Detection of heterozygous mutations in *BRCA1* using high density oligonucleotide arrays and two-colour fluorescence analysis

Joseph G. Hacia[1], Lawrence C. Brody[1], Mark S. Chee[2], Stephen P. A. Fodor[2] & Francis S. Collins[1]

The ability to scan a large gene rapidly and accurately for all possible heterozygous mutations in large numbers of patient samples will be critical for the future of medicine. We have designed high-density arrays consisting of over 96,600 oligonucleotides 20-nucleotides (nt) in length to screen for a wide range of heterozygous mutations in the 3.45-kilobases (kb) exon 11 of the hereditary breast and ovarian cancer gene *BRCA1*. Reference and test samples were co-hybridized to these arrays and differences in hybridization patterns quantitated by two-colour analysis. Fourteen of fifteen patient samples with known mutations were accurately diagnosed, and no false positive mutations were identified in 20 control samples. Eight single nucleotide polymorphisms were also readily detected. DNA chip-based assays may provide a valuable new technology for high-throughput cost-efficient detection of genetic alterations.

While some genetic conditions such as sickle cell disease, achondroplasia, or the triplet repeat disorders are mutationally monomorphic, extensive heterogeneity of mutations amongst affected individuals is the more usual observation. This allelic heterogeneity presents a considerable challenge to the development of high-throughput cost-effective analytical methods for mutation detection. All possible mutations must be detectable, and in diploid organisms a mutation in an autosomal gene must routinely be identifiable in the context of wild-type sequence from the other allele.

A case in point is the familial early onset breast cancer gene, *BRCA1*. Germline mutations in *BRCA1* are present in 50–60% of kindreds with breast and ovarian cancer, and may account for approximately 2–5% of all breast cancer cases in the general population[1-4]. Heterozygotes are markedly predisposed to early onset breast and ovarian cancer, and are also at moderately increased risk of developing colon and prostate cancer[5]. The protein coding region of *BRCA1* contains 5,592 basepairs (bp) in 22 coding exons spread over 100 kb of genomic DNA[6]. Over 111 unique *BRCA1* mutations distributed throughout the gene have been described[7,8]. Most of these are frameshift, nonsense, or splice mutations resulting in a disruption of the normal reading frame. Except for the Ashkenazi Jewish population, where two mutations account for the majority of *BRCA1* alterations[9–14], allelic heterogeneity confounds the ability to identify *BRCA1* mutation carriers by methods (such as allele-specific oligonucleotide (ASO) hybridization) which detect only a finite set of previously described mutations.

A variety of protocols have been used to screen for all

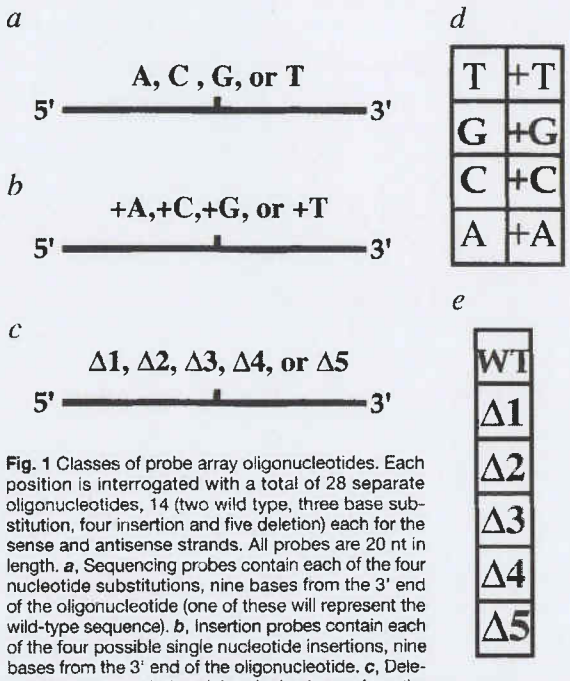

Fig. 1 Classes of probe array oligonucleotides. Each position is interrogated with a total of 28 separate oligonucleotides, 14 (two wild type, three base substitution, four insertion and five deletion) each for the sense and antisense strands. All probes are 20 nt in length. ***a***, Sequencing probes contain each of the four nucleotide substitutions, nine bases from the 3' end of the oligonucleotide (one of these will represent the wild-type sequence). ***b***, Insertion probes contain each of the four possible single nucleotide insertions, nine bases from the 3' end of the oligonucleotide. ***c***, Deletion probes have 1–5 nt deleted, nine bases from the 3'–end of the oligonucleotide. ***d***, Sequencing and insertion probes are tiled adjacent to one another in the upper portion of the array. ***e***, Deletion probes along with the corresponding wild-type probe are tiled in the lower portion of the array.

[1]*National Center for Human Genome Research, Building 49/3A14, National Institutes of Health, Bethesda, Maryland 20892, USA*
[2]*Affymetrix, 3380 Central Expressway, Santa Clara, California 95051, USA*

Correspondence should be addressed to F. S. C.

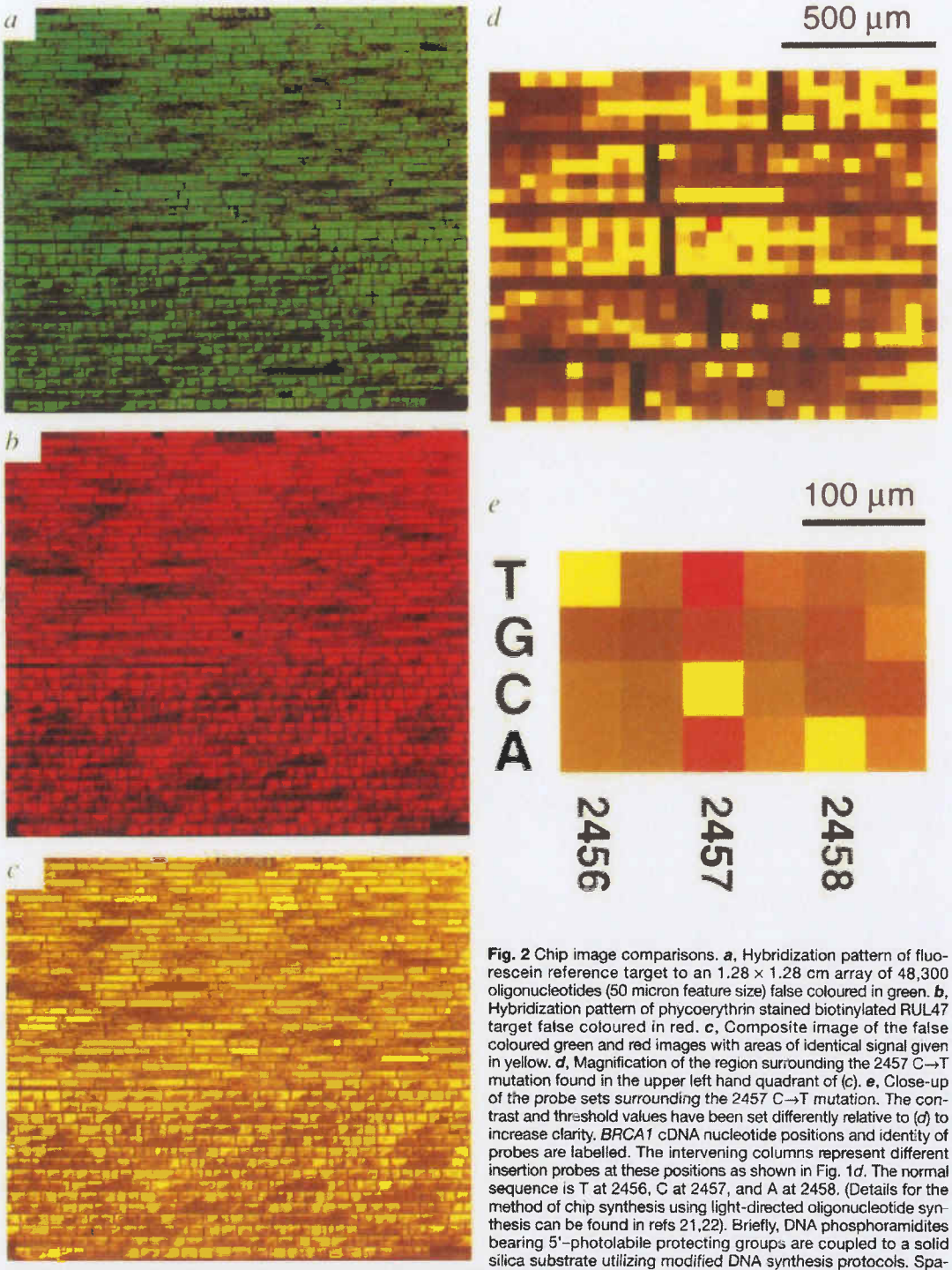

Fig. 2 Chip image comparisons. ***a***, Hybridization pattern of fluorescein reference target to an 1.28 × 1.28 cm array of 48,300 oligonucleotides (50 micron feature size) false coloured in green. ***b***, Hybridization pattern of phycoerythrin stained biotinylated RUL47 target false coloured in red. ***c***, Composite image of the false coloured green and red images with areas of identical signal given in yellow. ***d***, Magnification of the region surrounding the 2457 C→T mutation found in the upper left hand quadrant of (c). ***e***, Close-up of the probe sets surrounding the 2457 C→T mutation. The contrast and threshold values have been set differently relative to (*d*) to increase clarity. *BRCA1* cDNA nucleotide positions and identity of probes are labelled. The intervening columns represent different insertion probes at these positions as shown in Fig. 1*d*. The normal sequence is T at 2456, C at 2457, and A at 2458. (Details for the method of chip synthesis using light-directed oligonucleotide synthesis can be found in refs 21,22). Briefly, DNA phosphoramidites bearing 5'–photolabile protecting groups are coupled to a solid silica substrate utilizing modified DNA synthesis protocols. Spatially addressable synthesis of oligonucleotide species is obtained through photolithographic techniques where selected oligonucleotides are photodeprotected on the chip surface for each coupling cycle. Combinatorial synthesis strategies may yield up to 2^n different oligonucleotide species in n synthesis cycles allowing the described array to be manufactured in 80 such cycles. In the present scheme, 30 identical high density array chips were simultaneously produced in a single 3 h synthesis.

possible *BRCA1* germline mutations, virtually all of which begin with amplification of individual exons by DNA PCR or of the transcript by RT-PCR. These include the single-strand conformation polymorphism assay (SSCP), manual or automated direct DNA sequencing, clamped denaturing gel electrophoresis (CDGE), heteroduplex analysis, and the protein truncation assay[15–20]. All of these require gel electrophoresis, seriously complicating the challenge of scale-up, automation, and reduction in cost.

Advances in light-directed combinatorial chemical synthesis have made manufacturing of high-density

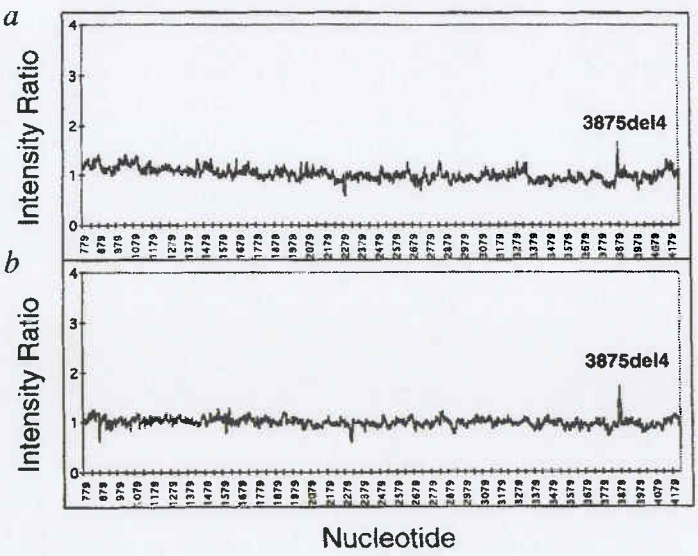

Fig. 3 Two-colour loss of signal assay for a deletion. Fluorescein-labelled reference and biotinylated targets were co-hybridized to the array. To correct for reproducible differences in the hybridization efficiencies of reference and test targets, the ratio of fluorescein to phycoerythrin signal at each wild-type position was normalized against ratios derived from a separate chip co-hybridization experiment. Five data point moving averages of sense and antisense strand-corrected ratios are plotted against nucleotide position. ***a***, Sense strand ratios from 185-F15 (3875del4). ***b***, Antisense strand ratios from 185-F15 (3875del4). A peak at the position of mutation is present on both strands.

oligonucleotide probe arrays on solid surfaces possible[21–22]. Oligonucleotides are generated *in situ* on a silicon surface by combining standard DNA synthesis protocols with phosphoramidite reagents modified with photolabile 5'–protecting groups. Spatially addressable synthesis is accomplished through selective photodeprotection of chip areas, utilizing a photolithographic mask set in a process similar in principle to that utilized in computer microchip manufacture. These deprotected areas are activated for chemical coupling. Selective deprotection of multiple areas containing a distinct oligonucleotide sequence (such as all those having an adenine as the next residue) allows for the simultaneous step-wise synthesis of numerous different oligonucleotide species. In one manufacturing protocol, there are four reactions performed for each coupling step corresponding to the separate photodeprotection and coupling to those oligonucleotides in areas needing incorporation of an A, C, G or T residue. The remainder of each synthesis cycle, including oxidation and capping steps, are performed simultaneously for all oligonucleotides in the array. Multiple arrays consisting of many thousands distinct oligonucleotides can be reproducibly manufactured in several hours' time. They can be designed to provide sequence information of any known gene. By analysing the hybridization pattern of fluorescent-labelled nucleic acid target to such arrays, sensitive high-throughput assays have been developed to screen for mutations in the cystic fibrosis (*CFTR*) gene[23], the HIV-1 reverse transcriptase and protease genes[24–25], the β-globin gene[26], and the mitochondrial genome[27]. None of these prior applications, however, have tested the ability of the method to detect all possible heterozygous mutations in a large gene at high sensitivity and specificity.

Design of oligonucleotide array

We examined the ability of a DNA-chip based assay to detect heterozygous mutations in the 3.45-kb *BRCA1* exon 11, which contains approximately 60% of the *BRCA1* coding region, including 10-bp of flanking intronic sequence. Families of over 96,600 oligonucleotides were designed to detect all possible single base substitutions, single base insertions, and 1–5-bp deletions on both strands. Four 20-nt sequencing probes, substituted with one of the four nucleotides in the central position, interrogate the identity of each nucleotide (Fig. 1*a*,*d*). Four 20-nt insertion probes containing the possible single base insertions at the central position query for the presence and identity of an insertion (Fig. 1*b*,*d*). Likewise five 20-nt long deletion probes query for the presence and identity of all possible 1–5-bp deletions (Fig. 1*c*,*e*). All probes were selected to be 20-nt long as this length gave the optimal signal intensity and specificity under the tested experimental conditions (data not shown). Allele-specific oligonucleotides complementary to other described mutations not included in the above classes may be easily incorporated.

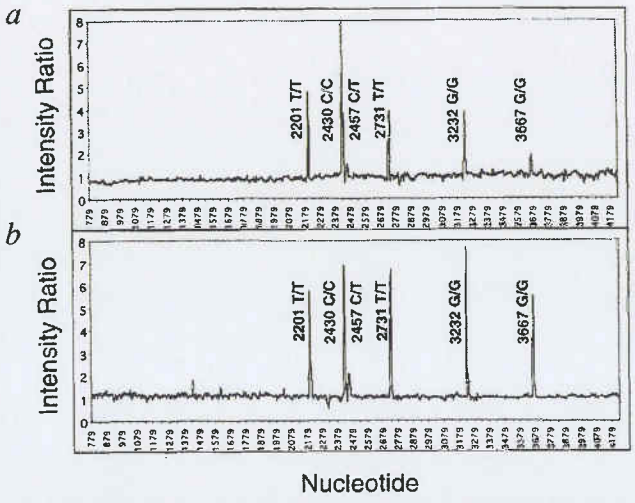

Fig. 4 Two-colour loss of signal assay for a nonsense mutation. ***a***, Sense strand ratios from RUL47 (2457 C→T). ***b***, Antisense strand ratios from RUL47 (2457 C→T). The heterozygous nonsense mutation and five homozygous polymorphisms appear as distinct peaks on both strands.

Table 1 Sensitivity of mutation detection in patient samples with germline mutations in *BRCA1* exon 11

Mutation	Sample	Gain of Signal[a]		Loss of Signal[b]		Mutation Identification[c]
		Coding	Noncoding	Coding	Noncoding	
1128insA	ST750	–	–	–	–	No
1294del40	624–F32	na	na	+	+	Yes
1323delG	3295	–	+	+	–	Yes
2294delG	ST755	–	–	+	+	Yes
2314del5	RUL57	+	+	+	+	Yes
2457C–>T	RUL47	+	+	+	+	Yes
2804delAA	MOC 52	+	+	+	+	Yes
3121delA	ENG9	+	+	–	+	Yes
3286delG	ENG7	+	+	+	+	Yes
3452del4	ENG5	–	–	+	+	Yes
3600del11	3265	–	–	+	+	Yes
3867G–>T	808–F161	+	+	+	+	Yes
3875del4	185–F15	–	–	+	+	Yes
3937insG	RUL77	+	+	+	+	Yes
3986delAA	ENG3	+	+	+	+	Yes

[a]'+'indicates that the mutant probe has an intensity 1.2× or greater than the corresponding wild-type probe. '–' indicates that the mutant probe has an intensity less than a factor of 1.2× to the corresponding wild-type probe; na indicates data not available. [b]'+' indicates a distinct peak at the mutant position. '–' indicates the absence of a distinct peak at the mutant position. [c]mutation detection algorithm defined in text and Fig. 5.

This chip design provides redundant information which contributes to sensitivity and specificity. Ideally, a heterozygous mutation in a patient sample should result in (i) a 'gain of signal' increase in hybridization to an oligonucleotide representing a perfect match to the mutant sequence, provided it is represented on the chip and (ii) a 50% 'loss of signal' intensity (relative to a normal control) for the family of wild-type oligonucleotide probes that query the position of the mutation. Under ideal circumstances, a true positive should appear on both stands. Due to the complexity of the hybridization reaction, specific mutations may only fulfill a subset of these ideal criteria. One of our goals was to define an algorithm which maximizes sensitivity and specificity in the analysis of this intentionally redundant data.

Advantages of a two-colour analysis system

Two-colour assays have been used to great advantage in comparative genomic hybridization (CGH) as a means of measuring two-fold or greater differences in copy number between samples such as occur with chromosomal deletions, duplications, and amplifications[28]. This approach is based upon measuring the relative ratios of fluorescence from reference and test targets, labelled with different fluorophores and bound to normal chromosomal spreads after a competitive hybridization step. Based on the success of this approach, we investigated the use of two-colour assay systems to provide an analytical means of detecting heterozygous sequence differences between samples at the nucleotide level. This allows a focus on the differences between a wild-type reference standard and the unknown patient sample, rather than requiring *de novo* determination of the sequence with each patient sample. In this approach, wild-type fluorescein-labelled ('green') reference and biotinylated (stained with a phycoerythrin-streptavidin 'red' conjugate after hybridization) *in vitro* transcribed RNA test targets were competitively co-hybridized to the array with the relative binding to all probes measured[27]. The ratio of reference and test targets occupancy to each of the 96,600 oligonucleotides in the array was used to detect sequence differences between the two samples. RNA-based targets were used in favor of DNA-based targets as they showed superior hybridization fidelity and signal strength in this system (data not shown).

Gain of signal analysis

Mutant substitution, insertion, and deletion probes should detect sequence changes through a gain of hybridization signal in the test target, as the wild-type target generally should not hybridize strongly to them. We used this analysis to detect heterozygous base substitutions (Fig. 2). The hybridization pattern of the reference is shown in green (Fig. 2*a*) while the hybridization pattern of sample RUL47 containing a 2457 C->T nonsense mutation is shown in red (Fig. 2*b*). These images were superimposed with the areas of identical signal given in yellow (Fig. 2*c*–e). Only the wild-type allele 2457 'C' probe hybridized with the reference target. In contrast, the heterozygous mutant target

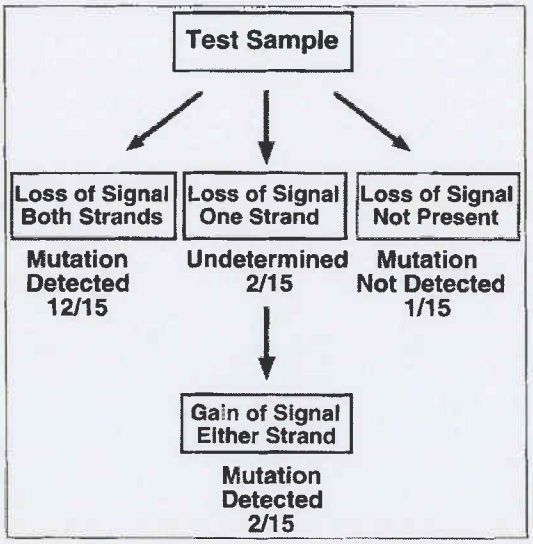

Fig. 5 Data analysis algorithm flowchart. The number of analysed patient samples containing exon 11 mutations which fulfill the indicated criteria of loss-of-signal or gain-of-signal assays is given. Fourteen of fifteen patient samples fulfill the requirements of mutation detection given in the text.

hybridized to both the wild-type allele 2457 'C' probe as well as the mutant allele 2457 'T' probe, resulting in a red signal at the position of the mutation.

Loss of signal analysis

Two-colour loss of signal assays are a second analytical means of detecting sequence differences between samples. Localized changes in hybridization signal ratios may reflect different stoichiometries of wild-type alleles in the hybridization solution, and thus potential heterozygous sequence differences between the two samples. A corrected ratio (see Fig. 3 legend) of reference sample and test sample binding to wild-type sequencing probes for each strand can be plotted against nucleotide position. Regions of identical sequence should be close to a value of 1.0, while regions with sequence differences should show a peak centered near the point of mutation. For a point mutation, this width should be about 20 bp wide; for a deletion of *n* bp it is expected to be approximately (*n*+20)-bp in width. This 'width property' of true peaks helps to distinguish them from noise (Fig. 3*a*,*b*). Ideal heterozygous mutations should produce peaks with a value of about 2.0, reflecting the two wild-type alleles in the reference compared to the single wild-type allele in the mutant heterozygote sample. Any cross-hybridization of the mutant allele to the wild-type probe will reduce this ratio closer to 1.0. In practice, a cut-off of 1.2 was found to represent a good threshold.

Relative to the heterozygous state, homozygous sequence differences will produce larger peaks due to the absence of signal from a wild-type allele. In these cases the theoretical peak height is infinite although in practice cross-hybridization usually reduces these signals to be within a value of 10. For example, six separate peaks were observed for each strand of sample RUL47 (Fig. 4*a*,*b*). One of these (the smallest) detects the heterozygous 2457 C→T nonsense mutation. The other five strong peaks correspond to five polymorphisms 2201 T/T, 2430 C/C, 2731 T/T, 3232 G/G, and 3667 G/G found in sample RUL47 in the homozygous state. These variants have been described and are in strong disequilibrium with each other[16,29]. Polymorphic signals will not be observed when the test sample has the same genotype as the reference.

Application to multiple *BRCA1* mutations

A summary of the results of two-colour gain and loss of signal analysis experiments for 15 known exon 11 *BRCA1* mutations is given in Table 1, representing all of the genomic DNA samples available to us which contain known alterations in this exon. In addition, twenty control samples from individuals without a known family history of breast cancer were evaluated to ascertain the specificity of the assay. Seven reported polymorphisms were detected along with a previously unreported heterozygous base substitution 1606 G→A (Arg496His) found in sample RUL57, which was confirmed by dideoxysequencing. Seven of the pathologic mutations (2314del5, 2457C→T, 2804delAA, 3286delG, 3867G→T 3937insG, and 3986delAA) were detected in completely optimal fashion with clear gain of signal and loss of signal results on each strand. For the other mutations, the sensitivity of any particular assay on one or both strands was imperfect. For example, while the 3875del4 mutation was readily detected with the loss of signal assay (Fig. 3*a*,*b*), the gain of signal assay failed to give a distinct signal, presumably because of strong wild-type target cross-hybridization at this location. Gain of signal assays based upon insertion and deletion probes were also capable of generating a significant number of false positive signals using this criteria (data not shown), whereas the loss of signal assay was much more robust.

Discussion

Our observations suggest the following procedure for interpreting chip hybridization data (Fig. 5). First examine the loss of signal data. If a peak of width >20 bp is found in the same position on both strands, a sequence alteration is almost certainly present. We encountered no false positives of this sort in examining 3.45 kb of *BRCA1* exon 11 sequence in 15 patients and 20 control samples (a total of ≥120 kb of *BRCA1* sequence). Twelve of the 15 patient mutations in Table 1 were immediately detectable by this strategy, and the precise mutation could then be identified in seven of them by examining the gain of signal data. In ambiguous cases where there is a loss of signal on one strand but not the other, the gain of signal data can still lead to accurate mutation detection. Two samples show loss of signal (of ≥20 bp) on one strand but not the other. In one of these (ENG9, 3121delA) there is a specific gain of signal for the appropriate oligonucleotide on both strands, whereas the other (3295, 1323delG) reveals this gain of signal above threshold on one strand only. In 20 control samples, a specific loss of signal on one strand was never accompanied by a confirmatory gain of signal on either strand. As this criterion does not appear to result in false positives, we scored both ENG9 and 3295 as having been correctly identified as mutation-bearing. Only one of the fifteen samples (ST750, 1128insA), which showed neither a specific gain or loss of signal on either strand, would be scored as a false negative in the current assay. That mutation results from the expansion of a poly (dA) • (dT) tract from 7–8 nts in length, and would be predicted to be particularly difficult to detect. Interestingly, two other samples that putatively contained the mutations 2086insG and 2035 T→A did not generate specific loss or gain of signal on either strand. Dideoxysequencing analysis confirmed that these samples were of wild-type sequence in this region. Thus with the current algorithm the sensitivity of the method is 93% and the specificity is 100%.

There are a number of possible mechanisms for false negative and false positive mutation detection results. Because the RNA targets are fragmented to an average length of 50–100 nt in order to minimize secondary structure formation prior to hybridization, the differing lengths of RNA species between reference and test targets may affect the result. Sequence changes that enhance or decrease hybridization due to intramolecular or intermolecular target structure may also confound analysis. Certain 20-mer probes may adopt secondary structures that could inhibit target hybridization. Short repetitive sequences (as is the 1128insA case) and duplications will pose a serious challenge to any hybridization based assay. In these cases, there is an increased potential for cross-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.