

# Targeted comparative genomic hybridization array for the detection of single- and multiexon gene deletions and duplications

Marwan K. Tayeh, PhD, Ephrem L. H. Chin, BSc, CLSp(MB), Vanessa R. Miller, MS, CGC, Lora J. H. Bean, PhD, FACMG, Bradford Coffee, PhD, FACMG, and Madhuri Hegde, PhD, FACMG

**Purpose:** To develop a high resolution microarray based method to detect single- and multiexons gene deletions and duplications. **Methods:** We have developed a high-resolution comparative genomic hybridization array to detect single- and multiexon deletions and duplications in a large set of genes on a single microarray, using the NimbleGen 385K array with an exon-centric design. **Results:** We have successfully developed, validated, and implemented a targeted gene comparative genomic hybridization arrays for detecting single- and multiexon deletions and duplication in autosomal and X-linked disease-associated genes. **Conclusion:** The comparative genomic hybridization arrays can be adopted readily by clinical molecular diagnostic laboratories as a rapid, cost-effective, highly sensitive, and accurate approach for the detection of single- and multiexon deletions or duplications, particularly in cases where direct sequencing fails to identify a mutation. *Genet Med* 2009;11(4):232–240.

**Key Words:** CGH array, single-exon and multiexon deletions/duplications, comprehensive molecular testing, two-step approach

Clinical molecular testing is getting more comprehensive, yet more complicated, with the burgeoning number of known disease-causing genes. Most clinical laboratories develop assays for mutation detection in disease-associated genes based on the reported mutation spectrum (Fig. 1). The mutation spectrum associated with genetic disease ranges from point mutations and small deletions and insertions, which can be detected by direct DNA sequencing, to single- and multiexon deletions and duplications. The gene mutation spectrum and comprehensive genetic testing vary between genes; for instance, >65% of Duchenne and Becker muscular dystrophy cases are due to deletions or duplications of single or multiple exons of the dystrophin gene.<sup>1–3</sup> In contrast, >90% of cystic fibrosis cases are due to single base-pair substitutions, microinsertions, or microdeletions in the *CFTR* gene<sup>4,5</sup>; ~80% of the familial adenomatous polyposis cases are due to point mutations, whereas, only ~20% are due to *APC* gene deletions or duplications<sup>6</sup>; and around 75% of the Sotos syndrome cases are due to point mutations in the *NDS1* gene, while 25% of cases are due to deletions and duplications.<sup>7</sup>

The current available techniques for detecting single- and multiexon deletions and duplications in a single gene include multiplex polymerase chain reaction (PCR), quantitative PCR, Southern blotting,<sup>8</sup> multiplex ligation-dependent probe amplification (MLPA),<sup>9</sup> detection of virtually all mutations-SSCP (*DOVAM-S*),<sup>10,11</sup> and single condition amplification/internal primer sequencing (SCAIP).<sup>12</sup>

Clinical assays based on multiplex PCR and quantitative real-time PCR have been used traditionally to detect intragenic deletions and duplications for a small number of genes. Importantly, multiplex PCR can detect deletions but cannot detect duplications accurately or examine multiple genes simultaneously. Quantitative real-time PCR methods are more accurate than multiplex PCR, but detection of intragenic duplications is still a challenge.<sup>13,14</sup> In addition, PCR efficiency and PCR product saturation may adversely impact the sensitivity and accuracy of real-time fluorescent PCR, whereas spectral overlap limits the use of real-time fluorescent PCR in multiplex assays.<sup>14,15</sup> Both multiplex PCR and quantitative real-time PCR, which use a single-probe hybridization that usually targets a common deleted region within a specific gene, can be hindered by single nucleotide polymorphisms, and the boundaries of deletions and duplications cannot be determined.

Southern blotting can detect relatively large deletions and duplications within a specific gene, however, Southern blotting is labor-intensive, time- and DNA-consuming, requires hazardous reagents, and is not designed for simultaneous examination of multiple genes.<sup>16</sup> MLPA, on the other hand, is a high-throughput technique in which multiple segments can be screened simultaneously to detect intragenic deletions and duplications; however, MLPA assay development is laborious, and occasionally gives false-positive and false-negative results, especially with diploid genes. MLPA false-negative and false-positive findings are caused largely by the presence of single nucleotide polymorphisms in the targeted regions. Single-exon dosage changes (duplications in particular) detected by MLPA require further confirmation by a second method (such as multiplex PCR or sequencing), which adds to both the complexity and cost of MLPA assays.<sup>14</sup>

*DOVAM-S* is a high-throughput multiplexed scanning assay that uses single-strand conformational polymorphism to detect virtually all point mutations and small deletions and duplications.<sup>10,11</sup> Sequencing of all variant bands is required to distinguish polymorphisms from disease-causing mutations, which does decrease the efficiency of this assay. In contrast, SCAIP was developed as a rapid, accurate, and economical direct sequencing technique for exons and flanking intronic sequences of large, multiexon genes.<sup>12</sup> Despite the improved detection of deletions and point mutations, SCAIP does not detect intragenic duplications. SCAIP can be labor-intensive, as it requires PCR amplification of all exons individually as the first step.

From the Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322.

Madhuri Hegde, Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Whitehead Biomedical Research Building, Atlanta, GA 30322. E-mail: mhedge@emory.edu.

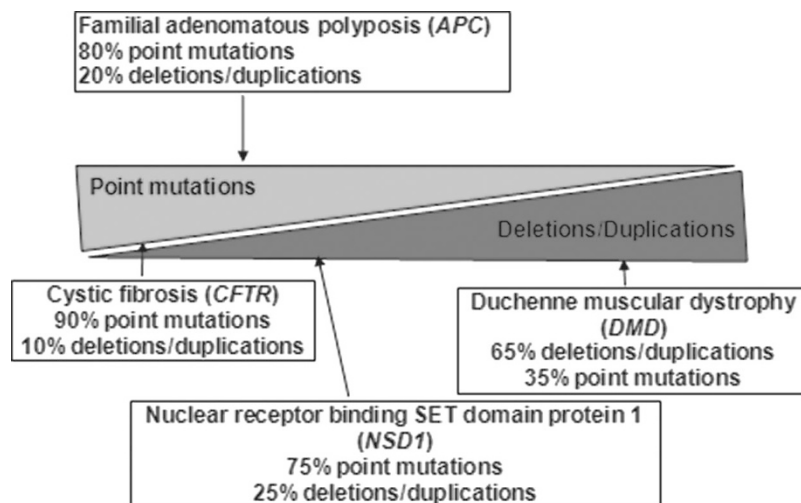
Disclosure: The authors declare no conflict of interest.

The first two authors contributed equally to this work.

Submitted for publication August 18, 2008.

Accepted for publication November 18, 2008.

DOI: 10.1097/GIM.0b013e318195e191



**Fig. 1.** Reported mutation spectrums for various genes. Mutation spectrum varies from point mutations and small deletions and duplications to intragenic single- and multiexon deletion and duplication mutations. Approximately 90%, 80%, and 75% (respectively) of the reported mutations in the cystic fibrosis (*CFTR*), familial adenomatous polyposis (*APC*), and nuclear receptor binding SET domain (*NSD1*) genes are point mutations, whereas ~65% of the reported mutations in the dystrophin gene are deletion and duplication mutations.

Recently, comparative genomic hybridization (CGH) using oligonucleotide arrays has been implemented in cytogenetic and molecular diagnostic laboratories as a robust, rapid, sensitive, and relatively inexpensive assay for detecting known and new microdeletion or microduplication syndromes,<sup>17</sup> and targeted gene deletions.<sup>18–20</sup> Recently, we developed a two-step high-resolution approach using CGH and resequencing microarrays to comprehensively detect mutations in the dystrophin gene.<sup>21</sup>

Here we describe the development, validation, and implementation of a targeted, high-density oligonucleotide CGH microarray in a clinical laboratory. Our targeted gene CGH array permits a high-resolution analysis on a common platform of 71 genes involved mainly in lysosomal storage diseases and metabolic disorders with equal sensitivity and accuracy (Table 1). The array was developed to detect deletions and duplications for all genes sequenced by the Emory Genetics Lab (EGL) to meet our goal of offering truly comprehensive molecular testing. The sensitivity and specificity of our targeted CGH array allows accurate detection of single- and multiexon deletions and duplications in autosomal and X-linked genes in both males and females. Our data demonstrate the advantage of using a high-resolution targeted gene CGH array to complement direct gene sequence analysis, particularly when direct sequencing fails to identify two mutations in autosomal recessive cases with an established clinical or biochemical diagnosis.

## MATERIALS AND METHODS

### CGH array design

The targeted gene high-resolution oligonucleotide CGH array was custom designed on a NimbleGen (Madison, Wisconsin) 385K platform to detect deletions and duplications in 71 genes associated with various genetic disorders. Long oligonucleotides (~45–60 mer) were used to achieve isothermal  $T_m$  across the array, with repeat sequence masking implemented to ensure greater sensitivity and specificity. Our targeted gene CGH array has 389,587 unique sequence probes tiled on the array; the average spacing is 2 bp within coding regions and 25

bp within intronic regions, with repeat sequence masking. Use of intronic oligonucleotide probes allows us to detect dosage changes within the entire genomic region of the gene and determine the approximate breakpoints.

### CGH protocol and analysis

DNA was extracted using the Qiagen Puregene DNA extraction kit (Valencia, CA) according to the manufacturer's recommendation. Patient samples were tested against sex-matched controls. Control DNA from males and females was obtained from Promega (Pittsburgh, PA). DNA samples (2  $\mu$ g) were sonicated to obtain a fragment size between 500–2000 bases, as verified on a 1% agarose gel. We used Klenow enzyme (NEB, Ipswich, MA) and Cy3 or Cy5 9mer wobble primers (Trilink Technologies, Nashua, NH) to label patient and control DNA samples, respectively. Labeled samples were purified by isopropanol precipitation. Thirteen micrograms of labeled DNA from each patient and sex-matched control were combined and desiccated in a vacufuge (GMI, Ramsey, MN), then resuspended in appropriate hybridization buffer mixed with Cy3 and Cy5 control CPK6 50mer oligonucleotides. Arrays were hybridized between 16 and 20 hours at 42°C in a microarray user interface (MAUI) hybridization system (BioMicro Systems, Salt Lake City, UT), then washed according to the manufacturer's recommendation and immediately scanned on a GenePix 4000 scanner (Molecular Devices, Sunnyvale, CA). Next, data were extracted from images using manufacturer-provided software (NimbleScan, NimbleGen Systems). Normalized log (2) ratio data were analyzed using two different analysis programs: (SegMNT or DNA copy) NimbleScan.<sup>22</sup> Both software programs graphically display results in a bar graph, where the y-axis indicates gain or loss of material (1 = gain, 0 = normal, -1 = loss) and the x-axis indicates the position of each feature on the chromosome.

### Assessment of array quality

To assess quality on each array, we used three control resequencing oligonucleotides, which correspond to synthetic se-

**Table 1** A list of genes represented on our targeted comparative genomic hybridization (CGH) array

Metabolic disorder genes	Lysosomal storage genes
Argininosuccinate lyase deficiency ( <i>ASL</i> )	Alpha-mannosidosis ( <i>MAN2B1</i> )
Biotinidase deficiency ( <i>BTD</i> )	Beta-mannosidosis ( <i>MANBA</i> )
Carnitine/acylcarnitine translocase deficiency ( <i>SLC25A20</i> )	Citrullinemia type 1 ( <i>CTLN1</i> )
Carnitine palmitoyltransferase II deficiency ( <i>CPT II</i> )	Fabry disease alpha-galactosidase ( <i>GLA</i> )
Carnitine uptake defect ( <i>OCTN2</i> )	Gaucher disease ( <i>GBA</i> )
Epimerase UDP-galactose-4-epimerase ( <i>GALE</i> )	Glycogen storage disease type II (Pompe disease) ( <i>GAA</i> )
Glucose-6-phosphate dehydrogenase ( <i>G6PD</i> )	Krabbe disease ( <i>GALC</i> )
Galactokinase ( <i>GALK1</i> )	Metachromatic leukodystrophy ( <i>ARSA</i> )
Galactosemia ( <i>GALT</i> )	Mucopolysaccharadosis type I ( <i>IDUA</i> )
Glutaric aciduria type I (GA-1) ( <i>GCDH</i> )	Mucopolysaccharadosis type II ( <i>IDS</i> )
Glutaric aciduria type II (Multiple acyl-CoA dehydrogenase deficiency ( <i>ETFA</i> , <i>ETFB</i> , and <i>ETFDH</i> ))	Mucopolipodosis type II ( <i>GNPTAB</i> )
Homocystinuria ( <i>CBS</i> )	Mucopolysaccharadosis type IIIA ( <i>SGSH</i> )
3-Hydroxy-3-methylglutaryl CoA lyase ( <i>HMGCL</i> )	Mucopolysaccharadosis type IIIB ( <i>NAGLU</i> )
Isovaleric acidemia ( <i>IVD</i> )	Mucopolysaccharadosis type IIIC ( <i>TMEM76</i> )
Long-chain hydroxy acyl-CoA dehydrogenase deficiency (LCHAD) ( <i>HADHA</i> )	Mucopolysaccharadosis type IIID ( <i>GNS</i> )
Maple syrup urine disease—BCKD complex ( <i>DBT</i> , <i>BCKDHA</i> and <i>BCKDHB</i> )	Mucopolysaccharadosis type IVA ( <i>GALNS</i> )
MSUD: Dihydroipoamide dehydrogenase ( <i>E3</i> )	Mucopolysaccharadosis type IVB ( <i>GLB1</i> )
Medium-chain acyl-CoA dehydrogenase ( <i>ACADM</i> )	Mucopolysaccharadosis type VI ( <i>ARSB</i> )
3-Methylcrotonyl-coA carboxyl deficiency (3-MCC)–( <i>MCCC1</i> and <i>MCCC2</i> )	Mucopolysaccharadosis type VII ( <i>GUSB</i> )
Methylmalonic aciduria ( <i>MMAA</i> and <i>MMAB</i> )	Nieman pick A/B ( <i>SMPD1</i> )
Methylmalonyl CoA mutase ( <i>MUT</i> )	Nieman pick C ( <i>NPC2</i> )
Mitochondrial acetoacetyl-coA thiolase (3-ketothiolase) ( <i>ACAT1</i> )	Sandhoff disease ( <i>HEXB</i> )
Multiple Co-A carboxylase ( <i>HLCS</i> )	Tay Sachs disease ( <i>HEXA</i> )
Phenylalanine hydroxylase ( <i>PAH</i> )	
Propionic acidemia ( <i>PCCA</i> and <i>PCCB</i> )	
Ornithine transcarbamylase deficiency ( <i>OTC</i> )	Other disorders
Trifunctional protein deficiency ( <i>HADHA</i> and <i>HADHB</i> )	Alpha thalassemia (Alpha globin gene)
Tyrosinemia type I ( <i>FAH</i> )	Autosomal dominant optic atrophy ( <i>OPA1</i> and <i>OPA3</i> )
Very long-chain acyl-CoA dehydrogenase ( <i>ACADVL</i> )	Chromosome 9q subtelomeric deletion syndrome ( <i>EHMT1</i> )
	Congenital adrenal hyperplasia ( <i>CYP21A2</i> )
	Congenital heart defects ( <i>NKX2.5</i> )
	Connexin 26–( <i>GJB2</i> )
	Connexin 30–( <i>GJB6</i> )
	Cystic fibrosis ( <i>CFTR</i> )
	Fragile X syndrome ( <i>FMRI</i> )
	Mowat-wilson syndrome ( <i>ZEB2</i> )
	Rett syndrome ( <i>MeCP2</i> )

quence designed with no cross-hybridization potential with any known sequence. This sequence was designed to have three distinct sequencing domains (A, B, and C), each with unique characteristics. The "A" domain contains long runs of G nucleotides that can be difficult to synthesize. The "B" domain contains a large perfect hairpin sequence. The "C" domain contains a straightforward domain that should always resequence. Resequencing of both the forward and the reverse strands generates six scores for the Cy3 channel and six scores for the Cy5 channel: A-forward and A-reverse, B-forward and B-reverse, C-forward and C-reverse. Failure of domain "C" indicates a catastrophic failure. A score from 0 to 100% is obtained that correlates with the overall performance of a microarray experiment.

### Data masking feature

Data files (.gff) generated from different averaging windows using NimbleScan software were parsed using a custom program (Nimkit) that was developed in-house. Nimkit enables the laboratory to select and analyze only the gene of interest. Nimkit generates a gene-specific report summarizing breakpoints detected in the gene of interest, the respective log (2) ratios, and the exons present at each region. All other regions are masked and not analyzed by Nimkit, preventing genetic analysis of genes for which clinical testing was not requested.

### Patients

DNA samples from clinically and biochemically diagnosed patients referred for molecular diagnostic testing were obtained from the EGL. Deletions and duplications in 13 previously characterized samples (V1–V13) were used for validation. In addition, deletion and duplication mutations were characterized in six autosomal cases (T1–T6) and one X-linked case (T7), where direct sequencing failed to identify a mutation. Targeted CGH array findings were confirmed by sequencing across the breakpoints or on the EmArray Cyto6000 oligonucleotide array.

## RESULTS

A custom CGH array was designed to detect deletions and duplications in 71 genes involved mainly in lysosomal storage and metabolic diseases (Table 1). DNA samples from 13 patients (V1–V13; Table 2) were used to validate the targeted gene CGH array. Deletions or gross genomic changes in these samples were previously characterized by direct sequencing, allele-specific PCR, fluorescence in situ hybridization, or the EmArray Cyto6000 oligonucleotide array. Table 2 summarizes the disorder, the gene involved, and previously identified mutations for each of the samples tested. There was 100% concordance between our targeted gene CGH array analysis and the previously characterized deletion or duplication. Furthermore, we detected deletions/duplications in six autosomal cases (T1–T6) and one X-linked case (T7), where direct sequencing failed to identify a mutation (Table 2). Patients V3, V7, V10, T1, and T5 are discussed in further detail below.

### Patient V3

A sample from this patient was referred to the EGL for molecular testing for maple syrup urine disease (MSUD) to confirm a clinical diagnosis. MSUD is an autosomal recessive disorder caused by deficient activity of the branched-chain alpha-keto acid dehydrogenase (BCKD) enzyme complex. The BCKD complex consists of two 2-oxoisovalerate dehydrogenase subunits, called alpha (*BCKDHA*, E1 alpha) and beta (*BCKDHB*, E1 beta), and a lipamide acyltransferase (*DBT*, E2).<sup>23–25</sup> Biochemical testing

demonstrated a low level of leucine decarboxylation activity. Direct sequence analysis failed to identify any mutations in the *BCKDHA*, *BCKDHB*, or *DBT* genes in this patient.

We were unable to PCR amplify exon 2 of the *DBT* gene with either of two independent sets of PCR primers. As exon 2 could not be amplified, a homozygous deletion of *DBT* exon 2 was assumed. Targeted gene CGH array analysis of the *DBT* gene detected a homozygous deletion of exon 2 in this patient. Approximate genomic break points at nucleotide positions g.100,484,550 in intron 1 and g.100,477,050 in intron 2 indicate a ~7500-bp deletion, which includes exon 2 (Fig. 2, C and D). Figure 2D displays a scatter plot, where the y-axis corresponds to signal intensity. A loss of genomic material is represented as a shift down of each feature, depicted on the x-axis. These data show a deletion of the corresponding genomic position, including exon 2 of the *DBT* gene in patient V3 (Fig. 2D).

### Patient V7

This patient was identified via newborn screening and diagnosed with Duarte variant galactosemia. Galactosemia is an autosomal recessive disorder caused by deficiency of the galactose-1-phosphate uridylyltransferase (*GALT*) enzyme that results from mutations within the *GALT* gene.<sup>26</sup> Galactosemia is a clinically heterogeneous disease with presentations ranging from the Duarte variant and patients who are apparently normal or have mild symptoms to classical severe galactosemia.<sup>26,27</sup> The Duarte allele is associated with the amino acid change asparagine to aspartic acid at position 314 (N314D).<sup>26</sup> Elsas et al.,<sup>28</sup> using G for the classic galactosemia allele, D for the Duarte allele, and N for the normal allele, proposed that the D/N, D/D, and D/G genotypes show ~75%, 50%, and 25% of normal *GALT* enzyme activity, respectively. Biochemical testing on individual V7 demonstrated low levels of *GALT* enzyme activity in red blood cells (~25%), consistent with the Duarte galactosemia variant (D/G). A sample from this patient was referred to the EGL for molecular testing. Upon direct sequencing of the *GALT* gene, an apparent homozygous nucleotide change c.2744A>G (p.N314D) in exon 10 was identified, consistent with a D/D phenotype. In addition, direct DNA sequencing on samples from the parents identified an N314D change in only one of the parents. Coffee et al.<sup>29</sup> have previously characterized two consecutive deletions in the *GALT* gene encompassing ~5.0 Kb, which results in the deletion of all exons of the *GALT* gene, with only 117 bp retained that correspond to portions of exon 8 and intron 8 of the *GALT* gene. To resolve the apparent discrepancy between the biochemical (D/G) and the apparent molecular (D/D) phenotype, allele-specific PCR analysis was performed. A ~5.0-Kb deletion was identified that retained 117 bp, including exon 8, within the *GALT* gene in this individual and in one of the parents. Thus, the actual molecular phenotype for this patient is D/G, consistent with the biochemical phenotype. In the process of validating the targeted gene CGH array, we analyzed this sample and confirmed the deletion of all exons, with the exception of exon 8 (~110-bp) of the *GALT* gene, which was retained (Fig. 2, A and B). Targeted CGH array analysis detected two consecutive deletions. The first deletion being at approximate genomic position g.34,636,633 upstream of the start site of the *GALT* gene to g.34,638,808 in intron 7. The second deletion starts at genomic position g.34,638,918 in intron 8 to g.34,641,648 extending beyond the end of the *GALT* gene. The deleted regions upstream and downstream of the *GALT* gene do not contain any known genes.

**Table 2** A list of patients, disorders, associate genes, and mutant alleles

Patient	Disorder	Gene	Allele 1	Allele 2
V1	Hearing loss	<i>GJB2/GJB6</i>	c.167delT <i>GJB2</i>	342-Kb deletion <i>GJB6</i>
V2	Mowat-Wilson syndrome	<i>ZEB2</i>	Deletion whole gene	—
V3	Maple syrup urine disease	<i>DBT</i>	Exon 2 deletion <sup>a,b</sup>	Exon 2 deletion <sup>a,b</sup>
V4	Krabbe disease	<i>GALC</i>	30-Kb deletion	30-Kb del
V5	Developmental delay, coarse features and failure to thrive	<i>EHMT1</i>	9q34.3 deletion	
V6	Dysmorphic features and language disorder		48, XXXY	
V7	Galactosemia	<i>GALT</i>	N314D	5-Kb deletion
V8	Encephalopathy, not elsewhere classified		4.4-Mb loss on 21q22.3 (including 53 genes)	
V9	No indication provided		8.4-Mb gain on 19q	2.8-Mb loss of 1 pter.
V10	Krabbe disease (carrier)	<i>GALC</i>	30-Kb del	—
V11	Krabbe disease (carrier)	<i>GALC</i>	30-Kb del	—
V12	Krabbe disease (carrier)	<i>GALC</i>	30-Kb del	—
V13	Krabbe disease (carrier)	<i>GALC</i>	30-Kb del	—
T1	Phenylketonuria	<i>PAH</i>	c.728G>A (p.R243Q)	Deletion, exon 5 to exon 6 <sup>b</sup>
T2	Sandhoff disease	<i>HEXB</i>	c.850C>T (p.R284X)	Deletion, exon 1 to exon 5 <sup>b</sup>
T3	Sandhoff disease (carrier)	<i>HEXB</i>	Deletion, exon 1 to exon 5 <sup>b</sup>	—
T4	Krabbe disease (carrier)	<i>GALC</i>	Duplication, exon 11 to exon 14 <sup>b</sup>	—
T5	Maple syrup urine disease	<i>BCKDHB</i>	c.752T>C (p.V251A)	Duplication, exon 7 to exon 9 <sup>b</sup>
T6	Maple syrup urine disease	<i>BCKDHB</i>	Deletion, exon 9 to exon 10 <sup>b</sup>	Deletion, exon 9 to exon 10 <sup>b</sup>
T7	Ornithine carbamoyltransferase deficiency (carrier)	<i>OTC</i>	Deletion of whole gene <sup>b</sup>	—

Thirteen patients (V1–V13) were involved in the targeted CGH array validation, where previously characterized deletion mutations were confirmed on targeted CGH array with a 100% concordance. Patients T1–T7 were tested on the targeted CGH array, to either identify a second mutation or when no mutation was identified by DNA direct sequencing.

<sup>a</sup>Direct sequencing PCR failed to amplify exon 2 of the *DBT* gene.

<sup>b</sup>Only identified by targeted CGH array analysis.

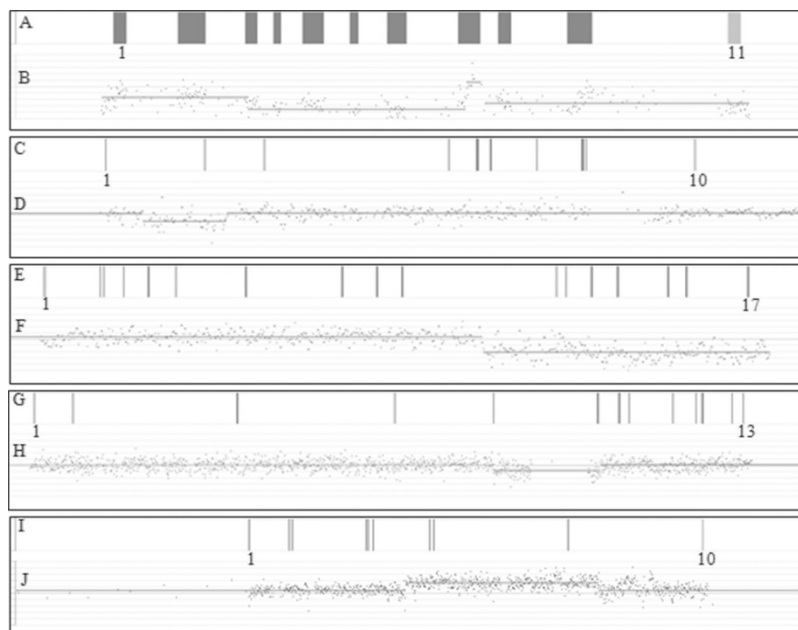
### Patient V10

A sample from this individual was referred to the EGL for molecular carrier testing for Krabbe disease. The individual's daughter was diagnosed with Krabbe disease, an autosomal recessive neurodegenerative disorder caused by deficiency of the enzyme galactocerebrosidase (*GALC*). Mutations within the *GALC* gene cause Krabbe disease.<sup>30</sup> A 30-Kb deletion starting in the large intron 10 and extending beyond the end of the *GALC* gene, encompassing exon 11 to exon 17, is the most common deletion observed in Krabbe patients.<sup>31</sup> Allele-specific PCR analysis for the 30-Kb deletion within the *GALC* gene detected a homozygous 30-Kb deletion in this individual's daughter.<sup>32</sup> Additionally, direct sequence analysis failed to identify any other mutations. Allele-specific PCR analysis for the 30-Kb deletion within the *GALC* gene also detected a heterozygous 30-Kb deletion. Consequently, targeted gene CGH array analysis for the *GALC* gene on this individual detected a deletion of ~32 Kb with approximate genomic breakpoints at nucleotide position g.87,492,950 in intron 10 and g.87,460,950 extending beyond the end of the *GALC* gene. Similarly, a heterozygous 30-Kb deletion within the *GALC* gene

was detected by allele-specific PCR analysis in this individual's partner and son (V11 and V12, respectively); these results were confirmed by targeted gene CGH analysis for the *GALC* gene (data not shown).

### Patient T1

A sample from this individual was referred to our laboratory for molecular testing for phenylketonuria (PKU) because of elevated plasma phenylalanine consistent with PKU/hyperphenylalaninemia. PKU is an autosomal recessive disorder, with ~97–99% of affected individuals having a deficiency of the enzyme phenylalanine hydroxylase (*PAH*). Mutations within the *PAH* gene cause PKU,<sup>33</sup> with the remaining 1–3% of PKU cases due to defects in the formation or recycling of tetrahydrobiopterin (BH4).<sup>34</sup> Direct DNA sequencing of the *PAH* gene identified one missense mutation (p.R243Q) in exon 7 of the *PAH* gene, but a second mutation in the *PAH* gene was not identified. Targeted gene CGH array analysis of the *PAH* gene detected an ~11.7-Kb deletion encompassing most of exon 5 and all of exon 6 of the *PAH* gene (Fig. 2, G and H), with approximate genomic nucleotide positions g.101,772,838 in



**Fig. 2.** Targeted gene CGH array for detecting single- and multiexon gene deletion and duplication mutations. A, Horizontal lines represent all 11 exons of the *GALT* gene. B, Bar graph displays loss of genomic material that corresponds to all exons of the *GALT* gene, except for exon 8. C, Horizontal lines represent all 10 exons of the *DBT* gene. D, Bar graph displays loss of genomic material that corresponds to exon 2 of the *DBT* gene. E, Horizontal lines represent all 17 exons of the *GALC* gene. F, Bar graph displays loss of genomic material that corresponds to exon 11 through exon 17 of the *GALC* gene. G, Horizontal lines represent all 13 exons of the *PAH* gene. H, Bar graph displays loss of genomic material that corresponds to exon 3 through exon 6 of the *PAH* gene. I, Horizontal lines represent all 10 exons of the *BCKDHB* gene. J, Bar graph displays a gain of genomic material that corresponds to exon 11 through exon 14 of the *BCKDHB* gene. Loss and gain of genomic material are depicted as a shift down and shift up of the features on the *x*-axis, respectively.

exon 5 and g.101,784,528 in intron 6 of the *PAH* gene. Junction fragment PCR analysis across the breakpoints indicated a deletion of 11,652 bp in size, with the 5' breakpoint in exon 5 and the 3' breakpoint in intron 6. Direct sequencing of the junction fragment product confirmed our targeted gene CGH array data, where most of exon 5 (with the exception of the first 11 nucleotides) and all of exon 6 are deleted in patient T1 (Fig. 3).

#### Patient T5

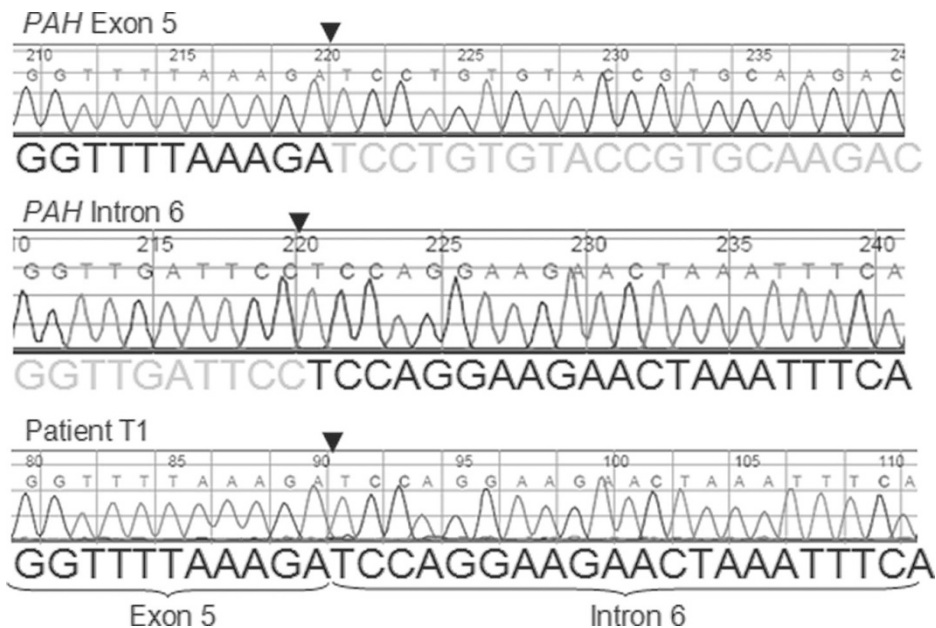
A sample from this patient was referred to the EGL for molecular testing for MSUD. Previous biochemical testing on this patient indicated elevation of the branched chain amino acids leucine, isoleucine, and valine, as well as alloisoleucine in plasma, consistent with a diagnosis of MSUD. Direct sequence analysis identified one copy of a c.752T>C (p.V251A) mutation in exon 7 of the *BCKDHB* gene in this patient. A second mutation was not identified. As with the previous patient, the availability of the targeted gene CGH array analysis for the *BCKDHB* gene as a clinical molecular test enabled the identification of a second mutation in this patient. Targeted CGH analysis for the *BCKDHB* gene identified a duplication mutation (~100-Kb) encompassing exon 7 to exon 9, with approximate genomic breakpoints at nucleotide positions g.80,955,500 in intron 6 and g.81,055,500 in intron 9 (Fig. 2, I and J).

### DISCUSSION

CGH arrays have been widely used to detect gross chromosomal changes, including large deletions and duplications

across the human genome; however, the application of targeted gene CGH arrays for detecting intragenic deletions or duplications is in its infancy. We designed, validated, and implemented a targeted gene CGH array analysis as a clinical molecular testing tool to simultaneously screen for deletions and duplications of single- and multiexon mutations in a large set of autosomal and X-linked disease-associated genes. In particular, we were interested in genes involved in lysosomal storage and metabolic disorders, and other genes for which the EGL offers direct DNA sequence analysis.

The targeted gene CGH array has several advantages over other available molecular methods, such as multiplex PCR, quantitative PCR, Southern blotting, MLPA, DOVAM-S, and SCAIP sequencing. These traditional methods are time-consuming, labor-intensive, complex, and fail to detect all mutations accurately. Duplication mutations in particular are very difficult to detect in both males and females in autosomal cases and in females in X-linked cases. In addition, mutations detected by the currently available techniques, especially single-exon deletions, require further costly confirmation by a different technique, which can affect the turnaround time for results. Because these more common techniques depend only on one or two probes for each targeted region, false-positive findings are known to occur. Furthermore, genomic positions of the breakpoints cannot be determined by traditional methods. In contrast, the targeted gene CGH array is a rapid, comprehensive, relatively inexpensive, highly sensitive, and accurate method for detecting single- and multiexon deletions and duplications in a large set of genes simultaneously on a common platform.



**Fig. 3.** Direct sequencing across the *PAH* gene break points in patient T1. Top histogram shows exon 5 control sequence (arrowhead represents 5' break point, and the deleted sequence is represented as gray letters below the histogram). Middle histogram represents the control sequence of intron 6 (arrowhead represents 3' break point, and the deleted sequence is represented as gray letters below the histogram). Bottom panel represents patient T1 sequence of a PCR-amplified product across the break-point junction (arrowhead represents the break-point junction).

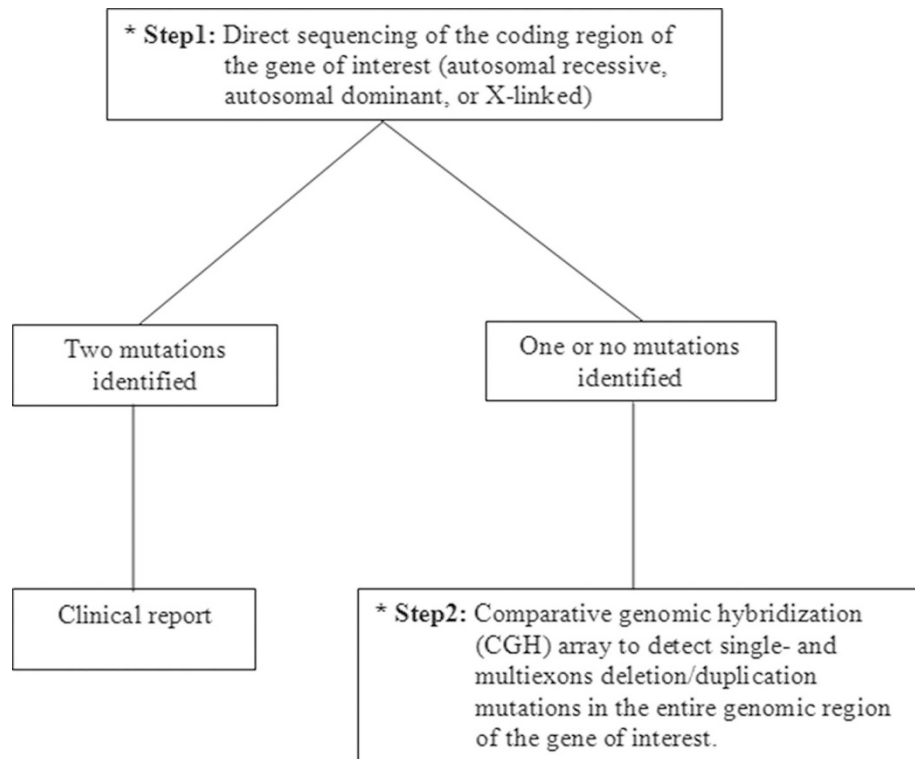
The array was designed as an exon centric array with an extremely dense spacing in the exons (2 bp) and more widely spaced oligos (25 bp) in the introns. The 385K Nimbelgen array format supported this design due to large excess of oligos. This array also lends the flexibility to add several additional genes. Addition of new genes may require reducing the density of the oligos in exons but this will not reduce the sensitivity of the array. From our prior work on development of a targeted array for the dystrophin gene<sup>21</sup> we have determined that it is important to have a minimum of four oligos for an exon size of 100 bp to allow for optimal detection of deletions and duplications. As gene CGH array technology is a recent technology to be used in the molecular setting and not yet a “Gold” standard, our laboratory is confirming all deletions/duplications by a secondary method. Deletions can be easily confirmed by designing primers across the breakpoints, whereas, duplications are relatively difficult to confirm using an alternative technology. Methods such as breakpoints PCR by taking into consideration the possibility of tandem duplication or inversion, southern blot, real time PCR, or confirmation on another array can be considered. In our experience, the targeted CGH array was able to detect the single-exon retention of exon 8 (~110-bp) in the middle of the large *GALT* gene deletion and multiexon duplications in the *GALC* and *BCKDHB* genes that had not been found before.

Targeted gene CGH array analysis has a higher sensitivity than the currently available methods for detecting deletions and duplications. Our targeted CGH array is based on the use of overlapping oligonucleotides representing the coding region of each gene, thereby allowing us to oversample and lower the likelihood of false-positive findings. Targeted CGH array analysis confirmed previously characterized deletion mutations in 13 cases with a 100% concordance (Table 2). Hence, targeted

gene CGH array analysis can determine approximate genomic breakpoints for all deletion and duplication mutations.

The need for comprehensive mutation detection assays with a reasonable turnaround time for results is a pressing necessity. Deletions and duplications constitute a significant portion of the mutation spectrum for some genes. The availability of comprehensive mutation detection assays that can accurately identify intragenic deletions and duplications in a single gene in a reasonable time will yield more accurate mutation detection across the mutation spectrum for disease-associated genes and will greatly improve clinical diagnosis, prenatal diagnosis, and genetic counseling. For example, intragenic deletion and duplication mutations had not been documented in MSUD patients previously. Molecular characterization of MSUD alleles has a ~95% mutation detection rate with a mutation spectrum of ~95% point mutations; the remaining 5% were considered uncharacterized mutations,<sup>35,36</sup> and large and single-exon deletions and duplications had not been seen before. Here we present for the first time two cases of multiexon deletion and duplication mutations in the *BCKDHB* gene, identified in patient T6 and T5, respectively. Furthermore, we also identified a multiexon duplication in the *GALC* gene (individual T4, Table 2), which had not been documented before. Such duplication mutations are difficult or impossible to detect by other methods.

Use of a targeted gene CGH array will improve the mutation detection rate for autosomal and X-linked genes, as intragenic single- and multiexon deletions and duplications can be identified easily. An improved mutation detection rate is critical for prenatal diagnosis and genetic counseling, as the identification of disease-causing mutations is important for better genetic counseling and a must for prenatal testing. Failure to identify disease-causing mutations would make prenatal testing for at-risk families impossible. In our experience, the targeted CGH



**Fig. 4.** A two-step approach to detect virtually all mutations by complementing gene sequencing analysis with targeted CGH array analysis based on the mutation spectrum of genes.

array detected intragenic single- and multiexon deletions and duplications in eight patients for whom DNA sequence analysis failed to identify a disease-causing mutation. Although the number of samples we have tested per gene remains small, the collective mutation detection rate for the genes that have been tested on our CGH array thus far is 13.3% (eight individuals out of 60). Hence, the combination of direct DNA sequence analysis with targeted gene CGH array analysis significantly improves the mutation detection rate for genes listed in Table 1 to ~98%, given that a rough estimate of the mutation detection rate for most genes is ~80–85%. Despite the maturity of sequencing and CGH, an estimated 1–2% of mutations located deep within intronic or promoter regions cannot be identified. The comprehensive testing strategy using the CGH array significantly increases the possibility of finding a mutation in a patient. This is a cost-effective approach that provides options for additional family members to be tested, prenatal testing and enrolment in clinical trials especially for metabolic disorders where the diagnosis has been confirmed by biochemical testing. Molecular tests can be expensive and the availability of a comprehensive testing strategy will reduce the burden for the payers and patients. Therefore, the targeted gene CGH array is appropriate to be applied for maximal mutation detection in a specific gene associated with a clinical or biochemical diagnosis in contrast to the genomic CGH arrays used in cytogenetic laboratories which involve testing for multiple loci across the genome involved in a range of phenotypes.

Here we propose a two-step approach based on the mutation spectrum of the gene of interest, wherein the CGH array complements the gene-sequencing assay. Our approach enables the detection of virtually all mutations in a large set of genes (Fig. 4), where we first sequence the coding region of the gene of

interest to identify two mutations. If direct DNA sequencing fails to identify two mutations, we will reflex the sample to targeted gene CGH array analysis for the gene of interest. These steps are flexible and interchangeable depending on the mutation spectrum for each gene.

To avoid testing for genes that are not requested and also to avert from providing unwanted medical information, we have developed a data masking feature that only allows the extraction of data corresponding to the gene being tested. In other words, the data for all other genes is masked and not extracted. We developed a custom program (Nimkit) to initially select the gene of interest before the data analysis. Nimkit analyzes data only for the gene being tested and generates a report specifying any breakpoints detected, the respective log (2) ratios, and the exons of the gene of interest. As Nimkit extracts data for the entire gene of interest deletions and duplications of the probes in the introns and noncoding exons will also be detected. Unless the identified intronic and noncoding region deletion or duplication has been previously identified and shown to affect gene function by methods such as cDNA analysis such events will be reported as variants of unknown clinical significance. Additional family studies to determine segregation of the variant with disease may help clarify the clinical significance of such variants but functional studies will need to be carried out to determine the effect of such intronic and noncoding region deletions or duplications.

The CGH array format allows simultaneous development of an assay for detecting deletions and duplications in a number of genes. This format is flexible to adding, removing, or masking genes on the array. The CGH array format and flexibility therefore results in significant savings for a clinical laboratory in test development costs. The major expenditure of this assay



is the cost of the CGH array. The reagents are relatively cheap and several arrays can be processed at the same time and data can be generated in two days. Thus, this assay requires significantly reduced labor and has fewer overheads in comparison to the more traditional methodologies.

In summary, targeted gene CGH array analysis will significantly improve the detection of single- and multiexon deletions and duplications in autosomal and X-linked disease-associated genes, especially when the targeted CGH array is designed to complement gene-sequencing analysis. Targeted gene CGH arrays will enable comprehensive clinical molecular testing, and effective genetic counseling. Finally, the targeted gene CGH arrays analysis can be easily adopted by the clinical molecular testing laboratories as a rapid, cost-effective, and highly sensitive and accurate approach for the detection of single- and multiexon deletion and duplication mutations.

## REFERENCES

- Mendell JR, Buzin CH, Feng J, et al. Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. *Neurology* 2001;57:645–650.
- Prior TW, Bridgeman SJ. Experience and strategy for the molecular testing of Duchenne muscular dystrophy. *J Mol Diagn* 2005;7:317–326.
- Norman AM, Thomas NS, Kingston HM, Harper PS. Becker muscular dystrophy: correlation of deletion type with clinical severity. *J Med Genet* 1990;27:236–239.
- Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073–1080.
- Bobadilla JL, Macek M Jr, Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations—correlation with incidence data and application to screening. *Hum Mutat* 2002;19:575–606.
- Hegde MR, Roa BB. Detecting mutations in the APC gene in familial adenomatous polyposis (FAP). *Curr Protoc Hum Genet* 2006;10:10.8.
- Visser R, Matsumoto N. Genetics of Sotos syndrome. *Curr Opin Pediatr* 2003;15:598–606.
- Devilee P, Cleton-Jansen AM, Cornelisse CJ. Ever since Knudson. *Trends Genet* 2001;17:569–573.
- Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002;30:e57.
- Liu Q, Feng J, Buzin C, et al. Detection of virtually all mutations-SSCP (DOVAM-S): a rapid method for mutation scanning with virtually 100% sensitivity. *Biotechniques* 1999;26:932, 936–938, 940–932.
- Buzin CH, Wen CY, Nguyen VQ, et al. Scanning by DOVAM-S detects all unique sequence changes in blinded analyses: evidence that the scanning conditions are generic. *Biotechniques* 2000;28:746–750, 752–743.
- Flanigan KM, von Niederhausern A, Dunn DM, Alder J, Mendell JR, Weiss RB. Rapid direct sequence analysis of the dystrophin gene. *Am J Hum Genet* 2003;72:931–939.
- Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res* 1996;6:986–994.
- Nguyen VQ, Liu Q, Sommer SS. A large-scale validation of dosage analysis by robust dosage-polymerase chain reaction. *Anal Biochem* 2007;371:37–42.
- Abbs S, Bobrow M. Analysis of quantitative PCR for the diagnosis of deletion and duplication carriers in the dystrophin gene. *J Med Genet* 1992;29:191–196.
- Charbonnier F, Raux G, Wang Q, et al. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Res* 2000;60:2760–2763.
- Gunn SR, Robetorye RS, Mohammed MS. Comparative genomic hybridization arrays in clinical pathology: progress and challenges. *Mol Diagn Ther* 2007;11:73–77.
- Wong LJ, Dimmock D, Geraghty MT, et al. Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. *Clin Chem* 2008;54:1141–1148.
- Dharm P, Coffey AJ, Abbs S, et al. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* 2005;76:750–762.
- Staa J, Tomgren T, Ramebeck E, et al. Detection and precise mapping of germline rearrangements in BRCA1, BRCA2, MSH2, and MLH1 using zoom-in array comparative genomic hybridization (aCGH). *Hum Mutat* 2008;29:555–564.
- Hegde MR, Chin EL, Mülle JG, Okou DT, Warren ST, Zwick ME. Microarray-based mutation detection in the dystrophin gene. *Hum Mutat* 2008; 29:1091–1099.
- Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 2004;20:3413–3422.
- Chuang JL, Fisher CR, Cox RP, Chuang DT. Molecular basis of maple syrup urine disease: novel mutations at the E1 alpha locus that impair E1(alpha 2 beta 2) assembly or decrease steady-state E1 alpha mRNA levels of branched-chain alpha-keto acid dehydrogenase complex. *Am J Hum Genet* 1994;55:297–304.
- Nellis MM, Danner DJ. Gene preference in maple syrup urine disease. *Am J Hum Genet* 2001;68:232–237.
- Nellis MM, Kasinski A, Carlson M, et al. Relationship of causative genetic mutations in maple syrup urine disease with their clinical expression. *Mol Genet Metab* 2003;80:189–195.
- Reichardt JK, Woo SL. Molecular basis of galactosemia: mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase. *Proc Natl Acad Sci U S A* 1991;88:2633–2637.
- Hammersen G, Houghton S, Levy HL. Rennes-like variant of galactosemia: clinical and biochemical studies. *J Pediatr* 1975;87:50–57.
- Elsas LJ, Dembure PP, Langley S, Paulk EM, Hjelm LN, Fridovich-Keil J. A common mutation associated with the Duarte galactosemia allele. *Am J Hum Genet* 1994;54:1030–1036.
- Coffee B, Hjelm LN, DeLorenzo A, Courtney EM, Yu C, Muralidharan K. Characterization of an unusual deletion of the galactose-1-phosphate uridylyl transferase (GALT) gene. *Genet Med* 2006;8:635–640.
- Sakai N, Inui K, Fujii N, et al. Krabbe disease: isolation and characterization of a full-length cDNA for human galactocerebrosidase. *Biochem Biophys Res Commun* 1994;198:485–491.
- Rafi MA, Luzi P, Chen YQ, Wenger DA. A large deletion together with a point mutation in the GALC gene is a common mutant allele in patients with infantile Krabbe disease. *Hum Mol Genet* 1995;4:1285–1289.
- Luzi P, Rafi MA, Wenger DA. Characterization of the large deletion in the GALC gene found in patients with Krabbe disease. *Hum Mol Genet* 1995; 4:2335–2338.
- Woo SL, Lidsky AS, Guttler F, Chandra T, Robson KJ. Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* 1983;306:151–155.
- Zurfluh MR, Zschocke J, Lindner M, et al. Molecular genetics of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. *Hum Mutat* 2008;29:167–175.
- Quental S, Macedo-Ribeiro S, Matos R, et al. Molecular and structural analyses of maple syrup urine disease and identification of a founder mutation in a Portuguese Gypsy community. *Mol Genet Metab* 2008;94:148–156.
- Rodriguez-Pombo P, Navarrete R, Merinero B, Gomez-Puertas P, Ugarte M. Mutational spectrum of maple syrup urine disease in Spain. *Hum Mutat* 2006;27:715.