Clinical Chemistry 53:12
2051–2059 (2007)

Development of a Focused Oligonucleotide-Array Comparative Genomic Hybridization Chip for Clinical Diagnosis of Genomic Imbalance

Yiping Shen,[1,2†] David T. Miller,[1,3,4†] Sau Wai Cheung,[5] Va Lip,[1] Xiaoming Sheng,[1] Keith Tomaszewicz,[1] Hong Shao,[1] Hong Fang,[1] Hung Siv Tang,[1] Mira Irons,[3,4] Christopher A. Walsh,[3,4,6] Orah Platt,[1,4] James F. Gusella,[2,4] and Bai-Lin Wu[1,4*]

Background: **Submicroscopic genomic imbalance underlies well-defined microdeletion and microduplication syndromes and contributes to general developmental disorders such as mental retardation and autism. Array comparative genomic hybridization (CGH) complements routine cytogenetic methods such as karyotyping and fluorescence in situ hybridization (FISH) for the detection of genomic imbalance. Oligonucleotide arrays in particular offer advantages in ease of manufacturing, but standard arrays for single-nucleotide polymorphism genotyping or linkage analysis offer variable coverage in clinically relevant regions. We report the design and validation of a focused oligonucleotide-array CGH assay for clinical laboratory diagnosis of genomic imbalance.**

Methods: **We selected >10 000 60-mer oligonucleotide features from Agilent's eArray probe library to interrogate all subtelomeric and pericentromeric regions and 95 additional clinically relevant regions for a total of 179 loci. Sensitivity and specificity were measured for 105 patient samples, including 51 with known genomic-imbalance events, as detected by bacterial artificial chromosome–based array CGH, FISH, or multiplex ligation-dependent probe amplification.**

Results: **Focused array CGH detected all known regions of genomic imbalance in 51 validation samples with 100% concordance and an excellent signal-to-noise ratio. The mean SD among $\log_2$ ratios of all noncontrol features without copy number alteration was 0.062 (median, 0.055). Clinical testing of another 211 samples from individuals with developmental delay, unexplained mental retardation, dysmorphic features, or multiple congenital anomalies revealed genomic imbalance in 25 samples (11.9%).**

Conclusions: **This focused oligonucleotide-array CGH assay, a flexible, robust method for clinically diagnosing genetic disorders associated with genomic imbalance, offers appreciable advantages over currently available platforms.**

Genomic imbalance causes a variety of human genetic disorders, ranging from imbalance of entire chromosomes, as in Down syndrome, to submicroscopic rearrangements, as in the 22q11 deletion that causes DiGeorge/velocardiofacial syndrome. Genomic imbalance also causes idiopathic mental retardation *(1, 2)* and is detectable in approximately 3%–4% of cases *(3)* by traditional cytogenetic methods, such as karyotype and fluorescence in situ hybridization (FISH)[7] analyses. These traditional cytogenetic methods are labor intensive, especially when multiple genomic regions are interrogated.

[1] Department of Laboratory Medicine, Children's Hospital Boston, Boston, MA.
[2] Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA.
[3] Division of Genetics, Children's Hospital Boston, Boston, MA.
[4] Departments of Medicine, Neurology, Pathology, Pediatrics, Harvard Medical School, Boston, MA.
[5] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.
[6] Howard Hughes Medical Institute, Chevy Chase, MD.
[†] These authors contributed equally to this work.
[*] Address correspondence to this author at: Departments of Laboratory Medicine and Pathology, Children's Hospital and Harvard Medical School, 300 Longwood Ave., Boston, MA 02115. Fax 617-730-0338; e-mail bailin.wu@childrens.harvard.edu.

Received April 20, 2007; accepted September 11, 2007.
Previously published online at DOI: 10.1373/clinchem.2007.090290

[7] Nonstandard abbreviations: FISH, fluorescence in situ hybridization; MLPA, multiplex ligation-dependent probe amplification; CGH, comparative genomic hybridization; BAC, bacterial artificial chromosome; CNV, copy-number variant; SNR, signal-to-noise ratio.

Molecular techniques such as multiplex ligation-dependent probe amplification (MLPA) and real-time PCR are alternatives to multiple FISH assays for evaluating the genomic copy number of multiple targets *(4)*. Microarray-based comparative genomic hybridization (CGH) offers the ability to interrogate many more genomic regions in a single assay. Early CGH arrays were composed of large-insert bacterial artificial chromosome (BAC) clones *(5)*. BAC-based arrays have revolutionized the detection of genomic imbalance in clinical cytogenetic laboratories *(6, 7)* but are challenging to develop, validate, and manufacture. The fact that BAC clones in standard libraries may be inaccurately mapped could lead to diagnostic errors without careful validation *(8)*. Additionally, BAC clone inserts average approximately 150 kb, limiting the resolution of detectable copy-number variants (CNVs) to the size of a BAC insert. Deletion breakpoints that extend beyond the BAC clone cannot be accurately determined. Once validated, BAC arrays are much more efficient than multiplex FISH analysis, but genetic information is constantly changing. Consequently, updates to a BAC-based array require successive rounds of extensive probe validation.

Oligonucleotide-based arrays offer advantages over BAC-based arrays, and many platforms are available. Oligonucleotide arrays designed for genotyping single-nucleotide polymorphisms may not provide uniform coverage at all sites of genomic imbalance *(9, 10)*. Custom oligonucleotide arrays that are based on libraries of validated synthetic probes can interrogate clinically relevant genomic regions without the need for large-insert clone libraries. We describe an array based on Agilent's eArray library, a large collection of 60-mer oligonucleotides specifically selected for robust copy-number analysis *(11)*. This targeted oligonucleotide-based array provides a flexible and adaptable method for CGH to detect genomic copy-number imbalance in the clinical diagnostic laboratory.

Materials and Methods

VALIDATION SAMPLES

DNA was obtained from the material remaining from 105 samples after previous clinical assays had been completed for patients who originally had been referred for genetic testing with BAC-based array CGH, FISH, karyotyping, or MLPA in the DNA Diagnostic Laboratory at Children's Hospital Boston and the Medical Genetics Laboratories at Baylor College of Medicine. Genomic imbalance was previously identified in 51 (49%) of the 105 samples. Samples with positive results from prior testing were assigned to a "validation set" and subjected to oligonucleotide-array CGH analysis in these 2 laboratories with the array platform described below. Laboratory personnel were blinded to prior testing results. The Children's Hospital Boston Institutional Review Board approved this project.

CLINICAL SAMPLES

After assay validation, we performed clinical array CGH testing of 211 consecutively submitted samples from presumably unrelated children. Samples were submitted after referral to specialists in the Divisions of Clinical Genetics and Developmental Medicine, and the Department of Neurology for clinical molecular-diagnostic testing. The referring diagnoses for these patients included developmental delay, mental retardation, dysmorphic features, or multiple congenital anomalies. All samples were compared with a reference sample for standard 2-color array CGH, either a 46,XY male or a 46,XX female sample. Reference DNA was purchased from Promega.

Genomic DNA was extracted from whole blood for all samples with a D50K PureGene DNA-isolation reagent set (Qiagen/Gentra) according to the manufacturer's instructions. All DNA was stored at −20 °C.

CHIP DESIGN

This focused oligonucleotide chip covers 179 clinically relevant regions of genomic imbalance, including all subtelomeric and pericentromeric regions, and 95 regions responsible for well-defined microdeletion/microduplication syndromes, mental retardation, and autism (for a summary of array coverage, see Supplemental Data 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue12). A total of 10 207 region-specific features and 603 quality-control and negative-control features selected from Agilent's eArray library *(11)* are randomly located on the array with an average spatial resolution of <35 kb within the targeted regions. Each subtelomeric region has a minimum coverage of 5 Mb. A subset of 660 features is duplicated on each block as a quality-control measure. Arrays were manufactured with Agilent's SurePrint Inkjet technology. In designing the targeted oligonucleotide-based array, we consulted the Database of Genomic Variants (http://projects.tcag.ca/variation/) to avoid CNVs with no apparent clinical relevance.

CGH

Oligonucleotide-array CGH was performed according to the manufacturer's Oligonucleotide Array–Based CGH for Genomic DNA Analysis protocol (version 3; Agilent Technologies; see Supplemental Data 2 in the online Data Supplement for a summary of the protocol).

Dye-swap verification was performed on all samples with positive findings. For other confirmation assays, we carried out BAC-array CGH and FISH confirmation as described previously *(12)*. MLPA confirmation was performed as described previously *(13)*. The MLPA oligonucleotides for the *CTNS*[8] (cystinosis, nephropathic) gene are as follows: exon 2, GTTTTCACACTGGGCGAAGG

[8] Human genes: *CTNS*, cystinosis, nephropathic; *NPHP1*, nephronophthisis 1 (juvenile); *CARKL*, carbohydrate kinase-like.

GAGGACT and CCTGAGCTCTGCCTCTTCCAGTAA CATTG; exon 6, CCGAGGATACGCTTTCTTGTGATCC and GCAGCAGCGCCATTAGCATCATAAACC; exon 12, CAACCAAGTTTGGACTCGGGGT and CTTCTC CATCGTCTT CGACGTCGTC.

DATA ANALYSIS

Scanned images were quantified with Feature Extraction software (version 9.0; Agilent Technologies). We used the signal-to-noise ratio (SNR) and the normalized $\log_2$ ratio (test:reference) with 40 nonpathologic individual DNA samples to evaluate the quality and variability of each feature/target. The SNR was calculated by dividing the mean signal intensity of each feature by the mean background signal intensity. Features with a mean SNR <4 or an SD of the $\log_2$ ratio >0.1 were considered to have poor signal quality and high variability and were filtered out before further analysis. These thresholds were chosen empirically and are similar to those used in similar studies *(1, 14)*.

We visualized the filtered data further with CGH Analytics software (version 3.4; Agilent Technologies) and evaluated the quality of each test with the quality-control metrics generated with CGH Analytics software. Copy-number aberration was indicated with the Aberration Detection Method 2 algorithm for the data that passed quality-control testing. The Aberration Detection Method 2 algorithm finds intervals of varying size with a consistent, appreciably low, or high $\log_2$ ratio. An aberration filter was set to indicate regions with at least 3 targets showing the same direction in copy-number change. The mean $\log_2$ ratio of each region of potential imbalance was calculated and compared with the SD for the whole dataset. A copy-number gain was called if the mean $\log_2$ ratio was greater than twice the SD of the whole dataset, and a loss was called if the mean was less than −2 SDs. These thresholds were chosen empirically and are similar to those used in other such studies *(15, 16)*. Cutoff values for genomic imbalance can be adjusted and set accordingly with the threshold function of CGH Analytics software, especially when a potential mosaic scenario is encountered. Variants not known to be pathogenic were compared with the Database of Genomic Variants (http://projects.tcag.ca/variation/) to facilitate interpretation.

Results

EVALUATION OF TARGET LOCI AND OVERALL CHIP PERFORMANCE

Forty sex-matched samples from healthy individuals were analyzed on the array, including 2 self-self hybridizations, to evaluate each feature on the array. Signal quality, $\log_2$ ratio variability, mean SNR, and SD of the $\log_2$ ratio were calculated for each noncontrol feature. The mean and median signal intensities of all the noncontrol features were 251 and 178, respectively. The mean and median values of the mean SNRs from all noncontrol features were 9.05 and 6.39, respectively. More than 91% of the target features had a mean SNR >4. Mean $\log_2$ ratios were distributed symmetrically around the zero value. Only a small fraction of features (52 of 10 025, 0.52%) exhibited mean values >0.1 or <−0.1. These features were excluded from the dataset before further analysis. The mean SD of the $\log_2$ ratio of all noncontrol features was 0.062 (median SD, 0.055).

We demonstrated a $\log_2$ ratio SD of >0.1 for 714 features (7.12%); we categorized these features as nonideal targets and excluded them from further analysis. More than 90% of the features passed the feature level filter criteria: an absolute mean $\log_2$ ratio <0.1, a $\log_2$ ratio SD <0.1, and a mean SNR >4. After excluding nonideal features, the dataset quality improved dramatically. For example, the SD of the $\log_2$ ratio dropped from >0.06 to <0.03. Features with a large $\log_2$ SD largely overlapped those with a low SNR, further validating this filtering approach. Because the nonideal targets are approximately evenly distributed across the target regions, the overall resolution of the chip is not appreciably affected.

Several key variables were used to evaluate chip quality and to describe the quality of the dataset as a whole. These variables included probe-to-probe $\log_2$ ratio noise (DLRSpread), the median signal intensity of both channels, background noise for both channels, and SNR. The following cutoffs were used to pass our quality-control testing: DLRSpread ≤0.25, median signal intensity ≥50, background noise ≤10, and SNR ≥15. None of the samples failed quality-control testing because of poor chip quality or problems with hybridization. Two samples failed testing because of DNA impurities; both samples passed quality-control tests after we repurified the DNA.

CHIP VALIDATION WITH BLINDED SAMPLES

We next blindly tested 65 samples for further chip validation. Genomic imbalance had previously been detected in 51 of the 65 samples by BAC-array CGH, FISH, karyotyping, or MLPA, or by some combination of these analyses, and these 51 samples served as positive controls for validation. The remaining 14 samples had previously been tested by targeted BAC-array CGH with nonpathologic results and thus served as negative controls for validation. All samples were traceable to the technologist who performed the hybridizations.

Of the 51 validation samples with a previously detected genomic imbalance (see Supplemental Data 3 in the online Data Supplement), the samples from 2 cases demonstrated aneuploidy for an entire chromosome, 3 cases involved unbalanced chromosomal rearrangements, 13 cases had subtelomeric deletions/duplications, 17 cases had interstitial deletions/duplications, and 15 cases were associated with known segmental aneuploidy regions, including Angelman/Prader–Willi syndrome (4 cases), atypical Angelman syndrome on 22q13.3 (1 case), an autism phenotype associated with duplication of 15q11-q13 (1 case), a velocardiofacial/DiGeorge syndrome re-

gion (5 cases with deletions and 2 with duplication), and Williams–Beuren syndrome (2 cases).

Across all samples, results from oligonucleotide-array CGH were consistent with the results obtained with the 4 prior methods, but oligonucleotide-based array CGH provided the most precise breakpoint boundaries. Fig. 1A shows a genomic-imbalance event (2q37.3 deletion) identified by oligonucleotide-based array CGH, and Fig. 1B shows the FISH confirmation of the 2q37.3 deletion.

Oligonucleotide-array CGH detected no appreciable imbalance events in any of the 14 negative controls, with the exception of several reported CNVs. The dye-swap scheme essentially eliminated false-positive results.

To further evaluate the confidence of each imbalance call by oligonucleotide-array CGH analysis, we calculated the mean $\log_2$ ratios for each detected imbalance region and compared them with the SD for the whole dataset. The value of the mean $\log_2$ ratio/SD indicates the separability of each imbalance event from the background noise of the whole dataset. For the majority of deletion events detected, the value was less than -3.2, whereas the value was >2.6 for the majority of gain events detected.

Genomic imbalance detected in clinical samples

We used focused oligonucleotide-array CGH to test 211 clinical samples that had been ascertained to have come from individuals with developmental delay, unexplained mental retardation, dysmorphic features, or multiple congenital anomalies. In this cohort, the detection rate for genomic imbalance was approximately 11.9% (25 of 211 samples). All abnormal findings were first verified with a dye-swap array CGH and then independently confirmed by either FISH or MLPA. All the genomic-imbalance events, including CNVs with unknown significance, were divided into 3 categories, as is described below. The imbalance events associated with known disorders or likely to cause disease are listed in Table 1.

Genomic imbalance associated with well-defined disorders

Table 1A lists 12 samples with 10 genomic imbalance events associated with known genetic disorders. In this group, 2 of the samples revealed a complex pattern involving both gain and loss on 2 different chromosomes (case 1) or on the same chromosome (case 2), 2 samples had a well-defined microdeletion syndrome (cases 3 and 4), 1 sample had a subtelomeric deletion (case 5), and 2 samples had whole-chromosome aneuploidy (cases 6 and 7). One dominantly inherited disorder could be diagnosed by the detection of haploinsufficiency for the relevant gene (cases 8 and 9), and 3 cases involved carriers of a recessive allele: deletion at the *NPHP1* [nephronophthisis 1 (juvenile); cases 10 and 11] and *CTNS* loci (case 12). The last 3 cases featured deletions of <100 kb, each of which covered defined disease genes.

Oligonucleotide-array CGH was able to detect cryptic rearrangements, submicroscopic alterations, and even single-gene deletions. The smallest imbalance event detected in this study was a heterozygous genomic deletion of 3 consecutive probes covering a minimal 23-kb interval. Fig. 2A shows the targeted array CGH data, and deletion of the 3 targets was confirmed by dye-swap hybridization (green in the forward hybridization and red in the reverse hybridization). In this case, we repeated the CGH analysis with Agilent's 244K whole-genome oligonucleotide array and confirmed the deletion, which includes the entire *CARKL* (carbohydrate kinase-like) gene and part of *CTNS* (Fig. 2B). The partial deletion of *CTNS* was independently confirmed by MLPA analysis. Fig. 2C shows a 1-copy deletion for *CTNS* exons 2 and 6 and a typical dosage for *CTNS* exon 12. To further characterize the deletion, we amplified the deletion junction by the PCR and confirmed by sequencing (Fig. 2D) that the deletion detected by array CGH is the common "European" deletion associated with cystinosis *(17)*. This case further demonstrated the excellent resolution and sensitivity of the custom oligonucleotide-array CGH method.

Genomic imbalance probably causing a disorder

Table 1B lists 5 samples with interstitial deletions/duplications that may be clinically relevant. The list includes 3 interesting cases: a 3.3 Mb duplication at 17p11.2, which is similar to that of a recently reported 17p11.2-duplication syndrome *(18)*, and 2 cases of a 546 kb de novo deletion at 16p11.2, which is within a region of frequently observed cytogenetic polymorphism but is not observed in the CNV database. Neither individual with the de novo 546-kb 16p11.2 deletion had specific dysmorphic features, but the absence of this deletion in the parents suggests that the deletion is responsible for the phenotype of developmental delay.

Genomic CNVs with unknown significance

Nine patients had unreported CNVs with relatively small genomic deletions of between 50 and 200 kb (data not shown). The clinical significance of these imbalance events is unclear. Although this custom oligonucleotide array was designed to avoid CNVs, many new CNV loci have been reported since the design of the array (http://projects.tcag.ca/variation/).

Discussion

Array CGH is a valuable clinical diagnostic assay for patients with mental retardation and other genetic conditions. Although high-resolution whole-genome oligonucleotide microarrays are commercially available for research, targeted array CGH offers several advantages in a clinical diagnostic laboratory *(8)*. We chose genomic regions with well-documented clinical relevance, analogous to those of the currently accepted BAC-based arrays designed by the leading array CGH laboratories *(19, 20)*.

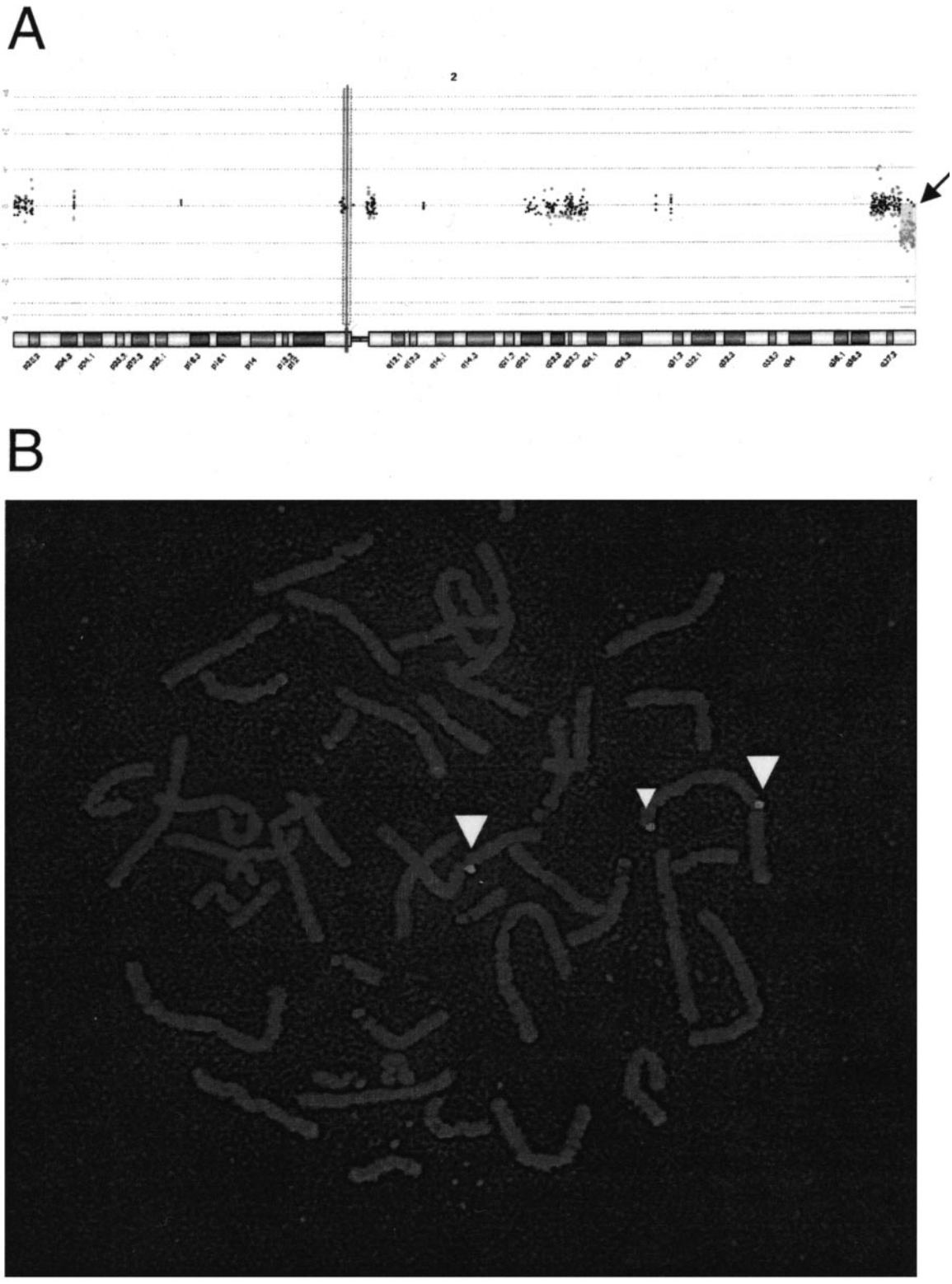

Fig. 1. Genomic-imbalance event (2q37.3 deletion) identified by oligonucleotide-based array CGH and FISH confirmation of the deletion.

(*A*), chromosome view of array CGH results showing a terminal deletion on 2q37.3. *Arrowhead* points to the targets with a downward-shifted $\log_2$ ratio. (*B*), 2q37.3 deletion confirmed by metaphase FISH. The *small arrowhead* indicates the signal of the region-specific BAC clone RP11-367H1, and the *larger arrowheads* indicate the signals of the chromosome 2 centromere–specific BAC clone.

Table 1. Genomic imbalance identified in clinical samples by oligonucleotide-array CGH and confirmed with alternative methods.

Genomic imbalance detected by focused oligonucleotide-array CGH	Cytogenetics (FISH/karyotype), MLPA, whole-genome CGH/targeted PCR	Confirmation
(A) Associated with disorders		
13qter gain, 20.7 Mb; 18qter loss, 5.6 Mb (partial trisomy 13q and partial monosomy 18q)	46,XX, add(18)(q2?1.3) ish der(18)t(13;18)(D13S327+, 18qtel11–)	Consistent
18pter-p11.21 loss, 13.4 Mb; 18p11.21 gain, 1.3 Mb (partial monosomy 18p and partial trisomy 18p)	46,XX ish 18pter(D18S552 × 1), 18p11.21 (RP11-720L3 × 3)	Consistent
17p11.2 loss, 3.6 Mb (Smith–Magenis syndrome)	46,XX, ish del(17)(p11.2 p11.2)	Consistent
1p36.21 loss, 1.8 Mb (1p36 deletion syndrome)	46,XY, ish del(1)(p36.2)	Consistent
4q35.2 loss, 1.1 Mb (autism spectrum disorder)	46,XY, ish del(4)(qter–)	Consistent
X gain (aneusomy X)	47,XXY	Consistent
Y gain (aneusomy Y)	47,XYY	Consistent
Yp11.2 loss, 2.7 Mb: 2 cases	ish Yp11.2 (RP11-115H13 × 0)	Consistent
2q13 loss, 100 kb (*NPHP1* deletion): 2 cases	Confirmed by whole-genome array CGH	Consistent
17p13 loss, 23 kb (*CARKL* and *CTNS* deletion; familial)	Confirmed by whole-genome array CGH, MLPA, parental array CGH, and PCR flanking the deletion	Consistent
(B) Likely clinically relevant		
17p11.2 gain, 3.3 Mb	46,XY, nuc ish 17p11.2 (RP11-363P3 × 3)	Consistent
16p11.2 loss, 546 kb (de novo): 2 cases	46,XY, ish del(16)(p11.2 p11.2)	Consistent
15q13.3 gain, 1.5 Mb	nuc ish 15q13.3 (RP11-303I13 × 3)	Consistent
5q22.1–q23.1 loss, 8.5 Mb	Confirmed by whole-genome array CGH	Consistent

Genomic imbalance identified on the targeted oligonucleotide array can be verified with existing FISH or MLPA probes, whereas secondary methods are not readily available for a whole-genome array.

Uniformly manufactured arrays facilitate reproducible results in the clinical laboratory. On BAC-based arrays, each BAC clone may include both unique and repetitive sequences. Each BAC must be propagated in culture, introducing possible variability in BAC inserts or DNA contamination between batches and potentially affecting the reproducibility and consistency of the manufactured arrays. In contrast, 60-mer oligonucleotide probes are synthesized robotically in situ and have a fixed GC content and melting temperature that facilitate uniform hybridization. Although all targets in Agilent's eArray library were designed in silico with narrow melting temperature ranges, the actual performance of each target needed to be evaluated in practice. Some features exhibited large variability, possibly due to properties of these unique sequences. Genomic CNVs unquestionably contribute to the variability in the performance of some probes.

The sensitivity and specificity of the oligonucleotide-array platform were excellent because of the high SNRs and low SDs. We observed the SDs of all targets to be consistently <0.1—with the majority being <0.08—with this oligonucleotide-based array. Given the nice separability of imbalance events above the baseline, we can identify genomic imbalance events with a high level of confidence. In addition to the 100% concordance between the oligonucleotide-array CGH results and the results generated by other methods for both positive and negative samples, we were able to identify smaller imbalance events (unreported CNVs; data not shown) that were not detected with other methods.

Oligonucleotide platforms can quickly and easily accommodate changes in genomic coverage. Manufacturing costs are not prohibitive and the list of available probe sequences extends the length of the genome. Thus, updates to oligonucleotide-based arrays can be accomplished more quickly and with less postproduction validation than BAC-based arrays, which require new BAC clones to be individually validated and DNA to be prepared from each clone before chips can be manufactured.

This focused oligonucleotide-based array CGH platform detected all genomic imbalance events in the 65 validation samples, with 100% concordance with BAC-based array CGH, FISH/karyotyping, or MLPA. Coverage of clinically relevant loci (see Supplemental Data 1 in the online Data Supplement) is equivalent to other BAC-based targeted array CGH platforms *(19, 20)*. The enhanced sensitivity and specificity of oligonucleotide-array CGH compared with other methods are attributable to better resolution and the custom design, respectively.

Despite the superior technical performance of oligonucleotide-based array CGH, clinical interpretation can be challenging. We encountered several scenarios: (*a*) known genomic-imbalance events with well-documented clinical relevance (Table 1A) that were verified with dye-swapping and either FISH or MLPA (on smaller regions); (*b*) known genomic-imbalance events with possible clinical relevance (Table 1B) that were verified by dye-swapping and the 244K whole-genome array for further characterization of the genomic-imbalance events; and (*c*) novel

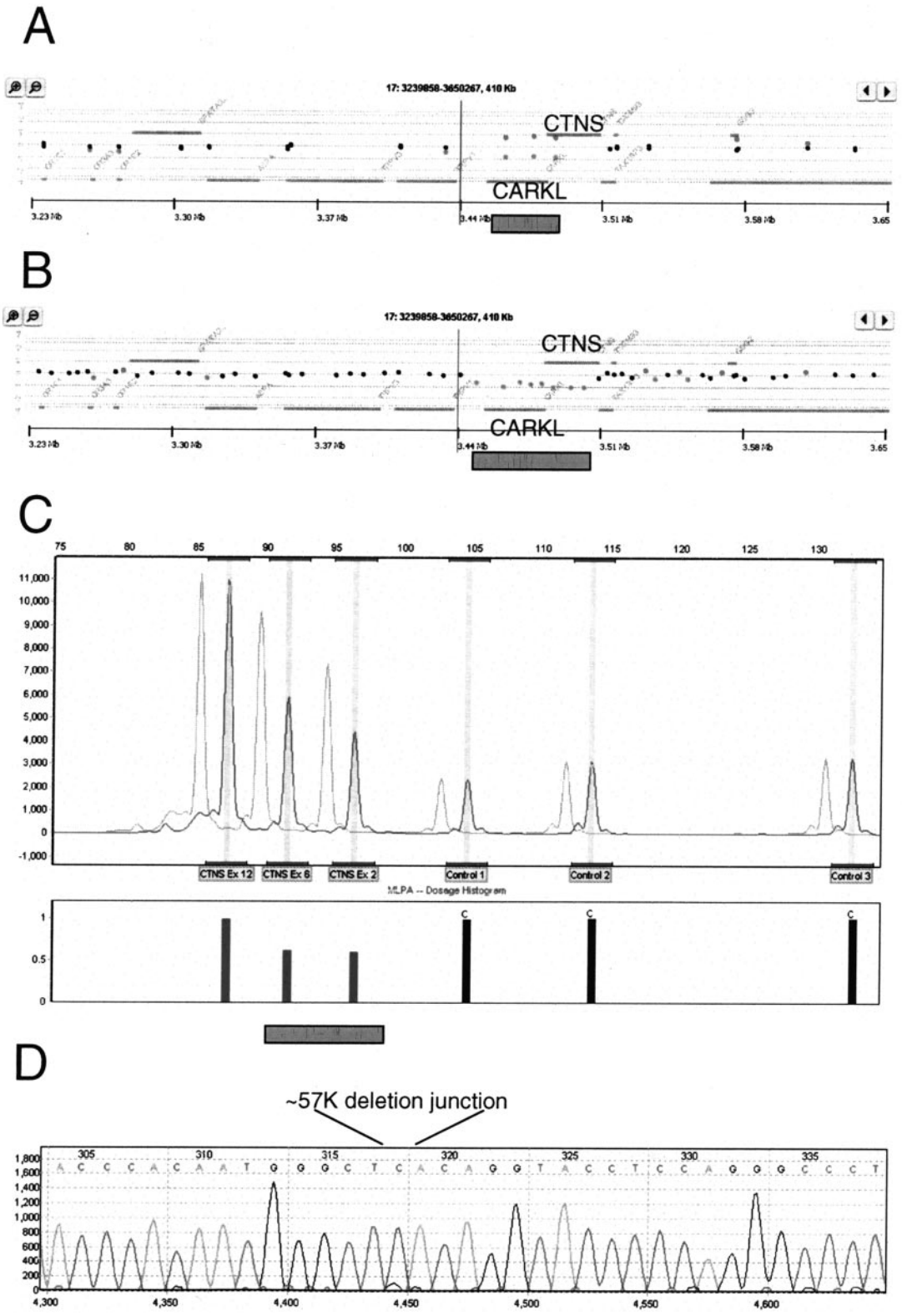

Fig. 2. Oligonucleotide-array CGH detection of a heterozygous genomic deletion of 3 consecutive probes covering a minimal 23-kb interval.

(*A*), the deletion of 3 consecutive targets (*underlined* with *rectangular bar*) was confirmed by dye-swap array CGH. Note the symmetrical opposite ratios between forward-labeled (downward-shift) and reverse-labeled (upward-shift) array CGH. (*B*), 244K whole-genome oligonucleotide-array CGH confirmation of the deletion. The *rectangular bar underlines* the deleted region. Note the entire *CARKL* gene deletion and the partial *CTNS* gene deletion (forward labeling only). (*C*), MLPA confirmation of the deletion. The *trace* in *lighter gray* is a sample from a healthy control individual; the *trace* in *darker gray* is the patient sample. The dosage of *CTNS* exons 2 and 6 appears reduced by half (*underlined* with the *rectangular bar*), whereas the dosage of *CTNS* exon 12 is unchanged. (*D*), a segment of the sequencing trace around the deletion junction of the common European deletion. One base (C) overlaps.

genomic imbalance with uncertain clinical relevance, cases of which were verified by dye-swapping, 244K whole-genome array analysis, and examination of parental samples to determine whether the imbalance was a de novo event.

Adding to the complexity of interpretation is that an apparent familial variant can actually be pathogenic. For example, a deletion variant that appears in a healthy parent and an affected child could be acting as a recessive allele in the parent. Likewise, a de novo region of genomic imbalance could be a previously undescribed nonpathogenic CNV or a case of nonpaternity. Dozens of CNVs are present in every human genome *(21, 22)*.

Although this custom oligonucleotide array has an average resolution of approximately 35 kb within the targeted regions, the fact that array formats continue to evolve provides ample opportunities to increase resolution through the selection of additional probes from available oligonucleotide databases or the design of custom oligonucleotide arrays. The flexibility of the oligonucleotide-array design allows individual laboratories to interrogate different genomic regions at the most biologically relevant level of resolution. For example, arrays can be designed to detect single-exon deletions and duplications in a gene of interest.

Resolution beyond that provided by this custom oligonucleotide array may not be necessary or desirable in all cases, however. A majority of the CNVs detected on high-resolution oligonucleotide platforms are <150 kb *(23, 24)*, and the detection of these CNVs complicates the clinical interpretation. CNV databases are highly valuable research tools, but they have limitations in the clinical setting. CNVs in databases are often determined by a variety of methods that may not provide accurate breakpoints. These databases are often not annotated with clinical information and thus do not indicate genotype-phenotype associations.

As the knowledge of genotype-phenotype associations in disorders of genomic imbalance accumulates, however, precise and accurate determination of the genomic sequences involved is critically important in the clinical diagnostic setting. FISH and BAC clones can detect imbalance equal to or larger than the size of clone insert, typically approximately 150 kb; thus, smaller deletions or duplications can be missed on a BAC-based array. Even genome-wide tiling BAC arrays do not have a resolution of less than approximately 46 kb *(25, 26)*.

Oligonucleotide-based arrays provide an unprecedented degree of resolution for determining both the size and breakpoints of regions of genomic imbalance. Our current design was able to precisely define a heterozygous deletion of 23 kb, a level of resolution not possible on a BAC-based array. This improvement in resolution could be important in the clinical interpretation of cases in which the inclusion or exclusion of certain genes may predict clinical features of the condition.

In conclusion, the custom oligonucleotide array that we have described offers a sensitive and specific clinical assay for detecting genomic imbalance. The platform can easily accommodate new information to keep pace with research advances. This methodology offers a proof of principle that well-designed oligonucleotide-based arrays offer substantial technical advantages over currently available platforms and at a comparable cost. We anticipate widespread adoption of this methodology in the clinical diagnostic setting.

Grant/funding support: This work was supported in part by National Institutes of Health (NIH) Grant NS24279 (to J.F.G.) and National Institute of Neurological Disorders and Stroke Grant R37 NS35129 (to C.A.W.). Y.S. is the recipient of a Young Investigator Award from the Children's Tumor Foundation. B.-L.W. received support from the NIH-CETT Program and the MDA Foundation for the development of cost-effective assays for genetic diagnostics.
Financial disclosures: None declared.
Acknowledgments: We thank the referring physicians and genetic counselors at the Children's Hospital Boston for their cooperation, Dr. Arthur L. Beaudet of Baylor College of Medicine for thoughtful discussion of the study, and Dr. Arthur R. Brothman and Emily Aston of the University of Utah for FISH confirmation of clinical cases.

References

1. Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, et al. 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 2004;41:241–8.
2. Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). J Med Genet 2005;42:699–705.
3. Shevell M, Ashwal S, Donley D, Flint J, Gingold M, Hirtz D, et al. 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 2003;60:367–80.
4. de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, et al. Diagnostic genome profiling in mental retardation. Am J Hum Genet 2005;77:606–16.
5. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20:207–11.
6. Bejjani BA, Shaffer LG. Application of array-based comparative genomic hybridization to clinical diagnostics. J Mol Diagn 2006;8:528–33.
7. Bejjani BA, Theisen AP, Ballif BC, Shaffer LG. Array-based comparative genomic hybridization in clinical diagnosis. Expert Rev Mol Diagn 2005;5:421–9.
8. Bejjani BA, Saleki R, Ballif BC, Rorem EA, Sundin K, Theisen A, 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.

9. Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L, et al. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 2006;79:500–13.

10. Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res 2006;34:445–50.

11. Agilent Technologies. eArray 4.5. http://earray.chem.agilent.com/earray/login.do (accessed August 2006).

12. Cheung SW, Shaw CA, Yu W, Li J, Ou Z, Patel A, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. Genet Med 2005;7:422–32.

13. Shao H, Lip V, Wu BL. Effectiveness of multiplex ligation-dependent probe amplification assay used for detecting deletion of Prader-Willi syndrome. Beijing Da Xue Xue Bao 2005;37:64–7.

14. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, et al. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 2007;80:91–104.

15. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, et al. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 2006;79:275–90.

16. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 2006;38:1038–42.

17. Forestier L, Jean G, Attard M, Cherqui S, Lewis C, van't Hoff W, et al. Molecular characterization of CTNS deletions in nephropathic cystinosis: development of a PCR-based detection assay. Am J Hum Genet 1999;65:353–9.

18. Potocki L, Bi W, Treadwell-Deering D, Carvalho CM, Eifert A, Friedman EM, et al. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. Am J Hum Genet 2007;80:633–49.

19. Baylor College of Medicine. Chromosomal Microarray Analysis. http://www.bcm.edu/cma (accessed August 2006).

20. Signature Genomic Laboratories. Home page. http://www.signaturegenomics.com (accessed August 2006).

21. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet 2004;36:949–51.

22. Eichler EE. Widening the spectrum of human genetic variation. Nat Genet 2006;38:9–11.

23. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006;444:444–54.

24. Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J, et al. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. Genome Res 2006;16:1575–84.

25. Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, et al. A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 2001;11:483–96.

26. Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, et al. A set of BAC clones spanning the human genome. Nucleic Acids Res 2004;32:3651–60.