Oligonucleotide Microarrays for Clinical Diagnosis of Copy Number Variation and Zygosity Status

David T. Miller,[1,2,3] Yiping Shen,[1,3] and Bai-Lin Wu[1,3]

[1]Department of Laboratory Medicine, Children's Hospital Boston, Boston, Massachusetts
[2]Division of Genetics, Children's Hospital Boston, Boston, Massachusetts
[3]Harvard Medical School, Boston, Massachusetts

ABSTRACT

Detection of submicroscopic genomic copy number variation is now considered the first-tier clinical test—in place of standard G-banded karyotyping—in the evaluation of children with unexplained developmental delay, intellectual disability, autism spectrum disorders, or congenital anomalies. Fluorescence in situ hybridization (FISH) was the first molecular method for detection of submicroscopic genomic copy number variants (CNVs), but microarray-based comparative genomic hybridization (array CGH) has a much higher diagnostic yield for these patients when compared to traditional cytogenetic methods such as karyotype and FISH. This unit focuses on oligonucleotide arrays, including updated information about detection of long contiguous stretches of homozygosity (LCSH) through inclusion of single-nucleotide polymorphism (SNP) probes. Most clinical laboratories now offer arrays with some level of probe coverage throughout the genome, and many are offering detection of LCSH. Updated guidelines for array design and result interpretation are reviewed. *Curr. Protoc. Hum. Genet.* 74:8.12.1-8.12.17. © 2012 by John Wiley & Sons, Inc.

Keywords: copy number variant • CNV • genomic imbalance • molecular diagnostics • array comparative genomic hybridization • CGH • aCGH

BASIC PROTOCOL

Array comparative genomic hybridization (aCGH) is a method for comparing copy number of genomic loci between test and reference samples. Many types of genomic copy number variation cause recognizable syndromes ranging from imbalance of entire chromosomes (aneuploidy), as in Down syndrome, to submicroscopic imbalance, as in 22q11.2 deletion causing DiGeorge/velocardiofacial syndrome (VCFS). Detection of genomic copy number variation is now considered the standard of care in the evaluation of children with unexplained developmental delay, intellectual disability, autism spectrum disorders, or congenital anomalies (Manning and Hudgins, 2010; Miller et al., 2010; Shen et al., 2010). Submicroscopic genomic imbalance is a frequent cause of idiopathic intellectual disability, with microscopically visible abnormalities accounting for at least 3% to 4% of cases (Shevell et al., 2003), and many of these cases are recognizable syndromes such as Down syndrome. In cases with a normal karyotype, the addition of subtelomeric (ST) FISH testing achieves a diagnosis in another ∼2.4% of cases (Ravnan et al., 2006). Traditional cytogenetic methods are labor intensive, especially when multiple genomic regions are interrogated by ST-FISH, and cannot approach the coverage and yield of aCGH. Studies based on earlier BAC aCGH and higher density oligo aCGH suggest that the yield is generally in the range of 10% to 15% (Stankiewicz and Beaudet, 2007; Baldwin et al., 2008; Sagoo et al., 2009, Miller et al., 2010).

aCGH provides a way to interrogate many sites in the genome in a single experiment on a standard 1 × 3–in. glass slide. In clinical use, the widest application has been in

Current Protocols in Human Genetics 8.12.1-8.12.17, July 2012
Published online July 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142905.hg0812s74

the diagnostic evaluation of children with developmental delay/intellectual disability, dysmorphic features, multiple congenital anomalies, and autism (Vissers et al., 2003; Shaw-Smith et al., 2004; Jacquemont et al., 2006; Sebat et al., 2007; Shaikh, 2007; Weiss et al., 2008; Shen et al., 2010). Regardless of content, most aCGH platforms rely on the general principles of (1) fluorescent labeling of sample and reference DNA, (2) hybridization of labeled DNA to the array, (3) scanning to detect fluorescence intensity at all sites of hybridization, and (4) data analysis. This protocol is specific to Agilent aCGH using a two-color system to compare sample and reference DNA. Several other commercially printed oligonucleotide arrays are available and can be used with proper optimization of conditions.

Two types of oligonucleotide probe designs have been typical for aCGH. Agilent and NimbleGen platforms use long (50 to 70 mer) oligonucleotide CGH arrays and focus on high-quality copy number detection. Affymetrix and Illumina genotyping arrays use short (25 mer) oligonucleotide probes for genome-wide SNP genotyping, as well as copy number detection (Shen and Wu, 2009). The former use a two-color system to label DNA prior to hybridization, while the latter use a one-color system and infer copy number by comparing to a reference dataset. Long-oligo and short-oligo platforms showed functional convergence in recent years, as exemplified by the introduction of Affymetrix's hybrid array (Affy 6.0 array with both SNP and CNV probes, and the more recent Cytoscan HD) and Agilent's CGH+SNP array. Now both types of array platforms can detect copy number variation, as well as copy neutral zygosity status for any chromosome segment (a.k.a. LCSH: long contiguous stretches of homozygosity). This facilitates clinical applications beyond constitutional copy number analysis of patients with developmental disabilities, including analysis of loss of heterozygosity (LOH) in tumors and constitutive absence of heterozygosity (AOH) that occurs with consanguinity or uniparental disomy (UPD).

The first step in aCGH is labeling of genomic DNA from test and reference samples with different color fluorophores (e.g., Cy5 and Cy3). Labeled DNA from test and reference samples is combined in an equimolar ratio and co-hybridized to an array of DNA probes corresponding to informative regions of the genome. During hybridization, sequences from the test and reference DNA compete for probe binding. All probes are validated empirically to hybridize under the same conditions so that any differences in signal intensity are reflective of differences between test and reference DNA copy number. After washing to eliminate nonspecific hybridization, the relative intensity of each fluorophore at a particular probe coordinate is read by a laser scanner. Software applications for imaging (Feature Extraction) and informatics (CGH Analytics, now part of CytoGenomics) quantitate the color ratio (Cy5/Cy3) for each probe coordinate. This ratio reflects the relative copy number of the corresponding genomic DNA between test and reference samples. Each array feature is mapped to the human genome reference sequence and the results are indicated in a graphical display.

The new Agilent CGH+SNP array uses a similar protocol to earlier Agilent arrays, with the additional step of DNA sample digestion with restriction enzymes, which allows genotyping of SNPs located at the enzymes' recognition sites. The procedure is outlined in Figure 8.12.1. For genotyping arrays, such as Affymetrix or Illumina, only a single test DNA is labeled and hybridized to the array. The genotypes are determined by evaluating the signal intensity between alternative probes. The copy number variations are inferred from comparing the signal intensity to signal of previous run controls (McCarroll et al., 2008), but detailed workflows for these arrays are not covered in this unit. The Agilent CGH+SNP arrays use the same Agilent CGH workflow as a CGH-only array. The copy number status is determined by the log2 ratios of long oligonucleotide CGH probes between sample and reference DNAs. In addition, a set of SNP probes (a total

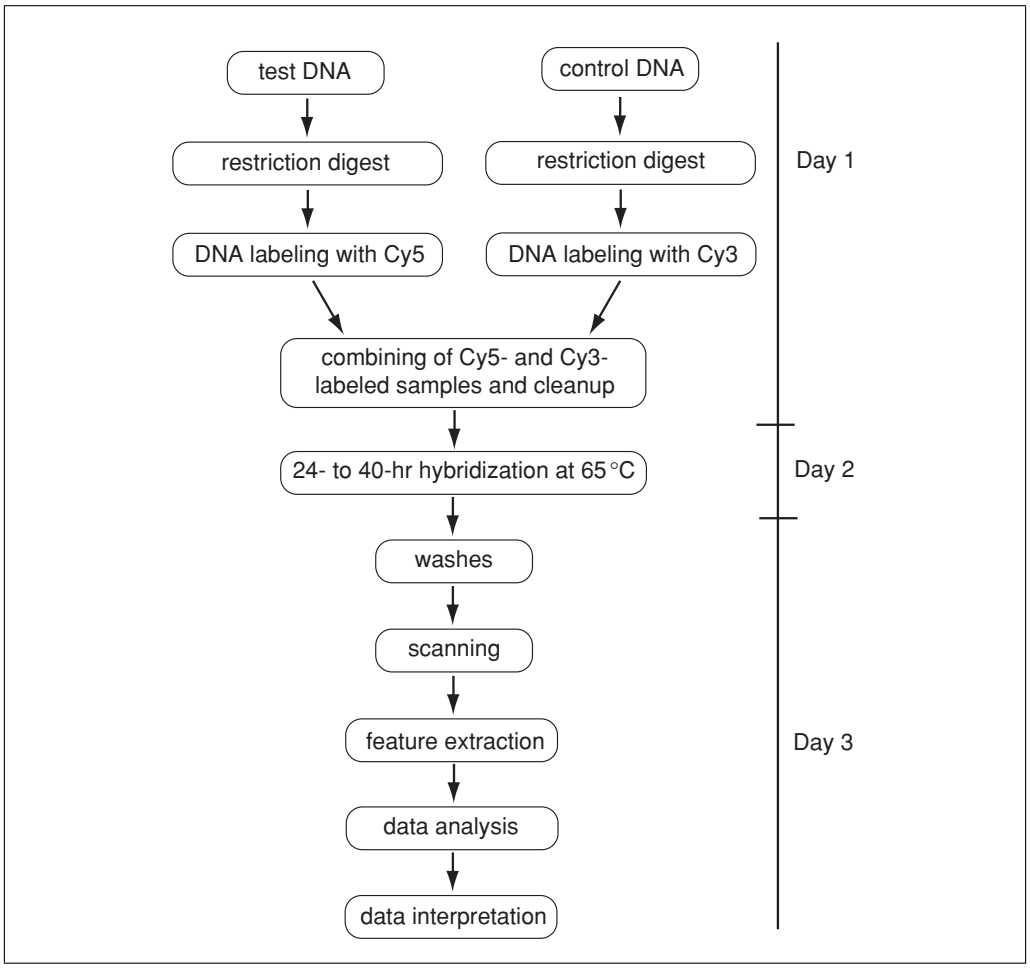

Figure 8.12.1 Basic steps of oligo array CGH with timeline.

of 60 K SNP probes on Agilent 4X180K CGH+SNP catalog array) is selected for genotyping purposes. Probes are designed to include restriction digestion sites (*Alu*I/*Rsa*I) that overlap with known SNP sites. Both sample and reference DNAs are digested by the restriction enzymes before labeling. If both alleles in the target DNA are cut (0 uncut alleles), the resulting signal of the SNP probe will be low. If both alleles are uncut, the signals of the SNP probe will be high. For a heterozygous SNP locus, there will be one cut and one uncut allele, and the resulting signal intensity of the SNP probe will be intermediate.

The copy number of one allele at each SNP site is determined by measuring the relative signal intensity between the sample and a reference. The zygosity status is determined by the number of uncut alleles: 0 uncut (AA); 1 uncut (AB); 2 uncut (BB). Regions of homozygosity are located by finding genomic regions with a statistically significant scarcity of heterozygous calls. Since the genotype of the reference is known, the SNP calls can be made for each individual SNP locus.

In a clinical diagnostic setting, data reliability can be ensured through several quality control steps. There are several data quality parameters, such as DLR spread (discussed below) that help determine overall data quality. Even if overall data quality is good, characteristics of each aberration must be considered. For example, clinical guidelines require that a certain number of successive probes show a consistent signal in order to make a call, such as five probes for a deletion and ten for a duplication. These requirements are generally more stringent for gains versus losses because the probe

ratios are closer to the reference for a single copy gain as compared to a single copy loss (Kearney et al., 2011a). As a result, aCGH resolution depends on feature density. Whole genome high-resolution arrays are capable of detecting very small aberrations throughout the genome (e.g., <10 kb), but current clinical guidelines suggest that detection of all copy number gains and losses >400 kb is sufficient (Miller et al., 2010; Kearney et al., 2011a). Because the whole genome is not evenly covered by the oligo probes, the actual detectability depends on the location of the imbalance events in the genome and the performance of the hybridization (i.e., the quality of the dataset). In some cases, a second hybridization with dye reversal to eliminate false-positive results from differences in fluorophore intensity can be considered.

Interpreting aCGH data for clinical diagnosis is challenging because copy number variants (CNVs) are now recognized as an extremely common type of genetic variant. Some behave in a Mendelian fashion, while others may contribute to complex traits or have no effect at all. The ability to distinguish pathogenic from benign copy number variation is a challenge for any clinical laboratory, and key factors are discussed in Background Information. The reader is referred to recent expert reviews and published guidelines addressing issues related to clinical aCGH interpretation (Lee et al., 2007; Vermeesch et al., 2007; Kearney et al., 2011b).

Materials

100 to 150 ng/μl genomic DNA in TE buffer:
  Sample (test) DNA (isolated from blood, see *UNIT 14.4*)
  Control (reference) DNA (e.g., normal male or female):
  NA18507 (Yoruban Male)
  NA18517 (Yoruban Female)
  NA12891 (European Male)
  NA12878 (European Female)
  NA18579 (Chinese Female)
Nuclease-free water
10× reaction buffer C (provided with *Rsa* I; Promega)
10 U/μl *Alu* I (Promega)
10 U/μl *Rsa* I (Promega)
QIAprep Spin Miniprep Kit (Qiagen; cat. nos. 27104, 27106) with Buffers PB, PE, and EB
Ethanol (denatured; Fisher)
BioPrime Array CGH Genomic Labeling System (Invitrogen; cat. no. 18095011), with:
  2.5× Random primers solution (octamers)
  40 U/μl exo-Klenow fragment
  Stop buffer
10× dUTP (nucleotide mix; PerkinElmer)
2′-Deoxyuridine conjugated with cyanine 3 and cyanine 5 (Cy3-dUTP and Cy5-dUTP; PerkinElmer)
1× TE buffer, pH 8.0 (Promega)
Oligonucleotide aCGH Hybridization kit, large volume (Agilent, 5188-5380), containing:
  10× Blocking Agent
  2× Hybridization buffer
Cot-1 DNA (Invitrogen)
Oligonucleotide aCGH wash buffers 1 and 2 (Agilent)

Spectrophotometer (e.g., NanoDrop ND-1000, Thermo Fisher Scientific)
Thermal cycler

Centrifuge
1.5-ml tubes
Speedvac Concentrator
37°C incubator
Microcon YM-30 Centrifugal Filter Kit (Millipore; cat. no. 42410)
Microcentrifuge
Oligonucleotide microarray, e.g., Human Genome 244 k array (Agilent Technologies, G4411B)
Hybridization Oven (Agilent Technologies)
Coplin jar
Ozone detector (e.g., Teledyne Model 400E), optional
Ozone converter (e.g., Ozone Solutions "Ozone Interceptor"), optional
Array scanner (e.g., Agilent G2565AA or 2565 BA)
Computer workstation (e.g., Dell Optiplex GX745)
Feature Extraction software (included with purchase of array scanner)
CGH Analytics software (Agilent)

Digest DNA samples

1. Obtain test and reference DNA samples and check purity on a NanoDrop ND-1000 UV-VIS spectrophotometer (*APPENDIX 3D*).

   All aCGH platforms are sensitive to the quality and quantity of prepared DNA. For clinical samples, DNA extracted from peripheral blood will yield the highest and most reproducible quality of DNA. Buccal swab samples are not as reliable for this application and should be discouraged.

2. Prepare a master mix (50 µl/reaction) for DNA digestion. Combine reagents in the order indicated:

   30 µl nuclease-free water
   10 µl 10× reaction buffer C (provided with *Rsa* I)
   5 µl 10 U/µl *Alu* I
   5 µl 10 U/µl *Rsa* I.

3. Prepare separate tubes containing 3 µg test DNA and 3 µg reference DNA in 50 µl TE buffer. Add 50 µl digestion master mix (final 100 µl/reaction), mix thoroughly, and incubate for 2 hr at 37°C in a thermal cycler. Place the tubes on ice following digestion.

Clean up digested DNA

4. Add 500 µl of 5× Buffer PB to each digested 100 µl sample.

   Buffers listed in this section refer to manufacturer's protocol.

5. Transfer to a QIAprep Spin Miniprep column and centrifuge 60 sec at 17,900 × *g*, room temperature. Discard flowthrough.

6. Add 750 µl Buffer PE (with ethanol added; see QIAprep Miniprep Handbook). Centrifuge 60 sec at 17,900 × *g*, room temperature, and discard flowthrough.

7. Repeat spin for 60 sec at the same speed and temperature and transfer the column into a clean 1.5-ml tube.

8. Elute DNA with 50 µl buffer EB, let sit 60 sec, and then centrifuge 60 sec at 17,900 × *g*, room temperature.

9. Concentrate the DNA to a volume of <21 µl using a Speedvac Concentrator.

10. Store the samples at –20°C or continue with labeling.

    Avoid drying the sample completely during concentration. Samples can be stored for several weeks at –20°C prior to labeling.

Label genomic DNA

11. Dispense 20 µl of 2.5× random primers solution into the 21 µl DNA sample.

    Perform this and all subsequent work in a room without an outside window or cover any outside windows. Standard-intensity fluorescent room lighting will decay the fluorophore signal within several seconds. Samples are most vulnerable when not covered (e.g., while pipetting hybridization solution onto microarray slides). These steps should be performed as quickly as possible. Samples should not be susceptible to photobleaching while in the covered water bath or in the hybridization oven.

12. Denature at 95°C in a heat block for 5 min and then set on ice for 5 min.

13. Prepare the exo-Klenow master mix in the order shown (9 µl/reaction):

    5 µl 10× dUTP
    3 µl Cy3-dUTP (for reference sample) or Cy5-dUTP (for test sample)
    1 µl exo-Klenow fragment.

    The exo-Klenow fragment of DNA polymerase I is a mutant of the large fragment of the DNA polymerase I holoenzyme that has both 5′-to-3′ and 3′-to-5′ exonuclease activity removed. Random octamers provide priming sites for the exo-Klenow enzyme. Fluorescently modified nucleotides are incorporated as the polymerase extends from the priming sites.

14. Add 9 µl mix to the DNA samples and incubate at 37°C in a water bath with a stainless steel cover for 2 hr.

15. Add 5 µl stop buffer. Store reactions at –20°C or continue with clean up.

    The labeled samples must be purified to remove contaminants prior to denaturing and hybridization to a microarray.

    Although labeled samples can be stored for several days at –20°C, proceed to the next step as soon as possible to achieve the highest signal strength during scanning. In typical laboratories, labeled samples would be stored overnight in the dark at –20°C.

Clean up labeled DNA

16. Combine Cy5 and Cy3 samples (control and test) for a total of 110 µl.

17. Add 400 µl of 1× TE, pH 8.0, and transfer to a MicroCon YM-30 filter in a 1.5-ml collection tube.

18. Centrifuge 7 min at 7000 × *g*, room temperature, and discard flowthrough.

19. Add 480 µl of TE, centrifuge again (use same conditions from step 18), and discard flowthrough.

20. Invert the filter into a fresh 1.5-ml tube and centrifuge 1 min at 700 × *g*, room temperature.

21. Bring the total volume of the sample to 70 µl with 1× TE, pH 8.0.

22. Store the sample at –20°C or continue with hybridization.

    Again, although labeled samples can be stored for several days at –20°C, proceed to the next step as soon as possible to achieve the highest signal strength during scanning.

    Pay attention to the size and color of the pellet. The pellet should have a purple hue, which indicates balanced labeling between Cy5 and Cy3. An off-color pellet (too blue or too red) may suggest labeling failure.

Perform hybridization

23. Add 1250 µl water to lyophilized Blocking Agent to make a 10× solution, and let sit at room temperature for 60 min to reconstitute.

    The 10× Blocking Agent can be prepared in advance and stored at –20°C.

24. Add the following components to each tube containing 70 µl purified labeled sample (total 300 µl/reaction):

    50 µl Cot-1 DNA
    30 µl Agilent 10× blocking Agent
    150 µl Agilent 2× hybridization buffer.

25. Transfer the tubes to 95°C for 3 min for denaturation.

26. Transfer the tubes to 37°C for 30 min for prehybridization.

27. Microcentrifuge the tubes for 5 min at top speed, room temperature, to pellet any precipitates.

28. CAREFULLY collect the supernatant without disturbing the bottom of the tube and apply the supernatant onto the gasket of the microarray. Assemble the gasket.

    Make sure the active side of the chip (labeled "Agilent") is used for hybridization. The numeric barcode is on the inactive side. The hybridization mixture should be applied directly to the gasket slide and not to the active side of the array slide. Tighten the gasket without allowing any liquid to leak out.

    After assembling the gasket, rotate the unit to make sure all areas are covered by the hybridization solution and all bubbles move freely. It is acceptable to have some bubbles in the hybridization, provided that the solution covers the entire surface area of the printed array. The array must not be allowed to dry out during hybridization or subsequent washes.

29. Hybridize at 65°C for 40 hr at rotation setting 20 in an Agilent hybridization oven.

    Samples at this stage can be stored for several days at –20°C, but proceeding to the next step as soon as possible will achieve the highest signal strength during scanning. Any remaining sample stored in this manner could be used for a repeat experiment if problems were encountered during scanning. However, samples showing low signal intensity during scanning would require repeat labeling.

Wash the samples

30. Disassemble the gasket in wash buffer 1.

31. Wash the array 5 min in wash buffer 1 in a Coplin jar with stirring.

32. Wash the array 1 min in prewarmed wash buffer 2 at 37°C in a Coplin jar with stirring.

33. Slowly take the slide out of wash buffer 2 and immediately scan the slide.

    Limit light exposure to the slide during and after washing and proceed to scanning as soon as possible after washing to avoid dye decay.

    After washing, the slides are very sensitive to photobleaching from ozone exposure. Ozone levels are higher in the daytime and during hot weather. Try to avoid prolonged exposure of the slides (<1 min) to room air when ozone concentration is >10 ppb. Washing the slides in batches of four slides helps to avoid delays between the washing and scanning steps.

    The standard protocol involves acetonitrile washes in a vented hood. Acetonitrile creates a physical barrier over the slide. The need for a vented hood can be circumvented by monitoring ozone levels and using an ozone converter.

Scan array and perform feature extraction

34. Perform scanning according to the manufacturer's protocol with the following settings:

    PMT power 100%
    Resolution 5 μM.

35. Save the image in the appropriate folder as raw data.

36. Perform feature extraction according to manufacturer's instructions, using TIFF image files.

    Exact procedures for feature extraction vary according to the manufacturer's software. Feature extraction requires comparison of data points on the assay to a design file containing coordinates for all features (probes) on the array. Processing of this data will follow a standard protocol designed by the array manufacturer. Perform manual gridding if necessary, and repeat feature extraction using the corrected grid file. Review the QC report generated by the feature extraction software. Using the derivative log ratio spread (dLRsd) as the main quality index, a dLRsd measure of <0.25 can be considered acceptable.

Analyze the data

37. Import the FE data file (.txt) to CGH analytics for data visualization and aberration detection.

    We use Aberration Detection Method algorithms (ADM-2) as the statistical method for aberration detection and use an aberration filter to identify aberrations covered by more than a predetermined number of probes.

38. Visually inspect any detected potential imbalance loci. Report the size and location of the imbalanced region.

39. Determine if the imbalance events are true for both forward (Cy5 for test sample, Cy3 for reference) and reverse (Cy3 for test sample, Cy5 for reference) labeling.

40. Determine if the imbalance is a reported copy number variant by comparison to publicly available databases and the internal laboratory database.

41. Determine if the imbalance event is associated with any reported genetic disorder.

42. Depending on the results, consider FISH or other forms of confirmatory testing.

    Recommendations may include parental testing to determine if the identified variant is a de novo imbalance event. Subsequent testing may include standard cytogenetics to look for a balance rearrangement in a parent.

43. Review and amend the original report when results of additional testing become available.

COMMENTARY

Background Information

General considerations for oligonucleotide array CGH

Array CGH (aCGH) is a convenient method for detecting submicroscopic copy number variation at many sites throughout the genome in a single experiment. aCGH can detect many types of variation—such as deletions, duplications, and chromosomal aneuploidy—at higher resolution than traditional cytogenetic methods. It is a particularly convenient platform for identification of material that is visible but not cytogenetically identifiable, such as chromosomal rearrangements or marker chromosomes. The increased resolution of detection results in a significant improvement in diagnostic yield, leading to current recommendations that aCGH be considered the first test—in place of G-banded karyotype—for patients with unexplained ID/DD, MCA, and

ASD (Manning and Hudgins, 2010; Miller et al., 2010; Shen et al., 2010). Traditional cytogenetic methods, such as G-banded karyotype, are still appropriate in certain situations, such as suspicion of Down syndrome, Turner syndrome, or Klinefelter syndrome, or when there is a family history of balanced rearrangement or multiple miscarriages of pregnancy. In addition, FISH provides the combination of copy number detection and spatial location of the imbalance, and is useful in certain cases to rule out insertional translocation of a copy number gain, or to further evaluate cases with multiple areas of imbalance, which might represent a translocation or other rearrangement.

Oligonucleotide arrays offer several advantages over large-insert bacterial artificial chromosome (BAC) arrays, and have essentially replaced BAC arrays for both clinical and research applications due to their many advantages. Probe sequences for oligonucleotide arrays are based on the reference human genome sequence, allowing the user to identify any sequence of interest as a potential target (Shen et al., 2007), offering advantages in assay design and ease of manufacturing. While individual probes on an oligo array will not perform as reliably as individual BAC probes, a given BAC region could be covered by several oligo probes. This redundancy within the assay lends confidence to interpretation. Oligo arrays offer much higher resolution, down to the level of individual exons, if desired. In the past, the versatility of array design led to variable coverage of the genome among laboratories performing aCGH. Currently, however, most clinical aCGH provide the recommended whole-genome coverage with the ability to detect changes 400 kb or greater, and differences in coverage between laboratories are not as important as they had been.

Array platforms and coverage

Many platforms are available for oligonucleotide-based arrays. Custom oligonucleotide arrays, based on libraries of validated synthetic probes with fixed GC content and melting temperature to facilitate uniform hybridization, can interrogate clinically relevant genomic regions without the need for large-insert clone libraries. Oligo arrays are flexible in terms of content, and can be customized for specific applications, e.g., using probes from Agilent's eArray library (*https://earray.chem.agilent.com/earray/* accessed August 2011), a large collection of 60-mer oligonucleotides specifically selected for robust copy number analysis (Fan et al., 2007; Shen et al., 2007). Targeted oligo-based arrays can be used for specific categories of disorders (e.g., arrays to provide dense coverage of the Duchenne muscular dystrophy gene). In general, clinical oligo aCGH for most patients is not designed for detection of copy number variation of single genes or exons because the number of probes available on typical whole-genome arrays would not provide sufficient coverage density to detect small intragenic events, and because such high-resolution coverage throughout the genome would result in detection of many variants of uncertain significance (VUS). For typical clinical testing of patients with unexplained DD/ID, MCA, or ASD, whole-genome coverage at a level of resolution around 400 kb is desirable. The International Standards for Cytogenomic Arrays (ISCA) Consortium has advocated for this level of uniform whole-genome coverage (Miller et al., 2010), and the clinical genetics community has agreed (Kearney et al., 2011a). As a result, all oligonucleotide aCGH platforms are capable of providing good coverage for copy number detection throughout the genome

Clinical validation

Clinical laboratories choose from a variety of oligo aCGH platforms with different levels of resolution and from various manufacturers. Regardless of which array a particular laboratory chooses, clinical applications demand rigorous validation. There is no specific FDA guidance about clinical validation of aCGH as a laboratory-developed test (LDT), and individual laboratories may vary in their approach. Validation through testing of known positive controls is not feasible for a test that can detect changes throughout the genome. General recommendations for clinical validation have been published by the American College of Medical Genetics Laboratory Quality Assurance Committee (Kearney et al., 2011a).

Confirmatory testing

With proper validation of each oligo aCGH platform design, confirmation of positive results through a secondary method is not a requirement in a clinical laboratory setting. The specifications for this type of validation are outlined in the American College of Medical Genetics (ACMG) guideline for the design and performance expectations for clinical aCGH (Kearney et al., 2011a). With proper validation, very low false-positive rates (<1%) can be achieved, and confirmatory testing of every result would be prohibitive in terms of

time and expense. Confirmatory testing purely for the purpose of determining copy number should not be confused with additional testing, such as FISH, to provide information about chromosomal rearrangements. For molecular laboratories performing aCGH, it is important that they establish a close working relationship with a cytogenetics laboratory in order to facilitate this additional testing. Deciding when FISH testing is required to further characterize a copy number imbalance requires the professional expertise inherent to a laboratory with clinical certification to perform genetic testing, and is beyond the scope of this unit.

Variants of unknown significance and parental/family testing

Identification of novel variants is common in aCGH testing, and more common with whole-genome arrays. Dozens of CNVs are present in every human genome (Iafrate et al., 2004; Eichler, 2006; Itsara et al., 2009), and the clinical significance of many CNVs is poorly understood. A discussion of the full breadth of expertise required to identify and interpret the clinical significance of CNVs is beyond the scope of this unit, but excellent professional guidelines have been published (Kearney et al., 2011b). This is perhaps the most challenging issue in clinical aCGH testing, especially for arrays with high-density whole-genome coverage. Database of Genomic Variants (*http://projects.tcag.ca/variation/*) has information about population frequency of CNVs. Clinical interpretations of CNVs can be found in public databases such as the International Standards for Cytogenomic Arrays (ISCA) Consortium (*https://www.iscaconsortium.org/*) and DECIPHER Database (*http://www.sanger.ac.uk/PostGenomics/decipher/*; Van Vooren et al., 2007; *UNIT 8.14*). Most clinical laboratories also maintain internal databases.

CNV interpretation is critically important, and has been addressed in published guidelines (Manning and Hudgins, 2007; Shaffer et al., 2007; Kearney et al., 2011b). However, these discussions assume that individual clinical laboratories take the final responsibility for accurately interpreting CNVs, because published guidelines will never be able to address all possible scenarios. For example, a deletion variant that appears in a healthy parent and an affected child could be acting as a recessive allele in the child. Likewise, a de novo region of genomic imbalance could be a previously undescribed nonpathogenic CNV or a case of nonpaternity. Each laboratory is ultimately responsible for understanding the limitations of aCGH and clearly stating those limitations in the report.

A common approach to the interpretation of variants of unknown significance involves offering similar testing for the parents to determine if the variant is de novo. The presence of similar variation in a parent may lead to a conclusion that identified variants are benign or pathogenic depending on the presence or absence of similar clinical symptoms in the parent. Appropriate methods for parental testing may vary depending on the aCGH finding in the proband. Other clues to pathogenicity include the gene content of genomic variants. Variants including known disease genes are more likely to explain a clinical phenotype. Even if a variant contains a known gene, the type of variant may affect pathogenicity. Deletions are generally considered more likely to be pathogenic than duplications, although this is not a universal rule. The size of genomic variants also influences pathogenicity, although relatively large variants ($>$500 kb) may be familial, and many relatively small variants may be pathogenic. A majority of CNVs detected on high-resolution oligonucleotide platforms are $<$150 kb (Komura et al., 2006; Redon et al., 2006) and detection of these CNVs complicates clinical interpretation. Overall, each of these factors must be considered before rendering a clinical interpretation of genomic CNVs, as reviewed in greater detail elsewhere (Lee et al., 2007; Kearney et al., 2011b).

Critical Parameters

Following the steps outlined in this protocol will produce successful aCGH data for the majority of samples tested, especially when good laboratory practices are followed. In general, the array slides must be handled with extreme care to avoid damaging the array. Always handle glass slides on their edges while wearing powder-free gloves. Never touch the surface of the array slides.

DNA quality

DNA purity and integrity are particularly important factors in aCGH performance in comparison with other molecular biology techniques. As noted in the protocol, checking DNA purity and concentration on a NanoDrop spectrophotometer, or equivalent, is a critical step. Genomic DNA integrity can be assessed by agarose gel electrophoresis, but this is not routinely performed on samples

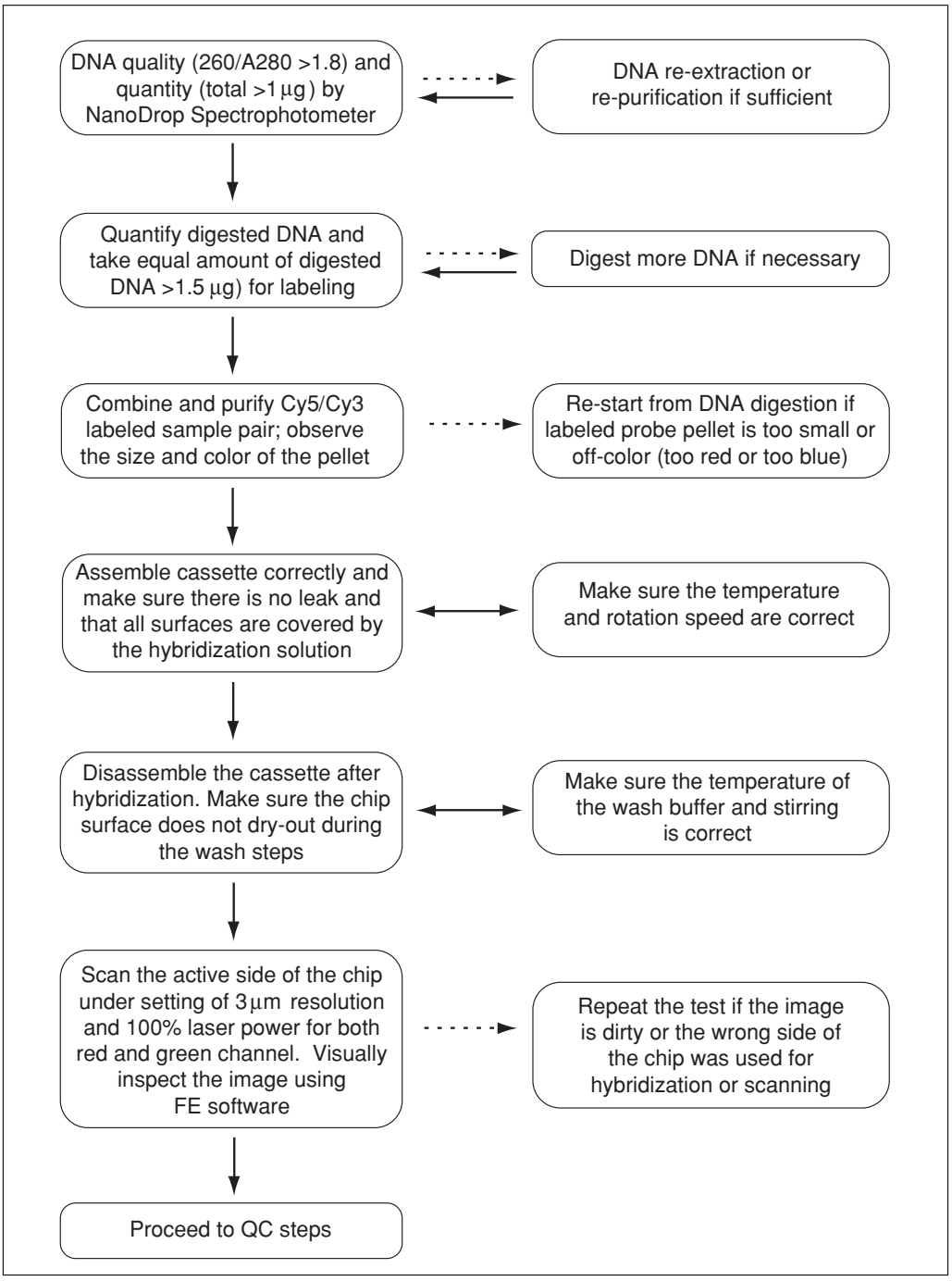

Figure 8.12.2 Troubleshooting flowchart for aCGH assay.

processed in most clinical laboratories with a validated DNA preparation protocol. Equimolar ratios of labeling are critically dependent on DNA quality. Quantity is not as important; as little as 500 ng DNA can be used as starting material. Quality is more important; poor-quality DNA will manifest as poor labeling with low signal intensity and high background noise. With high-quality DNA samples, the SNP call rate is greater than 95% with a greater than 99% accuracy. Copy-neutral aberrations as small as 5 Mb can detected.

Reference samples

When processing SurePrint G3 CGH+SNP microarrays, the reference needs to be DNA isolated from a single genotyped individual. One can use any of the following supported HapMap samples: NA18507 (Yoruban Male), NA18517 (Yoruban Female), NA12891 (European Male), NA12878 (European Female), or NA18579 (Chinese Female). The HapMap samples can be ordered from the Coriell Institute for Medical Research. Alternatively, one can genotype one's own reference

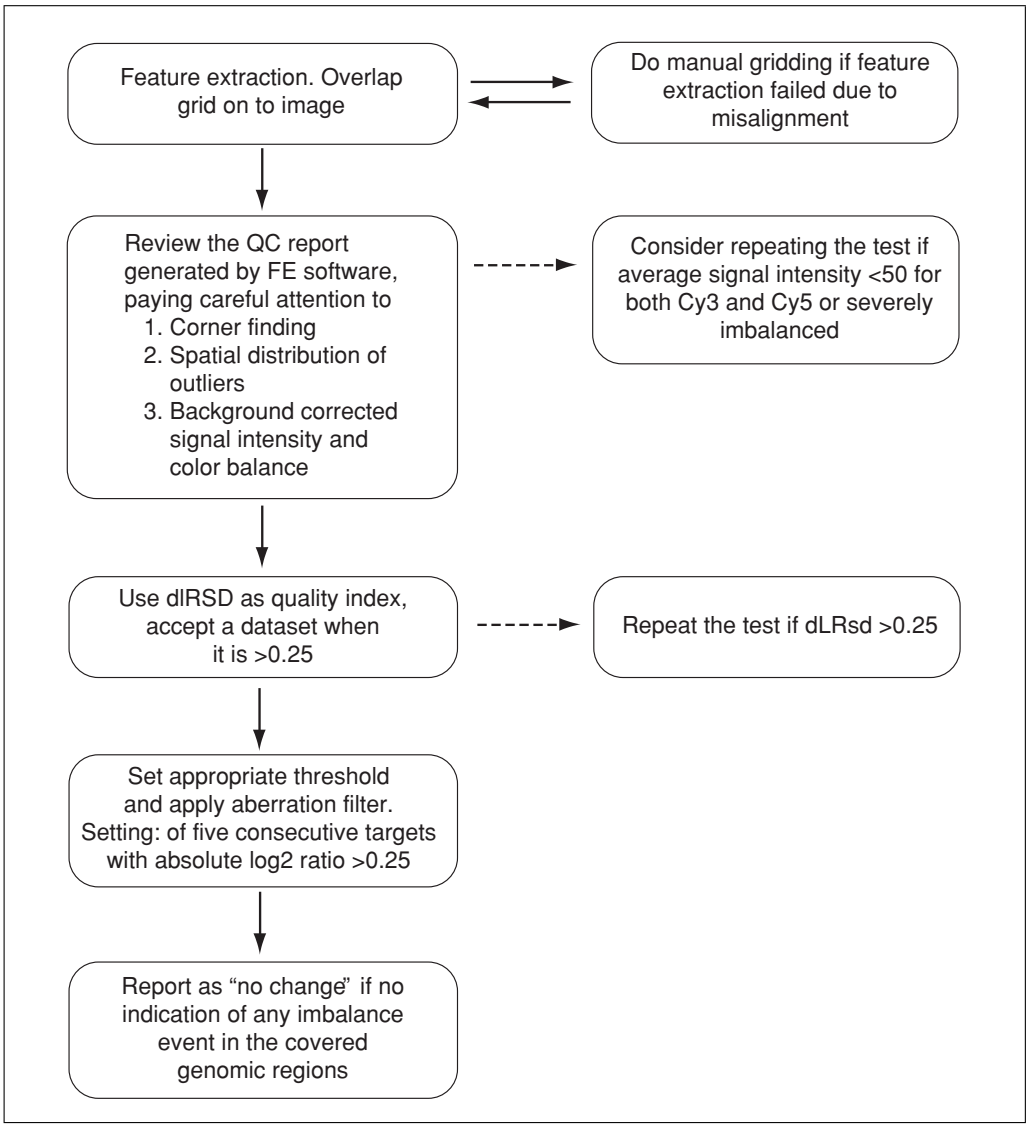

Figure 8.12.3 Troubleshooting flowchart for aCGH analysis.

isolated from a single individual by hybridizing it against all five supported HapMap samples on the Agilent CGH+SNP microarrays. This experiment only needs to be done once. The input amount of DNA for the experimental labeling reaction must be the same as for the reference sample labeling reaction. Inaccurate DNA quantitation can lead to different DNA inputs into the experimental and reference labeling reactions, which increases assay noise (DLRSD). Different DNA isolation methods can create different quantification artifacts, so a higher risk of assay noise exists when the experimental and reference DNA samples are isolated from different sources (e.g., experimental DNA isolated from blood and reference DNA obtained from a Coriell cell line). To minimize assay noise, especially when experimental and reference samples are isolated from different sources, measure the DNA concentration with both a spectrophotometer (e.g., Nanodrop) and a fluorometer (e.g., Qubit) to give two independent methods of measurement.

DNA labeling

Double enzymatic restriction digestion by *Alu*I/*Rsa*I is followed by Klenow-based enzymatic labeling. Efficient DNA labeling is critical, and the methods outlined in this protocol are generally reliable. Careful treatment of the cyanine dyes involves minimizing exposure to ambient light sources throughout the labeling procedure. These dyes are also sensitive to multiple freeze-thaw cycles. Generally, the labeling step is efficient and robust; however, difficulties may be encountered if the arrays are not handled properly during the hybridization and washing steps.

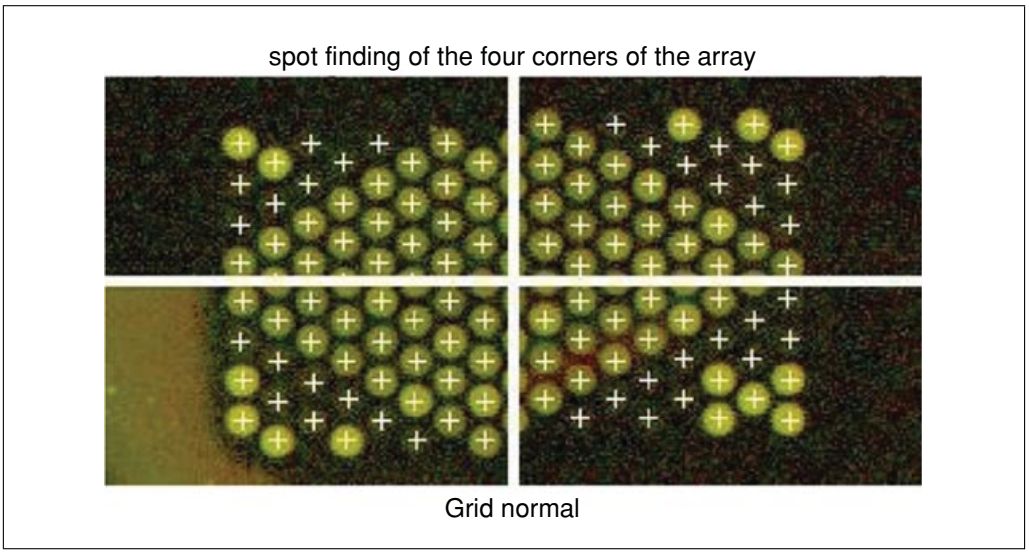

Figure 8.12.4 aCGH Grid Alignment showing spot finding of the four corners of the array. The words "Grid Normal" appear below the grid image indicating proper alignment. For the color version of this figure see *http://www.currentprotocols.com/protocol/hg0812*.

Hybridization

Proper assembly of the Agilent Microarray Hybridization Chamber is the first step in a successful hybridization. As noted in the protocol, the hybridization mixture should be applied directly to the gasket slide and not to the active side of the array slide. The slides must never be allowed to dry out during the hybridization and washing steps. As noted in the protocol, exposure to ambient light and ozone causes photobleaching. The fluorescent dyes are especially vulnerable during hybridization and washing steps. Minimization of ambient light and use of an ozone detector and converter offers the simplest solution to this problem. Keeping wash times to <1 min through small batch sizes (e.g., four slides) is optimal. In the subsequent cleanup of labeled genomic DNA, Cy3- and Cy5-labeled tubes must be kept separate to avoid signal contamination.

Scanning

With properly maintained equipment, scanning is typically straightforward; however, subsequent data analysis can present a challenge to the inexperienced user. Microarrays are scanned at 3 microns using the Agilent DNA High-Resolution Microarray Scanner (G2565CA). Improper array alignment interferes with accurate data interpretation. This will be evident by checking the "four corners" in the quality control report. This is not a common problem with standard arrays, but can occur more frequently with custom arrays. Improper alignment may require manual grid alignment, which can take $\sim$10 min per sample. CGH Analytics software includes measures for quality control to ensure the best possible data. Although absolute threshold may vary from one laboratory to another, when the standard deviation (SD) of the $\log_2$ ratio is >0.25, the chance of missing a true aberration rises to an unacceptable level, especially in a clinical laboratory setting. A flowchart for troubleshooting aCGH assay steps is shown in Figure 8.12.2, and for QC/analysis in Figure 8.12.3.

Anticipated Results

Measures of quality control are the first stage of CGH data analysis. QC procedures involve similar components regardless of array manufacturer, but the following comments refer specifically to Agilent arrays. Agilent's Feature Extraction software generates an automatic QC report, and important features of the report are discussed here. First, manually check grid alignment on the QC report. Proper alignment is indicated by gridmarks that line up exactly over the array spots, and should be checked at the four corners to assure proper alignment; the words "Grid Normal" will appear below the grid image (Fig. 8.12.4). If an array is not properly aligned, manual readjustment of the grid must be performed and the command "Evaluate Grid" will appear. Further analysis is possible after manual grid realignment. Note that custom oligo arrays may require more manual grid realignment than standard arrays.

Next on the QC Report, check the derivative of log ratio (DLR spread), which assesses

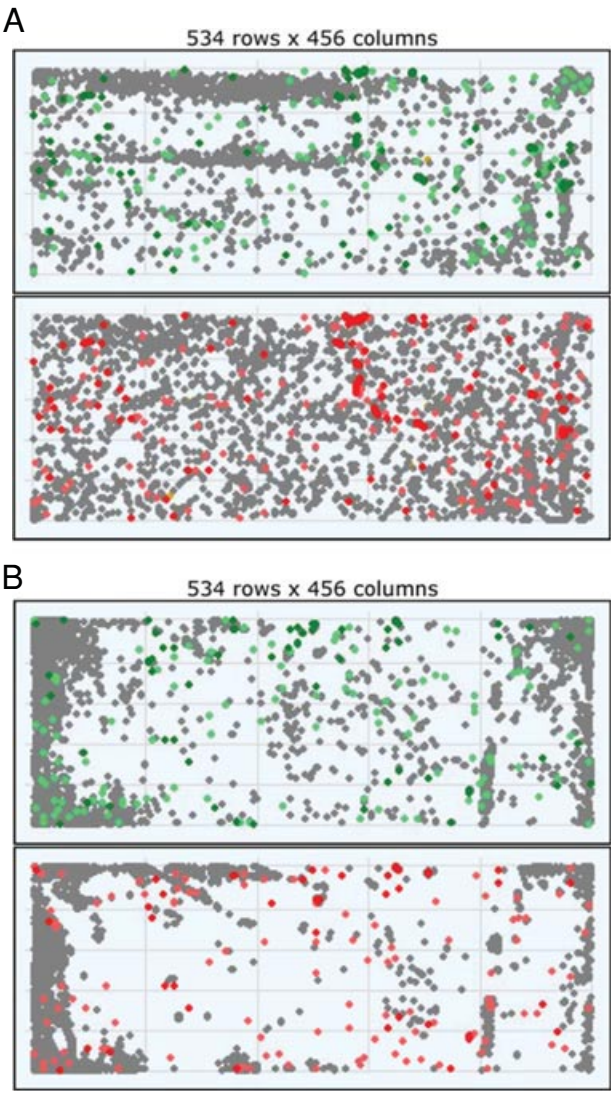

Figure 8.12.5 Spatial distribution of all outliers on the array (534 rows × 456 columns): (**A**) evenly distributed, (**B**) unevenly distributed.

the variability in signal intensity across the array. DLR spread should be <0.25 to ensure accurate and reliable detection of genomic imbalance events. Higher DLR spread lowers the sensitivity of the array to detect both gains and losses. If the DLR spread is >0.25, the entire process, or certain steps, may need to be repeated (see Fig. 8.12.3). The QC package graphically displays the location of outliers on the array as a function of individual features and by gene. In each case, such outliers are expected to appear as random on the array (Fig. 8.12.5A). Clustering of outliers in one physical location on the array is suggestive of technical artifact (Fig. 8.12.6B).

Next, check the intensity of "Red and Green Background Corrected Signals" on the QC Report. These intensities are displayed graphically with red and green on the vertical and horizontal axis, respectively. Ideally, the signal intensities should be equivalent, resulting in all data points falling close to a line with a slope of 1 (Fig. 8.12.6A). Clusters of data points that deviate from this line reflect differing signal intensities and thus poor data that could result in false-positive or false-negative results (Fig. 8.12.6B).

After checking these QC parameters, the data is ready to import into CGH Analytics. The main adjustments within CGH Analytics are the Aberration filter and the threshold, which can be adjusted to determine the sensitivity of detecting an aberration. These adjustments must be made in each laboratory

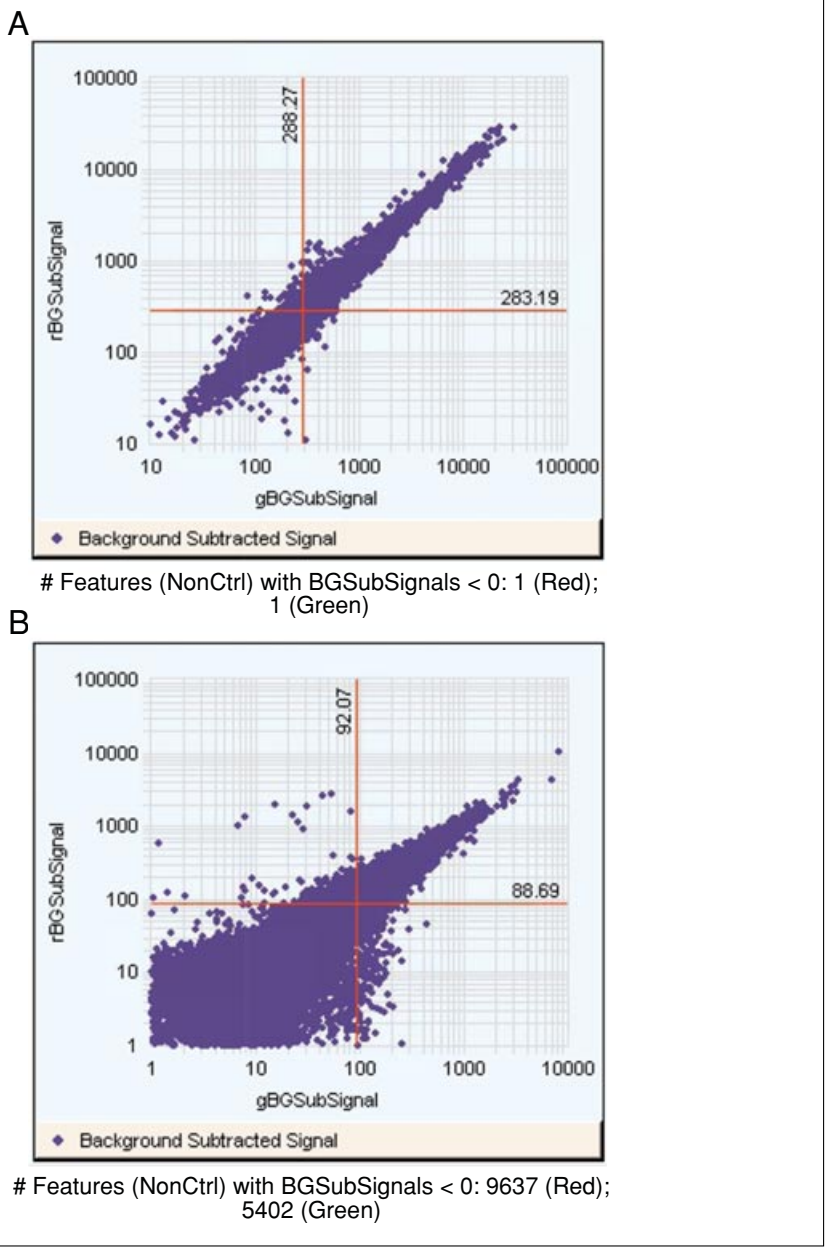

Figure 8.12.6 Plot of signal intensities showing Red and Green Background Corrected Signals (noncontrol inliers): (**A**) consistent results, (**B**) inconsistent results. The information below each panel indicates the number of noncontrol features with values <0 for the following calculation: intensity values for each array feature are subtracted from the median of the log summarized values of all control features on the array. Values <0 indicate high nonspecific hybridization signals.

based on comparing results to known samples (validation). Most laboratories expect five consecutive probes in order to call.

CNVs are abundant in the human genome, and analysis of each sample will identify multiple CNV regions. The process of determining which variants are clinically significant is not standardized and is still open to debate. The number of regions detected will vary depending on coverage density. Most clinical laboratories using oligonucleotide arrays would not consider a region of imbalance to be real if fewer than five consecutive probes show consistent results.

Time Considerations

Times required for each step of oligo aCGH are approximate. Times will vary depending on the capacity of available equipment, array formats (single versus multiple arrays per

slide), and the number of available technologists. The following discussion assumes a single technologist processing samples on an array platform where each sample will be hybridized to a separate slide. Presumably, any clinical laboratory already has a validated DNA extraction protocol, and the oligo aCGH protocol begins without considering the time required to extract genomic DNA. For a single technologist, it is reasonable to process the samples in batches of 24, with genomic DNA digestion, cleanup, and labeling on day 1 (Fig. 8.12.1). Labeled genomic DNA would then be stored overnight at –20°C. On day 2, the samples would be prepared for hybridization that lasts 40 hr at 65°C, thus including all of day 3. Note that the standard capacity of a single hybridization oven is 24 slides. Washing and scanning are performed on day 4. The scanner can hold 48 slides simultaneously.

Literature Cited

Baldwin, E.L., Lee, J.Y., Blake, D.M., Bunke, B.P., Alexander, C.R., Kogan, A.L., Ledbetter, D.H., and Martin, C.L. 2008. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet. Med.* 10:415-429.

Eichler, E.E. 2006. Widening the spectrum of human genetic variation. *Nat. Genet*. 38:911.

Fan, Y.S., Jayakar, P., Zhu, H., Barbouth, D., Sacharow, S., Morales, A., Carver, V., Benke, P., Mundy, P., and Elsas, L.J. 2007. Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization. *Hum. Mutat.* 28:1124-1132.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36:949-951.

Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., Mefford, H., Ying, P., Nickerson, D.A., and Eichler, E.E. 2009. Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* 84:148-161.

Jacquemont, M.L., Sanlaville, D., Redon, R., Raoul, O., Cormier-Daire, V., Lyonnet, S., Amiel, J., Le Merrer, M.,Heron, D., de Blois, M.C., Prieur, M., Vekemans, M., Carter, N.P., Munnich, A., Colleaux, L., and Philippe, A. 2006. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J. Med. Genet.* 43:843-849.

Kearney, H.M., South, S.T., Wolff, D.J., Lamb, A., Hamosh, A., and Rao, K.W. 2011a. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genet. Med.* 13:676-679.

Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., and South, S.T. 2011b. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* 13:680-685.

Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E., Lee, C., Scherer, S.W., Jones, K.W., Shapero, M.H., Huang, J., and Aburatani, H. 2006. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res.* 16:1575-1584.

Lee, C., Iafrate, A.J., and Brothman, A.R. 2007. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat. Genet.* 39:S48-S54.

Manning, M. and Hudgins, L. 2007. Use of array-based technology in the practice of medical genetics. *Genet. Med.* 9:650-653.

Manning, M., and Hudgins, L. 2010. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet. Med.* 12:742-745.

McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., and Altshuler, D. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40:1166-1174.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., Faucett, W.A., Feuk, L., Friedman, J.M., Hamosh, A., Jackson, L., Kaminsky, E.B., Kok, K., Krantz, I.D., Kuhn, R.M., Lee, C., Ostell, J.M., Rosenberg, C., Scherer, S.W., Spinner, N.B., Stavropoulos, D.J., Tepperberg, J.H., Thorland, E.C., Vermeesch, J.R., Waggoner, D.J., Watson, M.S., Martin, C.L., and Ledbetter, D.H. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* 86:749-764.

Ravnan, J.B., Tepperberg, J.H., Papenhausen, P., Lamb, A.N., Hedrick, J., Eash, D., Ledbetter, D.H., and Martin, C.L. 2006. Subtelomere FISH analysis of 11 688 cases: An evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J. Med. Genet.* 43:478-489.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho,

E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nisimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valesia, A., Woodward, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengo, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. 2006. Global variation in copy number in the human genome. *Nature* 444:444-454.

Sagoo, G.S., Butterworth, A.S., Sanderson, S., Shaw-Smith, C., Higgins, J.P., and Burton, H. 2009. Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet. Med.* 11:139-146.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C, Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimaki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.-C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., and Wigler, M. 2007. Strong association of de novo copy number mutations with autism. *Science* 316:445-449.

Shaffer., L.G., Beaudet, A.L., Brothman, A.R., Hirsch, B., Levy, B., Martin, C.L., Mascarello, J.T., and Rao, K.W. 2007. Microarray analysis for constitutional cytogenetic abnormalities. *Genet. Med.* 9:654-662.

Shaikh, T.H. 2007. Oligonucleotide arrays for high-resolution analysis of copy number alteration in mental retardation/multiple congenital anomalies. *Genet. Med.* 9:617-625.

Shaw-Smith, C., Redon, R., Rickman, L., Rio, M., Willatt, L., Fiegler, H., Firth, H., Sanlaville, D., Winter, R., Colleaux, L., Bobrow, M., and Carter, N.P. 2004. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.* 41:241-248.

Shen, Y. and Wu, B.L. 2009. Microarray-based genomic DNA profiling technologies in clinical molecular diagnostics (Review) *Clin. Chem.* 55:659-669.

Shen, Y., Irons, M., Miller, D.T., Cheung, S.W., Lip, V., Sheng, X., Tomaszewicz, K., Shao, H., Fang, H., Tang, H.S., Irons, M., Walsh, C.A., Platt, O., Gusella, J.F., and Wu, B.L. 2007. Development of a focused oligonucleotide-array comparative genomic hybridization chip for clinical diagnosis of genomic imbalance. *Clin. Chem.* 53:2051-2059.

Shen, Y., Dies, K.A., Holm, I.A., Bridgemohan, C., Sobeih, M.M., Caronna, E.B., Miller, K.J., Frazier, J.A., Silverstein, I., Picker, J., Weissman, L., Raffalli, P., Jeste, S., Demmer, L.A., Peters, H.K., Brewster, S.J., Kowalczyk, S.J., Rosen-Sheidley, B., McGowan, C., Duda, A.W. 3rd, Lincoln, S.A., Lowe, K.R., Schonwald, A., Robbins, M., Hisama, F., Wolff, R., Becker, R., Nasir, R., Urion, D.K., Milunsky, J.M., Rappaport, L., Gusella, J.F., Walsh, C.A., Wu, B.L., Miller, D.T.; Autism Consortium Clinical Genetics/DNA Diagnostics Collaboration. 2010. Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics* 125:e727-e735.

Shevell, M., Ashwal, S., Donley, D., Flint, J., Gingold, M., Hirtz, D., Majnemer, A., Noetzel, M., and Sheth, R.D. 2003. Practice parameter: Evaluation of the child with global developmental delay: Report of the Quality Standards Subcommittee of the American Academy of Neurology and The Practice Committee of the Child Neurology Society. *Neurology* 60:367-380.

Stankiewicz, P. and Beaudet, A.L. 2007. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr. Opin. Genet. Dev.* 17:182-192.

Van Vooren, S., Coessens, B., De Moor, B., Moreau, Y., and Vermeesch, J.R. 2007. Array comparative genomic hybridization and computational genome annotation in constitutional cytogenetics: Suggesting candidate genes for novel submicroscopic chromosomal imbalance syndromes. *Genet. Med.* 9:642-649.

Vermeesch, J.R., Fiegler, H., de Leeuw, N., Szuhai, K., Schoumans, J., Ciccone, R., Speleman, F., Rauch, A., Clayton-Smith, J., Van Ravenswaaij, C., Sanlaville, D., Patsalis, P.C., Firth, H., Devriendt, K., and Zuffardi, O. 2007. Guidelines for molecular karyotyping in constitutional genetic diagnosis. *Eur. J. Hum. Genet.* 15:1105-1114.

Vissers, L.E., de Vries, B.B., Osoegawa, K., Janssen, I.M., Feuth, T., Choy, C.O., Straatman, H., Van Der Vliet, W., Huys, E.H., van Rijk, A., Smeets, D., van Ravenswaaij-Arts, C.M., Knoers, N.V., Van Der Burgt, I., de Jong, P.J., Brunner, H.G., van Kessel, A.G., Schoenmakers, E.F., and Veltman, J.A. 2003. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am. J. Hum. Genet.* 73:1261-1270.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T., Platt, O.S., Ruderfer, D.M., Walsh, C.A., Altshuler, D., Chakravarti, A., Tanzi, R.E., Stefansson, K., Santangelo, S.L., Gusella, J.F., Sklar, P., Wu, B.L., and Daly, M.J. 2008. Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N. Engl. J. Med.* 358:667-675.