

# Multiplexed Genetic Analysis Using an Expanded Genetic Alphabet

SCOTT C. JOHNSON,<sup>1</sup> DAVID J. MARSHALL,<sup>1</sup> GERDA HARMS,<sup>1</sup> CHRISTIE M. MILLER,<sup>1</sup>  
CHRISTOPHER B. SHERRILL,<sup>1</sup> EDWARD L. BEATY,<sup>1</sup> SCOTT A. LEDERER,<sup>1</sup> ERIC B. ROESCH,<sup>1</sup>  
GARY MADSEN,<sup>1</sup> GARY L. HOFFMAN,<sup>2</sup> RONALD H. LAESSIG,<sup>2</sup> GREG J. KOPISH,<sup>2</sup>  
MEI WANG BAKER,<sup>2</sup> STEVEN A. BENNER,<sup>4</sup> PHILIP M. FARRELL,<sup>3</sup> and JAMES R. PRUDENT<sup>1\*</sup>

**Background:** All states require some kind of testing for newborns, but the policies are far from standardized. In some states, newborn screening may include genetic tests for a wide range of targets, but the costs and complexities of the newer genetic tests inhibit expansion of newborn screening. We describe the development and technical evaluation of a multiplex platform that may foster increased newborn genetic screening.

**Methods:** MultiCode<sup>®</sup> PLx involves three major steps: PCR, target-specific extension, and liquid chip decoding. Each step is performed in the same reaction vessel, and the test is completed in ~3 h. For site-specific labeling and room-temperature decoding, we use an additional base pair constructed from isoguanosine and isocytidine. We used the method to test for mutations within the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The developed test was performed manually and by automated liquid handling. Initially, 225 samples with a range of genotypes were tested retrospectively with the method. A prospective study used samples from >400 newborns.

**Results:** In the retrospective study, 99.1% of samples were correctly genotyped with no incorrect calls made. In the perspective study, 95% of the samples were correctly genotyped for all targets, and there were no incorrect calls.

**Conclusions:** The unique genetic multiplexing platform was successfully able to test for 31 targets within the

CFTR gene and provides accurate genotype assignments in a clinical setting.

© 2004 American Association for Clinical Chemistry

Few multiplexed technologies with sufficient specificity to identify small changes within the human genome are available for clinical use. Line probe or linear array technology uses nitrocellulose paper strips as the support matrix (1). A benefit of the line probe is that expensive instrumentation is not required. A technology termed "oligonucleotide ligation assay" uses target-specific ligation to generate mutation-specific products that are analyzed by gel electrophoresis (2). Other genetic analysis systems are coupled to microparticle flow cytometry (3–8). Limitations of these methods include complicated procedures that require extensive training (washings, centrifugations, and transfers), long completion times, expensive instrumentation, and technician-dependent result analysis. In addition, these technologies are not easily amenable to automation. Many hospitals and clinics currently send molecular diagnostics tests to large reference laboratories. This is in part attributable to the limitations described. Because there are benefits in having local laboratories perform diagnostic tests (9–12), there appears to be a need for molecular testing methods that circumvent the limitations posed by current testing methods.

We have constructed a three-step platform technology called MultiCode<sup>®</sup> PLx that eliminates the most technically challenging steps involved in multiplexed genetic analysis and compared it with the line probe assay in a clinical setting. MultiCode PLx uses one additional nucleobase pair constructed from the complementary nucleobases 2'-deoxy-isoguanosine (diG)<sup>5</sup> and 5-Me-isocytosine.

<sup>1</sup> Eragen Biosciences, Inc., Madison, WI.

<sup>2</sup> Wisconsin State Laboratory of Hygiene, Madison, WI.

<sup>3</sup> Department of Pediatrics, University of Wisconsin, Madison, WI.

<sup>4</sup> Department of Chemistry, University of Florida, Gainesville, FL.

\*Address correspondence to this author at: Eragen Biosciences, Inc., 918 Deming Way, Madison, WI 53717. Fax 608-662-9004; e-mail jprudent@eragen.com.

Received March 19, 2004; accepted July 21, 2004.

Previously published online at DOI: 10.1373/clinchem.2004.034330

<sup>5</sup> Nonstandard abbreviations: diG, 2'-deoxy-isoguanosine; iC, 5-Me-isocytosine; TSE, target-specific extension; diGTP, 2'-deoxy-isoG triphosphate; CF, cystic fibrosis; CFTR, cystic fibrosis transmembrane conductance regulator; and MFI, median fluorescence intensity.

cytosine (iC). These nucleobases specifically recognize each other based on a different pattern of hydrogen bondings. We chose this pair because no other such pair is available commercially and their chemistries have been well explored. For example, iG and iC have been successfully used for both molecular recognition (13–15) and for site-specific incorporation (16–18). In addition, because iG:iC recognition is “orthogonal” to the naturally occurring nucleobase pairs, a strand of DNA that contains several iC/iG components can be constructed so that it will not hybridize to natural DNA. In complex reactions in which high concentrations of natural DNAs of known or unknown sequence exist, this orthogonal attribute allows specific molecular recognition to take place without interference. These additional nucleobases are used in each step of the PLx process: PCR, extension labeling, and liquid decoding. PCR primers are first designed to be target-specific and contain single iCs. The amplicons act as labeling templates for the target-specific extension (TSE) step. During that step, labels attached to diG triphosphate (diGTP) are incorporated site specifically into coded target-specific extenders (Fig. 1). The tags are short sequences (typically 10 nucleotides in length) assembled by use of a mixture of natural and nonnatural bases. The tags are designed to hybridize only to their perfect complements encoded on color-addressed microspheres. In the final step, decoding of the extension reactions is accomplished at room temperature by capture of the coded extenders on the addressed microspheres and reading of the reporter signal on a Luminex<sup>100</sup> instrument. All steps are carried out in the same reaction vessel without transfers or washings.

To demonstrate the approach, we developed a screening test for carriers of cystic fibrosis (CF) that tests for all 25 mutations and the reflex targets in the widely used panel (19) and also tests for 394delTT and 3199del6. 394delTT is the second most prevalent mutation in populations with Nordic heritage and may be important in Wisconsin, where testing was performed. 3199del6 was added because several lines of evidence indicate that 3199del6 is a disease-causing mutation and may replace I148T (20).

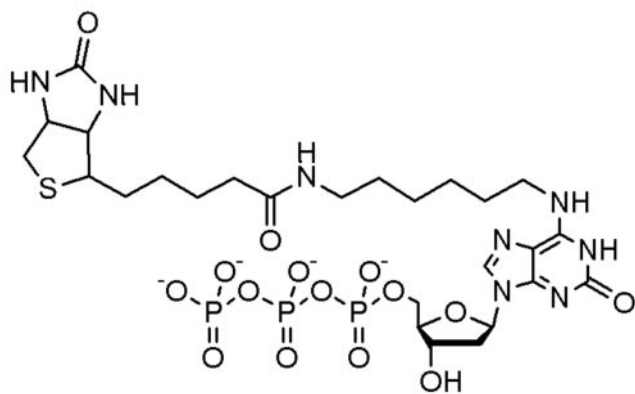


Fig. 1. Molecular structure of diGTP-biotin used in the described method.

The Wisconsin State Laboratory of Hygiene, in conjunction with the University of Wisconsin, was chosen as the major testing site because of its long history of screening newborns for CF (21) and study of screening methods (22–25). In March 2002, DNA testing of the top 4% of the daily immunoreactive trypsinogen results expanded to include all 25 mutations recommended by the American College of Medical Genetics, the American College of Obstetricians and Gynecologists, and the NIH.

In this report, we first present MultiCode PLx testing data on a variety of genomic DNAs with various genotypes prepared from several sample types for the presence or absence of mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Finally, data are presented on a prospective study comparing testing results obtained for freshly prepared blood spots by the PLx assay and the commercially available CF Gold Linear Array<sup>TM</sup> system (Roche Molecular).

## Materials and Methods

### SAMPLES AND DNA EXTRACTION

DNA samples were obtained from the Wisconsin State Laboratory of Hygiene (The Wisconsin Newborn Screening Laboratory, Madison, WI), Duke University Medical Center (Department of Pathology, Durham, NC), National Centre for Medical Genetics (Dublin, Ireland), and Laboratoire de Genetique Moleculaire et d'Histocompatibilite (Brest, France). DNA was extracted from whole blood and dried blood spots on Guthrie cards. The dried blood spots were stored at room temperature for 2–3 days after receipt until routine testing was completed. For long-term storage, dried blood spots were stored at 4 °C. DNA extraction from Guthrie cards was performed with the Generation Capture Card Kit (Gentra Systems) according to the manufacturer's guidelines for DNA purification and elution. DNA was extracted from whole blood by use of the Puregene DNA Isolation Kit (Gentra Systems) or the Whatman FTA system (Whatman Inc., Clifton, NJ). All genomic DNAs from these preparations were diluted 1:10 in water, and 1  $\mu$ L was added to each PCR reaction (1–50 ng of total DNA). As a cross-reference and in the prospective study, genotypes were also determined by CF Gold linear array or DNA sequencing if discrepancies were detected.

### SYNTHESIS OF OLIGONUCLEOTIDES

PCR and TSE oligonucleotides were manufactured in-house with a 48-column DNA Synthesizer (Northwest Engineering; now the 3900 from Applied Biosystems Inc.), using standard  $\beta$ -cyanoethyl phosphoramidite chemistry (for sequence information, see Tables 1, 2, and 3 in the Data Supplement accompanying the online version of this article at <http://www.clinchem.org/content/vol50/issue11/>). Synthesis of sequences that contain iC and iG can also be purchased from IDT, Eurogentec, or EraGen. A demonstration assay encompassing this specific system can be purchased from EraGen Biosciences.

Amino-labeled EraCode oligonucleotides were manufactured with an ABI 394 DNA Synthesizer (Applied Biosystems) using 3'-PT-Amino-Modifier C6 CPG (Glen Research). Target-specific extender primers were manufactured with the spacer phosphoramidite C3 (Glen Research) placed between the EraCode complement and the target-specific region. The isoG and isoC phosphoramidites [*N*<sup>6</sup>-(dimethylaminomethylidene)-5'-*O*-(dimethoxytrityl)-*O*<sup>2</sup>-(diphenylcarbamoyl)-2'-deoxyisoguanosine, 3'-*O*-cyano-ethoxydiisopropyl-phosphoramidite (Glen Research) and *N*<sup>2</sup>-(dimethylamino)methylidene-5'-*O*-dimethoxytrityl-5-methyl-2'-deoxycytidine, 3'-*O*-cyano-ethoxydiisopropyl-phosphoramidite (Glen Research), respectively] were coupled and deprotected under the conditions used for the natural base phosphoramidites. Postsynthesis work-up consisted of ammonium hydroxide deprotection followed by ethanol precipitation (PCR and TSE oligonucleotides) or 2-propanol precipitation (EraCode oligonucleotides).

#### EraCode BEAD COUPLING

For each conjugation reaction, a mixture (400  $\mu$ L) containing  $\sim 5 \times 10^6$  carboxylated microspheres (Luminex Corporation) was centrifuged at 8000g for 1 min. The supernatant was removed, and the beads were resuspended in 50  $\mu$ L of 0.1 mol/L MES, pH 4.5. After resuspension, 3.3  $\mu$ L of 300  $\mu$ mol/L 5'-amine-modified EraCode DNA oligonucleotide and 5.0  $\mu$ L of freshly prepared 10 g/L EDC (Pierce Chemical) were added. The reaction was mixed thoroughly and incubated at room temperature in the dark for 30 min. A second 5.0  $\mu$ L of 10 g/L EDC was added, and the mixture was incubated for 30 min. EraCode microspheres were washed by the addition of 1.0 mL of 10 mmol/L Tris (pH 8.0) containing 0.2 mL/L Tween 20 and pelleted by centrifugation at 8000g for 1 min; the supernatant was then removed. EraCode microspheres were washed a second time with 1 g/L sodium dodecyl sulfate in 10 mmol/L Tris (pH 8.0) and resuspended in 100  $\mu$ L of bead buffer (10 mmol/L MOPS, pH 7.5; 1 mmol/L EDTA, 0.5 g/L sodium dodecyl sulfate, 200 mmol/L sodium chloride, and 0.1 g/L sonicated herring sperm DNA). EraCoded microspheres were combined and diluted in bead buffer to generate a 50-code mixture at a concentration of 800 microspheres/ $\mu$ L for each class of microsphere.

#### EraCode DESIGN AND BEAD VALIDATION

Capture at room temperature of a tag to its specific complementary EraCode was tested by use of a model extension system where one target with a 5'-iC was constructed and 50 tags were interchanged onto a target-specific extender sequence. EraCodes were constructed by use of 9–10 bases comprising 2–3 iCs. In an effort to minimize interference from naturally occurring sequences, EraCodes were designed to have no more than four naturally occurring bases in a row. In an effort to minimize code-to-code cross-hybridization, no two codes

were allowed to contain the same series of more than four nucleobases in a row.

Fifty EraCodes were validated by use of 50 extenders, all containing the identical 3' target-specific sequence. The extenders were individually labeled using the target, standard concentrations of all four naturally occurring deoxynucleotide triphosphates, TiTaq, and diGTP-biotin at 1.5  $\mu$ mol/L. Each EraCode sequence was separately coupled to a specific color-coded fluorescent microsphere type (Luminex Corp). Each microsphere type contains a unique color or signature that can be determined by the Luminex<sup>100</sup>. After coupling, each EraCode bead conjugate was pooled to create an equalized bead master mixture (bead array: 50 different EraCode-labeled beads). The bead array was distributed into 50 separate reaction vessels. To each of the vessels, a specific extension reaction from above was added at room temperature along with streptavidin-phycoerythrin and was read on the Luminex<sup>100</sup>.

In a 10- $\mu$ L extension reaction, an excess of a template oligonucleotide (5'-iC-CGTATGGGCTCACCTCGCTGTG-3') at 150 nmol/L was combined with a control extension sequence at 100 nmol/L in EraGen MC TSE solution 35 (prod. no. PN1236; EraGen Biosciences, Inc.) with 100  $\mu$ mol/L each deoxynucleotide triphosphate, 1.5  $\mu$ mol/L diGTP-biotin, and 1 $\times$  Titanium Taq (Clontech). Each control extension oligonucleotide was composed of a 5' sequence complementary to each of the EraCode sequences, a three-carbon spacer, and a 3' sequence complementary to the template oligonucleotide (5'-CACAGC-GAGGTGAGCCCA-3'). Reactions were subjected to the following thermal cycling protocol: 30 s at 95  $^{\circ}$ C and 1 min at 65  $^{\circ}$ C. Each 10- $\mu$ L extension reaction was then hybridized to the EraCode bead mixture according to the described hybridization protocol. In addition, to mimic MultiCode conditions, unextended TSEs to the other EraCodes were also added.

Hybridization of extension products in each 10- $\mu$ L reaction was performed by mixing 40  $\mu$ L of EraGen MC Hyb Solution A (prod. no. PN1237; EraGen Biosciences) and 0.25  $\mu$ L of EraCoded bead mixture. To each hybridization reaction, 0.5  $\mu$ L of 2 g/L streptavidin-R-phycoerythrin conjugate (Prozyme) plus 39.5  $\mu$ L of sheath fluid (Luminex Corp.) were added, and 65  $\mu$ L of the final reaction was injected into the Luminex<sup>100</sup>.

#### PCR

PCR primers were designed using Primer 3 ([http://www.broad.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi)) or Hyther (<http://ozone2.chem.wayne.edu>). Primer pairs were designed to flank the polymorphism(s) of interest and have a melting temperature of 53–57  $^{\circ}$ C (see Table 2 in the online Data Supplement). A single iC was added to the 5' end of the PCR primers.

PCR reactions were performed in a final volume of 8.0  $\mu$ L; each reaction contained EraGen MC PCR solution 35 (prod. no. PN1235 or PN1305 for the 579T system;

EraGen), 0.4  $\mu\text{L}$  of 20 $\times$  EraGen CFTR PCR primer mixture, 1 $\times$  Titanium Taq, and 1  $\mu\text{L}$  of diluted genomic target DNA. Reactions were cycled according to the following profile: 1 cycle of 30 s at 95  $^{\circ}\text{C}$ , and 40 cycles of 1 s at 95  $^{\circ}\text{C}$ , 10 s at 55  $^{\circ}\text{C}$ , and 30 s at 72  $^{\circ}\text{C}$ . Each of the primer mixtures contained forward and reverse primers for a subset of the mutations at concentrations of 100–200 nM.

#### TSE

TSE was used to determine the exact mutations present. TSE requires TiTaq, diGTP-biotin, and extenders that contain 3' target-specific bases and 5' codes specific for the EraCode beads described above. During TSE, TiTaq incorporates labels on extenders only when specific targets having 5' iCs are present. The nonnatural base pair formed by iG and iC has been successfully used for both molecular recognition (13–15) and for site-specific enzymatic incorporation (16–18). To date, we have covalently coupled a series of different reporter groups to diGTP, using succinimidyl ester chemistry and confirmed enzymatic incorporation into nucleic acids by use of several commercially available polymerases (data beyond the scope of this report). In many cases, single amplicons spanning regions that incorporated multiple mutations were used as templates for multiple extenders. After TSE, EraCode-labeled beads and streptavidin-phycoerythrin were added, and the reactions were analyzed on the Luminex<sup>100</sup>. Data are reported in median fluorescence intensity (MFI).

For each polymorphism, pairs of TSE oligonucleotide primers were designed by use of Primer 3 and Hyther (see Table 3 in the online Data Supplement). Pairs were designed to have balanced melting temperatures ranging from 63 to 67  $^{\circ}\text{C}$ . Each TSE oligonucleotide was composed of three components: a 5' sequence tag complementary to one of the EraCodes; a three-carbon spacer; and a 3' target-specific sequence fully complementary to the allele of interest. The polymorphic sequence for each pair of TSE oligonucleotides was the 3' base.

Extension reactions were performed by bringing the 8- $\mu\text{L}$  PCR reactions to 10  $\mu\text{L}$  with addition of 2.0  $\mu\text{L}$  of EraGen MC TSE solution 35 (prod. no. PN1236; EraGen Biosciences) including primers for the 579T system (prod. no. PN1310), and 0.4  $\mu\text{L}$  of 20 $\times$  EraGen CFTR TSE primer mixture; final concentrations of TSEs were 25–100 nmol/L (see Table 3 in the online Data Supplement). Reactions were cycled according to the following profile: 1 cycle of 30 s at 95  $^{\circ}\text{C}$ ; 5 cycles of 1 s at 95  $^{\circ}\text{C}$  and 2 min at 65  $^{\circ}\text{C}$ .

#### DATA ANALYSIS

For this report, an assignment is defined as a genotype for a given target within the *CFTR* gene made for any given sample. Each sample tested will have 30 possible assignments because 30 targets for each sample are analyzed. To rapidly analyze the data obtained, we developed software to present the raw data from the instrument. The software directly imports the Luminex<sup>100</sup> data and organizes them

by target mutation. Fractions of wild-type to wild-type-plus-mutant and raw MFI signals can be visualized graphically (Fig. 2). The program sets default fraction windows at 1.0–0.8 for wild-type samples, 0.6–0.4 for heterozygous samples, and 0.0–0.2 for homozygous mutant samples. Default windows can be changed by the end user. The area between the windows is designated as the no-call window. The program also sets no-amplification windows where wild-type and mutant MFIs are <200.

Template set-up within the Luminex LabMap software was required, and importation to analysis software was achieved by dragging the Luminex file folder into the Analysis Software. Data files were parsed, and the resulting raw MFI values were organized by target and sample. After data acquisition from all 225 clinical samples tested, default cutoff windows for each target were empirically determined and set in a blinded fashion. Once determinations were made, reports were generated for offline analysis. The analysis software was written in Java and is compatible with Windows, Mac OS X, and Linux operating systems. Data importation is achievable with Luminex LabMap 100 (software versions 1.7, 2.1, and 2.2) and Bio-Rad BioPlex 3.0. The software displays scatter plots and fraction plots of each given target for all samples.

#### AUTOMATION

Automation was performed on a Tecan Genesis Model 200 robotic workstation (TRP) starting from either the PCR amplicon formation or from isolated genomic DNA and continuing through data analysis. Samples used were a subset of the 225 used in the initial manual testing. Instructions for the automation program were written with Gemini software (Tecan). Initialization and launch of the bead-reading protocol on the Luminex<sup>100</sup> was performed automatically by use of a custom image-recognition and command-scripting program written in Java. After the read, the results were remotely analyzed by MultiCode Analysis software. Genomic DNA was purified manually before automation as described above. The TRP was fitted with the low-volume eight-channel dispense pipette system, TeStack, to supply disposable tips, a MJ Research thermocycler equipped with remote Alpha docks with Power Bonnet, and a robotic manipulator arm. Liquid reagents and disposable tips were also placed on the deck before runs were started. Reactions and additions for the automated process were the same as those described for the manual method unless noted. Reactions were performed in low-profile hard-skirted polypropylene well microplates, which allowed the robotic arm to grip and move the plates as needed. Each reaction was overlaid with 15  $\mu\text{L}$  of light mineral oil (Sigma) to prevent evaporation during thermal cycling. The robotic manipulator was used to transfer the plates to and from the thermocycler block and the Luminex<sup>100</sup> mounted off deck. Initialization and launch of the bead-reading protocol on the Luminex<sup>100</sup> were performed automatically by use of a custom image-recognition and command-scripting pro-

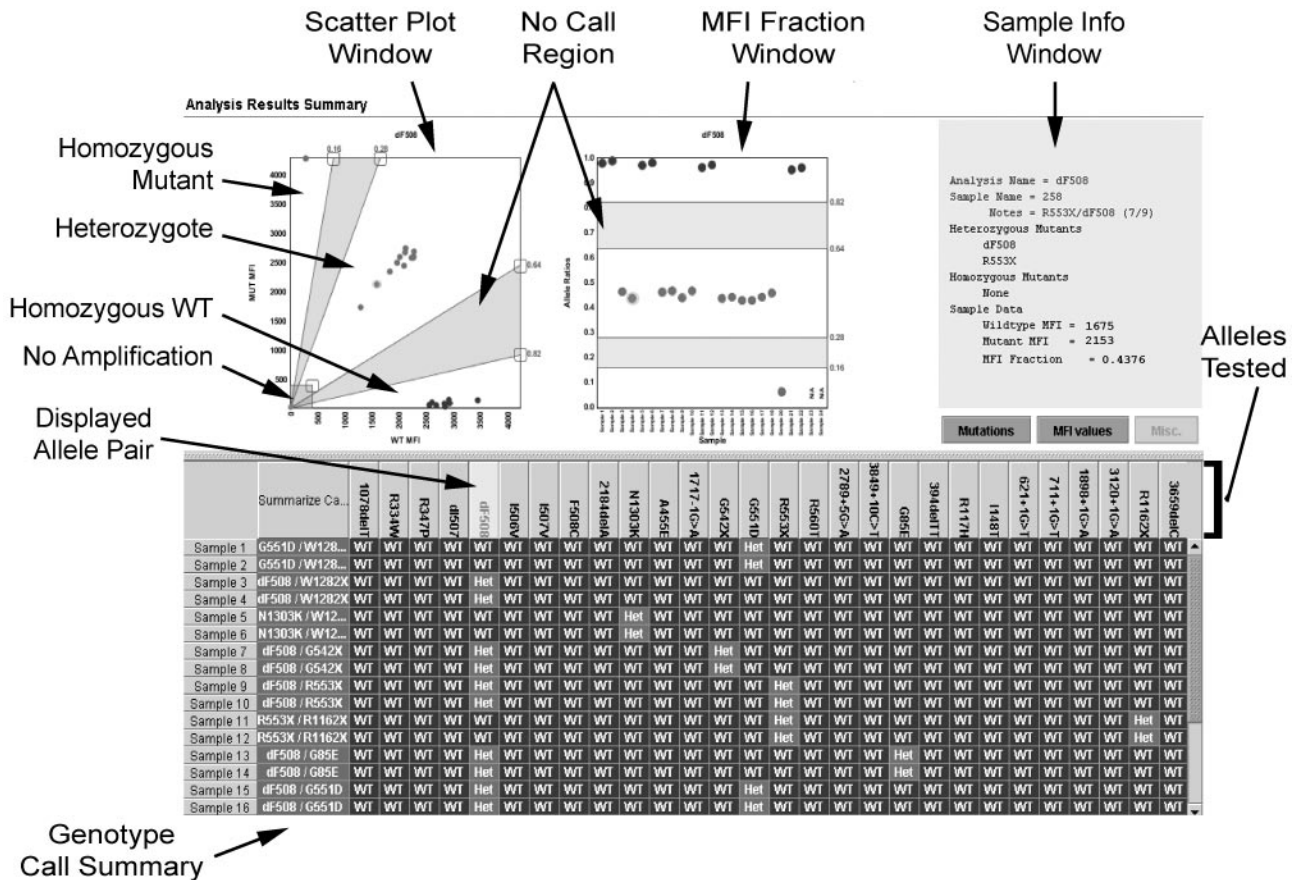


Fig. 2. Visual interface using the MultiCode analysis software developed for the *CFTR* mutation analysis test. The data are displayed in various forms: a scatter plot with raw MFI (top left); MFI fractions (top middle); an individual sample summary, which can include mutation determination and specific numerical MFI (top right); and graphical determinations for all samples and targets analyzed (bottom). WT, wild type.

gram written in Java. After bead-reading of each reaction, the results were analyzed by the MultiCode Analysis software as described above.

**Results**

**PRINCIPLES OF THE MULTICODE SYSTEM**

A schematic diagram of the MultiCode system is shown in Fig. 3. The MultiCode process has three steps. For the first step, regions of the target of interest are amplified by PCR using primers that contain a single 5'-iC. In the second step, the amplicons from the first step are used as templates for the TSEs, and iC directs specific diGTP-biotin incorporation. The TSEs are coded on their 5' ends with sequences that specifically recognized short complementary sequences (EraCodes) attached to an addressed solid matrix. The molecular recognition of any given code with its EraCode is specific enough so that increased temperatures, incubation steps, or washings are not required (typical requirements for hybridization capture using only naturally occurring base chemistries). During this second step, labels are placed site specifically on the TSEs. This process requires the presence of iC-containing amplicons that are complementary to the TSE. When the correct target is available, the extension reaction continues

to the 5' end of the amplicon, where the diGTP-biotin is used to insert the biotin into the TSE. Specificity and signal consistency arise when coded TSEs are covalently attached to a single biotin reporter in the presence of the correct target. In the final step, the TSEs are captured on the solid matrix and detected. The presence of the target is then determined by where the reporters are located spatially on the liquid matrix (created by the differentially colored microspheres).

**VALIDATION OF THE EraCode MOLECULAR RECOGNITION SYSTEM AND diGTP-BIOTIN INCORPORATION**

Because each EraCode should be specific only for its complementary tag, only the bead that contains the specific EraCode should report a high MFI. Results show that for the correct complementary code, specific signal was between 2130 and 4357 MFI, with a mean (SD) MFI of 3040 (532; see Table 4 in the online Supplemental Data). Signals generated on microspheres that contained EraCodes that were not completely complementary varied from 0 to 320 MFI with a mean (SD) MFI of 92 (28). This set gave signal-to-noise ratios that ranged from 32:1 and 10:1 for the worst cases (i.e., where the specific signal

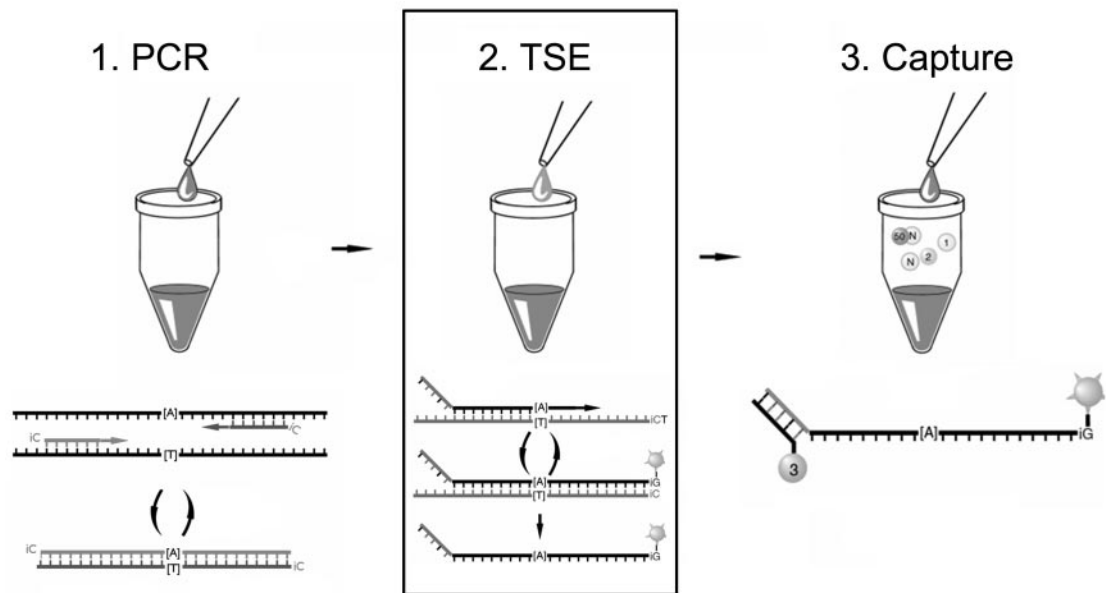


Fig. 3. Schematic of MultiCode PLx platform format.

(Step 1), target regions are amplified by PCR to produce amplicons with iC attached to their 5' ends. (Step 2), tagged target-specific extenders are site-specifically labeled with diGTP-label when the correct amplicon is present. For this study, the label was biotin, and streptavidin-phycoerythrin was used for signaling purposes. (Step 3), the tags are captured on EraCode-modified Luminex microspheres. The captured reactions are analyzed with the Luminex<sup>100</sup> system.

was compared with the nonspecific signal from the microsphere presenting the highest noise). To our understanding, there have been no reports of a genetic-based molecular recognition system that was successfully implemented at room temperature without washing.

#### MUTATION ANALYSIS USING THE MANUAL CF MULTICODE SYSTEM

PCR amplicons were generated for 17 regions within the *CFTR* gene, and 63 tagged extenders were used to analyze 27 mutations and 4 reflex targets. To improve the robustness of the overall results, the test was divided into four reactions with one additional test targeting the 5T7T9T reflex target. Two additional targets were included late in the development phase to include 394delTT and 3199del6. The test was transferred to the Wisconsin State Laboratory of Hygiene, where newborn *CFTR* genetic screening

has been ongoing since 1994 and multimutation testing used since 2002.

For the *CFTR*-specific PLx assay, positive signal intensities varied between 2000 and 5000 MFI and background signal varied from 50 to 1500 MFI. The variation in signal depended strictly on the target sequence. We observed that determination windows varied from 0.99 to 0.7 for wild-type samples, from 0.4 to 0.6 for heterozygous samples, and from 0.0 to 0.2 for homozygous mutant samples.

The calling results are summarized in Table 1 [also see Table 4 in the online Data Supplement for additional details]. The 225 samples included 65 wild-type samples; 93 heterozygous samples representing 22 types; 36 compound heterozygous samples for 9 types; and 31 homozygous mutant samples for 6 types. Of the 225 samples tested manually, 203 (90.2%) samples reported all correct assignments. The failure rate of 10% was most likely a

**Table 1. Summary of the data obtained in this study using the MultiCode PLx *CFTR* Multi-Mutation Analysis System.**

	<b>A</b>	<b>1</b>	<b>2</b>	<b>B</b>
Total samples tested	225	20	66	419
Wild-type samples	65	2	16	343
Heterozygous samples/types	93/22	11/10	27/16	72/9
Compound heterozygous samples/types	36/9	6/6	19/17	4/3
Homozygous mutant samples/types	31/6	1/1	4/4	0
Samples showing 100% correct assignments	203 (90.2%)	0 (0.0%)	64 (97.0%)	398 (95%)
Incorrect assignments	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Number of correct overall assignments	6704 (99.3%)	600 (100%)	1974 (99.7%)	12 521 (99.6%)

<sup>a</sup> The assay was performed either manually at the Wisconsin State Laboratory of Hygiene (columns A and B) or at EraGen Biosciences using complete automation (column 1) or automation started after the PCR step (column 2). The data were obtained by use of a set of banked samples (columns A, 1, and 2) or fresh samples (column B).

product of DNA quality and not assay sensitivity. As stated in the *Materials and Methods*, the DNA samples were not fresh, and signal intensities indicated gross failure consistent with PCR failure. There were no discrepancies for genotype classification. For these 225 samples, there were a total of 6750 potential assignments ( $225 \times 27$  mutations + 3 reflex tests; 5T7T9T was not included). Correct assignments were made 99.3% (6704 of 6750) of the time. For the separate 5T7T9T-specific reflex test, 194 samples were tested and included at least 10 separate samples from each of the six possibilities (see Table 4 and Fig. 1 in the online Data Supplement). Test results showed that there were zero incorrect determinations and one no amplification failure.

#### MUTATION ANALYSIS USING THE AUTOMATED CF MULTICODE SYSTEM

The final automated processes for the *CFTR* gene were conducted on a total of 86 samples. The results are reported from two separate processes: (1) PCR through data acquisition (20 samples) and (2) TSE through data acquisition (66 samples; Table 1; also see Table 5 in the online Data Supplement for further sample and testing information). Because workstation space is limited on the Tecan Genesis Model 200, tip usage is high when disposable tips are implemented, and amplicon carryover may be a concern. Process 1 may be applied for higher throughput scenarios where the PCR step could be implemented on yet another workstation or manually. For process 1, 20 (100%) samples reported calls correct. For process 2, 64 (97%) samples reported all correct assignments. There were no incorrect determinations.

#### COMPARATIVE ANALYSIS

After assay validation, which included the testing of various mutations and sample types, the PLx assay for *CFTR* gene analysis was directly compared with the CF Gold linear array method. Using freshly obtained blood cards from newborns that had immunoreactive trypsinogen scores in the upper 4%, we used both tests to analyze *CFTR* for the recommended core panel of mutations and reflex targets. The testing was performed on 419 samples, and 100% concordance was recorded (no incorrect assignments; Table 1, column B). Of the samples tested, there were 21 samples (5%) in which the complete lists of calls could not be obtained with the PLx system. The failure rate of 5% was documented as errors resulting from pipetting errors. These samples were retested, and correct assignments were made. Of the possible 12 570 potential assignments, 12 521 (99.6%) were correct. Run time after sample preparation for the CF Gold test was  $\sim 7$  h compared with  $< 3$  h for the PLx system. Total hands-on time for the CF Gold test was  $\sim 3.5$  h compared with 0.5 h for the PLx system. The data output for the CF Gold test was visual inspection of the nitrocellulose strips, whereas the PLx system has a computerized calling system as described.

#### Discussion

We have described a novel high-throughput platform that incorporates nucleic acid chemistries that simplify genetic analysis. PLx uses one additional base pair not found in nature in combination with microsphere flow cytometry technology to eliminate the most cumbersome steps commonly found in established multiplexed systems. The EraCode molecular recognition system described is particularly suited for the Luminex<sup>100</sup> instrumentation because the solid support can be dispensed into the complex reaction and decoding can be done at room temperature without washing, creating a true "Liquid Chip" type of system. Our study indicates that the chemistry is transferable to a clinical setting. In addition, although not required, the process can be performed with a robotic workstation for higher throughput.

We demonstrated the principles of the chemistry by use of an essential target and selected mutations found within the *CFTR* gene recommended by leading medical organizations. Data from the Wisconsin State Laboratory of Hygiene confirmed that the chemistry is accurate. For example, there were no incorrect determinations, and the percentage of correct determinations was high. The data presented here were tabulated by use of a broad sample set that included every mutation possible within the panel targeted with the exception of the 3199del6. For 24 samples, the entire manual process takes  $\sim 3$  h to obtain final results. An upgraded prototype version of the *CFTR* test where all mutations are analyzed in a single well takes  $\sim 2.5$  h to complete analysis of 96 patient samples (validation data not yet available). To demonstrate the system in a fully automated mode, we elected to use the Tecan genesis liquid-handling system. The entire automated process also takes  $\sim 3$  h from prepared sample to final read for an entire 96-well plate or 24 samples. We presented data from two possibilities (when automation started at the PCR step and when it started at the TSE step) because there may be a need to perform PCR offline.

We believe that the expanded genetic alphabet enabled us to eliminate many of the steps that are required in technologies based on just the two base pairs provided by nature. The steps eliminated include solid support washings, transfers, aspirations, and high-temperature hybridizations. Specifically, we believe that there are many reasons that the extra base pair is beneficial. For example, we used short code sequences in the capture step to decode the complex extension mixtures. This may be possible with code sets constructed from only A, G, C, or T, but optimization would surely be more difficult because of cross-hybridization with sequence complements or close complements found in nature. The additional informational content that iC:iG builds into DNA also allows for greater diversity when compared with their all-natural counterparts. Specifically, if the four naturally occurring bases are used to make a library of 10mers, there are  $4^{10}$ , or  $10^6$  possibilities. In contrast, a six-letter system generates  $6^{10}$ , or  $10^8$  unique 10mers. As demon-

strated here, specific assembly can occur at conditions such as room temperature, which is in effect more amenable to and simplifies molecular diagnostics.

We also demonstrated for the first time that biotin labels attached to diGTP could be site specifically incorporated into extension primers. More generally, we demonstrated that labeled diGTP can be incorporated enzymatically into DNA. This could allow for broad use of such compounds in molecular biology. Whenever a reporter group is required at a specific site within an amplicon, transcript, or extension product, the use of an additional base pair may be helpful.

Recently we reported that site-specific enzymatic incorporation of a quencher attached to diGTP can be used to create a new and highly sensitive real-time PCR system (26). Polymerase incorporation of labels only at sites where expanded genetic alphabet bases are positioned will also have other uses, such as 3' end labeling. Polymerase incorporation of EraCodes within PCR amplicons has now also been reported (27). The EraCode system that uses additional nonnatural base pairing may have advantages over other DNA-based molecular recognition systems (28–30). These short DNA codes are simple to produce, specifically hybridize at room temperature, do not require wash steps, can be amplified with standard polymerases, and do not cross-hybridize with natural DNA. Taken together, we believe that the expansion of the genetic alphabet is a true paradigm shift to the ways molecular diagnostics will be developed in the future.

We thank the National Institute of Diabetes and Digestive and Kidney Diseases for supporting this work (DK62743). We thank Dr. David Barton at the National Centre for Medical Genetics (Dublin, Ireland) and Dr. Claud Ferec at the Laboratoire de Genetique Moleculaire et d'Histocompatibilite (Brest, France) for supplying us with samples and invaluable assistance for this project. We also thank Dr. Susan H. Bernacki of Duke University Medical Center (Department of Pathology, Durham, NC) for providing additional samples. Finally, we would like to thank Dr. Michael Moser for reviewing the manuscript and providing assistance with figures and tables.

## References

1. Servais J, Lambert C, Fontaine E, Plessier JM, Robert I, Arendt V, et al. Comparison of DNA sequencing and a line probe assay for detection of human immunodeficiency virus type 1 drug resistance mutations in patients failing highly active antiretroviral therapy. *J Clin Microbiol* 2001;39:454–9.
2. Day NS, Tadin M, Christiano AM, Lanzano P, Piomelli S, Brown S. Rapid prenatal diagnosis of sickle cell diseases using oligonucleotide ligation assay coupled with laser-induced capillary fluorescence detection. *Prenat Diagn* 2002;22:686–91.
3. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR, Jr. Advanced multiplexed analysis with the FlowMetrix system. *Clin Chem* 1997;43:1749–56.
4. Smith PL, WalkerPeach CR, Fulton RJ, DuBois DB. A rapid, sensitive, multiplexed assay for detection of viral nucleic acids using the FlowMetrix system. *Clin Chem* 1998;44:2054–6.
5. The San Diego Conference. *Nucleic Acid Technology: The Cutting Edge of Discovery*. November 6–8, 1997. Abstracts. *Clin Chem* 1997;43:2213–24.
6. Wedemeyer N, Potter T, Wetzlich S, Gohde W. Flow cytometric quantification of competitive reverse transcription-PCR products. *Clin Chem* 2002;48:1398–405.
7. Dunbar SA, Jacobson JW. Application of the Luminex LabMAP in rapid screening for mutations in the cystic fibrosis transmembrane conductance regulator gene: a pilot study. *Clin Chem* 2000;46:1498–500.
8. Colinas RJ, Bellisario R, Pass KA. Multiplexed genotyping of  $\beta$ -globin variants from PCR-amplified newborn blood spot DNA by hybridization with allele-specific oligodeoxynucleotides coupled to an array of fluorescent microspheres. *Clin Chem* 2000;46:996–8.
9. Ross JS. Financial determinants of outcomes in molecular testing. *Arch Pathol Lab Med* 1999;123:1071–5.
10. Lee DS, Rosenberg MA, Peterson A, Makhholm L, Hoffman G, Laessig RH, et al. Analysis of the costs of diagnosing cystic fibrosis with a newborn screening program. *J Pediatr* 2003;142:617–23.
11. Poon LL, Chan KH, Wong OK, Yam WC, Yuen KY, Guan Y, et al. Early diagnosis of SARS coronavirus infection by real time RT-PCR. *J Clin Virol* 2003;28:233–8.
12. Firmani MA, Broussard LA. Molecular diagnostic techniques for use in response to bioterrorism. *Expert Rev Mol Diagn* 2003;3:605–16.
13. Kern D, Collins M, Fultz T, Detmer J, Hamren S, Peterkin JJ, et al. An enhanced-sensitivity branched-DNA assay for quantification of human immunodeficiency virus type 1 RNA in plasma. *J Clin Microbiol* 1996;34:3196–202.
14. Moser MJ, Marshall DJ, Grenier JK, Kieffer CD, Killeen AA, Ptacin JL, et al. Exploiting the enzymatic recognition of an unnatural base pair to develop a universal genetic analysis system. *Clin Chem* 2003;49:407–14.
15. Sherrill CB, Marshall DJ, Richmond CS, Scherrer CW, Prudent JR. Molecular recognition on demand. *Nanotech* 2003;1:52–4.
16. Piccirilli JA, Krauch T, Moroney SE, Benner SA. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 1990;343:33–7.
17. Switzer CY, Moroney SE, Benner SA. Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* 1993;32:10489–96.
18. Moser MJ, Prudent JR. Enzymatic repair of an expanded genetic information system. *Nucleic Acids Res* 2003;31:5048–53.
19. Grody WW, Cutting GR, Klinger KW, Richards CS, Watson MS, Desnick RJ. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* 2001;3:149–54.
20. Rohlfes EM, Zhou Z, Sugarman EA, Heim RA, Pace RG, Knowles MR, et al. The 1148T CFTR allele occurs on multiple haplotypes: a complex allele is associated with cystic fibrosis. *Genet Med* 2002;4:319–23.
21. Farrell PM, Mischler EH, Fost NC, Wilfond BS, Tluczek A, Gregg RG, et al. Current issues in neonatal screening for cystic fibrosis and implications of the CF gene discovery. *Pediatr Pulmonol Suppl* 1991;7:11–8.
22. Farrell PM, Kosorok MR, Rock MJ, Laxova A, Zeng L, Lai HC, et al. Early diagnosis of cystic fibrosis through neonatal screening prevents severe malnutrition and improves long-term growth. Wisconsin Cystic Fibrosis Neonatal Screening Study Group. *Pediatrics* 2001;107:1–13.



23. Farrell PM, Aronson RA, Hoffman G, Laessig RH. Newborn screening for cystic fibrosis in Wisconsin: first application of population-based molecular genetics testing. *Wis Med J* 1994;93:415–21.
24. Fost N, Farrell PM. A prospective randomized trial of early diagnosis and treatment of cystic fibrosis: a unique ethical dilemma. *Clin Res* 1989;37:495–500.
25. Farrell PM, Kosorok MR, Laxova A, Shen G, Koscik RE, Bruns WT, et al. Nutritional benefits of neonatal screening for cystic fibrosis. Wisconsin Cystic Fibrosis Neonatal Screening Study Group. *N Engl J Med* 1997;337:963–9.
26. Sherrill CB, Marshall DJ, Moser MJ, Larsen CA, Daude-Snow L, Prudent JR. Nucleic acid analysis using an expanded genetic alphabet to quench fluorescence. *J Am Chem Soc* 2004;126:4550–6.
27. Johnson SC, Sherrill CB, Marshall DJ, Moser MJ, Prudent JR. A third base pair for the polymerase chain reaction—inserting isoC and isoG. *Nucleic Acids Res* 2004;32:1937–41.
28. Li Y, Tseng YD, Kwon SY, D'Espaux L, Bunch JS, McEuen PL, et al. Controlled assembly of dendrimer-like DNA. *Nat Mater* 2004;3:38–42.
29. Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol* 1999;292:251–62.
30. Kool ET. Replacing the nucleobases in DNA with designer molecules. *Acc Chem Res* 2002;35:936–43.