Resequencing and mutational analysis using oligonucleotide microarrays

Joseph G. Hacia

Genetics and Molecular Biology Branch, National Human Genome Research Institute, Building 49/3A14, National Institutes of Health, Bethesda, Maryland 20892, USA. e-mail: jhacia@nhgri.nih.gov

Oligonucleotide microarray (DNA chip)-based hybridization analysis is a promising new technology which potentially allows rapid and cost-effective screens for all possible mutations and sequence variations in genomic DNA. Here, I review current strategies and uses for DNA chip-based resequencing and mutational analysis, the underlying principles of experimental designs, and future efforts to improve the sensitivity and specificity of chip-based assays.

As the Human Genome Project contemplates completion of a reference sequence of the human genome, increasing attention is being paid to uncovering DNA sequence variations among groups of individuals as well as between different human populations[1,2] (see page 56 of this issue (ref. 3)). Identifying and cataloguing these variations is a critical part of positional and candidate gene approaches that seek to identify the genetic basis for predisposition and resistance to disease. These sequence variations will serve as genetic markers in studies of diseases and traits with complex inheritance patterns and strong environmental interactions[4,5]. Large-scale sequence analysis is needed for population-based genetic risk assessment and diagnostic tests once mutations have been identified, and traditional technologies cannot easily meet the demands for rapid and cost-effective large-scale comparative sequence and mutational analysis.

Microarray (or 'chip')-based hybridization analysis is a newly emerging technology with diverse applications[6,7]. Two commonly used types of DNA chips differ only in the size of the arrayed nucleic acid components (see page 10 of this issue (ref. 8)): those sporting relatively large nucleic acid components (usually longer than 100 nt) are often used in RNA expression analysis[9] and those with short nucleic acids (oligonucleotides up to 25 nt) can be used for both RNA expression[10] and sequence analysis.

Early proposals for large-scale, hybridization-based sequence analysis involved combinatorial arrayed libraries of all oligonucleotide species N (typically eight) nucleotides in length to interrogate target DNA of interest[11–14]. In theory, such universal arrays could be used to obtain sequence information for virtually any sequence as all possible complementary sequences are present in the array. However, imperfect hybridization specificity and difficulties in analysing sequences containing short repetitive elements and repeated octamer sequences in multiple isolated locations pose problems, and application to large-scale sequence determination has not been reported. In a related approach, exons 5–8 of *TP53* were PCR-amplified from 12 samples, spotted onto nylon filters and individually probed with 8,192 noncomplementary, radiolabelled, 7-base oligonucleotides[15]. All 13 distinct homozygous or heterozygous mutations in these samples were detected by determining which oligonucleotide probes hybridized to the immobilized targets. Although promising, the need to perform thousands of separate hybridization reactions to evaluate each sample makes this approach more amenable to a large-scale clinical diagnostic lab than a common research laboratory setting.

Most chip-based sequence analyses have been carried out with arrays designed to evaluate specific sequences. The current scientific literature largely centres on arrays (called GeneChips®) manufactured using photolithographic-based methodologies developed by Affymetrix (ref. 16; see also, page 20 of this issue (ref. 17)). Although we will focus here on published data, it should be noted that numerous unpublished DNA chip-based strategies are currently under investigation, primarily in the private sector[18,19]. Several of these technologies, such as mass spectroscopy-based hybridization detection, could have an important role in coming years.

Screening for previously characterized sequence variants

Oligonucleotide array-based detection of known genomic DNA sequence variations was first reported in 1989 (ref. 20). Probes complementary to six *HLA-DQA* alleles as well as nine mutations in *HBB* (encoding β-globin) were spotted onto nylon filters and incubated with biotin-labelled PCR products. Genotype was determined by hybridization signal intensity produced by a colourimetric assay at each allele-specific probe—demonstrating the application of the 'reverse dot blot' approach to a simple system relevant to medical genetics. Advanced oligonucleotide array manufacturing processes have opened the way to evaluating more complex systems. Arrays of 1,480 oligonucleotide probes synthesized *in situ* by photolithographic-based processes were designed to detect 37 known mutations in the coding region of *CFTR*, as well as all possible single-nucleotide substitutions[21]. In a blinded study, ten genomic DNA samples were successfully genotyped by characterizing fluorescent hybridization signals from test and wild-type reference samples at mutation-specific probes relative to those from wild-type samples. In a separate study, arrays of six oligonucleotide probes, generated by spotting oligonucleotides onto activated surfaces, were used to detect three different mutations in *HBB* (ref. 22).

Screening for all possible sequence variants

The complex mutation spectrum of most highly penetrant disease genes in the general population, with the possible exception of alleles associated with more common but less penetrant diseases,

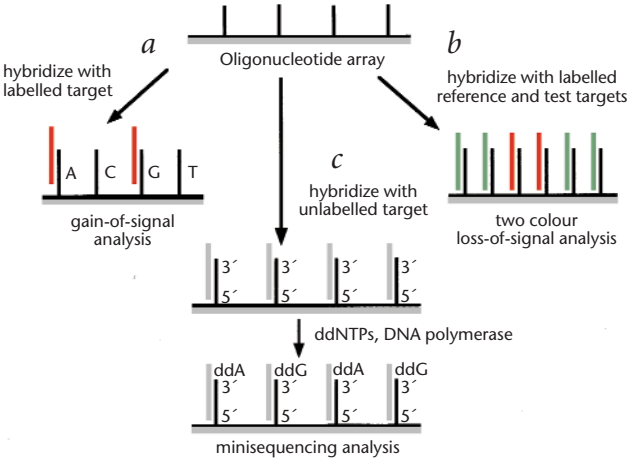

Fig. 1 Three experimental strategies for array-based sequence analysis. In this example, a target heterozygous for both G and A residues at a specific nucleotide position is subject to hybridization-based analysis using three different approaches. In the gain of hybridization signal approach (***a***), relative hybridization to allele-specific probes complementary to each of the four possible nucleotides at the interrogated nucleotide position is used for genotype analysis. In the loss of hybridization signal approach (***b***), decreased hybridization of red-labelled test target relative to green-labelled reference target to perfect match probes interrogating the area of interest indicates the presence of a sequence change. In the minisequencing approach (***c***) unlabelled target is hybridized to perfect match probes (attached through a 5′ linkage to the array to leave an exposed 3′-OH group) interrogating the nucleotide position of interest. Fluorescently tagged ddNTPs are used in subsequent enzymatic primer extension reactions to extend the hybridized primers. The identity of the extended ddNTP is used in genotype analysis.

reflects the allelic heterogeneity of the human genome[3,23,24]. A pertinent example is the hereditary breast and ovarian cancer gene *BRCA1*, for which over 400 distinct mutations have been reported (listed at http://nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/). As a consequence of this diversity, thorough comparative sequencing and mutation analysis must detect all possible changes in the heterozygous state.

Here, we will focus on two hybridization-based, array-based approaches and one enzyme-dependent approach to scanning a sequence of interest for all possible changes using oligonucleotide arrays (Fig. 1). The first strictly hybridization-based approach uses probes complementary to a significant subset of sequence changes of interest, such as those described for *CFTR* (ref. 21), and measures gain of hybridization signal to these probes in tests relative to reference samples. Relative 'gain' of signal by these probes indicates a sequence change similar to that of conventional dot blot analysis (Fig. 1*a*). The second hybridization-only approach analyses selective 'loss' of hybridization signal to perfect-match probes fully complementary to wild-type sequence in a manner analogous to comparative genomic hybridization studies[25,26] (Fig. 1*b*). The enzyme-dependent minisequencing approach[27] analyses primer extension reactions to arrays of perfect match probes hybridized to the target of interest (Fig. 1*c*).

Gain of hybridization signal approach

The gain-of-signal approach allows for a partial scan of a DNA segment for all possible sequence variations (Fig. 2). An array designed to interrogate both strands of a target of length N for all possible single nucleotide substitutions would consist of 8N probes, usually 20–25 nt. There are 4N probes (often termed 'A', 'C', 'G' and 'T' probes) complementary to each of the four possible nucleotides present on a single strand at a particular location. Probes are designed such that the location of the interrogated target base is in the centre-most position of the potential target/probe duplex, and thus provides the best discrimination for hybridization specificity (see page 22 of this issue (ref. 17); ref. 21). To interrogate both target strands for all possible deletions of a specific length, 2N probes are needed. However, interrogating both target strands of length N for all possible insertions of length X requires $2(4^X)N$ probes. When it is necessary to screen both strands of a 10-kb DNA target, 80,000 probes are needed to interrogate for all possible single nucleotide substitutions. An additional 100,000 probes are needed to screen for all 1–5-bp deletions, whereas 27,280,000 probes are needed to screen for all 1–5-bp insertions. The use of shorter probe lengths and analytical schemes proposed for universal combinatorial library arrays could reduce the number of probes necessary to interrogate for these complex sequence changes, but at the expense of proven sensitivity and specificity. Thus, despite recent advances in array manufacturing technology, the search for insertions of more than a single base pair is usually not a practical alternative.

The entire 297-bp HIV-1 protease gene coding sequence from 167 viral isolates was screened for all single nucleotide changes using a gain-of-signal approach[28]. Fluorescently tagged RNA-based targets were hybridized to a pair of arrays consisting of 12,224 oligonucleotides. Nucleotide identities were assigned on the basis of hybridization signals from single-nucleotide substitution probes. Dideoxysequencing and array-based analysis of 33,858 bp of HIV-1 *pr* sequence from 114 samples demonstrated a 98.26% concordance between the data sets. From a biophysical viewpoint, this is functionally equivalent to screening for homozygous changes in genomic DNA sequences.

Loss-of-signal approach

In the loss-of-signal approach, sequence variations are scored by quantitating relative losses of hybridization signal to perfect match oligonucleotide probes in test samples relative to wild-type reference targets. Ideally, a homozygous sequence change results in a complete loss of hybridization signal to perfect match probes interrogating the region surrounding the sequence change (Fig. 3). For heterozygous sequence variations, a 50% loss of signal intensity relative to the wild-type target would ideally be found for perfect match probe signals interrogating and flanking the sequence change.

Loss-of-signal analysis allows for a practical screen to be set up for virtually any sequence variation. With this approach, an array designed to interrogate both target strands of N bp, for all possible sequence changes, minimally consists of 2N overlapping probes. For example, an array of 11,000 oligonucleotides would

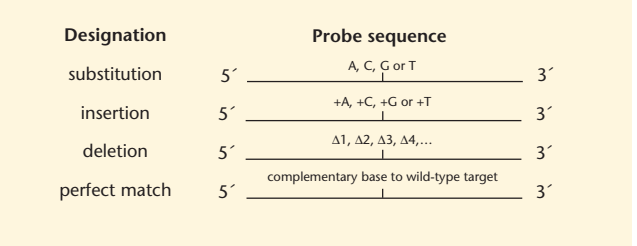

Fig. 2 Typical oligonucleotide probes used in gain-of-signal sequence analysis. Substitution probes interrogate for the presence of all possible single nucleotide substitution sequence variations. Insertion probes interrogate for the presence of inserted sequences, with single nucleotide insertions being the most feasible to completely represent on the array. Deletion probes interrogate for the presence of all possible nucleotide deletion lengths up to a certain size. Perfect-match probes are fully complementary to wild-type sequence. One of these will be represented in the substitution set, but the presence of additional perfect match probes over a range of sizes facilitates the quantitative robustness of the loss-of-signal analysis.

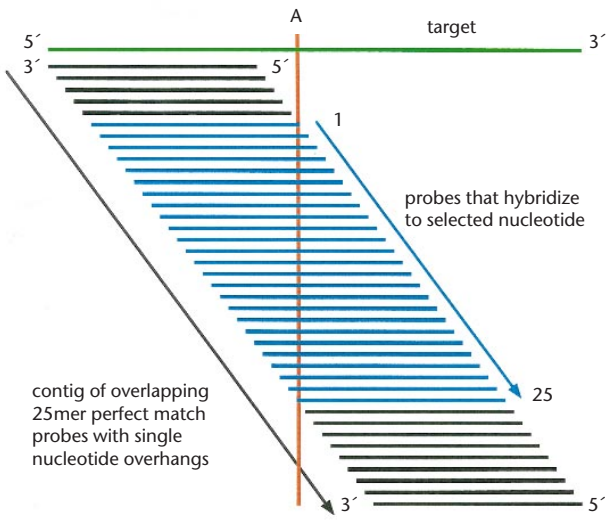

Fig. 3 Oligonucleotide probes used in loss of hybridization signal sequence analysis. Ideally each target nucleotide position contributes to hybridization to a set of N overlapping N-base perfect match probes in an oligonucleotide array. In this example, hybridization to 25 overlapping 25-base probes are affected by changes in a single target nucleotide.

be needed to screen the 5.5 kb of *BRCA1* for all possible sequence variations. Multiple probes (N probes, N nucleotides in length) contribute to detecting a sequence variation, which minimizes random sources of error caused by hybridization signal fluctuations. These fluctuations can bias gain-of-signal analysis, which relies on only two mutation-specific probes (one per strand). Using the loss-of-signal approach, the presence of sequence variations can be displayed by plotting the ratio of perfect-match probe signal intensities from test and reference samples. True losses of hybridization signal are shown by peaks with distinct width and height properties. A disadvantage of the loss-of-signal analysis is that the mutation cannot be discerned; the identity of the sequence change must be established by subsequent dideoxysequencing of the region surrounding the loss of signal signature.

This approach can be further improved by using internal standards in 'two-colour' assays. In this scheme, reference targets of known sequence can be co-hybridized to the arrays along with the test target of interest. By labelling each target with a different fluorophore, a direct comparison of hybridization signals from these two targets can be made. Two reports have appeared using two-colour assays to analyse large DNA segments of hemizygous (mitochondrial DNA[25]) and heterozygous (*BRCA1*) sequence content[26].

Similarly, a pair of oligonucleotide arrays consisting of over 135,000 probes was used to interrogate the entire 16.6-kb human mitochondrial genome from ten samples[25]. By simple gain-of-signal analysis, 99% of the genome could be read correctly, as suggested by parallel dideoxysequencing analysis. In a more thorough analysis of a 2.5-kb mitochondrial sequence from 12 samples, 179 of 180 polymorphisms were detected using gain- and loss-of-signal analysis. Only 2% of this sequence, approximately 50 bp scattered throughout the sequence, required dideoxysequencing analysis, with the remainder unambiguously assigned. Although this was the first study to analyse large targets for all possible sequence variants, it did not test the ability of the approach to scan for heterozygous base changes.

To compare gain- and loss-of-signal analyses, we scanned the 3.43-kb *BRCA1* exon 11 sequence for heterozygous sequence variations[26]. Over 96,000 oligonucleotides in a pair of arrays directly interrogated for all single nucleotide substitutions, single base-pair insertions and 1–5-bp deletions using the gain-of-signal approach. Evaluation of both loss and gain of signal led to the development of a two-tiered algorithm for mutational analysis, detecting 14 of 15 heterozygous mutations scattered throughout the exon. Single nucleotide substitutions generally produced more robust loss-of-signal signatures than small insertions and deletions. The mutation not detected was an insertion into a repetitive sequence in which a poly$(dA)_7$ tract was expanded to a poly$(dA)_8$ tract. We found it necessary to analyse data from both target strands; some sequence changes were more readily discernable on one strand than the other. The loss-of-signal assay demonstrated increased sensitivity and specificity compared with the gain-of-signal assay. The latter is insensitive to the presence of larger deletions and single-base insertions (relative to single nucleotide substitutions) due to increased cross-hybridization of wild-type target to these probes. Furthermore, care should be taken when using mutation-specific probes to identify specific sequence changes. A target containing a single-nucleotide deletion could cross-hybridize to a single-nucleotide substitution probe of similar sequence. Misidentifying a sequence change could dramatically alter the assessment of its biological impact, but misidentification is much less likely when evaluating common sequence variants with known hybridization patterns. Keeping these caveats in mind, gain-of-signal analysis can be used in a complementary manner to validate questionable loss of hybridization signal signatures.

Recent research in our laboratory has focused on scanning the entire coding sequence of *BRCA1* for all possible heterozygous sequence changes using high density oligonucleotide arrays (GeneChips®) supplied by Affymetrix. Two-colour co-hybridization experiments are performed on test and reference targets representing all 22 *BRCA1* coding exons. As with the *BRCA1* exon-11 assay, we have used a combination of loss- and gain-of-

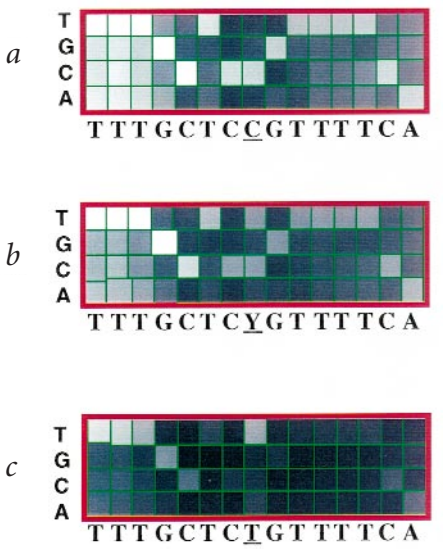

Fig. 4 Hybridization patterns to *BRCA1* oligonucleotide arrays. Magnified digitized greyscale images, 50 μm feature size, showing hybridization patterns of fluorescently labelled *BRCA1* test targets to oligonucleotide arrays evaluating sense-strand target nt 2724–2738 all at comparable contrast levels. Hybridization patterns of nearby mismatches are difficult to visualize due to limitations in printing technology as well as in the linear range of the human eye for detecting subtle greyscale changes. Nucleotide identities are given under the respective columns with the polymorphic 2731 position underlined. *a*–*c*, Depiction of a tiling pattern based on a 2731 C/C reference sequence. Hybridization patterns of samples with 2731 C/C (***a***), 2731 C/T (***b***) and 2731 T/T (***c***) genotypes are shown. The probes interrogating the nucleotides flanking position 2731 in (*c*) are dim due to the presence of single nucleotide mismatches in all target-probe duplexes.

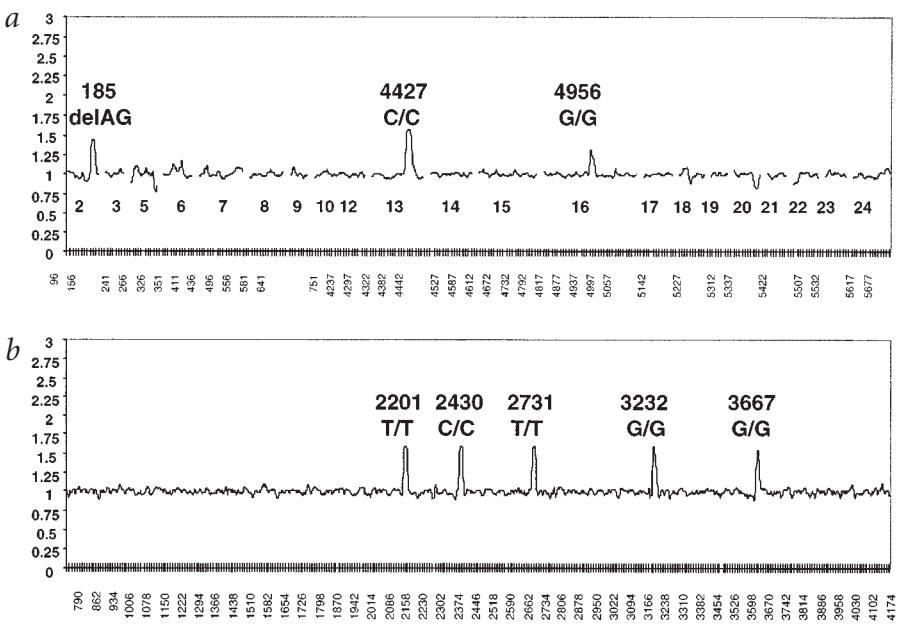

Fig. 5 Detection of the 185delAG *BRCA1* mutation and sequence variations using loss of signal analysis. Fluorescein-labelled (green) reference and a biotinylated (red) test targets representing the entire 5,382-bp *BRCA1* coding sequence were co-hybridized to oligonucleotide arrays interrogating *BRCA1* sense and antisense strand sequence, respectively. Averaged sense and antisense strand reference/test sample processed hybridization signal ratios are shown with the identity of each exon listed below the appropriate data points. Results from all coding exons except 11 (***a***), as well as results from the 3.43-kb *BRCA1* exon 11 (***b***) are shown. The labelled peak in (*a*) is at the site of the heterozygous 185delAG mutation found in the test sample. The other labelled peaks found in (*a*) and (*b*) correspond to seven common homozygous polymorphisms also found in this sample.

signal approaches (Figs 4,5). Preliminary blinded analysis of samples with known *ATM* and *BRCA1* mutations suggests that greater than 90% mutation detection sensitivity can be achieved using oligonucleotide microarrays in these systems (ref. 29; J.G.H and F. Collins, manuscript in in preparation).

Oligonucleotide arrays are now used in large-scale identification and genotyping of single nucleotide polymorphisms (SNPs) in the human genome—149 chip designs, each containing 150,000–300,000 oligonucleotides have been used in one published screen[29]. A total of 2.3 Mb of human sequence was screened and 3,241 candidate SNPs were found through both gain- and loss-of-signal schemes, without the use of a two-colour co-hybridization approach. A separate chip containing a simple tiling scheme for scoring SNP alleles on both target strands was used for the simultaneous genotyping of 500 markers. SNPs involving base substitutions lend themselves beautifully to array technology, as they are generally easier to detect than insertions and deletions, and because SNP screens do not have to be exhaustive to be successful.

In loss-of-signal assays without internal two-colour hybridization standards, it is critical that all perfect match probe hybridization signals be linear with respect to target concentration. Although true for probes of preselected sequence content as shown by quantitative RNA expression experiments[10], we have not found this to be true for all perfect match probes tiling across the entire *BRCA1* gene. Robust loss-of-signal hybridization signatures, such as those frequently generated from single nucleotide substitutions, can be detected using single-colour approaches without internal standards. However, less robust signatures derived from many deletion and insertion mutations could not be detected in our system without a co-hybridized internal standard. Significantly smoother baseline ratio values essential for rigorous heterozygous *BRCA1* mutation scanning applications are provided by two-colour co-hybridization experiments (J.G.H. and F. Collins, unpublished data).

Minisequencing-based assays

Minisequencing assays provide a powerful means to scan for all possible sequence variations. They couple target hybridization with enzymatic primer extension reactions[27,30–32]. Oligonucleotides are typically tethered to the surface via 5′-end linkage to leave an exposed free 3′-OH group, or placed into a microtitre dish format. Labelled dideoxyribonucleoside triphosphates are used in primer extension reactions in which hybridized target and oligonucleotide probes serve as template and primer, respectively. A mixture of all four dideoxyribonucleoside triphosphates, each labelled with a different fluorescent dye, can be used in the primer extension reactions[32]. The identity of the extended dideoxyribonucleotide is determined through fluorescent microscopy and used to assign the identity or identities of the target nucleotide extended from the 3′ end of each probe. Similar to the loss-of-signal approach, a minisequencing array designed to interrogate both target strands of N bp for all possible sequence changes minimally consists of 2N overlapping probes. Single nucleotide substitutions in the heterozygous state will produce two signals corresponding to the identity of the two alleles. Although the exact nature of target insertions and deletions cannot always be determined, their end points can be identified. Thus far, the most complex nucleotide tract analysed by this approach for all possible sequence changes is a 33-bp region of *TP53* (ref. 32).

To test the minisequencing approach in more complex systems, high density oligonucleotide arrays with the appropriate surface linkages must become accessible. Commercially available Affymetrix arrays are not suited to this task, as their oligonucleotides are tethered through 3′-end linkages to the surface. Modified chemistries should allow for other types of arrays to be manufactured using photolithography[33]. Alternative processes, such as the ink jet[34] and spotting techniques[35], can also be used[36].

Future challenges

Among the greatest strengths of array-based mutational analysis is the ability to detect specific sequence changes of interest. Once specific hybridization patterns or 'signatures' of large numbers of mutant alleles of interest are known, it will be possible to search for those signatures in many different samples simultaneously. Screening for all possible sequence changes, however, is more difficult. Similar to other mutation screening procedures, the array-based system is more sensitive in the detection of homozygous base changes, but there are only a limited number of applications[25,37].

Although it is possible to use oligonucleotide microarrays in heterozygous mutation screens, accuracy must be improved. Similar to almost all mutation screening technologies (with the possible exception of rigorously designed dideoxysequencing

analysis), the array-based approach is not yet ready for use in applications needing more than 98% heterozygous mutation detection in complex autosomal sequences. The current methodologies are more suitable under conditions where a modest false negative error rate (5–10%) is permitted, such as in polymorphism screens in which it is not necessary to generate an exhaustive collection of single nucleotide polymorphism markers or in some population genetics studies in which individual genotypes are not returned to study participants. The biggest challenge in hybridization-based mutational analysis is in decreasing the false negative rate. Variations in target nucleotide sequence composition, including the presence of repetitive sequence elements, dramatically influence the sensitivity of hybridization-based sequence analysis. The same experimental approach applied to any two sequences can yield results with vastly different accuracy. For example, it would not be reasonable to apply DNA chip technology to genotyping genes involved in triplet repeat-based diseases, such as the Huntington disease gene[38], where it is crucial to know the exact number of tandem repeat units in a sequence of interest and a simple gel-based analysis system is currently available.

Intra- and intermolecular structures, such as hairpin and G-quartets, present in either target or probe can make target hybridization less predictable (ref. 39; see also, page 5 (ref. 40)). Variant sequences that disrupt secondary structures could actually have higher affinity to 'perfect match' probes than wild-type sequences containing these structures. Furthermore, it may not be possible to fully interrogate all target sequence tracts due to poor hybridization signals caused by secondary structure. Sequence variations involving insertions and deletions pose a special problem as they produce targets that may form duplexes containing bulged nucleotides with wild-type probes. In some repetitive contexts, such as homopolymeric nucleotide tracts, bulged duplex formation can be energetically favoured over formation of duplexes containing centrally-placed, single base-pair mismatches. Such factors make detecting insertions and deletions by hybridization-based methodologies generally more challenging than single nucleotide substitutions.

Several specialized problems affecting almost all mutation detection schemes also apply to array technology. All PCR-based assays are prone to errors associated with sequence changes being present under a primer-binding site which allow only one target allele to be represented in the analysis. Another specialized problem exists when attempting to detect sequence changes present in a trace population. This may exist when analysing tumour sample DNA or trace viral species. In these cases, advances in target preparation, such as laser capture microdissection[41], may be necessary to improve target preparation.

A number of approaches have been made to normalize hybridization signals in all the arrayed probes, which would increase the diversity of sequences that can be interrogated at one time. For example, despite the use of buffer systems such as those containing tetramethylammonium chloride[25,26], it is especially difficult to simultaneously assay sequence tracts with localized regions of high G/C and A/T content. Suboptimal hybridization conditions must be used to allow sufficient hybridization signal to be gathered at low stability A/T-rich duplexes and allow single mismatch discrimination at more highly stable G/C-rich duplexes. Properties of chemically modified oligonucleotides should be exploited[42]. A pertinent example is the development of peptide nucleic acid arrays which could increase hybridization sensitivity and specificity[43,44]. Furthermore, incorporating modified uridine ribonucleoside triphosphates in RNA-based targets has proven useful in *BRCA1* analysis[45], and other modified triphosphates could have beneficial effects as well[46]. Hybridization signal can be enhanced through the use of energy transfer dyes with optimized spectral properties and enzymatic rolling-circle amplification assays[47,48]. Improved synthesis protocols that increase the quality as well as the number of oligonucleotides in each array may increase the robustness of the assays as well. It is critical that arrays produced within, as well as between, manufacturing lots have reproducible hybridization properties. A sophisticated means of normalizing hybridization signal involves coupling electrophoresis-based technology with hybridization[49,50]. Target affinity to specific probes can be altered by electrical fields at specific probe locations.

Other proposals have been made to enhance the specificity of array hybridization. One involves coupling enzymatic ligation reactions to the hybridization scheme to maximize the destabilizing effect of mismatches at the end of duplexes[51,52]. Another involves the hybridization of unlabelled target to the oligonucleotide array and the subsequent hybridization of sets of fluorescently tagged 5-base oligonucleotides to the immobilized duplex. In this 'contiguous stacking hybridization' approach, sequence information is derived from the identity of the short oligonucleotides which hybridize to the target-probe complex[22,53]. These must be tested on more complex systems before their practical utility can be evaluated.

Although hybridization-based sequence analysis faces technical hurdles, there is room for optimism. There is the potential for unprecedented throughput in mutational analysis with a high degree of accuracy. Given that it took more than ten years before dideoxysequencing analysis evolved into a high throughout sequencing tool[54,55], hybridization-based sequencing assays are in their infancy. Further improvements will ensure that it matures into an even more powerful analytical tool for accurate and high-throughput resequencing and mutational analysis.

Acknowledgements

I thank L. Brody, F. Collins, K. Edgemon, A. Mayer and B. Sun from NIH for helpful suggestions and discussion. I also thank M. Chee at Illumina, M. Cronin at Protogene, S. Fodor at Affymetrix and N. Hunt at GeneTrace for comments and discussion.

1. Collins, F.S., Guyer, M.S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
2. Nickerson, D.A. *et al.* DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233–240 (1998).
3. Chakravarti, A. Population genetics—making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
4. Moldin, S.O. & Gottesman, I.I. At issue: genes, experience and chance in schizophrenia—positioning for the 21st century. *Schizophr. Bulletin* **23**, 547–561 (1997).
5. Zhang, H., Zhao, H. & Merikangas, K. Strategies to identify genes for complex diseases. *Ann. Med.* **29**, 493–498 (1998).
6. Southern, E.M. DNA chips: analysing sequence by hybridization to oligonucleotides on a large scale. *Trends Genet.* **12**, 110–115 (1996).
7. Ramsay, G. DNA chips: state-of-the art. *Nature Biotechnol.* **16**, 40–44 (1998).
8. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10–14 (1999).
9. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
10. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
11. Bains, W. & Smith, G. A novel method for nucleic acid sequence determination. *J. Theor. Biol.* **135**, 303–307 (1988).
12. Khrapko, K.R. *et al.* An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* **256**, 118–122 (1989).
13. Drmanac, R., Labat, I., Brukner, I. & Crkvenjakov, R. Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics* **4**, 114–128 (1989).
14. Southern, E.M., Maskos, U. & Elder, J.K. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* **13**, 1008–1017 (1992).
15. Drmanac, S. *et al.* Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nature Biotechnol.* **16**, 54–58 (1998).

16. Fodor, S.P.A. *et al.* Light-directed spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
17. Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. & Lockhart, D.J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, 20–24 (1999).
18. Marshall, A. & Hodgson, J. DNA chips: an array of possibilities. *Nature Biotechnol.* **16**, 27–31 (1998).
19. Schena, M. *et al.* Microarray: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**, 301–306 (1998).
20. Saiki, R.K., Walsh, P.S., Levenson, C.H. & Erlich, H.A., Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl Acad. Sci. USA* **86**, 6230–6234 (1989).
21. Cronin M.T. *et al.* Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum. Mutat.* **7**, 244–255 (1996).
22. Yershov, G. *et al.* DNA analysis and diagnostics on oligonucleotide microchips. *Proc. Natl Acad. Sci. USA* **93**, 4913–4918 (1996).
23. Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S. & Schmidtke, J. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**, 201–205 (1985).
24. Hofker, M.H. *et al.* The X chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. Hum. Genet.* **39**, 438–451 (1986).
25. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
26. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P. & Collins, F.S. Detection of heterozygous mutations in *BRCA1* using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genet.* **14**, 441–447 (1996).
27. Pastinen, T., Kurg, A., Metsapalu, A., Peltonen, L. & Syvanen, A.C. Mini-sequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* **7**, 606–614 (1997).
28. Kozal, M. *et al*. Extensive polymorphisms observed in HIV-1 cladeB protease gene using high-density oligonucleotide arrays. *Nature Med.* **2**, 753–759 (1996).
29. Hacia, J. *et al.* Strategies for mutational analysis of the large multi-exon *ATM* gene using high-density oligonucleotide arrays. *Genome Res.* (in press).
30. Wang, D.G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
31. Shumaker, J.M., Metspalu, A. & Caskey, C.T. Mutation detection by solid phase primer extension. *Hum. Mutat.* **7**, 346–354 (1996).
32. Nikiforov, T.T. *et al.* Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res.* **22**, 4167–4175 (1994).
33. Head, S.R. *et al.* Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumor suppressor gene. *Nucleic Acids Res.* **25**, 5065–5071 (1997).
34. McGall, G. *et al.* Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl Acad. Sci. USA* **93**, 13555–13560 (1996).
35. Blanchard, A.P., Kaiser, R.J. & Hood, L.E. High-density oligonucleotide arrays. *Biosens. Bioelectron.* **11**, 687–690 (1996).
36. Proudnikov, D., Timofeev, E. & Mirzabekov, A. Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. *Anal. Biochem.* **259**, 34–41 (1998).
37. Hacia, J.G. *et al*. Evolutionary sequence comparisons using high-density oligonucleotide arrays. *Nature Genet.* **18**, 155–158 (1998).
38. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
39. Milner, N., Mir, K.U. & Southern, E.M. Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nature Biotechnol.* **15**, 537–541 (1997).
40. Southern, E. Mir, K. & Shchepinov, M. Molecular interactions on microarrays. *Nature Genet*. **21**, 5–9 (1999).
41. Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R. & Liotta, L.A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet.* **14**, 272–276 (1998).
42. Hoheisel, J.D. Sequence-independent and linear variation of oligonucleotide DNA binding stabilities. *Nucleic Acids Res.* **24**, 430–432 (1996).
43. Weiler, J., Gausepohl, H., Hauser, N., Jensen, O.N. & Hoheisel, J.D. Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays. *Nucleic Acids Res.* **25**, 2792–2799 (1997).
44. Arlinghaus, H.F., Kwoka, M.N. & Jacobson, K.B. Analysis of biosensor chips for identification of nucleic acids. *Anal. Chem.* **15**, 3747–3753 (1997).
45. Hacia, J.G. *et al*. Enhanced high density oligonucleotide array- based sequence analysis using modified nucleoside triphosphates. *Nucleic Acids Res.* **26**, 4975–4982 (1998).
46. Nguyen, H.K. *et al.* The stability of duplexes involving AT and/or G4EtC base pairs is not dependent on their AT/G4EtC ratio content. Implication for DNA sequencing by hybridization. *Nucleic Acids Res.* **26**, 4249–4258 (1998).
47. Hacia, J.G. *et al*. Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes. *Nucleic Acids Res.* **26**, 3865–3866 (1998).
48. Lizardi, P.M. *et al*. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genet.* **19**, 225–232 (1998).
49. Edman, C.F. *et al.* Electric field directed nucleic acid hybridization on microchips. *Nucleic Acids Res.* **25**, 4907–4914 (1997).
50. Sosnowski, R.G., Tu, E., Butler, W.F., O'Connell, J.P. & Heller, M.J. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc. Natl Acad. Sci. USA* **94**, 1119–1123 (1997).
51. Broude, N.E., Sano, T., Smith, C.L. & Cantor, C.R. Enhanced DNA sequencing by hybridization. *Proc. Natl Acad. Sci. USA* **91**, 3072–3076 (1994).
52. Dubiley, S., Kirillov, E., Lysov, Y. & Mirzabekov, A. Fractionation, phosphorylation and ligation on oligonucleotide microchips to enhance sequencing by hybridization. *Nucleic Acids Res.* **25**, 2259–2265 (1997).
53. Parinov, S. *et al*. DNA sequencing by hybridization to microchip octa- and decanucleotides extended by stacked pentanucleotides. *Nucleic Acids Res.* **24**, 2998–3004 (1996).
54. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. USA* **74**, 5463–5467 (1977).
55. Wilson, R.K., Chen, C., Avdalovic, N., Burns, J. & Hood, L. Development of an automated procedure for fluorescent DNA sequencing. *Genomics* **6**, 626–634 (1990).