C.; Eng, C. M.; Zachary, J. M.; Savage, M.; Platt, L. D.; Saltzman, D.; Grobman, W. A.; Klugman, S.; Scholl, T.; Simpson, J. L.; McCall, K.; Aggarwal, V. S.; Bunke, B.; Nahum, O.; Patel, A.; Lamb, A. N.; Thom, E. A.; Beaudet, A. L.; Ledbetter, D. H.; Shaffer, L. G.; Jackson, L. Chromosomal Microarray versus Karyotyping for Prenatal Diagnosis. *N. Engl. J. Med*. **2012**, *367* (23), 2175–2184.

3. Fiorentino, F.; Napoletano, S.; Caiazzo, F.; Sessa, M.; Bono, S.; Spizzichino, L.; Gordon, A.; Nuccitelli, A.; Rizzo, G.; Baldi, M. Chromosomal Microarray Analysis as a First-Line Test in Pregnancies with a Priori Low Risk for the Detection of Submicroscopic Chromosomal Abnormalities. *Eur. J. Hum. Genet*. **2013**, *21* (7), 725–730.

4. Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P. Integrative Genomics Viewer. *Nat. Biotechnol*. **2011**, *29* (1), 24–26.

5. Mason-Suares, H.; Kim, W.; Grimmett, L.; Williams, E. S.; Horner, V. L.; Kunig, D.; Goldlust, I. S.; Wu, B.-L.; Shen, Y.; Miller, D. T.; Martin, C. L.; Rudd, M. K. Density Matters: Comparison of Array Platforms for Detection of Copy-Number Variation and Copy-Neutral Abnormalities. *Genet. Med*. **2013**, *15* (9), 706–712.

6. Reiner, J.; Karger, L.; Cohen, N.; Mehta, L.; Edelmann, L.; Scott, S. A. Chromosomal Microarray Detection of Constitutional Copy Number Variation Using Saliva DNA. *J. Mol. Diagn*. **2017**, *19* (3), 397-403

7. Agilent Application Note: Use of Spike-ins for Sample Tracking in Agilent Array CGH. (5991-6661EN)

www.agilent.com

Agilent
Trusted Answers