Counting Chromosomes to Exons: Advances in Copy Number Detection

Swaroop Aradhya · Athena M. Cherry ·
Santhosh Girirajan

Published online: 12 March 2013
© Springer Science + Business Media New York 2013

Abstract Abnormal copy number of chromosomes, genes, or individual exons can have deleterious effects that lead to recognizable genetic disorders. Until recently, the traditional methods of karyotyping, fluorescence in situ hybridization, and rudimentary PCR-based assays were the only choices available to detect copy number abnormalities. The advent of chromosomal microarrays and next-generation sequencing has now dramatically improved our ability to detect deletions or duplications with superior resolution compared to that possible with previous methods. Each method has inherent properties and variations that provide advantages in detecting mutations in specific genomic environments, but those same properties can be limiting in other regions as well as in scalability. Improvements in bioinformatics algorithms predict a complete shift toward next-generation sequencing for detecting the entire range of copy number abnormalities, although microarray and other technologies will continue to be useful as confirmatory tests and for investigating complex structural rearrangements and non-unique sequences.

Keywords Copy number · Methods · Array · Sequencing · Probes

Introduction

Copy number variation in the human genome can be associated with severe clinical disorders, or can represent a benign polymorphism [1•–3]. The methods to detect this variation have evolved from simple single-locus interrogations, initially, to genome-wide high-resolution surveys in recent years (Table 1). Aneuploidy was perhaps the most commonly recognized form of copy number abnormality in the human genome with a clinical outcome, but studies in the past decade have made it evident that >1 kb submicroscopic deletions and duplications are frequent in our genome and now qualify as the most common form of copy number variation in humans [4]. The extent of copy number variation in the 100 bp–1 kb range remains poorly understood due to the absence of robust technologies to detect this variation on a whole genome scale. Larger copy number variants (CNVs) were initially characterized by fluorescence in situ hybridization and other methods, but detection of these CNVs genome-wide and en masse only occurred with the recent advent of high-resolution DNA microarrays. Similarly, deletions and duplications of small sequences within genes (e.g., exons) were initially investigated by Southern blot or quantitative PCR, but more sophisticated methods such as multiplex ligation-dependent amplification (MLPA) and exon-focused microarrays have brought significantly higher resolution and flexibility. We now recognize the entire gamut of copy number variation, from whole chromosomes to complete genes to individual exons, and each category is associated with a plethora of genetic disorders. Most methods for copy number detection

S. Aradhya (✉)
GeneDx, 207 Perry Parkway, Gaithersburg, MD 20877, USA
e-mail: swaroop@genedx.com

A. M. Cherry
Clinical Cytogenetics, Stanford University, 3375 Hillview
Avenue, Palo Alto, CA 94304, USA

S. Girirajan
Departments of Biochemistry and Molecular Biology and
Anthropology, Pennsylvania State University, 205A Life
Sciences Building, University Park, PA 16803, USA

are rigid and not scalable, preventing the development of a single platform for genome-wide and high-resolution analysis of copy number. While microarrays somewhat surpass this limitation, it is really exome and whole-genome sequencing that offer a robust remedy because of the single nucleotide resolution. However, unlike with previous methods, complicated algorithms are necessary for accurate copy number detection from next-generation sequencing (NGS) data. As exome and whole-genome sequencing become standard methods in clinical labs in the next 5 years, the algorithms to calculate copy number will improve significantly and leave other methods to be used for more specialized purposes, such as interrogating structural rearrangements and assessing copy number in complex sequences.

Traditional Cytogenetics

For the last 45 years karyotyping was the gold standard for studying chromosomes [5]. A relatively inexpensive method, karyotyping has been routinely used in postnatal and prenatal genetic testing to detect aneuploidies, translocations, inversions, supernumerary chromosomes, and large deletions and duplications. With some variation in the chromosome staining method used to identify the hetero- and euchromatic bands, karyotyping has been valuable in addressing many concepts in classical cytogenetics, including the frequencies of aneuploidies, mechanisms of unbalanced rearrangements, and meiotic dispersal of abnormal chromosomes. Despite its significant contributions to clinical genetics, karyotyping is limited to detecting copy number changes that extend from approximately 5 Mb to the full length of a chromosome (Table 1).

In some cases, traditional chromosomal analysis can detect copy number changes less than 5 Mb in size, dependent upon the karyotypic resolution and specific genomic location (e.g., the 17p11.2 deletion causing Smith–Magenis syndrome). However, it was not until fluorescence in situ hybridization (FISH) was developed that deletions or duplications in the range of 200 kb–5 Mb could be routinely investigated [5]. The basic concept behind FISH usage in clinical diagnostics is the hybridization of large DNA fragments (e.g., BAC clones) to chromosome preparations from cells isolated from a clinically affected individual to identify deletions, duplications, amplifications, or copy-neutral rearrangements. FISH can be used on cells in metaphase captured with appropriate cell culture treatments; this approach is sensitive for detecting deletions larger than 200 kb and for duplications larger than 1 Mb [6, 7]. FISH on cells in interphase is more suitable for duplications between 500 kb and 2 Mb, but these can be difficult to detect when they are in tandem. A modified protocol called fiber FISH can be used to detect

small tandem duplications [8, 9], but this is not typically available in clinical settings. Recent developments have replaced the large insert clones that are typically used as FISH probes with synthetic oligonucleotides, which provide advantages in terms of specificity and the number of targets [10•].

FISH has a variety of important applications, including detecting recurrent CNVs associated with microdeletion syndromes, confirming array CGH findings and determining the nature of complex rearrangements (e.g., mapping unbalanced translocations or identifying marker chromosomes), and investigating genome-wide copy number changes. This last application is based on a modified method called spectral FISH that evolved from comparative genomic hybridization (CGH), an approach used to map all >20 Mb deletions and duplications in a clinically affected genome tested against a normal reference genome [11]. Spectral FISH and CGH have been largely used for research in cancer cytogenetics but did not find broad application in a clinical setting [12, 13]. However, CGH did provide the conceptual basis for a dramatic advance in genome-wide copy number detection based on DNA microarrays, described in detail in a later section.

Molecular PCR-Based Approaches

While aneuploidies and >200 kb microdeletions and duplications can be interrogated by cytogenetic methods, deletions or duplications that span a few hundred nucleotides to a few kilobases are amenable to study via a variety of molecular methods, including Southern blotting, quantitative PCR, MLPA, and digital droplet PCR (Table 1). This review will not discuss Southern blotting since it is largely obsolete in clinical laboratories and restricted to very specialized purposes, such as confirming expanded triplet repeats. Readers are referred to previously published articles for more detail [14]. More recent technologies, such as molecular inversion probe assays (MIP), multiplex amplicon quantitation (MAQ), quantitative oligonucleotide ligation assay (qOLA), invader assay, and pyrosequencing, are now available and well-suited for analyzing small deletions, duplications, and even amplifications.

Quantitative real-time PCR (qPCR) has been widely used in research but less often in clinical settings. It was initially used for gene expression studies but was eventually adopted for assaying copy number in genomic DNA. There are variations of qPCR methods with commercially available kits, and one of the most robust involves a *Taqman* assay involving sequence-specific PCR primers and fluorescence-tagged hydrolysis probes using fluorescence resonance energy transfer (FRET) [15]. Briefly, a short hydrolysis probe with a fluorescent reporter dye on the 5' end and a quencher dye on the 3' end hybridizes to a target

Table 1 Methods and contexts for copy number detection

Method	Appropriate context for use					Features of each method			Comments	
	CNVs >5 Mb	CNVs <5 Mb	Single gene CNVs	Exonic CNVs	Structural rearrangements ^a	Resolution	Specificity	Scalability of targets in a single assay		Typical use dynamic range
Quantitative PCR	•	•	•	•		100 bp	High	Individual loci	0–4 copies	Requires custom probes for each target locus.
Multiplex ligation-dependent amplification (MLPA)			•	•		100 bp	High	≤40 loci	0–4 copies	Resolution related to target loci and intended use ^d
										MLPA kits available only for specific genomic regions.
Digital droplet PCR			•	•		100 bp	High	≤50 loci	0–10 copies	Resolution related to target loci and intended use ^d
Molecular inversion probe (MIP)			•	•		Single nucleotide	High	≤50 loci	0–60 copies	Resolution related to target loci and intended use ^d
Multiplex amplicon quantification (MAQ)			•	•		100 bp	High	≤50 loci	0–4 copies	Resolution related to target loci and intended use ^d
Quantitative oligonucleotide ligation assay (qOLA)			•	•		Single nucleotide	High	≤50 loci	0–6 copies	Resolution related to target loci and intended use ^d
			•	•		Single nucleotide	High	≤50 loci	0–6 copies	Resolution related to target loci and intended use ^d
Invader assay			•	•		Single nucleotide	High	≤50 loci	0–6 copies	Resolution related to target loci and intended use ^d
Oligonucleotide exon-focused microarray CGH			•	•		200 bp	Moderate to high	Thousands	0–4 copies	Use depends on array design/probe coverage of target exonic regions
Oligonucleotide chromosomal microarray CGH	•	•	•	•	• ^b	100 kb	Moderate to high	Thousands	0–4 copies	For single gene CNVs, depends on array design/probe coverage

Table 1 continued

Method	Appropriate context for use				Features of each method				Comments	
	CNVs >5 Mb	CNVs <5 Mb	Single gene CNVs	Exonic CNVs	Structural rearrangements ^a	Resolution	Specificity	Scalability of targets in a single assay		Typical use dynamic range
BAC chromosomal microarray CGH	•	•			• ^b	250 kb	Moderate	Hundreds	0–4 copies	For single gene CNVs, depends on array design/probe coverage
SNP bead microarrays	•	•			• ^b	100 kb	Moderate to high	Thousands	0–4 copies	
SNP short probe microarray	•	•			• ^b	100 kb	Moderate to high	Thousands	0–4 copies	
Multiplex amplifiable probe hybridization (MAPH)			•	•		400–600 bp	High	≤50 loci	0–4 copies	Appropriate for interrogating specific targets
Next-generation capture sequencing	•	•	•	•	• ^c	Single nucleotide	Moderate to high	Millions	0–4 copies	Dependent on length of paired-end reads
FISH—deletions	•	•			•	250 kb	High	Individual loci	0–2 copies	Can detect single gene deletions if ≥100 kb
FISH—duplications	•	•			•	1 Mb	High	Individual loci	Qualitatively up to many copies	Small duplications (<1 Mb) may be difficult to detect by metaphase FISH and require interphase FISH
Karyotyping	•				•	5 Mb	High	Good for large CNVs	Qualitatively up to many copies	

^a Microarrays can detect unbalanced rearrangements but cannot provide detail on the nature and location of the DNA material present in abnormal copy number (e.g., markers or translocated material)

^b While DNA microarrays can identify abnormal copy number associated with unbalanced rearrangements (e.g., translocations or marker chromosomes), they cannot provide structural information about the aberrations

^c Next-generation sequencing can theoretically define the breakpoints of translocations, inversions, and other rearrangements but requires long paired-end reads (>250 bp)

^d While some molecular methods can theoretically be used to detect >1 Mb CNVs, they are not typically used for that purpose as there are better suited methods, such as DNA microarrays, FISH, or custom qPCR

sequence. Subsequent PCR amplification with flanking primers cleaves the probe due to the 5' exonuclease activity of the *Taq* polymerase. Separation of the reporter and quencher causes activation of the reporter, which can be measured by a fluorescence-detecting instrument. Quantification of the template copy number is based on the number of PCR cycles (the C_t value) required to bring the fluorescence to an arbitrary point within the exponential phase of amplification. The C_t value of the test sample is determined relative to that of a control sample for which the copy number is known. While real-time qPCR has the advantage of being used for virtually any gene in the genome, it is laborious because it requires careful primer selection, optimization of primers over a standard curve, and testing of multiple controls to ensure reproducibility.

MLPA is a widely used technology to evaluate copy number at up to forty genomic loci [16]. The method is based on using two probes that anneal to specific sequences adjacent to each other. In the subsequent step, DNA ligase joins the two probes if the hybridization is perfect and there is no mismatch at the end of each probe, thereby creating one single large fragment. Each probe consists of two parts—a sequence complementary to a target and a universal primer sequence. Amplification of ligated products with a single dye-labeled primer set and subsequent capillary electrophoresis provides copy number data. The number of ligation products serves as a direct measure of the copy number of the target sequence. Using probes of different lengths to separate products during electrophoresis allows multiplexing to simultaneously interrogate multiple loci within the same gene as well as in different genes. A modified MLPA protocol is also available to detect differentially methylated sequences, enabling diagnosis of imprinting disorders such as Prader–Willi and Angelman syndromes [17].

A modification of traditional qPCR, digital droplet PCR is a recent development that allows more sensitive quantitation of template DNA. Where traditional PCR is sometimes unreliable for quantitation because it can be compromised by low template concentrations, non-reproducible amplification at the exponential phase, and inconsistency in the number of cycles required to reach the plateau phase, digital PCR circumvents these limitations because it occurs in individual compartments that each contain a single template molecule. Emulsion droplets are one example of these compartments, and several commercially available kits based on this technology are available. Integration of *Taqman* chemistry into the digital droplet PCR protocol provides a sensitive platform for assaying copy number at predetermined targets [18•]. The concentrations of amplified product are calculated based on the number of fluorescent droplets produced by primers at the target locus and compared to that produced by primers at a reference locus.

MIPs offer another method to evaluate copy number at specific loci [19]. The technology is based on using a pair of probes that hybridize to a target sequence and are separated by a single nucleotide, in the same manner as MLPA probes. However, MIP probes also contain other sequences to allow circularization of the perfectly hybridized probe pair and subsequent PCR amplification. The amplified products are hybridized to single nucleotide polymorphisms (SNP) microarrays to detect the genotype and copy number of target loci. The advantages of this technology include high specificity, low template DNA amount, a large dynamic range to accurately count up to 50 copies, and scalability across many targets. MIP has also been adapted for obtaining sequence data [20].

Less commonly used molecular methods for copy number detection include MAQ, qOLA, invader assay, and pyrosequencing. MAQ is a multiplex PCR assay with isothermal primers that amplify up to 50 targets from both test loci and copy number-stable control loci in a single reaction [21–23]. The copy number is determined by electrophoresis and fluorescence-based quantitation of amplified products from the test loci in relation to those from the control loci. qOLA is a variation of MLPA and has been used to quantify zero to six copies of target loci and can also be used to genotype alleles [24–26]. Different targets are identified by varying amplicon lengths. The Invader assay is a commercially available technology that uses a probe containing a target-specific region and a 5' flap sequence. An invader oligonucleotide that binds the sequence adjacent to the target and has a one-base overlap with the probe elongates in a PCR and cleaves the 5' flap sequence. The released flap fragment binds as an invader oligo on a synthetic target that contains a FRET probe. Cleavage of the FRET probe releases a fluorescent signal that is quantified. The Invader assay is highly scalable and specific, and can be used to interrogate multiple loci simultaneously and also genotype target alleles [27, 28]. Lastly, pyrosequencing is a technology based on detecting released pyrophosphates that are used to generate ATP for a luciferase reaction. Pyrosequencing can be used to assay for specific single nucleotides but also quantify target loci up to six copies [29–31]. However, since each of these methods require careful and laborious probe and primer optimization and calibration of reaction conditions across many control samples, they are not routinely used in the clinical setting.

DNA Microarrays

Initial phases of the Human Genome Project focused on building physical maps using large insert clones, such as bacterial artificial chromosomes (BACs). This was essential to not only create a map to position known genomic

landmarks but also to use the clones themselves as templates for sequencing. The availability of these clones, the completion of the human genome sequence, and the development of glass surface-based nucleotide arrays together led to the first DNA microarrays that could be used to evaluate copy number of sequences in the human genome [32]. Taking a cue from traditional CGH used in cancer cytogenetics, BAC arrays were deployed in clinical settings to detect copy number abnormalities first at specific targets but eventually across the whole genome [33, 34]. DNA microarray CGH (array CGH) technology essentially compares copy number at specific loci in a patient sample in relation to the copy number in a co-hybridized and differently labeled reference sample. Unlike traditional CGH, which hybridizes labeled test and reference genomes to a metaphase cell spread, array CGH utilizes a set of DNA fragments on a glass surface. The singular advantage of array CGH is the ability to select which portions of the genome to target and thereby define the resolution of detection. The methodology has been described in detail elsewhere [13, 35].

While BAC arrays set the stage for whole genome copy number analysis in clinical diagnostics [33, 34], it was oligonucleotide arrays that firmly established this technology as the standard for evaluating the human genome for CNVs [36, 37]. Oligonucleotide arrays provide superior resolution and quality when compared to BAC arrays and have been used extensively in the last 5 years to generate a high-resolution map of pathogenic and benign copy number variation in the human genome [2, 3, 38–40]. BAC microarrays offer resolution down to 150–250 kb whereas oligonucleotide arrays refine that resolution to less than 1 kb.

The availability of CGH oligonucleotide-based microarrays also enabled copy number analysis of single exons and genes, essentially replacing MLPA and qPCR, because the array offered better resolution across every exon and data from a large numbers of genes [41–43]. These ultra high-resolution arrays can detect deletions as small as 200 bp at virtually every exon of targeted genes. This same methodology can be expanded to cover the entire exome to complement whole genome or exome sequencing and will likely be available for broad clinical use in the near future.

Oligonucleotide microarrays with probes containing SNPs offer an alternative to CGH microarrays [44, 45]. SNP microarrays were initially developed to genotype specific alleles for association studies and similar investigations, but evolved into ones that can survey genotypes as well as copy number [46, 47]. This is accomplished by qualitatively detecting which probe is bound by labeled DNA fragments from the tested genome (hence identifying the alleles in that genome) and also quantitating the fluorescence of the bound DNA. Probes representing each

allele at a polymorphic locus are represented on the microarray. A reference genome is not co-hybridized. Unlike CGH arrays, because SNP arrays provide not only copy number data but also genotype information, they allow efficient detection of long stretches of homozygosity that can indicate uniparental disomy or identity by descent. The methodology and the advantages of SNP microarrays are described in detail elsewhere [45, 48].

Multiplex amplifiable probe hybridization (MAPH) is a technology that utilizes a set of probes that are hybridized to a test genome on a nylon filter and then recovered for quantitative amplification. A more robust version of this method, replacing gel electrophoresis in the last steps with oligonucleotide microarrays, significantly increases the number of targets assayed [49]. MAPH provides a much higher signal-to-noise ratio because of the specific probe targeting and PCR amplification instead of a whole genome hybridization as performed in microarray CGH, and therefore can be used even within complex genomic sequences that are usually difficult to assay by other methods.

Next-Generation Sequencing

While DNA microarray technologies have shown considerable promise in high-throughput and cost-efficient diagnostic studies, their ability to accurately determine the length of an aberration can be limited by the density of oligonucleotide probes or SNPs within a target region [50]. The robustness of any CNV detection methodology or analysis algorithm not only lies in accurate delineation of the breakpoints and precise estimation of size but also in estimating the absolute change in copy number of a genomic region and delineating different classes of variants. NGS offers the advantage of a whole genome approach and the ultimate resolution of a single nucleotide. Sophisticated computational algorithms are necessary to extract copy number data from NGS data by aligning sequence reads against a reference genome for comparison. Algorithms to detect CNVs were first developed for read-pairs from BAC clone end sequences generated from the breast cancer cell line MCF-7 [51], and subsequently adapted to detect variation from fosmid paired-end sequences and from next generation paired-end sequence data [52, 53]. These algorithms were developed to assess genome-wide copy number changes using whole-genome sequencing data. It is important to note that these algorithms are confounded by exome data sets because of the hybridization biases and uneven coverage throughout the genome.

The read-pair approach takes into account the span and orientation of sequence reads in comparison to a reference genome. This approach is based on the principle of

identifying discordant signatures of sequence content and orientation, which may be diagnostic of different patterns of structural variation. Read pairs whose sequenced ends anchor to the reference genome are considered discordant if the mapping distance between them varies from the expected length [54]. The accuracy of the read-pair approach depends on the read length, insert size, and physical coverage of the genome. This approach also allows for the detection of inversions (discordant for orientation) in addition to deletions and duplications. Read-depth approaches estimate copy number by quantifying the mapping depth of sequence reads that are assumed to be in a random Poisson distribution. Duplications and deletions are discovered based on the deviation (higher or lower depth) compared to known diploid regions of the genome. Modifications to this strategy, including incorporation of robust statistical parameters and single unique nucleotide identifiers, have improved the sensitivity and accuracy of CNV detection. The read-depth approach, however, cannot detect inversions and tandem duplications. Statistical and algorithmic modifications to the read-depth approach have been used to normalize depth-of-coverage counts from exome sequencing data. For example, singular value decomposition (SVD) normalization for CoNIFER [55], principle component analysis and hidden Markov model for XHMM [56], Geary–Hinkley transformation for ExomeCNV [57], or circular binary segmentation algorithm for VarScan2 [58], have been added to read-depth data for improved CNV calling. Split-read approaches were devised to detect exact breakpoint locations based on the broken reads or gaps among the reads, and can also be extended to identify mobile-element insertions, paralogous repeats, and pseudogenes [59]. For example, Karakoc et al. recently devised an algorithm based on split reads to discover smaller insertion-deletions from exome sequencing data from individuals with autism [60]. Optimum use of this method requires longer reads and higher coverage of the genome. Moreover, instead of mapping reads to a reference, *de novo* assembly of sequences would provide a more accurate estimate of copy, content, and structure of genomic regions [50, 61]. However, this approach requires long and high-quality sequence reads. While local sequence assemblies from fosmid clones have been systematically used to discover CNVs, sequences are now assembled *de novo* as well as locally and then compared to a high-quality reference genome.

Despite the promise held by NGS-based copy number detection methods, current algorithms are limited by the fact that the sensitivity and specificity vary with each approach and a significant fraction of variants are unique to a specific approach [50]. While the read-depth approach can detect absolute copy number using single unique nucleotide identifiers in duplicated regions of the genome,

it cannot resolve breakpoints accurately and cannot identify structural differences such as tandem duplications or inversions. Read-pair and split-read approaches do not reliably resolve copies within repetitive regions. In addition, these strategies are not sufficiently powered to detect CNVs generated due to a potential fork-stalling template switching mechanism or translocation events [62]. Recent studies have also shown that pair-wise comparisons of assembled genomes can be biased against repetitive or multi-copy regions [50]. One solution to increase the sensitivity and specificity of CNV detection is to combine computational methodologies. Algorithms that combine two orthogonal approaches for better CNV detection are now available (e.g., SPANNER [63], CNVer [64], and Genome STRiP [65]). Only generation of longer sequence reads with adequate genomic coverage will mitigate most limitations associated with detection of CNVs from NGS data.

General Considerations

There are several factors that impact all copy number detection methods (Table 1). For instance, an important consideration is the ability to avoid pseudogenes with high homology to the target of interest. There are more than 10,000 pseudogene sequences in the human genome and these exist as either processed or non-processed (segmental duplication) sequences [66]. DNA microarrays are not suitable for assaying copy number at loci that have pseudogene copies elsewhere because it is difficult to discern the location of deleted or duplicated material. NGS based on capture methods is also limited in its ability to quantify copy number at loci that have one or more homologous sequences scattered in the genome. Methods that can take advantage of sequence differences between the functional gene and its pseudogene copy are typically very effective at detecting copy number changes at the target locus (functional gene), and these include MLPA, MIP, qOLA, invader assay, or pyrosequencing.

A second consideration for copy number detection methods is scalability—the number of target loci that can be simultaneously interrogated (Table 1). NGS and DNA microarrays are of course the best options for large numbers of targets. NGS offers the ultimate solution for scalability, but even a single microarray design is sufficient to target the whole genome and even include the mitochondrial genome. Even with exon-focused microarrays, the whole exome can be targeted on a single high-resolution oligonucleotide array (SA, unpublished data). Most PCR-based molecular methods can only interrogate anywhere from one to forty targets and are appropriate for frequently used clinical tests aimed at a limited number of loci.

The resolution of detection is a consideration mainly for microarrays since the other methods are essentially designed with a known target and NGS has single nucleotide-level resolution. Oligonucleotide CGH microarrays can detect deletions and duplications from 200 bp to the full length of a chromosome. SNP microarrays have less resolution (10–15 kb at a minimum) but can identify individual nucleotide variation qualitatively. Molecular methods in general can detect deletions or duplications that are 100 bp or larger and contain the target sequence for the probes or primers used in the assay, and some of these can also qualitatively identify single nucleotides.

The dynamic range of detectable copy number varies among the described methods. While all of them can distinguish between zero and four copies, some are limited in their abilities to detect amplifications of DNA material (present in four to many tens of copies). Interphase FISH is used to qualitatively detect oncogene amplifications in cancer, such as HER-2 amplifications in breast cancer [67]. However, detecting large numbers of a genomic locus quantitatively is challenging and only some methods are capable of this, such as MIP (Table 1).

Finally, it is important to consider which method is appropriate for specific clinical contexts. Each method has advantages as well as limitations that define its suitability for use in diagnostic testing. For example, the traditional cytogenetic methods are more appropriate and sensitive for investigating balanced rearrangements and for detecting mosaicism. Similarly, when results are needed quickly, as in prenatal testing, locus-specific FISH and PCR-based methods are better suited compared to NGS and array CGH. However, once the locus of interest narrows to a small region or even a single gene, molecular methods rather than FISH or karyotyping are required to obtain the necessary data, such as genotype, intragenic deletions, and submicroscopic multi-gene CNVs. In contrast, high-resolution surveys of the whole genome call for DNA microarrays or NGS-based approaches to address complex genetic conditions involving developmental delay, intellectual disability, and congenital anomalies [36]. Therefore, while the availability of a myriad copy number detection methods provides the flexibility to address a wide range of situations in genetic testing, the appropriate methods have to be chosen carefully depending on the clinical context.

Conclusions

Copy number detection has become easier, more accurate, and highly scalable. With methods such as NGS, microarrays, and MLPA, a wide variety of needs can be met in a clinical testing environment. While some disorders, such as

holoprosencephaly or aniridia, require a simpler method with limited targets but high specificity, other disorders that are more genetically heterogeneous have to be addressed at a whole genome scale. With rapidly dropping costs, increasing efficiency, and more accurate and reproducible data, NGS promises to replace most copy number methods and restrict them to specialized needs, such as determining the structural nature of chromosomal rearrangements, surveying complex sequences, or providing a high throughput option for a small target gene list. The transition to NGS-based copy number detection is largely dependent on the development of robust algorithms and end-user software, and many efforts are underway in both the public and private sectors to address this need. Identifying deletions or duplications from a the whole genome scale and down to a single nucleotide resolution, and analysis of thousands of individuals in the coming years, will significantly complement sequence variation data and contribute to our understanding of the magnitude of all variation in our genome.

Disclosure S. Aradhya is employed by GeneDx, Inc.; A. M. Cherry and S. Girirajan have no disclosures.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
 - Of major importance
1. •• Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437–55. *This is a comprehensive and updated review of copy number variation in the human genome. It discusses CNVs specifically in the disease context.*
 2. Johansson AC, Feuk L. Characterization of copy number-stable regions in the human genome. *Hum Mutat*. 2011;32:947–55.
 3. Pang AW, MacDonald JR, Pinto D, et al. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol*. 2010;11:R52.
 4. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451–81.
 5. Emanuel BS, Saitta SC. From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. *Nat Rev Genet*. 2007;8:869–83.
 6. Cannizzaro LA. Fluorescent in situ hybridization of DNA probes in the interphase and metaphase stages of the cell cycle. *Methods Mol Biol*. 2013;946:61–83.
 7. Bayani J, Squire JA. Fluorescence in situ hybridization (FISH). *Curr Protoc Cell Biol*. 2004; Chapter 22:Unit 22 24.
 8. Shimojima K, Imai K, Yamamoto T. A de novo 22q11.22q11.23 interchromosomal tandem duplication in a boy with developmental delay, hyperactivity, and epilepsy. *Am J Med Genet A*. 2010;152A:2820–6.
 9. Gervasini C, Bentivegna A, Venturin M, et al. Tandem duplication of the NF1 gene detected by high-resolution FISH in the 17q11.2 region. *Hum Genet*. 2002;110:314–21.

10. • Beliveau BJ, Joyce EF, Apostolopoulos N, et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci USA*. 2012;109:21301–6. *This is a novel application of FISH methodology suited for specific genomic targets that are not easily assayed using the traditional approach of hybridizing large clones.*
11. Pinkel D, Seagraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 1998;20:207–11.
12. Lapierre JM, Tachdjian G. Detection of chromosomal abnormalities by comparative genomic hybridization. *Curr Opin Obstet Gynecol*. 2005;17:171–7.
13. Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet*. 2005;6:782–92.
14. Brown T. Southern blotting. *Curr Protoc Mol Biol*. 2001; Chapter 2:Unit 2 9A.
15. Mouritzen P, Nielsen PS, Jacobsen N, et al. The ProbeLibrary: expression profiling 99 % of all human genes using only 90 dual-labeled real-time PCR probes. *Biotechniques*. 2004;37:492–5.
16. Schouten JP, McElgunn CJ, Waaij R, et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*. 2002;30:e57.
17. Procter M, Chou LS, Tang W, et al. Molecular diagnosis of Prader–Willi and Angelman syndromes by methylation-specific melting analysis and methylation-specific multiplex ligation-dependent probe amplification. *Clin Chem*. 2006;52:1276–83.
18. • Hindson BJ, Ness KD, Masquelier DA, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem*. 2011;83:8604–10. *This paper describes the use and utility of digital droplet PCR and its applications. It provides a description of using this technology in a high-throughput fashion.*
19. Wang Y, Moorhead M, Karlin-Neumann G, et al. Analysis of molecular inversion probe performance for allele copy number determination. *Genome Biol*. 2007;8:R246.
20. O’Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*. 2012;338:1619–22.
21. Kumps C, Van Roy N, Heyrman L, et al. Multiplex amplicon quantification (MAQ), a fast and efficient method for the simultaneous detection of copy number alterations in neuroblastoma. *BMC Genomics*. 2010;11:298.
22. Van Den Bossche MJ, Johnstone M, Strazisar M, et al. Rare copy number variants in neuropsychiatric disorders: specific phenotype or not? *Am J Med Genet Part B Neuropsychiatr Genet*. 2012; 159B:812–22.
23. Multiplicon. 2013. <http://www.multiplicon.com>. Accessed Jan 2013.
24. Seo BY, Park EW, Ahn SJ, et al. An accurate method for quantifying and analyzing copy number variation in porcine KIT by an oligonucleotide ligation assay. *BMC Genet*. 2007;8:81.
25. Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. *Science*. 1988;241:1077–80.
26. Barany F. Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc Natl Acad Sci USA*. 1991;88: 189–93.
27. Mast A, de Arruda M. Invader assay for single-nucleotide polymorphism genotyping and gene copy number evaluation. *Methods Mol Biol*. 2006;335:173–86.
28. Hosono N, Kubo M, Tsuchiya Y, et al. Multiplex PCR-based real-time invader assay (mPCR-RETINA): a novel SNP-based method for detecting allelic asymmetries within copy number variation regions. *Hum Mutat*. 2008;29:182–9.
29. Pielberg G, Olsson C, Syvanen AC, Andersson L. Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics*. 2002;160:305–11.
30. Zackrisson AL, Lindblom B. Identification of CYP2D6 alleles by single nucleotide polymorphism analysis using pyrosequencing. *Eur J Clin Pharmacol*. 2003;59:521–6.
31. Liu Z, Schneider DL, Kornfeld K, Kopan R. Simple copy number determination with reference query pyrosequencing (RQPS). *Cold Spring Harb Protoc*. 2010; 2010(9):pdb prot5491.
32. Ishkanian AS, Malloff CA, Watson SK, et al. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet*. 2004;36:299–303.
33. Bejjani BA, Saleki R, Ballif BC, et al. Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more? *Am J Med Genet A*. 2005;134:259–67.
34. Cheung SW, Shaw CA, Yu W, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med*. 2005;7:422–32.
35. Shinawi M, Cheung SW. The array CGH and its clinical applications. *Drug Discov Today*. 2008;13:760–70.
36. • Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*. 2010;86:749–64. *This landmark paper describes a meta-analysis of chromosomal variation and justifies rethinking of the 40-year-old paradigm of using karyotyping as a first-line test for developmental delay, intellectual disability, and/or congenital anomalies.*
37. Aradhya S, Cherry AM. Array-based comparative genomic hybridization: clinical contexts for targeted and whole-genome designs. *Genet Med*. 2007;9:553–9.
38. Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43:838–46.
39. Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005;77:78–88.
40. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–54.
41. Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat*. 2010;31:1326–42.
42. Saillour Y, Cossee M, Leturcq F, et al. Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method. *Hum Mutat*. 2008;29:1083–90.
43. • Aradhya S, Lewis R, Bonaga T, et al. Exon-level array CGH in a large clinical cohort demonstrates increased sensitivity of diagnostic testing for Mendelian disorders. *Genet Med*. 2012; 14:594–603. *This paper describes the first large clinical dataset for exon-level copy number analysis and shows the frequency and types of exonic deletions and duplications at specific genes in the genome.*
44. • Kearney HM, Kearney JB, Conlin LK. Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. *Clin Lab Med*. 2011;31:595–613. *This paper describes the use of SNP arrays in a clinical setting and the complicated interpretations arising from observations of significant genome-wide homozygosity.*
45. Yau C, Holmes CC. CNV discovery using SNP genotyping arrays. *Cytogenet Genome Res*. 2008;123:307–12.
46. Bignell GR, Huang J, Greshock J, et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res*. 2004;14:287–95.
47. Huang J, Wei W, Zhang J, et al. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics*. 2004;1:287–99.
48. Schwartz S. Clinical utility of single nucleotide polymorphism arrays. *Clin Lab Med*. 2011;31:581–94.

49. Kousoulidou L, Mannik K, Sismani C, et al. Array-MAPH: a methodology for the detection of locus copy-number changes in complex genomes. *Nat Protoc*. 2008;3:849–65.
50. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. *Nat Methods*. 2011;8:61–5.
51. Volik S, Zhao S, Chin K, et al. End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci USA*. 2003;100:7696–701.
52. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. *Nat Genet*. 2005;37:727–32.
53. Korbel JO, Urban AE, Affourtit JP, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science*. 2007;318:420–6.
54. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods*. 2009;6:S13–20.
55. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res*. 2012;22:1525–32.
56. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*. 2012;91:597–607.
57. Sathirapongsasuti JF, Lee H, Horst BA, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*. 2011;27:2648–54.
58. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–76.
59. Ye K, Schulz MH, Long Q, et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25:2865–71.
60. Karakoc E, Alkan C, O’Roak BJ, et al. Detection of structural variants and indels within exome data. *Nat Methods*. 2012;9:176–8.
61. • Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12:363–76. *This is a useful review of the methods available to detect copy number variation and their limitations.*
62. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;131:1235–47.
63. Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470:59–65.
64. Medvedev P, Fiume M, Dzamba M, et al. Detecting copy number variation with mated short reads. *Genome Res*. 2010;20:1613–22.
65. Handsaker RE, Korn JM, Nemesh J, McCarroll SA. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat Genet*. 2011;43:269–76.
66. Pei B, Sisu C, Frankish A, et al. The GENCODE pseudogene resource. *Genome Biol*. 2012;13:R51.
67. Pathmanathan N, Bilous AM. HER2 testing in breast cancer: an overview of current techniques and recent developments. *Pathology*. 2012;44:587–95.