Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies


Chromosomal microarray (CMA) is increasingly utilized for genetic testing of individuals with unexplained developmental delay/intellectual disability (DD/ID), autism spectrum disorders (ASD), or multiple congenital anomalies (MCA). Performing CMA and G-banded karyotyping on every patient substantially increases the total cost of genetic testing. The International Standard Cytogenomic Array (ISCA) Consortium held two international workshops and conducted a literature review of 33 studies, including 21,698 patients tested by CMA. We provide an evidence-based summary of clinical cytogenetic testing comparing CMA to G-banded karyotyping with respect to technical advantages and limitations, diagnostic yield for various types of chromosomal aberrations, and issues that affect test interpretation. CMA offers a much higher diagnostic yield (15%–20%) for genetic testing of individuals with unexplained DD/ID, ASD, or MCA than a G-banded karyotype (~3%, excluding Down syndrome and other recognizable chromosomal syndromes), primarily because of its higher sensitivity for submicroscopic deletions and duplications. Truly balanced rearrangements and low-level mosaicism are generally not detectable by arrays, but these are relatively infrequent causes of abnormal phenotypes in this population (<1%). Available evidence strongly supports the use of CMA in place of G-banded karyotyping as the first-tier cytogenetic diagnostic test for patients with DD/ID, ASD, or MCA. G-banded karyotype analysis should be reserved for patients with obvious chromosomal syndromes (e.g., Down syndrome), a family history of chromosomal rearrangement, or a history of multiple miscarriages.

Introduction

Scope and Purpose
Clinical genetic testing, including chromosome analysis, is a standard practice for patients with diagnoses including unexplained developmental delay/intellectual disability (DD/ID), autism spectrum disorders (ASD), and multiple congenital anomalies (MCA). These categories of disorders account for the largest proportion of cytogenetic testing because of their high prevalence in the population. The incidence of DD/ID in the general population approaches 3%, and ASD affects ~1:150 individuals. Most patients lack sufficient specific history or features from physical examination to suggest a specific genetic (or non-genetic) cause. Published guidelines for testing such patients have emphasized (1) testing for chromosomal abnormalities by G-banded karyotyping and (2) testing for common single-gene disorders, such as fragile X syndrome.

Microarray-based genomic copy-number analysis is now a commonly ordered clinical genetic test for this patient population and is offered under various names, such as “chromosomal microarray” (CMA) and “molecular...
karyotyping. CMA, as used here, encompasses all types of array-based genomic copy number analyses, including array-based comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP) arrays. G-banded karyotyping allows a cytogenticist to visualize and analyze chromosomes for chromosomal rearrangements, including genomic gains and losses. CMA performs a similar function, but at a much higher resolution for genomic imbalances. G-banded karyotyping has been the standard first-tier test for detection of genetic imbalance in this population for more than 35 years, whereas CMA is not yet standard in all clinical settings.

Clinical Interpretation of CMA Results
Although clinical genetic laboratories are familiar with recurrent copy-number changes mediated by segmental duplication architecture, population studies suggest that the vast majority of copy-number variation is not recurrent. Determining the clinical significance of variants identified by CMA can be challenging. Although CMA offers the sensitivity of high-resolution genome-wide detection of clinically significant copy-number variants (CNVs), the additional challenge of interpreting variants of uncertain clinical significance (VOUS), the preferred terminology based on a recent study of variant terminology, can impose a burden on clinicians and laboratories. Furthermore, recent efforts to evaluate reporting of CNVs among clinical laboratories indicates variability of interpretation.

Lack of uniform guidelines for the expected clinical yield of CMA, in terms of resolution and coverage (i.e., whole genome or locus specific), impedes standardization of clinical practice for CMA testing. CMA is currently performed in many different laboratories with different technology platforms and different array design and content. Uniform interpretation of results is an equally formidable challenge to standardization. Most clinical laboratories maintain internal databases of pathogenic and benign CNVs, but they might not agree on interpretations. Although College of American Pathology (CAP) proficiency testing is now available for CMA, there is not yet an openly accessible centralized resource for comparing clinical interpretations for thousands of possible variants among dozens of laboratories performing this testing. These problems can be addressed through more uniform array content, a rational approach to variant interpretation, and increased data sharing among laboratories and clinicians. Sharing of information across laboratories will provide the most benefit to laboratories, clinicians, and—ultimately and most importantly—patients.

We evaluated the benefits and limitations of CMA, as compared to G-banded karyotyping, for detecting pathogenic genomic imbalances in patients with DD/DD, ASD, and/or MCA. Recommendations for CMA use and standardization as a first-tier genetic test in this patient population are provided.

Subjects and Methods
Assembly and Focus of the ISCA Consortium

The goals of the ISCA Consortium are analogous to the goals of the Newborn Screening (NBS) Expert Group convened by the ACMG in 2006 to address the lack of uniformity in state-based NBS programs. For many years, each state made independent decisions regarding which diseases to include in NBS, leading to a wide range of diseases tested by different states. An expert panel developed an evidence-based recommendation for a uniform, core panel of 29 well-defined disorders. Testing for the vast majority of these core conditions has now been implemented or mandated (but not yet implemented) in all 50 states.

The ISCA Consortium is focused on the clinical application of CMA as opposed to laboratory technical guidelines. This report is intended as a review of available data to determine whether CMA should be adopted as a first-tier test before routine G-banded karyotyping in this patient population (defined below). We reviewed previously published clinical-practice guidelines and laboratory guidelines for clinical CMA testing. CMA is a high-complexity assay with many technical considerations beyond the scope of this discussion. Suggested guidelines about specimen requirements, DNA preparation, labeling, analysis algorithms, and assay validation are beyond the scope of this discussion but have been outlined elsewhere.

Patient Population
The analysis and recommendations in this manuscript are focused solely on CMA for constitutional abnormalities in postnatal testing of patients with DD/DD, ASD, or MCA. These analyses and conclusions are not necessarily applicable to other CMA applications, such as prenatal testing, hematological malignancies, or other forms of cancer. Although the ISCA Consortium recognizes the potential for CMA to play a role in prenatal diagnosis, current evidence is not sufficient to allow recommendations regarding prenatal CMA, and traditional cytogenetic methods, such as G-banded karyotyping and fluorescence in situ hybridization (FISH), are still the standard for prenatal diagnosis, as supported in a recent American College of Obstetrics and Gynecology (ACOG) Statement. Multicenter studies making a direct comparison of the performance of CMA to conventional cytogenetic analysis in a prenatal setting are currently underway.

Systematic Literature Review
We conducted a systematic literature review focused on clinical CMA by searching the PubMed database of the National Center for Biotechnology Information (NCBI) and using the following
controlled vocabulary MeSH terms: (Microarray Analysis) AND ((chromosomal disorders) OR (chromosomal aberration)) AND ((mental retardation) OR (developmental disabilities) OR (Autism) OR (congenital abnormalities)). We included case series or cohort studies that were posted to PubMed prior to April 15, 2009 and that included either bacterial artificial chromosome (BAC) or oligonucleotide arrays. We considered studies with (1) a clear description of the patient population, including patient selection criteria; (2) a description of the CMA platform and resolution; and (3) a description of the process for interpretation of CMA results. We also included published studies that were not necessarily identified by this combination of search terms if they met the aforementioned criteria. We excluded studies that presented only validation samples for new techniques or platforms or were focused on a particular medical condition or syndrome (e.g., only one type of congenital anomaly). We identified 33 original reports (not reviews), including 21,698 patients for expert review and discussion by ISCA Consortium workshop participants.

Classification of CMA Variants
Typical strategies for interpreting pathogenic or benign status for CNVs (Table 1) have been outlined elsewhere. In general, copy-number variants are assigned the following interpretations (1) abnormal (e.g., well-established syndromes, de novo variants, and large changes); (2) VOUS; (3) likely benign (e.g., not previously reported but inherited from a healthy parent). Diagnostic yield was defined as the number of patients with abnormal variants divided by the total number of patients tested and was derived directly from each original study. Data were not systematically collected on the number of VOUS in each study.

Development of a CNV Database through the National Institutes of Health
The ISCA Consortium has initiated a new database for CNV and phenotype data generated from clinical CMA laboratories as a project within dbGaP (Database of Genotype and Phenotype at NCBI). This database is capable of receiving and managing raw data and normalized data files from all current CMA platforms, including both copy-number and SNP arrays. The common denominator for displaying and comparing data from different labs and platforms will be the use of genome sequence coordinates, based on the UCSC Genome Browser, for defining the start and stop points for any copy-number imbalance (loss or gain). To minimize barriers to data contribution by clinical laboratories and maximize the empirical data available on CNVs from this patient population, an opt-out mechanism to notify patients that their deidentified data will be contributed to the database has been developed and approved by the central institutional review board (IRB) at the National Institutes of Health and the IRBs at individual contributing centers. All participating centers will also be strongly encouraged to obtain a full informed consent and submit qualified cases to DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources) as well.

The opt-out mechanism is a relatively new concept that allows deidentified data collected as part of clinical testing to be used for multiple reasons, including improving the accuracy and quality of test results, test validation, educational purposes, or research. Patients are notified of the opt-out mechanism through various avenues, such as test requisition forms, clinical laboratory reports, and laboratory websites. It is also the responsibility of the ordering physician or other downstream providers who review requisitions and reports containing the opt-out language to inform patients how samples are managed and data are used and that they can opt out of inclusion at any time. Because this mechanism is a growing area that will be of great use to the clinical and research communities and will ultimately improve clinical care,
Table 2. Summary of Clinical CMA Studies on 21,698 Subjects

<table>
<thead>
<tr>
<th>Author</th>
<th>Location</th>
<th>Patients</th>
<th>Phenotype</th>
<th>Controls</th>
<th>Array Type</th>
<th>Resolution</th>
<th>Diagnostic Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vissers et al. (2003)</td>
<td>Netherlands and U.S.</td>
<td>20</td>
<td>ID, DF</td>
<td>4 (2M, 2F)</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>10.0%</td>
</tr>
<tr>
<td>Shaw-Smith et al. (2004)</td>
<td>U.K., France</td>
<td>50</td>
<td>ID (moderate to severe)</td>
<td>Pooled DNA (20M, 20F)</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>14.0%</td>
</tr>
<tr>
<td>de Vries et al. (2005)</td>
<td>Netherlands</td>
<td>100</td>
<td>ID</td>
<td>72 parents of probands</td>
<td>BAC</td>
<td>Whole genome (50 kb)</td>
<td>10.0%</td>
</tr>
<tr>
<td>Schoumans et al. (2005)</td>
<td>Sweden</td>
<td>41</td>
<td>ID (mild to severe)</td>
<td>Pooled DNA (10 subjects)</td>
<td>BAC</td>
<td>Whole genome (1.3 Mb)</td>
<td>9.8%</td>
</tr>
<tr>
<td>Tyson et al. (2005)</td>
<td>Canada</td>
<td>22</td>
<td>ID, DF</td>
<td>Pooled normal of four to six males or females (Promega)</td>
<td>BAC</td>
<td>Whole genome (1.3 Mb)</td>
<td>13.7%</td>
</tr>
<tr>
<td>Wong et al. (2005)</td>
<td>U.S.</td>
<td>102</td>
<td>ID</td>
<td>Normal controls</td>
<td>BAC</td>
<td>Telomere</td>
<td>18.6%</td>
</tr>
<tr>
<td>Ballif et al. (2006)</td>
<td>U.S.</td>
<td>3600</td>
<td>ID, DD</td>
<td></td>
<td>BAC</td>
<td>Targeted</td>
<td>5.1%</td>
</tr>
<tr>
<td>Friedman et al. (2006)</td>
<td>Canada</td>
<td>100</td>
<td>ID</td>
<td>8 unaffected siblings</td>
<td>SNP Oligo</td>
<td>Whole genome (30 kb)</td>
<td>11.0%</td>
</tr>
<tr>
<td>Krepski-Santos et al. (2006)</td>
<td>Brazil</td>
<td>95</td>
<td>ID, DF, MCA</td>
<td>100 control observations for each chromosome pair</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>16.8%</td>
</tr>
<tr>
<td>Menten et al. (2006)</td>
<td>Belgium</td>
<td>140</td>
<td>ID, MCA</td>
<td>Other patient samples in cohort</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>13.6%</td>
</tr>
<tr>
<td>Ming et al. (2006)</td>
<td>U.S.</td>
<td>10</td>
<td>MCA</td>
<td>128 (42 Caucasian, 42 African-American, 20 Asian, 20 NIGMS)</td>
<td>SNP Oligo</td>
<td>Whole genome (30 kb)</td>
<td>20.0%</td>
</tr>
<tr>
<td>Miyake et al. (2006)</td>
<td>Japan</td>
<td>30</td>
<td>ID, MCA</td>
<td>2 negative (1M, 1F), 1 positive</td>
<td>BAC</td>
<td>Targeted (1.4 Mb)</td>
<td>16.7%</td>
</tr>
<tr>
<td>Rosenberg et al. (2006)</td>
<td>Netherlands, Brazil, U.K.</td>
<td>81</td>
<td>ID (Mild to severe), DF</td>
<td>100 control observations for each chromosome pair</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>16.0%</td>
</tr>
<tr>
<td>Sharp et al. (2006)</td>
<td>U.K., U.S.</td>
<td>290</td>
<td>ID +/- DF; +/- MCA</td>
<td>316 controls from various ethnicities</td>
<td>BAC</td>
<td>Targeted</td>
<td>5.5%</td>
</tr>
<tr>
<td>Shaffer et al. (2006)</td>
<td>U.S.</td>
<td>1500</td>
<td>Various indications</td>
<td>Opposite gender control</td>
<td>BAC</td>
<td>Targeted</td>
<td>5.6%</td>
</tr>
<tr>
<td>Aradhya et al. I(2007)</td>
<td>U.S.</td>
<td>20 (12M, 8F)</td>
<td>ID or DD, +/- DF; +/- MCA, +/- GR</td>
<td>Normal male or female (Promega)</td>
<td>Oligo</td>
<td>Whole genome (70 kb)</td>
<td>35.0%</td>
</tr>
<tr>
<td>Batis et al. (2007)</td>
<td>U.S.</td>
<td>234</td>
<td>ID or DD, DF, MCA</td>
<td>50 normal controls (25M, 25F) and 36 patients with abnormal chromosome testing</td>
<td>BAC</td>
<td>Targeted</td>
<td>5.6%</td>
</tr>
<tr>
<td>Engels et al. (2007)</td>
<td>Germany</td>
<td>60</td>
<td>ID</td>
<td>Pooled DNA (10M, 10F)</td>
<td>BAC</td>
<td>Whole genome and Targeted (0.5 Mb)</td>
<td>10.0%</td>
</tr>
<tr>
<td>Fan et al. (2007)</td>
<td>U.S.</td>
<td>100</td>
<td>ID or DD</td>
<td>7M, 7F gender matched</td>
<td>Oligo</td>
<td>Whole genome (35 kb)</td>
<td>15.0%</td>
</tr>
<tr>
<td>Hoyer et al. (2007)</td>
<td>Germany</td>
<td>104</td>
<td>ID</td>
<td>Not specified</td>
<td>SNP Oligo</td>
<td>Whole genome (30 kb)</td>
<td>9.6%</td>
</tr>
<tr>
<td>Lu et al. (2007)</td>
<td>U.S.</td>
<td>2513</td>
<td>ID or DD, DF, MCA, ASD</td>
<td>Normal (1M, 1F)</td>
<td>BAC</td>
<td>Targeted</td>
<td>7.0%</td>
</tr>
<tr>
<td>Newman et al. (2007)</td>
<td>U.K.</td>
<td>36</td>
<td>DD, LD, DF</td>
<td>Not specified</td>
<td>BAC</td>
<td>Whole genome (1.0 Mb)</td>
<td>13.8%</td>
</tr>
<tr>
<td>Shaffer et al. (2007)</td>
<td>U.S. and “abroad”</td>
<td>8789</td>
<td>ID or DD, MCA</td>
<td></td>
<td>BAC</td>
<td>Targeted</td>
<td>6.9%</td>
</tr>
</tbody>
</table>
educational materials are also being developed for patients and healthcare providers. Our current NIH-funded grant supports an education and ethics working group comprised of experienced genetic counselors and clinicians who will continually address these important issues.

Working groups have been established to determine standard data formats and vocabularies for genotypic and phenotypic data, as well as to make recommendations for standard interpretation guidelines for CMA results. An expert curation process is being established so that discrepancies within the data can be identified and resolved, and public data releases are planned on a quarterly basis. These will be available to major genomics resources (e.g., UCSC Genome Browser, ENSEMBL, DECIPHER, Database of Genomic Variants [DGV], etc.) via the Database of Structural Variants (dbVar), but also to individual laboratories and commercial vendors for CMA software and database software products.

Results

Diagnostic Yield of CMA Is Greater than that of G-Banded Karyotyping
We reviewed 33 studies, together including 21,698 patients tested by CMA (Table 2). CMA detected pathogenic genomic imbalances with an average diagnostic yield of 12.2% across all studies in this patient population, about 10% more than G-banded karyotyping alone. We also reviewed studies related to potential limitations of CMA as a clinical test, especially with regard to detection of VOUS, balanced translocations, and low-level mosaicism. Our results also include an overview of the CNV Database developed through the ISCA Consortium.

Diagnostic yield has improved as the resolution of cytogenetic testing for patients with developmental disabilities has evolved. G-banded karyotyping can sometimes detect genomic imbalances as small as 3 Mb, but it can often miss genomic imbalances in the 5–10 Mb range, depending on the genomic region involved and/or conditions of the assay (Figure 1). If patients with Down syndrome (MIM 190685) are excluded, G-banded karyotyping of patients with ID has typically identified abnormalities in fewer than 3% of individuals.1,53–56 Karyotyping is also limited because it is based on the subjective assessment of gains and losses and is prone to considerable interpersonal and interlaboratory variation in detection rates.

The addition of subtelomeric fluorescence in situ hybridization (FISH) for identifying submicroscopic deletions and duplications of genomic regions proximal to the telomeres has been shown to almost double the diagnostic yield. In the largest analysis of subtelomeric FISH testing, presumed pathogenic changes were found in 2.6% of 11,688 unselected cases,57 and a recent review of another 7,000 cases found subtelomeric rearrangements in 2.4% of the patients.

Table 2. Continued

<table>
<thead>
<tr>
<th>Author</th>
<th>Location</th>
<th>Patients</th>
<th>Phenotype</th>
<th>Controls</th>
<th>Array Type</th>
<th>Resolution</th>
<th>Diagnostic Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shen et al. (2007)</td>
<td>U.S.</td>
<td>211</td>
<td>ID or DD, MCA</td>
<td>Normal (1M, 1F)</td>
<td>Oligo (Agilent)</td>
<td>Targeted (35 kb in target regions)</td>
<td>7.6%</td>
</tr>
<tr>
<td>Thuresson et al. (2007)</td>
<td>Sweden</td>
<td>48</td>
<td>ID, DF, MCA</td>
<td>8M, 8F gender matched</td>
<td>BAC</td>
<td>1 Whole genome (1.0 Mb)</td>
<td>6.3%</td>
</tr>
<tr>
<td>Wagenstaller et al. (2007)</td>
<td>Germany</td>
<td>67</td>
<td>ID (mild to severe)</td>
<td>44 unaffected parents and 4 children with known translocations</td>
<td>SNP Oligo (Affy 100k)</td>
<td>Whole genome (30 kb)</td>
<td>16.4%</td>
</tr>
<tr>
<td>Aston et al. (2008)</td>
<td>U.S.</td>
<td>1075</td>
<td>ID, DD, DF, MCA</td>
<td>Normal control of the opposite gender (Promega)</td>
<td>BAC</td>
<td>Whole genome (1.3 Mb)</td>
<td>5.3%</td>
</tr>
<tr>
<td>Baldwin et al. (2008)</td>
<td>U.S.</td>
<td>211</td>
<td>ID, DD, DF, MCA, ASD</td>
<td>Normal control of the opposite gender (Promega)</td>
<td>Oligo (custom Agilent 44k)</td>
<td>Whole genome and Targeted (75 kb)</td>
<td>15.6%</td>
</tr>
<tr>
<td>Pickering et al. (2008)</td>
<td>U.S.</td>
<td>1176</td>
<td>ID or DD</td>
<td>Normal (1M, 1F)</td>
<td>BAC</td>
<td>Targeted (n = 822) and whole genome (1 Mb; n = 354)</td>
<td>7.8%</td>
</tr>
<tr>
<td>Shevell et al. (2008)</td>
<td>Canada</td>
<td>94</td>
<td>ID, DD</td>
<td>Not specified</td>
<td>BAC</td>
<td>Targeted</td>
<td>6.4%</td>
</tr>
<tr>
<td>Xiang et al. (2008)</td>
<td>U.S.</td>
<td>50</td>
<td>ID, DD</td>
<td>Gender mismatched</td>
<td>Oligo (Agilent 44k)</td>
<td>Whole genome (35 kb)</td>
<td>18.0%</td>
</tr>
<tr>
<td>Nowakowska et al. (2008)</td>
<td>U.S.</td>
<td>91</td>
<td>ID, DF</td>
<td>Normal (1M, 1F)</td>
<td>BAC</td>
<td>Targeted</td>
<td>11.8%</td>
</tr>
<tr>
<td>Lu et al. (2008)</td>
<td>U.S.</td>
<td>638</td>
<td>MCA</td>
<td>Normal (1M, 1F)</td>
<td>BAC (n = 372) and Oligo (n = 266)</td>
<td>Targeted with backbone</td>
<td>17.1%</td>
</tr>
</tbody>
</table>

ASD = autism spectrum disorders; BAC = bacterial artificial chromosome; DD = developmental delay; DF = dysmorphic features; ID = intellectual disability; LD = learning disabilities; MCA = multiple congenital anomalies. Where an array is indicated as “targeted,” the resolution of regions outside the target area (e.g., backbone coverage) is not specified.
Most FISH assays used in a clinical cytogenetic setting detect submicroscopic changes in the range of 100 kb or larger. BAC-based arrays can detect genomic imbalances 100 kb and larger, depending on probe coverage. The increased yield of clinical genetic diagnoses with CMA parallels the evolution of array designs (Figure 2). Early clinical CMA tests included BAC clones and targeted subtelomeric and pericentromeric regions, “known” recurrent microdeletion and microduplication syndromes, or whole-genome coverage with clones at ~1 Mb genomic intervals (Figure 2), and diagnostic yields were 7%–11% for children with normal G-banded cytogenetic analysis. Subsequent versions of clinical CMA tests included additional probes, so-called genomic “backbone” coverage, allowing for the detection of pathogenic genomic imbalances outside of the well-described critical regions for microdeletion and microduplication syndromes. Rapid increases in the availability of genome-wide arrays with density sufficient to allow detection of copy-number changes of ~100 kb throughout the genome increased diagnostic yields to between 11% and 15% based on studies of DD/ID patients who had normal G-banded cytogenetic analysis; these yields are far higher than what was obtained via traditional cytogenetic methods.

CMA Platform Comparison
Most commercial and academic laboratories have now converted to oligonucleotide- or SNP-based arrays. Some clinical laboratories use oligonucleotide-based arrays that emulate the coverage of earlier versions of BAC-based arrays. Advantages of oligonucleotide-based arrays include more flexibility in terms of probe selection, thus facilitating higher probe density and customization of array content. Arrays designed specifically for copy-number analysis use longer oligonucleotide probes (~60-mer), which have been shown to provide better signal-to-noise ratios (SNRs) than the shorter oligonucleotide probes (~22-mer) used in some SNP-detecting array platforms. Longer oligonucleotides (50-mer) on some SNP-detecting arrays result in a better signal-to-noise ratio than short oligonucleotide SNP arrays. Also, SNP-detecting arrays now typically include a mixture of SNP and copy-number probes to address this issue.

In general, high specificity can be achieved by current BAC or oligonucleotide arrays, including SNP arrays, depending on coverage. BAC-based arrays employ large-insert genomic clones of 100–150 kb, and accurate copy-number determination can be made for individual clones. Single oligonucleotide probes do not provide accurate determination of copy number, and multiple consecutive probes

---

**Figure 1. Examples of Genomic Imbalances Detected by a CMA but Not by G-Banded Karyotyping**

(A) A 10.9 Mb deletion, including more than 60 genes. The deletion includes the Williams-Beuren syndrome region at chromosome region 7q11 but extends beyond the typical breakpoints for this syndrome. The arrow is pointing to the deleted chromosome that was observed by retrospective analysis of G-banded slides.

(B) A 7.2 Mb duplication on the long arm of chromosome 11. Again, the arrow is pointing to the chromosome that has the duplication shown by the darker G-positive band.

---

754 The American Journal of Human Genetics 86, 749–764, May 14, 2010
indicating the same copy-number change are required for accurately determining a gain or loss. The number of consecutive probes required may vary between arrays of long-oligonucleotide probes and SNP probes, and SNP arrays typically require more consecutive probes to achieve comparable specificity.

SNP arrays have the advantage of also detecting long stretches of homozygosity, which might represent uniparental disomy (UPD) or consanguinity not suspected on the basis of clinical history. However, routine CMA analysis with SNP-based platforms will detect heterodisomic UPD only in that portion of cases where there are blocks of isodisomy. UPD occurring by trisomy rescue usually contains blocks of heterodisomy and isodisomy but can result in complete heterodisomy.

Arrays based entirely on probes for SNPs are biased to certain genomic segments containing common SNPs. One earlier SNP-detecting array could detect only ~26% of the CNVs that were detected by fosmid end sequence mapping strategies. For this reason, newer SNP-detecting arrays also contain nonpolymorphic oligonucleotide probes exclusively designed for copy-number detection to provide more robust and uniform coverage. Overall, with sufficiently dense probe coverage, all current array platforms are able to provide sufficient sensitivity for clinical CMA testing.

**Recommendations for CMA Coverage and Resolution**

The clinically effective resolution of CMA represents a balance between sensitivity and specificity. Analytical sensitivity is primarily influenced by probe coverage, resolution, and genomic spacing of probes selected for the array. For CMA to identify genomic imbalances at a higher resolution than a conventional G-banded karyotype, the array must consistently detect clinically significant genomic imbalances smaller than 5 Mb. Clinical sensitivity of CMA should be considered in relation to G-banded karyotyping and not in comparison to gene-specific molecular-based assays for Mendelian disorders.

Available oligonucleotide platforms can now detect genomic imbalances as small as 500 bp, permitting the detection of genomic copy-number changes as small as 10–20 kb in many regions of the genome. Clinical arrays are typically designed to detect imbalances of 20–50 kb in targeted regions (e.g., within known Mendelian genes) and imbalances of 100–250 kb in nontargeted (backbone) regions of the genome. The ability to identify smaller deletions allows diagnosis of diseases caused by as few as one haploinsufficient gene, which increases the diagnostic capability of this testing. However, this level of resolution is not necessary as a genome-wide clinical test of genomic imbalance.

Most current clinical CMA platforms can detect copy-number changes with a lower limit of resolution of ~400 kb throughout the genome, representing a ≥10-fold improvement in resolution in comparison to G-banded karyotyping. This level of resolution will provide a broad genomic survey and reliably identify all known recurrent microdeletion and microduplication syndromes mediated by segmental duplication architecture and most nonrecurrent pathogenic imbalances that are unequivocally pathogenic. The smallest known recurrent microdeletion syndromes, such as deletions at 17q21.31 (MIM 610443), and 16p11.2 (MIM 611913), are ~500 kb or larger.

Although copy-number alterations smaller than ~400 kb may certainly be pathogenic and smaller disease-causing CNVs are likely to be identified on an ongoing basis, higher-resolution CMA detects an increasing proportion of benign CNVs. Decisions about clinically interpretable, effective resolution for CMA should therefore also consider empiric data about size distribution of CNVs in the human genome. Research-based tiling arrays used on HapMap samples (considered healthy individuals) revealed a median size for presumably benign CNVs of ~200 kb and showed that 90%–95% of these CNVs are less than 500 kb.

More recent data with higher-density arrays suggest that many more “small” CNVs exist, that there is an overall median CNV length of ~2.9 kb, and that 95% of CNVs are less than 100 kb. These smaller CNVs were previously undetectable because of technological limitations, and CNV sizes from previous studies have often been overestimated for the majority of CNVs. These recent studies demonstrate that only 1%–2% of all CNVs in normal individuals are larger than 1 Mb. A summary of recommendations appears in Table 3 and in the Supplemental Data.

**Impact of Balanced Rearrangements and Low-level Mosaicism on Clinical Sensitivity of CMA**

G-banded karyotyping is still important in the detection of balanced rearrangements and low-level mosaicism, both of which are not uniformly detectable by CMA. The clinical sensitivity of CMA depends on the proportion of...
potentially pathogenic balanced rearrangements, and mosaic chromosomal material, in this patient population. The available evidence suggests that (1) truly balanced rearrangements represent only a small proportion of clinically significant genomic events in this patient population and (2) many “apparently balanced” rearrangements detected by G-banding are not balanced at the DNA level.

Balanced rearrangements account for a minority of cytogenetically detectable events in patients with ID. G-banded karyotyping of large cohorts of individuals with ID shows apparently balanced structural abnormalities in 1/500 patients, and in 1/2,000 patients, these occur de novo.55 Balanced rearrangements make up only about 10% of cytogenetically visible abnormalities in patients with ID, meaning that only about 0.3% of patients with ID who are tested by karyotype analysis would have such changes.1,54,76 In the general population, balanced rearrangements are also quite common. A study of 377,357 amniocenteses showed a de novo reciprocal translocation in approximately 1/2,000 and a balanced Robertsonian translocation in 1/9,000.27

Moreover, although cytogenetic events might appear balanced at the microscopic level of resolution, many have a submicroscopic imbalance, especially among individuals with an abnormal phenotype.78,79 In one study of ten patients with abnormal phenotypes and chromosomal rearrangements that appeared balanced by conventional cytogenetic methods, six had submicroscopic imbalance according to CMA with 1 Mb resolution.78 In a larger study of 59 patients with apparently balanced rearrangements (41 de novo reciprocal translocations and 18 de novo complex chromosomal rearrangements), 27 patients were found to have an abnormal phenotype attributed to the chromosomal rearrangement itself. Analysis of these 27 patients with higher-resolution oligonucleotide arrays (Agilent 44k and 244k) revealed a submicroscopic imbalance in 11/27 patients, or 41%.79 A more recent study of 14 patients with an apparently balanced chromosomal rearrangement revealed a submicroscopic imbalance in 4/14 patients, or 29%.80

Another approach to this question is to consider the impact of balanced translocations on phenotypically normal individuals. Most apparently balanced reciprocal translocations in phenotypically normal individuals do not contain genomic imbalance.81 Among balanced rearrangements that interrupt a gene, approximately half (16/31) are found in healthy individuals.80 In the setting of multiple miscarriages, a balanced translocation in one of the parents could be the explanation for unbalanced offspring, and G-banded karyotyping should still be the standard of care for this indication. In some cases, family history can be a clue to segregation of a balanced...
translocation. In clinical settings, probands are more likely to carry an imbalance, and parents and other family members would subsequently be tested by traditional cytogenetic methods.

The incidence of low-level mosaicism is relatively low compared to the incidence of other detectable chromosomal imbalances in this patient population. Conlin et al. (2010) observed mosaic aneuploidy in 1% of 2,019 patients with a variety of diagnoses (the most common diagnosis was DD), and G-banded karyotyping of a standard 20 cells is estimated to detect such mosaicism at a level of 14%. Detection of mosaicism as low as 10% is possible with BAC-based arrays, but the trend among laboratories is for increasing use of oligonucleotide arrays. Although similar validation studies are needed for long-oligo-based arrays, clinical experience indicates that mosaicism for segmental aneuploidy is detectable to a level of about 20%–30% mosaic cells with most current platforms. SNP arrays can detect lower levels of mosaicism (below 5% in some cases) because of the increased resolution afforded by genotyping in conjunction with probe intensity. Many types of mosaicism are missed by karyotyping either because they are not present in the relevant lymphocytes or because the cells with the imbalance are unable to respond to the mitogen. Regardless of the method, ease of detection will be proportional to the size of mosaic chromosome fragments. At this time, CMA performs comparably to G-banded karyotyping for detection of mosaicism.

Development of an Open-Access CMA Database

Difficulties interpreting VOUS can be overcome with the cooperation of researchers, clinicians, clinical laboratory directors, and bioinformaticians.

The most comprehensive existing online database for presumed benign CNVs is DGV, but this valuable resource is based on testing of presumably healthy individuals and does not provide clinical information. A recently launched database of structural variation from both normal control populations and disease populations is dbVAR, developed and housed at NCBI within the NIH. Correlation of clinical phenotypes with specific genomic rearrangements requires databases that include data from individuals with DD/ID or MCA and is exemplified by the interactive web-based database DECIPHER, which incorporates a suite of tools designed to aid in the interpretation of submicroscopic chromosomal imbalances, inversions, and translocations. DECIPHER enables free public access to fully consented molecular and phenotypic data as well as password-protected protocols for individual labs to securely upload and manage patient information.

With more than 100 clinical CMA laboratories currently participating in the ISCA Consortium, it can be anticipated that this database will exceed several hundred thousand CMA cases within the next 2–3 years. Data at the individual level will be stored in a controlled-access environment in dbGaP, and curated summary data, including CNV region calls and phenotypic associations, will be available and visualized through various resources, including public databases and individual vendor software. This large dataset, in combination with DECIPHER, DGV, and dbVAR, should substantially reduce the current frequency of VOUS results. However, CNVs with incomplete penetrance and variable expression will remain significant challenges, particularly in the prenatal setting, unless the molecular basis of the phenotypic variation can be identified.

Discussion

Clinical geneticists, pediatric neurologists, and developmental pediatricians are increasingly ordering CMA to obtain a genetic diagnosis for their patients with unexplained DD/ID, ASD, and MCA. A specific genetic diagnosis facilitates comprehensive medical care and accurate recurrence risk counseling for the family. Similar to our results, a recent meta-analysis of CMA on 13,926 subjects with ID and/or MCA, most of whom had normal conventional cytogenetic studies, reported an overall diagnostic rate of 10% for pathogenic genomic imbalances. Another retrospective analysis of 36,325 patients with DD estimated that a pathogenic abnormality could be detected in ~19% of unselected DD/ID patients via genome-wide array-based assays with a 30–70 kb median probe spacing.

Our recommendation based on current evidence is to offer CMA as the first-tier genetic test, in place of G-banded karyotype, for patients with unexplained DD/ID, ASD, or MCA. Although others have proposed that CMA should be considered a first-tier test, this point of view is still not widely accepted. Except in special cases, such as those involving family history of multiple miscarriages, a karyotype is not cost effective in a child with DD/ID, ASD, or MCA and a negative array study. CMA testing is not inexpensive, but the cost is less than the cost of a G-banded karyotype plus a customized FISH test such as subtelomeric FISH, and the yield is greater.

G-banding has been available for more than 38 years and has the advantage of being a widely accepted and uniform technique with an international system of cytogenetic nomenclature (ISCN). By contrast, CMA is new and more diverse in terms of techniques used, coverage, and approach to data interpretation. Overcoming this lack of uniformity will allow clinicians to leverage the increased sensitivity and specificity of CMA with genome-wide coverage to detect genomic imbalances at higher resolution and with less subjectivity in clinical settings. In addition, the more precise delineation of deletions and duplications provides direct links to the genome content in these areas, allowing recognition of disease-causing genes.

CMA offers additional advantages beyond the ability to detect submicroscopic genomic imbalances. Although G-banding is better at detecting a small marker chromosome that contains exclusively pericentromeric repeat sequences, the clinical significance of such an event is negligible, and CMA is better than traditional cytogenetic
techniques for identifying the composition of small marker chromosomes when they contain sufficient euchromatic material. CMA is also superior to FISH for detecting submicroscopic duplications because of its higher resolution (multiple small oligonucleotide probes can recapitulate the coverage of a single BAC probe) and because of the technical difficulty of visualizing tandem duplications by metaphase FISH analysis. Clinically significant submicroscopic duplications, including the reciprocal duplication of known microdeletion syndromes such as the 7q11 Williams-Beuren syndrome region or the 17p11.2 Potocki-Lupski syndrome region, are more easily identified by CMA.

The ISCA Consortium proposes the clinical algorithm in Figure 3 to guide postnatal testing in this patient population. For clinical testing, traditional cytogenetic methods, such as FISH, might offer the best confirmation for certain abnormal findings. For example, terminal deletions or duplications are more likely than interstitial events to be involved in a rearrangement, especially when more than one deletion or duplication is identified in a single individual. Some labs might use other methods, such as quantitative PCR (qPCR) and multiplex ligation and probe amplification (MLPA). The need for confirmatory testing purely for copy-number determination is debatable in cases such as those involving very large deletions or duplications (typically involving dozens of consecutive probes).

In general, traditional cytogenetic methods are still needed for single-cell analysis. Other circumstances in which traditional cytogenetic methods are indicated include cases in which traditional cytogenetic methods are indicated instead of (or at least before) CMA include when the patient has a recognizable chromosomal syndrome such as trisomy 21, trisomy 13, Turner syndrome, or Klinefelter syndrome. For these circumstances, conventional cytogenetic analysis or interphase FISH analysis might provide a more rapid turn-around time, allow more sensitive detection of low-level mosaicism, and provide information regarding
position to distinguish free trisomy from translocation-associated trisomy. For couples with a history of recurrent miscarriage or other significant family history of reproductive loss, conventional chromosome analysis to identify potential balanced translocations would still be the most appropriate first-tier test.

Some might consider FISH follow-up testing to be important for detection of possible insertional translocations (IT). These events are rare at the karyotypic level; they occur in approximately 1:80,000 live births. IT might be more common at the submicroscopic level, but specific incidence rates have not been published. The potential implications of IT for genetic counseling about recurrence risk are one argument in favor of FISH testing after identification of a copy-number gain by CMA. Other evidence suggests that parental testing is best done by CMA because breakpoints may be unstable on transmission. However, on the basis of limited published studies and clinical experience, most CNVs have similar, if not identical, breakpoints between parent and child.

CMA testing of this patient population must provide whole-genome coverage in order to replace a karyotype, as outlined above. Clinical labs might increasingly utilize custom arrays to detect intragenic deletions and duplications in individual genes associated with Mendelian disorders, and we agree that first-tier clinical testing for most Mendelian disorders should be based on sequencing and/or platforms, such as custom arrays, with high resolution for small intragenic deletions. For CMA with resolution of ~400 kb, as proposed here, information about “targeted” Mendelian genes should clearly state that CMA is intended to detect large deletions or duplications that include part, or all, of the targeted gene but that it is not intended to replace complete gene sequencing or high-resolution array.

We acknowledge that concerns about identification of VOUS might also cause hesitation about adoption of CMA as a first-tier test, but we argue that this issue can be adequately addressed through (1) choosing a level of resolution that balances sensitivity and specificity; (2) increased data sharing through the established database; and (3) parental studies to determine whether CNVs are de novo or inherited (Figure 3). The fact that no single CMA platform has been found to be clearly superior to all of the others for clinical purposes and the absence of published clinical standards for coverage and resolution have led to a lack of uniformity in arrays offered in different laboratories. The variety of available tests with nonoverlapping lists of covered genomic regions is a potential barrier to the widespread adoption of CMA as the first-tier cytogenetic test, but our recommendations (summarized in Table 3 and the Supplemental Data) address this issue.

Even for arrays that contain comparable genomic content, accurate interpretation of test results requires increased sharing of primary data among clinical laboratories so that benign copy-number variants can be distinguished from pathogenic ones. New recommendations for CMA nomenclature outlined in ISCN 2009 should facilitate interlaboratory comparisons of copy-number changes, but they do not address the interpretation of clinical significance. Clinical utility will be improved through a standard approach to interpretation of copy-number variation applicable to all technology platforms, and especially through efforts such as the widely accessible ISCA database of pathogenic and benign copy-number changes for comparison of clinical interpretations.

Performing large numbers of parental samples is costly by any method (CMA, FISH, MLPA, or qPCR), but the need for parental testing will diminish with accumulating data about benign CNVs. Population studies suggest that >99% of all benign CNVs are inherited, and the vast majority of inherited CNVs are much smaller than 500 kb. Most pathogenic copy-number alterations are larger than 1 Mb, and most occur de novo. Inherited CNVs, such as those seen in 1q21.1 (MIM 612474 and 612475), 1q41q42, 3q29 (MIM 609425 and 611936), 15q11.2, 15q13.2q13.3 (MIM 612001), 16p11.2 (MIM 611913), 16p13.11 and 22q11.2 (MIM 188400 and 608363), can also be pathogenic, but they have incomplete penetrance and variable expressivity. Variable clinical implications also result from mechanisms of inheritance that influence the expression of a trait, for example when an inherited deletion uncovers an imprinted region or a pathogenic recessive allele on a homologous chromosome.

Most genomic copy-number changes associated with DD/DD are sporadic, but others can be inherited with a recurrence risk as high as 50%. More highly penetrant syndromes may be more easily recognized through history and clinical exam, but such syndromes are more likely to occur because of de novo changes that confer minimal recurrence risk. Thus, syndromes exhibiting nonpene trance in some individuals can be more challenging to diagnose, and the cost to the family for failure to diagnose might be higher because there might be more inherited variants for these types of syndromes and thus a significant recurrence risk. In general, VOUS within the resolution of CMA should be reported so that they can be interpreted clinically as the field advances.

An extensive discussion of ethical and legal issues associated with advances in CMA technology is beyond the scope of the current manuscript. Brief examples include the ability of SNP arrays to identify unsuspected consanguinity that might point to an autosomal-recessive disease. In some cases, the degree of consanguinity revealed may be indicative of parent-child incest, and there are potential ethical and legal implications of reporting of such results. Whole-genome analysis also raises important questions regarding intellectual-property issues around gene patents, leading some institutions to instruct their CMA laboratories to omit or mask coverage of such patented genes, to the detriment of patient care. Current legal challenges to gene patenting in the United States might alleviate this particular concern.

As a community of clinical and laboratory geneticists, we have an opportunity to increase the uniformity of CMA and
its clinical interpretation in laboratories around the world. Clinical geneticists have always worked closely with genetics laboratories to interpret results in the context of the patient’s clinical findings. In this regard, clinical geneticists must have or develop a basic understanding of CMA, including indications for testing, interpretation of results, and counseling of families. As more medical specialists utilize this technology, clinical and laboratory geneticists will play an important role in helping these specialists interpret the tests and communicate results to the families. Educational tools should be developed and utilized in the setting of ongoing continuing medical education for individuals in practice as well as for those in laboratory and clinical training programs.

Supplemental Data

Supplemental Data include general recommendations of the ISCA Consortium.

Acknowledgments

Workshops and conference calls related to the ISCA Consortium were funded through a grant from the ACMG Foundation and Luminex to D.H.L. and C.L.M., and ongoing standards and database development are funded in part by grants from the National Institutes of Health (RC2HD064525 and MH074090 to D.H.L. and C.L.M.). This work was also supported by Simons Foundation grant 137578 (to E.E.E.). E.E.E. is an investigator of the Howard Hughes Medical Institute. The opinions expressed here are those of the authors and the ISCA Consortium and do not necessarily reflect the views of the institutions to which the authors are affiliated, the ACMG, or the American Society of Human Genetics (ASHG). The authors would like to thank Dr. Omer Gokcumen for his assistance with summary of population CNV data. Regarding conflict of interest, the majority of authors are involved in CMA testing and/or cytogenetic testing, either as clinicians who order the testing, laboratory professionals who perform the testing, or researchers in the genomics field. S.A. is a director of clinical cytogenetics at GeneDx, a subsidiary of BioReference Laboratories; A.R.B. is a director of clinical cytogenetics and molecular cytogenetics at ARUP Laboratories; E.E.E. is a member of the Pacific Biosciences Scientific Advisory Board. S.W.S. is on the Scientific Advisory Board of Combimatrix Diagnostics and the Scientific Advisory Committee of Autism Speaks. J.T. is a director of clinical cytogenetics at Laboratory Corporation of America. J.V. is on the board of directors of Cartagenia. M.W. raises funds from industry to support the educational activities of the ACMG Foundation.

Received: February 15, 2010
Revised: April 12, 2010
Accepted: April 19, 2010
Published online: May 13, 2010

Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants (DGV), http://projects.tcag.ca/variation/
Database of Structural Variation, dbVAR ;http://www.ncbi.nlm.nih.gov/dbvar/
International Standard Cytogenomic Array Consortium, https://isca.genetics.emory.edu
UCSC Genome Bioinformatics Site, http://genome.ucsc.edu/

References


