Article No. jmbi.1999.3063 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 292, 251–262

JMB

Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations

Norman P. Gerry[1], Nancy E. Witowski[2], Joseph Day[1], Robert P. Hammer[3], George Barany[2] and Francis Barany[1]*

[1]*Department of Microbiology Hearst Microbiology Research Center, and Strang Cancer Prevention Center, Joan and Sanford I. Weill Medical College of Cornell University 1300 York Ave., Box 62, New York, NY 10021, USA*

[2]*Departments of Chemistry and Laboratory Medicine & Pathology, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455, USA*

[3]*Department of Chemistry Louisiana State University 232 Choppin Hall, Baton Rouge LA 70803, USA*

**Corresponding author*

Cancers arise from the accumulation of multiple mutations in genes regulating cellular growth and differentiation. Identification of such mutations in numerous genes represents a significant challenge in genetic analysis, particularly when the majority of DNA in a tumor sample is from wild-type stroma. To overcome these difficulties, we have developed a new type of DNA microchip that combines polymerase chain reaction/ligase detection reaction (PCR/LDR) with ''zip-code'' hybridization. Suitably designed allele-specific LDR primers become covalently ligated to adjacent fluorescently labeled primers if and only if a mutation is present. The allele-specific LDR primers contain on their 5′-ends ''zip-code complements'' that are used to direct LDR products to specific zip-code addresses attached covalently to a three-dimensional gel-matrix array. Since zip-codes have no homology to either the target sequence or to other sequences in the genome, false signals due to mismatch hybridizations are not detected. The zip-code sequences remain constant and their complements can be appended to any set of LDR primers, making our zip-code arrays universal. Using the K-*ras* gene as a model system, multiplex PCR/LDR followed by hybridization to prototype 3×3 zip-code arrays correctly identified all mutations in tumor and cell line DNA. Mutations present at less than one per cent of the wild-type DNA level could be distinguished. Universal arrays may be used to rapidly detect low abundance mutations in any gene of interest.

Keywords: zip-code addressing; DNA hybridization; thermostable DNA ligase; ligase detection reaction; single nucleotide polymorphism (SNP)

Introduction

Cancers arise from the accumulation of mutations in genes controling the cell cycle, apoptosis, and genome integrity. These mutations may be inherited or somatic, arising from exposure to environmental factors or from malfunctions in DNA replication and repair machinery (Fearon, 1997; Fearon & Vogelstein, 1990; Liu *et al*., 1996; Perera, 1997). Oncogenes may be activated by point mutations, translocation, or gene amplification, while tumor suppressor genes may be inactivated by point mutations, frameshift mutations and deletions (Bishop, 1991; Da Costa *et al*., 1996; Venitt, 1996). A major hurdle to detecting mutations in these genes is that, in primary tumors, normal stromal cell contamination can be as high as 70% of total cells, and thus a mutation present in only one of the two chromosomes of a tumor cell may represent as little as 15% of the DNA sequence present in a sample for that gene. Thus, there is an urgent need to develop technology that can identify accurately one or more low abundance mutations, at multiple adjacent, nearby, and distal loci in a large number of genes.

The advent of DNA arrays has resulted in a paradigm shift in detecting sequence variations and monitoring gene expression levels on a genomic scale (Beattie *et al*., 1995; Brown & Botstein, 1999; Chee *et al*., 1996; Cronin *et al*., 1996; DeRisi *et al*., 1996; Drobyshev *et al*., 1997; Eggers *et al*., 1994; Gunderson *et al*., 1998; Guo *et al*., 1994; Hacia, 1999; Hacia *et al*., 1996; Kozal *et al*., 1996;

Abbreviations used: LDR, ligase detection reaction; FAM, 6-carboxyfluorescein; Mes, 2-(*N*-morpholino) ethanesulfonic acid; SNP, single nucleotide polymorphism.

E-mail address of corresponding author: barany@mail.med.cornell.edu

0022-2836/99/370251–12 $30.00/0

Pease *et al*., 1994; Schena *et al*., 1996; Shalon *et al*., 1996; Southern *et al*., 1999; Yershov *et al*., 1996; Zhu *et al*., 1998). DNA chips designed to distinguish single nucleotide differences are generally based on hybridization of labeled targets (Beattie *et al*., 1995; Chee *et al*., 1996; Cronin *et al*., 1996; Drobyshev *et al*., 1997; Eggers *et al*., 1994; Guo *et al*., 1994; Hacia *et al*., 1996; Kozal *et al*., 1996; Parinov *et al*., 1996; Sapolsky *et al*., 1999; Wang *et al*., 1998; Yershov *et al*., 1996) or polymerase extension of arrayed primers (Lockley *et al*., 1997; Nikiforov *et al*., 1994; Pastinen *et al*., 1997; Shumaker *et al*., 1996). While DNA chips based on these two formats can confirm a known sequence, the similarities in hybridization profiles create ambiguities in distinguishing heterozygous from homozygous alleles (Beattie *et al*., 1995; Chee *et al*., 1996; Eggers *et al*., 1994; Kozal *et al*., 1996; Southern, 1996; Wang *et al*., 1998). To overcome this problem, several methods have been proposed, including the use of: (i) two-color fluorescence analysis (Hacia *et al*., 1996, 1998a); (ii) a tiling strategy that uses 40 overlapping addresses for each known polymorphism (Cronin *et al*., 1996); (iii) incorporation of nucleotide analogues in the array sequence (Guo *et al*., 1997; Hacia *et al*., 1998b); and (iv) adjacent cohybridized oligonucleotides (Drobyshev *et al*., 1997; Gentalen & Chee, 1999; Yershov *et al*., 1996). A recent side-by-side comparison revealed that the use of hybridization chips for nucleotide discrimination gave an order of magnitude higher background than was observed with the primer extension approach, resulting in an increased likelihood of false positive identifications (Pastinen *et al*., 1997). Nevertheless, solid-phase primer extension can also generate false positive signals from mononucleotide repeat sequences, template-dependent errors, and template-independent errors (Nikiforov *et al*., 1994; Shumaker *et al*., 1996). In addition, neither of these two types of arrays can detect cancer mutations when these are present in a minority of the total target DNA.

Over the past few years, our laboratories have pursued an alternate strategy in DNA array design. In concert with polymerase chain reaction/ligase detection reaction (PCR/LDR) assays carried out in solution (Barany, 1991a,b; Belgrader *et al*., 1996; Day *et al*., 1995, 1996; Khanna *et al*., 1999), our array concept allows for accurate identification of mutations and single nucleotide polymorphisms (SNPs). Primary PCR amplification of the gene of interest is followed by LDR, which uses a thermostable *Tth* DNA ligase that links two adjacent oligonucleotides annealed to a complementary target if and only if the nucleotides are perfectly base-paired at the junction (Figure 1(a)). Since a single-base mismatch prevents ligation, it is possible to distinguish mutations with exquisite specificity, even at low abundance (Khanna *et al*., 1999). Furthermore, such assays are ideal for multiplexing, since several primer sets can ligate along a gene without the interference encountered in polymerase-based assays (Belgrader *et al*., 1996; Day *et al*., 1995; Khanna *et al*., 1999). High-throughput detection of specific multiplexed LDR products is then achieved *via* divergent sequences termed ''zip-code'' complements which guide each LDR product to a designated zip-code address on a DNA array (Figure 1(b)). This concept is analogous to molecular tags developed for bacterial and yeast genetics (Hensel *et al*., 1995; Shoemaker *et al*., 1996). Based on recent multiplexed PCR/LDR results from our laboratory, the new approach should allow detection of: (i) dozens to hundreds of polymorphisms in a single-tube multiplex format; (ii) small insertions and deletions in repeat sequences; and (iii) low abundance mutations in a background of normal DNA (Khanna *et al*., 1999, and unpublished results).

Results and Discussion

Zip-code concept and design

Our approach uses microarrays of unique 24-base oligonucleotides that are coupled to a three-dimensional polymer at known locations. These 24-mers or zip-codes (Table 1) hybridize specifically to molecules containing sequences that are complementary to the zip-codes. By linking the zip-code complements to fluorescent primers *via* a tandem PCR/LDR strategy, zip-code microarrays can be used to assess the presence and abundance of mutations in biological specimens. Importantly, because the zip-codes represent unique artificial sequences, zip-code microarrays can be used as a universal platform for molecular recognition simply by changing the gene-specific sequences linked to the zip-code complements.

Each zip-code sequence is composed of six tetramers (designed as described below) such that the full-length 24-mers have similar t_m values. The 256 (4^4) possible combinations in which the four bases can be arranged as tetramers were reduced to a set of 36; these were chosen such that each tetramer differed from all others by at least two bases (Figure 2). Tetramer complements, as well as tetramers that would result in self-pairing or hairpin formation of the zip-codes, were eliminated. Furthermore, tetramers that were palindromic, e.g. TCGA, or repetitive, e.g. CACA, were excluded (diagonally hatched boxes in Figure 2). The indicated set of 36 tetramers represents just one of the possible sets that can be created; alternative sets can be developed by starting in any of the unused light gray boxes (Figure 2).

Six tetramers were chosen from the larger set of 36 for use in designing the zip-codes for the prototype array. These six tetramers were combined such that each zip-code differs from all others by at least three alternating tetramer units (Table 1). This ensures that each zip-code differs from all other zip-codes by at least six bases, thus preventing even the closest zip-code sequences from cross-hybridizing. The t_m values of correct hybridizations range from 70 °C to 82 °C and are at least 24 deg. C

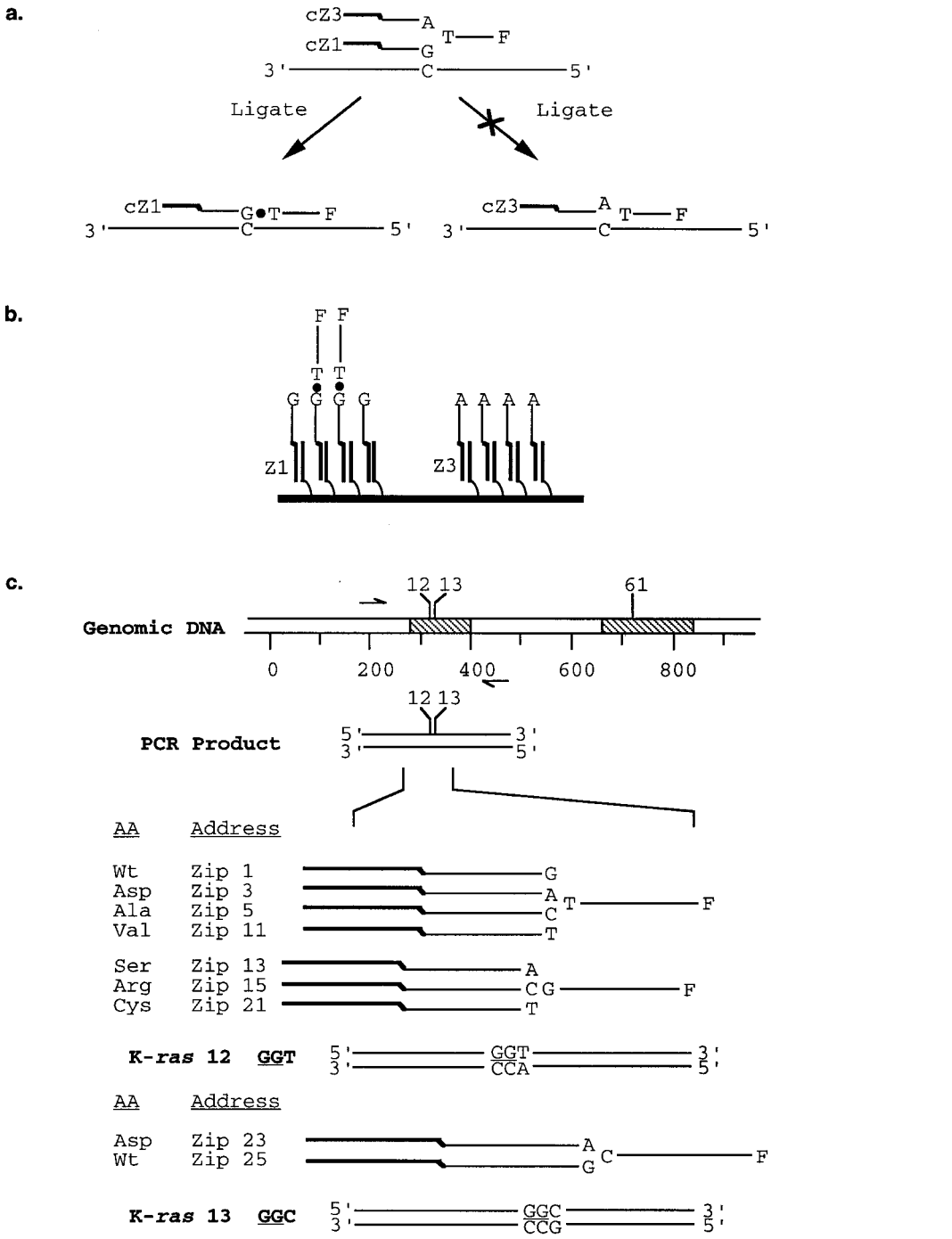

Figure 1. Scheme for PCR/LDR detection of mutations using an addressable array. (a) Schematic representation of LDR primers used to distinguish mutations. Each allele-specific primer contains an addressable sequence complement (cZ1 or cZ3) on the 5′-end and the discriminating base on the 3′-end. The common LDR primer is phosphorylated on the 5′-end and contains a fluorescent label on the 3′-end. The primers hybridize adjacent to each other on target DNA, and the nick will be sealed by the ligase if and only if there is perfect complementarity at the junction. (b) The presence and type of mutation is determined by hybridizing the contents of an LDR to an addressable DNA array. The zip-code sequences are designed to be sufficiently different, so that only primers containing the correct complement to a given zip-code will remain bound at that address. (c) Schematic representation of chromosomal DNA containing the K-*ras* gene. Exons are shaded and the positions of codons 12 and 13 are shown. Exon-specific primers were used to selectively amplify K-*ras* DNA flanking codons 12 and 13. Primers were designed for LDR detection of seven possible mutations in these two codons as described in (a).

higher than that of any incorrect hybridization (calculated using Oligo 6.0, Molecular Biology Insights, Inc., Cascade, CO). The concept of using alternating rows and columns of tetramer units may be extended to include all 36 tetramers, hence creating an array with 1296 divergent addresses.

Table 1. Zip-code sequences used in prototype array

Zip#	Tetramer order[a]	Zip-code sequence (5′ → 3′)[b]
Zip1	1-6-3-2-6-3	TGCG-ACCT-CAGC-ATCG-ACCT-CAGC-spacer-NH_2
Zip3	3-6-5-2-2-3	CAGC-ACCT-GACC-ATCG-ATCG-CAGC-spacer-NH_2
Zip5	5-6-1-2-4-3	GACC-ACCT-TGCG-ATCG-GGTA-CAGC-spacer-NH_2
Zip11	1-4-3-6-6-1	TGCG-GGTA-CAGC-ACCT-ACCT-TGCG-spacer-NH_2
Zip13	3-4-5-6-2-1	CAGC-GGTA-GACC-ACCT-ATCG-TGCG-spacer-NH_2
Zip15	5-4-1-6-4-1	GACC-GGTA-TGCG-ACCT-GGTA-TGCG-spacer-NH_2
Zip21	1-2-3-4-6-5	TGCG-ATCG-CAGC-GGTA-ACCT-GACC-spacer-NH_2
Zip23	3-2-5-4-2-5	CAGC-ATCG-GACC-GGTA-ATCG-GACC-spacer-NH_2
Zip25	5-2-1-4-4-5	GACC-ATCG-TGCG-GGTA-GGTA-GACC-spacer-NH_2

[a] Order of tetramer oligonucleotide segments in the corresponding zip-code sequence. Six tetramers were chosen from the full set of 36 to prepare the zip-codes for the prototype array. The six tetramers which were renumbered for ease of use are: 1, TGCG; 2, ATCG; 3, CAGC; 4, GGTA; 5, GACC; and 6, ACCT. Closely related sequences, (Zip1, 3, 5), (Zip11, 13, 15) and (Zip21, 23, 25) differ at the first, third, and fifth tetramer positions, but are identical at the second, fourth, and sixth tetramer positions.

[b] spacer-NH_2 = -$O(PO_2)O$-$(CH_2CH_2O)_6$-PO_2-$O(CH_2)_3NH_2$.

Array preparation

Numerous types of two and three-dimensional matrices were examined with respect to: (i) ease of preparation of the surface; (ii) oligonucleotide loading capacity; (iii) stability to conditions required for coupling of oligonucleotides, as well as for hybridization and washing; and (iv) compatibility with fluorescence detection. Our currently favored methodology to construct zip-code arrays involves initial creation of a lightly crosslinked film of acrylamide/acrylic acid copolymer on a glass solid support; subsequently, the free carboxyl groups dispersed randomly throughout the polymeric surface are activated with *N*-hydroxysuccinimide, and amine terminated zip-code oligonucleotide probes are added to form covalent amide linkages (Figure 3(a)). The described coupling chemistry is rapid, straightforward, efficient, and amenable to both manual and robotic spotting. Both the activated surfaces and the surfaces with attached oligonucleotides are stable to long-term storage.

Optimization of hybridization conditions

Hybridizations of a fluorescently labeled 70-mer probe onto model zip-code arrays were measured as a function of buffer, metal cofactors, volume, pH, time, and the mechanics of mixing (Table 2). Even with closely related zip-codes, cross-hybridization was negligible or non-existent, with a signal-to-noise ratio of at least 50:1. Our experiments suggest that different zip-codes hybridize at approximately the same rate, i.e. the level of fluorescent signal is relatively uniform when normalized for the amount of oligonucleotide coupled per address (data not shown). Magnesium ion was obligatory to achieve hybridization, and less than 1 fmol of probe could be detected in the presence of this divalent cation (Table 2 and Figure 4). The hybridization signal was doubled upon lowering the pH from 8.0 to 6.0, most likely due to masking of negative charges (hence reducing repulsive interactions with oligonucleotides) arising from uncoupled acrylic acid groups in the bulk polymer

Table 2. Effect of hybridization conditions on hybridization signal

Hybridization buffer	Vol. (μl)	Mixing[a]	Time (minutes)	Relative signal
Buffer A	55	Inter.	30	1
Buffer A minus $MgCl_2$	55	Inter.	30	<0.01
Buffer A	20	Inter.	30	2.5
Buffer B	55	Inter.	30	2
Buffer B	20	Inter.	30	3
Buffer B	55	Cont.	30	4
Buffer B	55	Cont.	60	8
Buffer A + Capped Surface	55	Cont.	60	8
Buffer B minus $MgCl_2$	55	Cont.	60	<0.01
Buffer B	55	Cont.	180	10

Following general procedures described in Materials and Methods, hybridizations were carried out with 1 pmol of FAMcZip13-Prd and 3 × 3 manually spotted arrays. Buffers were: buffer A, 300 mM bicine (pH 8.0), 10 mM $MgCl_2$, 0.1% SDS; buffer B, 300 mM Mes (pH 6.0), 10 mM $MgCl_2$, 0.1% SDS.

[a] Mixing was as follows: intermittent (Inter.), manual mixing of the sample once every ten minutes; continuous (Cont.), mixing of sample at 20 rpm in a hybridization oven.

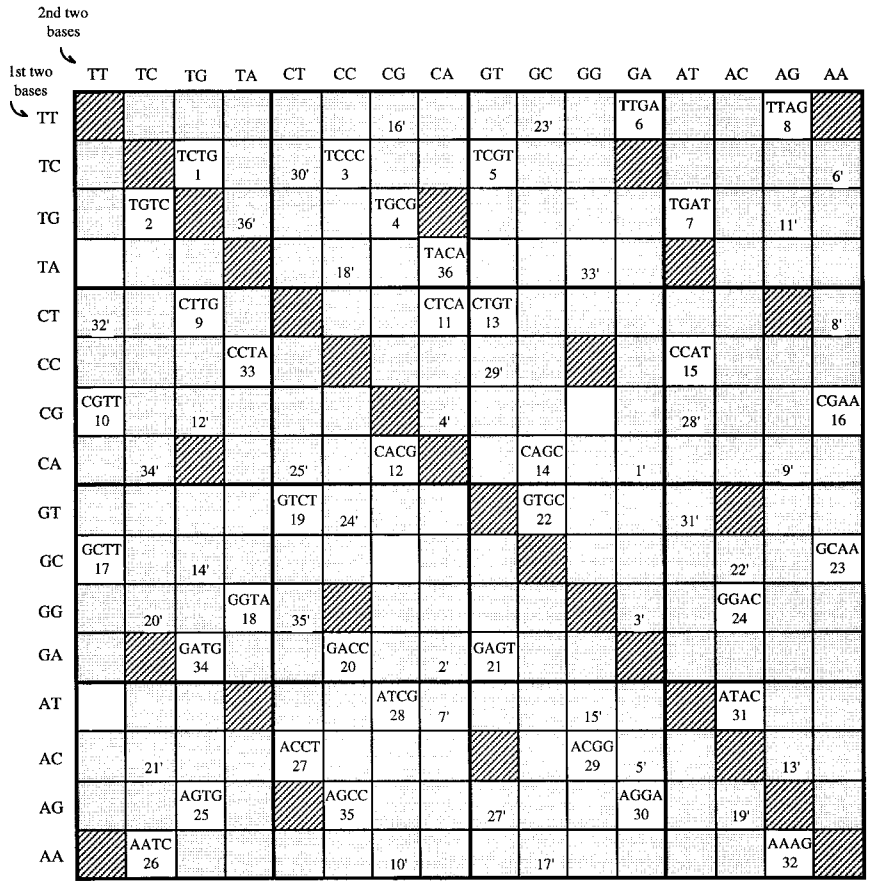

Figure 2. Design of tetramers for use in zip-code arrays. The checkerboard pattern shows all 256 possible tetramers. A given square represents the two bases on the left followed by the two bases on the top of the checkerboard. To be included, each tetramer must differ from all others by at least two bases, and be non-complementary. The chosen tetramers are shown in the white boxes, while their complements are listed as (number)′. Thus, as an example, the complementary sequences GACC (20) and GGTC (20′) are mutually exclusive in this scheme. In addition, tetramers that are palindromic, e.g. TCGA (off-diagonal hatched boxes) or repetitive, e.g. CACA (hatched boxes on diagonal from upper left to lower right) have been eliminated. All other sequences which differ from the 36 tetramers by only one base are shaded in light gray. Four potential tetramers were not chosen as they are either all A·T or G·C bases (open boxes).

matrix. To confirm this hypothesis, the free carboxyl groups on arrays to which zip-codes had already been attached were capped with ethanolamine under standard coupling conditions. Hybridizations of the capped arrays at pH 8.0 gave results comparable to hybridizations at pH 6.0 of the same arrays without capping. Continuous mixing proved to be crucial for obtaining good hybridization, and studies of the time-course led us to choose one hour at 65 °C as standard. Reducing the hybridization volume improved the hybridization signal due to the relative increase in probe concentration. Further improvements may be achieved using specialized small volume hybridization chambers that allow for continuous mixing.

Array hybridization of K-*ras* LDR products

PCR/LDR amplification coupled with zip-code detection on an addressable array was tested with the K-*ras* gene as a model system. Exon-specific PCR primers were used to selectively amplify K-*ras* DNA flanking codons 12 and 13. LDR primers were designed to detect the seven most common mutations found in the K-*ras* gene in colorectal cancer (Figure 1(c) and Table 3). For example, the second position in codon 12, GGT, coding for glycine, may mutate to GAT, coding for aspartate, which is detected by ligation of the allele-specific primer (containing a zip-code complement, cZip3, on its 5′-end, and a discriminating base, **A**, on its 3′-end) to a fluorescently labeled common primer (Figure 1(c)).

PCR/LDR was carried out on nine individual DNA samples derived from cell lines or paraffin-embedded tumors containing known K-*ras* mutations (as described in Materials and Methods). An aliquot (2 μl) was taken from each reaction and electrophoresed on a sequencing apparatus to confirm that LDR was successful (data not shown). Next, the different mutations were distinguished by hybridizing the LDR product mixtures on 3 × 3 addressable DNA arrays (each zip-code address was spotted in quadruplicate), and detecting the

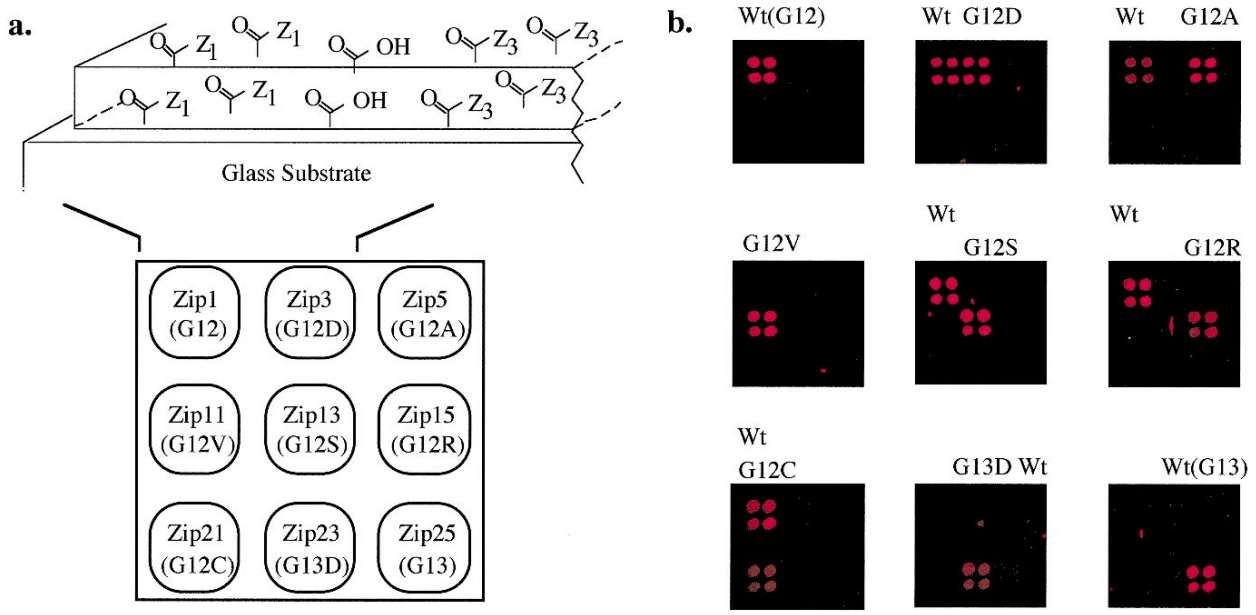

Figure 3. Detection of K-*ras* mutations on a DNA array. (a) Schematic representation of gel-based zip-code array. Glass microscope slides treated with γ-methacryloyloxypropyltrimethoxysilane are used as the substrate for the covalent attachment of an acrylamide/acrylic acid copolymer matrix. Amine-modified zip-code oligonucleotides are coupled to *N*-hydroxysuccinimide-activated surfaces at discrete locations (see Materials and Methods). Each position in the 3 × 3 grid identifies an individual zip-code address (and corresponding K-*ras* mutation or wild-type sequence). (b) Each robotically spotted array was hybridized with an individual LDR and fluorescent signal detected as described in Materials and Methods using a two second exposure time. All nine arrays identified the correct mutant and/or wild-type for each tumor (G12S, G12R, and G12C) or cell line sample (Wt, G12D, G12A, G12V, and G13D). The small spots seen in some of the panels, e.g. near the center of the panel containing the G13D mutant, are not incorrect hybridizations, but noise due to imperfections in the polymer.

positions of fluorescent spots (Figure 3(b)). The wild-type samples, Wt(G12) and Wt(G13), each displayed four equal hybridization signals at Zip1 and Zip25, respectively, as expected. The mutant samples each displayed hybridization signals corresponding to the mutant, as well as for the wild-type DNA present in the cell line or tumor. The sole exception to this was the G12V sample, which was derived from a cell line (SW620) homozygous for the G12V K-*ras* allele. The experiment was repeated several times, using both manually and robotically spotted arrays, and LDR primers labeled with either fluorescein or Texas Red. False-positive or false-negative signals were not encountered in any of these experiments. A minimal amount of noise seen on the arrays can be attribu-

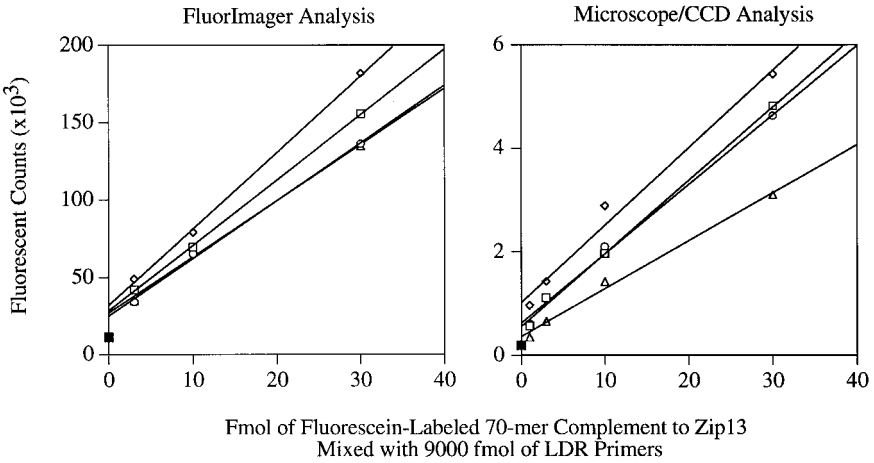

Figure 4. Determination of zip-code array capture sensitivity using two different detection instruments. Quadruplicate hybridizations were carried out on manually spotted arrays as described in Materials and Methods. The graphs depict quantification of the amount of captured 70-mer complement using either a fluorimager (left) or an epifluorescence microscope/CCD (right). Each symbol represents hybridizations to an individual array. The filled square on each graph is the average of the backgrounds from all four arrays.

Table 3. Primers designed for K-*ras* mutation detection by PCR/LDR/array hybridization

Primer	Sequence (5′ → 3′)
K-*ras* exon 1 forward	ATAAGGCCTGCTGAAAATGACTGAA
K-*ras* exon 1 reverse	CTGCACCAGTAATATGCATATTAAAACAAG
cZip1-K-*ras* c12.2WtG	**GCTGAGGTCGATGCTGAGGTCGCA**AAACTTGTGGTAGTTGGAGCTGG
cZip3-K-*ras* c12.2D	**GCTGCGATCGATGGTCAGGTGCTG**AAACTTGTGGTAGTTGGAGCTGA
cZip5-K-*ras* c12.2A	**GCTGTACCCGATCGCAAGGTGGTC**AAACTTGTGGTAGTTGGAGCTGC
cZip11-K-*ras* c12.2V	**CGCAAGGTAGGTGCTGTACCCGCA**AAACTTGTGGTAGTTGGAGCTGT
K-*ras* c12 Com-2	pTGGCGTAGGCAAGAGTGCCT-fluorescein pTGGCGTAGGCAAGAGTGCCT-Texas Red
cZip13-K-*ras* c12.1S	**CGCACGATAGGTGGTCTACCGCTG**ATATAAACTTGTGGTAGTTGGAGCTA
cZip15-K-*ras* c12.1R	**CGCATACCAGGTCGCATACCGGTC**ATATAAACTTGTGGTAGTTGGAGCTC
cZip21-K-*ras* c12.1C	**GGTCAGGTTACCGCTGCGATCGCA**ATATAAACTTGTGGTAGTTGGAGCTT
K-*ras* c12 Com-1	pGTGGCGTAGGCAAGAGTGCC-fluorescein pGTGGCGTAGGCAAGAGTGCC-Texas Red
cZip23-K-*ras* c13.4D	**GGTCCGATTACCGGTCCGATGCTG**TGTGGTAGTTGGAGCTGGTGA
cZip25-K-*ras* c13.4WtG	**GGTCTACCTACCCGCACGATGGTC**TGTGGTAGTTGGAGCTGGTGG
K-*ras* c13 Com-4	pCGTAGGCAAGAGTGCCTTGAC-fluorescein pCGTAGGCAAGAGTGCCTTGAC-Texas Red

The PCR primers were specifically designed to amplify exon 1 of K-*ras* without co-amplifying N and H-*ras*. The allele-specific LDR primers contained 24-mer zip-code complement sequences on their 5′-ends (bold) and the discriminating bases on their 3′-ends (underlined). The common LDR primers contained 5′-phosphate groups and either a fluorescein or a Texas Red label on their 3′-ends.

ted to dust, scratches, and/or small bubbles in the polymer. These flaws are readily recognized because they are weak and sporadic, rather than reproducing the quadruplicate spotting pattern; we expect such noise will be minimized with more stringent manufacturing conditions. Ultimately, these protocols are amenable to quantifying the relative amounts of each allele, and work is currently in progress to convert our quantitative PCR/LDR protocols for K-*ras* mutations from gel-based detection to array-based detection (unpublished results).

Array capture sensitivity

After an LDR, the successfully ligated and fluorescently labeled LDR product competes with an excess of unligated discriminating primer for hybridization to the correct zip-code address on the array. To determine capture sensitivity, DNA arrays were hybridized in quadruplicate, under standard conditions, with from 100 amol (= 1/90,000) to 30 (= 1/300) fmol of a labeled synthetic 70-mer, FAMcZip13-Prd (this simulates a full-length LDR product; see Materials and Methods for the sequence), in the presence of a full set of K-*ras* LDR primers (combined total of 9000 fmol of discriminating and common primers). Array analyses with a FluorImager (Figure 4, left-side) indicate that a signal-to-noise ratio of greater than 3:1 can be achieved when starting with a minimum of 3 fmol (= 1/3,000) of FAMcZip13-Prd-labeled probe in the presence of 4500 fmol of FAM-labeled LDR primers and 4500 fmol of zip-code complement primers in the hybridization solution. Results using microscope/CCD instrumentation to quantify fluorescence were even more striking: a 3:1 signal-to-noise ratio was maintained starting with 1 fmol (= 1/9,000) of labeled product (Figure 4, right-hand side) on three out of the four arrays; the signal to noise was 2:1 on the fourth array. For a given array, with fluorescence quantified by either instrument, the captured counts varied linearly with the amount of labeled FAMcZip13-Prd added. Rehybridization of the same probe, at the same concentration, to the same array, was reproducible within ±5 % (data not shown). Variations in fluorescent signal between arrays may reflect variations in the amount of zip-code oligonucleotide coupled, due to the inherent inaccuracies of manual spotting and/or variations in polymer uniformity.

Detection of low abundance mutations by PCR/LDR/array hybridization

To determine the limit of detection of low-level mutations in wild-type DNA using PCR/LDR/array hybridization, a dilution series was set up and analyzed. PCR-amplified pure G12V DNA was diluted into wild-type K-*ras* DNA in ratios ranging from 1:20 to 1:500. Duplicate LDRs were carried out on 2000 fmol of total DNA, using a two-primer set consisting of 2000 fmol each of the discriminating and common primers for the G12V mutation. It proved possible to quantify a positive hybridization signal at a dilution of 1:200 with a signal-to-noise ratio of 2:1 (Figure 5). A signal was distinguishable even at a dilution of 1:500, although noise levels due to dust or bubbles in the polymer prevented us from accurately quantifying the results. A control of pure wild-type DNA showed no hybridization signal. These results indicate clearly that zip-code array hybridization, when coupled with PCR/LDR, may detect polymorphisms present at less than 1 % of the total DNA. These results are consistent with our earlier work showing that PCR/LDR, using a 26-primer set and analyses based on gel electrophoreses of

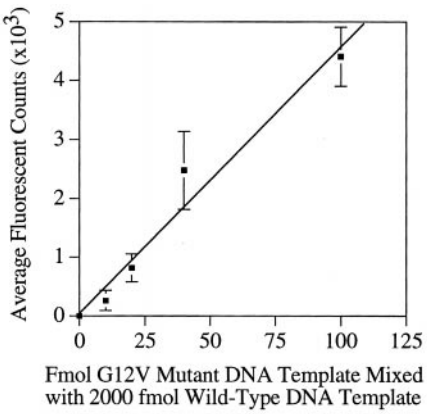

Figure 5. Detection of minority K-*ras* mutant DNA in a majority of wild-type DNA using PCR/LDR with zip-code array capture. DNA from cell line SW620, containing the G12V mutation, and DNA from normal lymphocytes were PCR amplified in exon 1 of the K-*ras* gene. Mixtures containing 10, 20, 40, or 100 fmol of G12V-amplified fragment plus 2000 fmol of PCR-amplified wild-type fragment were prepared, and the presence of mutant DNA determined by LDR using primers specific for the G12V mutation (2000 fmol each of discriminating and common primer). Images were collected by CCD using exposure times from five to 25 seconds. Data were normalized by dividing fluorescent signal intensity by acquisition time. Each data point represents the average hybridization signal from four independent robotically spotted arrays. The average background signal from all four spots at each address following hybridization of pure wild-type control (880 average fluorescent counts) was subtracted from the mutant signal.

products, can detect any K-*ras* mutation in the presence of up to a 500-fold excess of wild-type, with a signal-to-noise ratio of at least 3:1 (Khanna *et al*., 1999).

Comparison of universal array to gene-specific arrays

Our approach to mutation detection has three orthogonal components: (i) primary PCR amplification; (ii) solution-phase LDR detection; and (iii) solid-phase hybridization capture. Therefore, background signal from each step can be minimized and, consequently, the overall sensitivity and accuracy of our method are significantly enhanced over those provided by other strategies. For example, hybridization of labeled target methods require: (i) multiple rounds of PCR or PCR/T7 transcription; (ii) processing of PCR amplified products to fragment them or render them single-stranded; and (iii) lengthy hybridization periods (ten hours or more) which limits throughput (Chee *et al*., 1996; Cronin *et al*., 1996; Guo *et al*., 1994; Hacia *et al*., 1996; Schena *et al*., 1996; Shalon *et al*., 1996; Wang *et al*., 1998). Additionally, since the immobilized probes on the aforementioned arrays have a wide range of t_m values, it is necessary to perform the hybridizations at temperatures from 0 °C to 44 °C. The result is increased background noise and false signals due to mismatch hybridization and non-specific binding, for example, on small insertions and deletions in repeat sequences (Cronin *et al*., 1996; Hacia *et al*., 1996; Southern, 1996; Wang *et al*., 1998). In contrast, our approach allows multiplexed PCR in a single reaction (Belgrader *et al*., 1996), does not require an additional step to convert product into single-stranded form, and can readily distinguish all point mutations including slippage in repeat sequences (Day *et al*., 1995; Khanna *et al*., 1999). Alternative DNA arrays suffer from differential hybridization efficiencies due to either sequence variation or to the amount of target present in the sample. By using our approach of designing divergent zip-code sequences with similar thermodynamic properties, hybridizations can be carried out at 65 °C, resulting in a more stringent and rapid hybridization. The decoupling of the hybridization step from the mutation detection stage offers the prospect of quantification of LDR products, as we have already achieved using gel-based LDR detection (Khanna *et al*., 1999).

Arrays spotted on polymer surfaces provide substantial improvements in signal capture, as compared with arrays spotted or synthesized *in situ* directly on glass surfaces (Drobyshev *et al*., 1997; Parinov *et al*., 1996; Yershov *et al*., 1996). However, the polymers described by others are limited to using 8 to 10-mer addresses, while our polymeric surface readily allows 24-mer zip-codes to penetrate and couple covalently. Moreover, LDR products of length 60 to 75 nucleotide bases are also found to penetrate and subsequently hybridize to the correct address. As additional advantages, our polymer gives little or no background fluorescence and does not exhibit non-specific binding of fluorescently labeled oligonucleotides. Finally, zip-codes spotted and coupled covalently at a discrete address do not ''bleed over'' to neighboring spots, hence obviating the need to physically segregate sites, e.g. by cutting gel pads.

Summary and Conclusions

Here, we describe a strategy for high-throughput mutation detection which differs substantially from other array-based detection systems presented previously in the literature. In concert with a polymerase chain reaction/ligase detection reaction (PCR/LDR) assay carried out in solution, our array allows for accurate detection of single base mutations, whether inherited and present as 50% of the sequence for that gene, or sporadic and present at 1% or less of the wild-type sequence. We achieve this sensitivity because thermostable DNA ligase provides the specificity of mutation discrimination, while the divergent addressable portions (zip-codes) of our LDR primers guide each LDR product to a designated address on the DNA array. Since the zip-code sequences remain con-

stant and their complements can be appended to any set of LDR primers, our zip-code arrays are universal. Thus, a single array design can be programmed to detect a wide range of genetic mutations.

Robust methods for the rapid detection of mutations at numerous potential sites in multiple genes hold great promise to improve the diagnosis and treatment of cancer patients. Non-invasive tests for mutational analysis of shed cells in saliva, sputum, urine, and stool could significantly simplify and improve the surveillance of high risk populations, reduce the cost and discomfort of endoscopic testing, thus leading to more effective diagnosis of cancer in its early, curable stage. Although the feasibility of detecting shed mutations has been demonstrated clearly in patients with known and genetically characterized tumors (Caldas *et al*., 1994; Hasegawa *et al*., 1995; Nollau *et al*., 1996; Sidransky *et al*., 1992; Wu *et al*., 1994), effective presymptomatic screening will require that a myriad of potential low frequency mutations be identified with minimal false-positive and false-negative signals. Furthermore, the integration of technologies for determining genetic changes within a tumor with clinical information about the likelihood of response to therapy could radically alter how patients with more advanced tumors are selected for treatment. Identification and validation of reliable genetic markers will require that many candidate genes be tested in large-scale clinical trials. While costly microfabricated chips can be manufactured with over 100,000 addresses, none of them has as yet demonstrated a capability to detect low abundance mutations (Chee *et al*., 1996; Hacia *et al*., 1996; Kozal *et al*., 1996; Sapolsky *et al*., 1999; Wang *et al*., 1998), as required to accurately score mutation profiles in such clinical trials. The universal zip-code array approach introduced here has the potential to allow rapid and reliable identification of low abundance mutations in multiple codons in numerous genes. As new therapies targeted to specific genes or specific mutant proteins are developed, the importance of rapid and accurate high-throughput genetic testing will undoubtedly increase.

Materials and Methods

Oligonucleotide synthesis and purification

Oligonucleotides were obtained as custom synthesis products from IDT, Inc. (Coralville, IA), or synthesized in-house on an ABI 394 DNA Synthesizer (PE Biosystems Inc., Foster City, CA) using standard phosphoramidite chemistry. Spacer phosphoramidite 18, 3′-amino-modifier C3 CPG, and 3′-fluorescein CPG were purchased from Glen Research (Sterling, VA). All other reagents were purchased from PE Biosystems. Oligonucleotides with 3′-amino modifications and/or fluorescent labels were cleaved from the supports by treatment with concentrated aqueous NH_4OH for two hours at 25 °C, and deprotection continued in solution for 24 hours at 25 °C. Texas Red labeling was achieved by adding 150 μl of 0.2 M $NaHCO_3$ and 200 μg of oligonucleotide to tubes containing a solution of 500 μg of Texas Red-X succinimidyl ester (Molecular Probes; Eugene, OR) in 28 μl of anhydrous DMF. Following overnight stirring at 25 °C, the majority of unreacted label was removed by the addition of 20 μl of 3 M NaCl and 500 μl of cold ethanol, chilling in a dry ice/ethanol bath for 30 minutes, and centrifuging at 12,000 *g* for 30 minutes. The supernatants were removed, the pelleted oligonucleotides were washed with 100 μl of 70% ethanol, and dried. FAMcZip13-Prd, a fluorescein-labeled 70-mer that simulates a full-length LDR product containing the complementary sequence to Zip13, was synthesized on 1000 Å pore-size CPG. The sequence was: 5′-fluorescein-**CGCACGATAGG TGGTCTACCGCTG-**ATATAAACTTGTGGTAGTTGG-AGCTAGTGGCGTAGGCAAGAGTGCC-3′ (the zip-code complement is in bold).

Both labeled and unlabeled oligonucleotides were purified by electrophoresis on denaturing 12% polyacrylamide gels. Bands were visualized by UV shadowing, excised from the gel, and eluted overnight in 0.5 M NaCl, 5 mM EDTA (pH 8.0) at 37 °C. Oligonucleotide solutions were desalted on C18 Sep-Paks (Waters Corporation; Milford, MA) according to the manufacturer's instructions, following which the oligonucleotides were concentrated to dryness (Speed-Vac) and stored at −20 °C.

DNA extraction from cell lines

Cell lines of known K-*ras* genotype (HT29, wild-type; SW1116, G12A; LS180, G12D; SW620, G12V; DLD1, G13D) were grown in RPMI culture media with 10% fetal bovine serum. Harvested cells ($\sim 10^7$) were resuspended in DNA extraction buffer (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM EDTA (pH 8.0), 0.5% (w/v) SDS, 200 μg/ml proteinase K) and incubated at 37 °C for four hours. A 30% volume of 6 M NaCl was added and the mixture was centrifuged. The supernatant was transferred to a clean tube and the DNA was pelleted through the addition of three volumes of ethanol, chilling on dry ice, and centrifugation. The pellet was washed with 70% ethanol and resuspended in 10 mM Tris-HCl (pH 7.5), 2 mM EDTA (pH 8.0).

DNA extraction from paraffin sections

Tissue sections (10 μm) were cut from paraffin-embedded colon tumors. Samples were deparaffinized *via* sequential extraction with xylene, ethanol, and acetone, and dried under vacuum. The DNA in the pellets was purified using a QIAamp Tissue Kit (Qiagen, Chatsworth, CA).

Polymer coated slides

Microscope slides (Fisher Scientific, precleaned, 3 in × 1 in × 1.2 mm) were immersed in 2% γ-methacryloyloxypropyltrimethoxysilane, 0.2% triethylamine in $CHCl_3$ for 30 minutes at 25 °C, and then washed with $CHCl_3$ (two washes of 15 minutes). A monomer solution (20 μl of 8% acrylamide, 2% acrylic acid, 0.02% *N*,*N*′-methylene-bisacrylamide (500:1 ratio of monomers:cross-linker), 0.8% ammonium persulfate radical polymerization initiator) was deposited on one end of the slides and spread out with the aid of a cover-slip (24 mm × 50 mm) that had been previously silanized (5% $(CH_3)_2SiCl_2$ in $CHCl_3$). Polymerization was achieved

by heating the slides on a 70 °C hotplate for 4.5 minutes. Upon removal of the slides from the hotplate, the coverslips were immediately peeled off with the aid of a single-edge razor blade. The coated slides were rinsed with deionized water, allowed to dry in open atmosphere, and stored under ambient conditions.

Zip-code arrays

Polymer-coated slides were pre-activated by immersing them for 30 minutes at 25 °C in a solution of 0.1 M 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride plus 20 mM *N*-hydroxysuccinimide in 0.1 M K_2HPO_4/KH_2PO_4 (pH 6.0). The activated slides were rinsed with water, and then dried in a 65 °C oven; they were stable upon storage for six months or longer at 25 °C in a desiccator over Drierite.

For manual spotting, 0.2 μl aliquots were taken with a Rainin Pipetman from stock solutions (500 μM) of zip-code oligonucleotides in 0.2 M K_2HPO_4/KH_2PO_4 (pH 8.3), and deposited in a 3 × 3 array onto the pre-activated polymeric surfaces. The resulting arrays were incubated for one hour at 65 °C in humidified chambers containing water/formamide (1:1). For robotic spotting, 10-50 nL aliquots of zip-code oligonucleotides (1.5 mM in the same buffer) were deposited at 25 °C on the pre-activated surfaces by using a robot (PE Biosystems, ''in-house'' design) equipped with a quill-type spotter in a controled atmosphere chamber. Two pairs of 3 × 3 arrays were spotted on each slide, with addresses consisting of groups of four spots. Following spotting using either method, uncoupled oligonucleotides were removed from the polymer surfaces by soaking the slides in 300 mM bicine (pH 8.0), 300 mM NaCl, 0.1 % SDS, for 30 minutes at 65 °C, rinsing with water, and drying. The arrays were stored at 25 °C in slide boxes until needed.

PCR amplification of K-*ras* DNA samples

PCR amplifications were carried out under paraffin oil in 20 μl reaction mixtures containing 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 800 μM dNTPs, 2.5 μM forward and reverse primers (12.5 pmol of each primer; Table 3), and 1-50 ng of genomic DNA extracted from paraffin-embedded tumors or from cell lines. Following a two minute denaturation step at 94 °C, 0.2 unit of *Taq* DNA polymerase (PE Biosystems) was added. Amplification was achieved by thermally cycling for 40 rounds of 94 °C for 15 seconds and 60 °C for two minutes, followed by a final elongation at 65 °C for five minutes. Following PCR, 1 μl of proteinase K (18 mg/ml) was added, and reactions were heated to 70 °C for ten minutes and then quenched at 95 °C for 15 minutes. One microliter of each PCR product was analyzed on a 3 % agarose gel to verify the presence of amplification product of the expected size.

LDR of K-*ras* DNA samples

LDR was carried out under paraffin oil in 20 μl volumes containing 20 mM Tris-HCl (pH 8.5), 5 mM $MgCl_2$, 100 mM KCl, 10 mM DTT, 1 mM NAD^+, 8 pmol of total LDR primers (500 fmol each of discriminating primers + 4 pmol of fluorescently labeled common primers), and 1 pmol of PCR products from cell line or tumor samples. Two primer mixes were prepared, each containing the seven mutation-specific primers, the three common primers, and either the wild-type discriminating primer for codon 12 or that for codon 13 (Figure 1(c) and Table 3).

The reaction mixtures were pre-heated for two minutes at 94 °C, and then 25 fmol of wild-type *Tth* DNA ligase was added. The LDRs were cycled for 20 rounds of 94 °C for 30 seconds and 65 °C for four minutes. An aliquot of 2 μl of each reaction was mixed with 2 μl of gel loading buffer (8 % blue dextran, 50 mM EDTA (pH 8.0), formamide (1:5)), denatured at 94 °C for two minutes, and chilled on ice; 1 μl of each mixture was loaded onto a denaturing 10 % polyacrylamide gel and electrophoresed on an ABI 377 DNA sequencer at 1500 volts.

Hybridization of K-*ras* LDR products to DNA arrays

The LDRs (17 μl) were diluted with 40 μl of 1.4× hybridization buffer to produce a final buffer concentration of 300 mM Mes (pH 6.0), 10 mM $MgCl_2$, 0.1 % SDS, denatured at 94 °C for three minutes, and chilled on ice. Arrays were pre-incubated for 15 minutes at 25 °C in 1× hybridization buffer. Coverwells (Grace, Inc; Sunriver, OR) were filled with the diluted LDRs and attached to the arrays. The arrays were placed in humidified culture tubes and incubated for one hour at 65 °C and 20 rpm in a rotating hybridization oven. Following hybridization, the arrays were washed in 300 mM bicine (pH 8.0), 10 mM $MgCl_2$, 0.1 % SDS for ten minutes at 25 °C. Fluorescent signals were measured using a microscope/CCD (see below).

Hybridization of synthetic LDR products to DNA arrays

Quadruplicate hybridization mixtures were prepared containing 100 amol, 1fmol, 3 fmol, 10 fmol, or 30 fmol of FAMcZip13-Prd (a synthetic 70-mer LDR product complementary to zip-code 13) combined with 4500 fmol of total fluorescein-labeled common LDR primers and 9 × 500 fmol of each unlabeled, zip-code-containing discriminating LDR primer in 55 μl of 300 mM Mes (pH 6.0), 10 mM $MgCl_2$, 0.1 % SDS. Hybridizations were conducted according to the protocol described above, and FluorImager as well as epifluorescence microscopy data were acquired and analyzed (see below).

LDR and hybridization of G12V/G12 dilution series to DNA arrays

These experiments were carried out in a volume of 20 μl. The PCR-amplified SW620 cell line DNA containing the G12V mutation was diluted from 5 nM (100 fmol = 1/20) to 0.050 nM (1 fmol = 1/2000) in LDR mixtures containing 100 nM (2000 fmol) of wild-type (G12) DNA and 100 nM (2000 fmol) of both G12V-discriminating primer and Texas Red-labeled common primer. The LDR and hybridization proceeded as above, and imaging on the microscope/CCD was carried out as detailed below.

Image analysis

Arrays were imaged using a Molecular Dynamics FluorImager 595 (Sunnyvale, CA) or an Olympus AX70 epifluorescence microscope (Melville, NY) equipped with a Princeton Instruments TE/CCD-512 TKBM1 camera (Trenton, NJ). For analysis of fluorescein-labeled probes on the FluorImager, the 488 nm excitation was used with

a 530/30 emission filter. The spatial resolution of scans was 100 μm per pixel. The resulting images were analyzed using ImageQuaNT software provided with the instrument. The epifluorescence microscope was equipped with a 100 W mercury lamp, a FITC filter cube (excitation 480/40, dichroic beam splitter 505, emission 535/50), a Texas Red filter cube (excitation 560/55, dichroic beam splitter 595, emission 645/75), and a 100 mm macro objective. The macro objective allows illumination of an object field up to 15 mm in diameter and projects a 7 mm × 7 mm area of the array onto the 12.3 mm × 12.3 mm matrix of the CCD. Images were collected in 16-bit mode using the Winview32 software provided with the camera. Analysis was performed using Scion Image (Scion Corporation, Frederick, MD).

Acknowledgements

We thank Michael Wigler, Donald Bergstrom, Phillip Paty, Eric Spitzer, Leo Furcht, and Matthew Lubin for critical comments and helpful discussions, and express gratitude to our collaborators at PE Biosystems, Michael Albin, Emily Winn-Dean, and Andy Blasband for encouragement and for assistance with spotting arrays. Antonio Picone, Marilyn Khanna, and Monib Zirvi are acknowledged for providing tumor and cell line DNA samples, and for expert technical assistance, and Herman Blok and Maria Kempe are thanked for significant experimental contributions in exploratory stages of this project. Support for this work was provided by the National Institute of Standards and Technology (1995-08-0006F) and the National Cancer Institute (P01-CA65930).

References

Barany, F. (1991a). Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl Acad. Sci. USA,* **88**, 189-193.

Barany, F. (1991b). The ligase chain reaction (LCR) in a PCR world. *PCR Methods Appl.* **1**, 5-16.

Beattie, K. L., Beattie, W. G., Meng, L., Turner, S. L., Coral-Vazquez, R., Smith, D. D., McIntyre, P. M. & Dao, D. D. (1995). Advances in genosensor research. *Clin. Chem.* **41**, 700-706.

Belgrader, P., Marino, M., Lubin, M. & Barany, F. (1996). A multiplex PCR-ligase detection reaction assay for human identity testing. *Genome Sci. Technol.* **1**, 77-87.

Bishop, J. M. (1991). Molecular themes in oncogenesis. *Cell,* **64**, 235-248.

Brown, P. O. & Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, 33-37.

Caldas, C., Hahn, S. A., Hruban, R. H., Redston, M. S., Yeo, C. J. & Kern, S. E. (1994). Detection of K-*ras* mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Res.* **54**, 3568-3573.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science,* **274**, 610-614.

Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. & Miyada, C. G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Human Mutat.* **7**, 244-255.

Da Costa, L. T., Jen, J., He, T. C., Chan, T. A., Kinzler, K. W. & Vogelstein, B. (1996). Converting cancer genes into killer genes. *Proc. Natl Acad. Sci. USA,* **93**, 4192-4196.

Day, D. J., Speiser, P. W., White, P. C. & Barany, F. (1995). Detection of steroid 21 hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. *Genomics,* **29**, 152-162.

Day, D. J., Speiser, P. W., Schulze, E., Bettendorf, M., Fitness, J., Barany, F. & White, P. C. (1996). Identification of non-amplifying CYP21 genes when using PCR-based diagnosis of 21-hydroxylase deficiency in congenital adrenal hyperplasia (CAH) affected pedigrees. *Hum. Mol. Genet.* **5**, 2039-2048.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. & Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**, 457-460.

Drobyshev, A., Mologina, N., Shik, V., Pobedimskaya, D., Yershov, G. & Mirzabekov, A. (1997). Sequence analysis by hybridization with oligonucleotide microchip: identification of β-thalassemia mutations. *Gene,* **188**, 45-52.

Eggers, M., Hogan, M., Reich, R. K., Lamture, J., Ehrlich, D., Hollis, M., Kosicki, B., Powdrill, T., Beattie, K., Smith, S., Varma, R., Gangadharam, R., Mallik, A., Burke, B. & Wallace, D. (1994). A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups. *BioTechniques,* **17**, 516-525.

Fearon, E. R. (1997). Human cancer syndromes: clues to the origin and nature of cancer. *Science,* **278**, 1043-1050.

Fearon, E. R. & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell,* **61**, 759-767.

Gentalen, E. & Chee, M. (1999). A novel method for determining linkage between DNA sequences: hybridization to paired probe arrays. *Nucl. Acids Res.* **27**, 1485-1491.

Gunderson, K. L., Huang, X. C., Morris, M. S., Lipshutz, R. J., Lockhart, D. J. & Chee, M. S. (1998). Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res.* **8**, 1142-1153.

Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucl. Acids Res.* **22**, 5456-5465.

Guo, Z., Liu, Q. & Smith, L. M. (1997). Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nature Biotechnol.* **15**, 331-335.

Hacia, J. G. (1999). Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet.* **21**, 42-47.

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. & Collins, F. S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genet.* **14**, 441-447.

Hacia, J. G., Sun, B., Hunt, N., Edgemon, K., Mosbrook, D., Robbins, C., Fodor, S. P. A., Tagle, D. A. & Collins, F. S. (1998a). Strategies for mutational analysis of the large multiexon *ATM* gene using high-density oligonucleotide arrays. *Genome Res.* **8**, 1245-1258.

Hacia, J. G., Woski, S. A., Fidanza, J., Edgemon, K., Hunt, N., McGall, G., Fodor, S. P. A. & Collins, F. S. (1998b). Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates. *Nucl. Acids Res.* **26**, 4975-4982.

Hasegawa, H., Ueda, M., Watanabe, M., Teramoto, T., Mukai, M. & Kitajima, M. (1995). K-*ras* gene mutations in early colorectal cancer . . . flat elevated *vs* polyp-forming cancer. *Oncogene,* **10**, 1413-1416.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science,* **269**, 400-403.

Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P. & Barany, F. (1999). Multiplex PCR/LDR for detection of K-*ras* mutations in primary colon tumors. *Oncogene,* **18**, 27-38.

Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M. & Gingeras, T. R. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature Med.* **2**, 753-759.

Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nature Med.* **2**, 169-174.

Lockley, A. K., Jones, C. G., Bruce, J. S., Franklin, S. J. & Bardsley, R. G. (1997). Colorimetric detection of immobilised PCR products generated on a solid support. *Nucl. Acids Res.* **25**, 1313-1314.

Nikiforov, T. T., Rendle, R. B., Goelet, P., Rogers, Y. H., Kotewicz, M. L., Anderson, S., Trainor, G. L. & Knapp, M. R. (1994). Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucl. Acids Res.* **22**, 4167-4175.

Nollau, P., Moser, C., Weinland, G. & Wagener, C. (1996). Detection of K-*ras* mutations in stools of patients with colorectal cancer by mutant-enriched PCR. *Int. J. Cancer,* **66**, 332-336.

Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A. & Mirzabekov, A. (1996). DNA sequencing by hybridization to microchip oct- and decanucleotides extended by stacked pentanucleotides. *Nucl. Acids Res.* **24**, 2998-3004.

Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A. C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* **7**, 606-614.

Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P. & Fodor, S. P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA,* **91**, 5022-5026.

Perera, F. P. (1997). Environment and cancer: who are susceptible?. *Science,* **278**, 1068-1073.

Sapolsky, R. J., Hsie, L., Berno, A., Ghandour, G., Mittman, M. & Fan, J.-B. (1999). High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet. Anal.* **14**, 187-192.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA,* **93**, 10614-10619.

Shalon, D., Smith, S. J. & Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6**, 639-645.

Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450-456.

Shumaker, J. M., Metspalu, A. & Caskey, C. T. (1996). Mutation detection by solid phase primer extension. *Hum. Mutat.* **7**, 346-354.

Sidransky, D., Tokino, T., Hamilton, S. R., Kinzler, K. W., Levin, B., Frost, P. & Vogelstein, B. (1992). Identification of *ras* oncogene mutations in the stool of patients with curable colorectal tumors. *Science,* **256**, 102-105.

Southern, E. M. (1996). DNA chips: analysing sequence by hybridization to oligonucleotides on a large scale. *Trends Genet.* **12**, 110-115.

Southern, E., Mir, K. & Shchepinov, M. (1999). Molecular interactions on microarrays. *Nature Genet.* **21**, 5-9.

Venitt, S. (1996). Mechanisms of spontaneous human cancers. *Environ. Health Perspect.* **3**, 633-637.

Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T. & Hubbell, E., *et al*. (1998). Large-scale identificatin, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science,* **280**, 1077-1082.

Wu, S., Hoshino, B., Zhou, D. F. H., Liu, A., Aziz, D. C. & Barathur, R. R. (1994). Practical approaches to molecular screening of colon cancer. In *Early Detection of Cancer Molecular Markers* (Srivastava, S., Lippman, S. M., Hong, W. K. & Mulshine, J. L., eds), pp. 237-254, Futura Publishing Company, Inc., Armonk.

Yershov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S. & Mirzabekov, A. (1996). DNA analysis and diagnostics on oligonucleotide microchips. *Proc. Natl Acad. Sci. USA,* **93**, 4913-4918.

Zhu, H., Cong, J.-P., Mamora, G., Gingeras, T. & Shenk, T. (1998). Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc. Natl Acad. Sci. USA,* **95**, 14470-14475.

Edited by K. Yamamoto

(Received 16 April 1999; received in revised form 19 July 1999; accepted 20 July 1999)