TOWARDS AUTOMATION FOR MOLECULAR DIAGNOSIS OF CANCER

Maurizio Ferrari, Paola Carrera

Corresponding author's address:
Paola Carrera
Unit Genomics for Human Disease
Diagnosis and Diagnostica e Ricerca
San Raffaele SpA,
IRCCS Ospedale San Raffaele,
Via Olgettina, 60,
Milan, Italy, 20132

13.1 Mutation detection techniques and steps towards automation.

Mutation detection techniques can be divided into those which test for known mutations (genotyping) and those which scan for any mutation in a given target region (mutation scanning). Broader aspects of mutation detection include identification of gene dosage alterations, gross re-arrangements and methylation. The prime considerations in any approach to mutation detection are sensitivity (the proportion of mutations that can be detected) and specificity (the proportion of false positives). Cost per genotype and throughput are also important factors in service delivery. It is often difficult to assess these features accurately from the published scientific literature.

Automation represents a significant progress in the mutation detection field since it improves throughput, decreases costs per sample and allows simultaneous analyses. Major reaction principles for mutation detection, both for genotyping and mutation scanning, have been transferred from manual to semi-automated or full-automated formats.

Genotyping can be performed mainly for linkage analysis or for detection of SNPs. Million SNPs have been already identified and mapped in the human genome. They provide a rich source of information about the evolutionary history of human population. Human SNP map will be very useful in dissecting the contribution of individual genes to diseases that have a complex inheritance. Most human variation that is influenced by genes can be traced to SNPs, especially in such medically important traits as how likely you are to become afflicted with a particular disease, or how you might respond to a particular pharmaceutical treatment.

With such a scenario, it is possible to imagine an evolution and also a profound change in the way society approaches health problems in humans. The greatest development has just begun and involves the search for diagnostic methods. In fact, a new generation of automated analytical systems is absolutely necessary in order to carry out a great number of tests in a simple, efficient and economic way.

Genotyping methods include a wide range of techniques, most of which are PCR-based. We will focus particularly on existing methods recently developed or adapted to automated processes.

13.1.1 Minisequencing

Minisequencing, also referred to as single nucleotide primer extension and genetic bit analysis determines the base immediately 3' to a primer by extending the primer by one base only. Although the original report described detection from genomic DNA without amplification, all subsequent reports have used PCR amplification to prepare primer extension templates. Base extension can be monitored by gel electrophoresis and commercial kits are available to run these assays on DNA sequencers, for example "SnapShot" from Applied Biosystems. As with most genotyping assays, if the variant is present as a minority species (for example in a tumour or a germinal mosaic) the reliability of the assay declines, although increased sensitivity of detection by pre-treatment of a mixed population containing H-ras codon 61 mutants has been reported using the MutEx assay. Minisequencing is a flexible method that can operate using fairly basic equipment, or be adapted to highly automated systems.

13.1.2 Pyrosequencing

Pyrosequencing is a non-electrophoretic real-time DNA sequencing method using a unique approach to read small runs of bases. The luciferase-luciferin light release is a detection signal for nucleotide incorporation into target DNA. This method can be adapted for automated high throughput operation, and has the advantage of typing bases that flank the SNP for confirmation that the correct target is being analysed. Pyroseqeuncing of the human p53 gene using a nested multiplex PCR method for amplification of exons 5-8 has been described, reporting accurate detection of p53 mutations and allele distribution. If the current length of sequence limitation could be overcome, pyrosequencing has considerable potential as a highly automatable sequencing tool.

13.1.3 Invader

Invader technology uses a Flap Endonuclease (FEN) for allele discrimination and a universal fluorescence resonance energy transfer (FRET) reporter system. A study by Mein et al. genotyped three hundred and eighty-four individuals across a panel of 36 SNPs and one insertion-deletion (indel) polymorphism with Invader assays using PCR product as template. The average failure rate of 2.3% was mainly associated with PCR failure, and the typing was 99.2% accurate when compared with genotypes generated with established techniques. Semi automated data interpretation allows the generation of approximately 25.000 genotypes per person per week. Using an "Invader squared" method, factor V Leiden

genotyping has been achieved on genomic DNA samples without prior amplification, although at present most assays in routine use rely on the PCR to generate templates for genotyping.

13.1.4 Arrays

The idea of using arrays for high throughput genotyping has been in existence for many years: early arrays were two dimensional spots of DNA targets on nylon or nitrocellulose membranes and the method of detection was allele specific hybridization. This method still has utility and recent improvements in the oligonucleotide binding capacity of membranes could extend this further. However DNA arrays typically refer to glass, plastic or silicon supports with either oligonucleotide or cloned DNA attached by adhesion or covalent linkage. Arrays mechanically deposited onto a glass microscope slide have feature sizes of around 200 microns and are scanned at between 5-20 micron resolution. Such arrays can carry 10-15.000 features. Affymetrix manufacture high-density arrays by a proprietary photochemical oligonucleotide synthesis method than can result in a small (10 micron) feature size, enabling a large number of 20-24 base oligonucleotide probes to be packed into a small area. Whilst these arrays have had the most success in gene expression studies, they have not yet produced the anticipated breakthrough in DNA sequencing or mutation scanning, even though their use has been reported in CFTR, mitochondrial and BRCA1 mutation detection. The reason for the limited use of the Affymetrix system for mutation detection to date lies in its limited sensitivity. Di-deoxy sequencing of the p53 gene in 100 primary human lung cancer by cycle sequencing was compared with sequence analysis by using the p53 GeneChip assay. The GeneChip assay detected 46 of 52 missense mutations (88%) but 0 of 5 frameshift mutations. In another study, a 92% sensitivity for the detection of p53 mutations in a series of 108 ovarian tumours was reported, less than might be expected from a current mutation scanning tool such as DHPLC. Several recent studies have indicated that the use of primer extension (solid-phase minisequencing, arrayed primer extension-APEX and single base extension-SBE) or ligation or hybridization on microelectronic microchips can improve the specificity of mutation detection on arrays.

Mutation scanning is the search for novel sequence variants within a defined DNA fragment. Numerous methods, exploiting different physical, chemical, and biological consequences of DNA sequence variation have been developed to facilitate mutation scanning. The ideal mutation scanning method has been characterized as one which would screen kilobase lengths of DNA with 100% sensitivity and specificity and would completely define the mutation. It would be a simple, single step, non-electrophoretic protocol with high throughput and low cost; requiring no complex equipment and no harmful reagents. Cost and data analysis time continue to be major barriers to meeting the demand for genetic testing and no current method satisfies all of these criteria.

Most scanning methods do not identify the precise nature of the change to the DNA sequence, although some indicate the location of the mutation within the fragment analysed. Consequently the majority of methods are used as a first round screen to identify those samples which contain mutations and these samples are subsequently sequenced to define the mutations.

Several factors will influence the choice of scanning method:

Mutation detection sensitivity: in the clinical diagnostic setting, sensitivity should be as close to 100% as is reasonably practicable. Mutation scanning for other purposes such as candidate gene analysis may be able to tolerate a trade-off between a reduction in sensitivity and an increase in throughput. In practice, it is unlikely that any single technique will detect 100% of mutations. An awareness of the limitations of the technique selected is essential. Factors that influence sensitivity include fragment resolution, reactivity of any enzyme or chemicals used and template features such as sequence (e.g. G+C content), length and secondary structure. Measurement of sensitivity is empirical: the literature is replete with examples of non-blinded studies or studies using small series, from which it is difficult to draw general conclusions about assay performance.

Mutation detection specificity: in a pre-screening method, low specificity (large number of false positives) may generate excessive downstream analysis and reduce the advantage of pre-screening. Some regions of interest may be highly polymorphic, and generate many samples that require further analysis. Whilst there have been claims that common polymorphisms generate "characteristic" mobility shifts, for example in DHPLC analysis, these claims should be treated with caution in a diagnostic setting.

Suitability for proposed sample type: Current diagnostic practice is largely restricted to genomic DNA samples extracted from peripheral blood lymphocytes. Future developments are likely to include increasing analysis of DNA extracted from tumour samples, which present a number of problems not encountered when studying germline DNA. In germline samples, mutations can be present at 0% (homozygous or hemizygous wild type), 50% (heterozygous) or 100% (homozygous or hemizygous mutant) of the total DNA, depending on zygosity, unless mosaicism is present. In tumour samples, the mutation can be present at any proportion of the total DNA because of factors which include loss of heterozygosity; contamination of the tumour with surrounding wild type material and variable proportions of mutant cells in the tumour. Some methods such as DHPLC are able to detect mutations which are present as a minor fraction in the sample better than others. Many methods are dependent on the generation of heteroduplex DNA for the detection of mutations: depending on whether the expected mutation are likely to be homozygous, hemizygous or heterozygous it may or may not be necessary to add 50% wild type DNA to the samples.

Suitability for predicted mutation type: Some of the methods described here have limitations on the types of mutations they can detect. For instance, denaturing HPLC (DHPLC) cannot reliably detect homozygous mutations; heteroduplex analysis (HA) detects insertions/deletions with higher efficiency than substitutions and the protein truncation test (PTT) detects only polypeptide chain terminating mutations.

Where the nature of mutation is unknown, a detection method which is unbiased toward any type of mutation should be used. For conditions/genes where a single type of mutation predominates, it may be more appropriate to select a method designed to detect only that type of mutation.

Features of the DNA sequence analysed: knowledge of the presence of common polymorphisms in the fragment to be analysed may also affect the choice of method. With the exception of the scanning methods which unambiguously identify the mutation present in most cases the available information will be only that a mutation is present or absent. Some methods, for instance DHPLC and FSSCP, may produce a mutation profile, which superficially at least, appears characteristic for the mutation; but there is evidence to suggest that this may be unreliable. It would usually, therefore, be necessary to sequence all samples showing a change from the wild type pattern. Thus, in the presence of a common polymorphism, a large proportion of samples may require analysis by both a scanning

method and DNA sequencing and in these cases it may be that DNA sequencing alone would be a more suitable choice.

Health and safety considerations: both legislation and good practice require that, as far as reasonably practicable, where alternative techniques are available, the safer option should be chosen. Non-radioactive detection methods are thus preferable to radioactive detection; methods which avoid the use of toxic chemicals are preferable to those methods which are dependent on the use of toxic chemicals.

Expected requirements for sample throughput: as the expected throughput increases, it becomes necessary to increase automation; decrease analysis time and complexity; decrease the number of manipulations and to increase the level of multiplexing.

Capital equipment costs and ongoing running costs: DHPLC, microarrays and any technique requiring fluorescent labelling and detection requires significant investment in equipment before the technique can be established in a laboratory.

Requirement for post-PCR manipulation: it is usually advantageous to minimise the number of post-PCR manipulations for several reasons. The more stages involved in an assay, the greater the likelihood for operator error. Complex techniques are usually low throughput and less amenable to automation. Additionally, a requirement for post-PCR reactions will results in an increase in the cost per genotype.

There are many different mutation scanning methods, most can be fitted into one of four categories: physical methods (which depend upon the presence of a mutation changing the physical properties of the DNA molecule), cleavage methods (which identify the presence of a mutation by the differential cleavage of wild type and mutant DNA) and methods which detect the consequences of mutation in a protein molecule or a functional assay. Finally direct sequencing can itself be used to detect mutations.

13.2 Physical methods

For physical methods, the practical consequence of sequence variation is a differential physical property of wild type versus mutant DNA, for example gel mobility or homoduplex stability. Although physical methods typically require little post-PCR manipulation and can be performed in a low technology format using routine laboratory equipment, throughput and sensitivity have been enhanced by the utilisation of fluorescent labelling and automated detection.

13.2.1 Single Strand Conformation Polymorphism (SSCP)

ISingle stranded DNA in non- denaturing solution folds in a sequence- specific manner. A change in the DNA sequence causes a change in the folded structure which in turn alters the mobility of the conformer on a non-denaturing gel. The sensitivity reported for SSCP range between 35 and 100%, although the majority of studies detected more than 80% of mutations. Multiple conditions of analysis can be used to increase the sensitivity. One major limitation for SSCP is fragment size: a study by Sheffield reported that sensitivity varied dramatically with fragment size and that the optimum size was as little as about 150 bp. 300 bp is generally regarded as the upper limit on fragment size. Utilization of fluorescence and capillary electrophoresis (CE) technology has

resulted in higher sensitivities in blinded trials and may allow high sensitivity detection in larger fragments.

Dideoxy-fingerprinting (ddF) is an interesting variant of the SSCP method in which chain terminated products are analysed by SSCP, resulting in increased sensitivity, but a rather complex image to analyse. Very high sensitivity has been reported using ddF on a high through CE system, but the workload must start to approach that of sequencing, reducing the advantage of simplicity that may represent the major asset of SSCP.

13.2.2 Heteroduplex analysis (HA) and conformation sensitive gel electrophoresis (CSGE)

On electrophoresis in a non-denaturing gel, heteroduplexes have retarded mobility compared to homoduplexes. The technique was first described for insertion /deletion mutations but can also be applied to single base mismatches. HA has been successfully applied to fragments of >1kb in size, although evidence suggests that mutation detection efficiency may be reduced in larger fragments. Like SSCP, HA is a very simple technique, requiring no DNA labelling or specialist equipment and the two techniques can be run together on a single gel.

Conformation sensitive gel electrophoresis (CSGE) is a variant of the HA method, employing mildly denaturing gel conditions. For fragments in the size range of 200-800 bp, sensitivity of 88% has been detected; reduction in the maximum size of fragment has been associated with an increase in the detection rate close to 100%. Mutations within 50bp of the end of a fragment are not detected, presumably because the distortion of the duplex is not great enough to generate a significant mobility shift. Recent developments in CSGE include the application of fluorescent labelling and detection and capillary electrophoresis.

13.2.3 Denaturing gradient gel electrophoresis (DGGE)

In DGGE, duplex DNA electrophoresis is through a gradient of increasing denaturant concentration. At a characteristic point in this gradient, the duplex will become partially denatured and electrophoretic mobility retarded as a consequence. Staking forces make DNA denaturation highly sensitive to nucleotide sequence: a single nucleotide substitution significantly alters the melting properties and hence the mobility in DGGE. Separation of different homoduplex molecules can be achieved by DGGE although separation of homo- and hetero-duplex DNA is far greater. A major constraint on DGGE is that mutations can only be detected in the lowest melting domain of the fragment because complete denaturation of the molecule retards the mobility to the extent that no separation of mutant and wild type molecules occurs. To ensure that the region of interest forms the lowest melting domain, a GC clamp of 20-45- bp is usually added to one end of the fragment to be analysed. The sensitivity of DGGE is in the range of 95-99%, for fragments of up to 500 bp.

In classical DGGE, separation is achieved by electrophoresis through a polyacrylamide gel containing a chemical denaturant gradient. Variations on the principle of DGGE include temperature gradient gel electrophoresis and constant denaturant gel electrophoresis. CDGE has been adapted to a fluorescent CE format.

The principal disadvantages of DGGE are a relatively low throughput; complex primer design to include GC clamps in the optimum position and maintain the fragment to be scanned as a single melting domain and a requirement for extensive optimisation

for each analysis. Nevertheless, its high sensitivity has made it a relatively popular technique within the diagnostic setting.

A temperature gradient capillary electrophoresis technique works on the same principle as DGGE has recently been described. No prior labelling of the sample is required, and the technique is fully automated for high throughput.

13.2.4 Denaturing high performance liquid chromatography (DHPLC)

DHPLC, also known as temperature modulated heteroduplex analysis (TMHA) exploits the differential melting properties of homo- and heteroduplex DNA in order to detect mutations in a manner which has some similarities to DGGE. Differential retention on a chromatography column under conditions of partial thermal denaturation is the physical principle underlying DHPLC. DHPLC has become very popular and is widely used for both research and diagnostic applications.

Many studies have examined the sensitivity and specificity of DHPLC and it is clear from these that DHPLC is a highly sensitive (91-100% detection) and specific technique although analysis at multiple temperatures may be required for maximum detection. The principle advantages of DHPLC are its high sensitivity and high throughput, coupled with minimal post-PCR manipulation and no requirement for sample labelling, although a modification to utilise fluorescent detection has been introduced. Disadvantages include the high capital equipment cost and the need to predict a precise temperature for analysis of each fragment, although theoretical prediction from DNA sequence is possible.

13.3 Cleavage methods

Cleavage methods are able to scan larger fragments than most of the physical techniques and to identify the location of the mutation in the fragment. For most of the cleavage techniques, a single assay condition is applicable to the analysis of all fragments, whereas many of the physical assays require specific optimisation for each different fragment analysed. Cleavage techniques were originally devised for radioactive labelling, polyacrylamide gel electrophoresis (PAGE) and autoradiography, and can still be used in this format although non-radioactive and/or fluorescent versions of most methods have been described. None of the cleavage methods currently find widespread use probably because of the considerable amount of post-PCR manipulation required to generate data.

13.3.1 Chemical cleavage of mismatch (CCM)

Mismatched C and T bases can be chemically modified by hydroxylamine and osmium tetroxide and the modified duplex cleaved at the site of the modification. The sample to be tested is mixed with a labelled wild type probe to generate heteroduplexes. For maximum detection, both possible heteroduplexes should be investigated, as modification is restricted to mismatched C and T residues. Cleavage products are separated by electrophoresis, with the size of the cleaved product giving the approximate location of the mutation. CCM has an extremely high mutation detection rate of essentially 100%, although failure to detect T:G mismatches in some sequence contexts has been reported. CCM is applicable to DNA fragments of 1 kb or longer. However, it suffered from the disadvantage of being highly laborious and requiring radioactive labelling and highly toxic chemicals for DNA modification although

more recent adaptations to the protocol have addressed many of these problems.

13.3.2 Enzyme cleavage of mismatch (ECM)

The resolvase T4 endonuclease VII introduces double strand breaks into duplex DNA at the site of single base mismatches and small loops. This activity is used for mutation in the enzyme cleavage of mismatch assay (EMC), also developed commercially as Enzyme Mismatch Detection (EMD). T7 endonuclease I has also been tested in EMC assays.

Although T4 endonuclease VII shows variable reactivity with different types of mismatch and loop and is also dependent on sequence context, the mutation detection rate of EMC is high; in the range of 91-100%. Like CCM, EMC performs well on fragments of over 1 kb. One drawback of EMC is non-specific background cleavage, which can complicate interpretation and may obscure genuine results.

More recently, the use of a plant endonuclease, CEL I, in a similar type of assay has been reported. Initial results were promising, and suggested that compared to T4 endonuclease VII, CEL I has more even activity with different mismatches and less non-specific activity. A high throughput mutation screening assay utilizing CEL I has recently been described. It seems that so far, the ideal mismatch-cleavage enzyme has not been identified, although the recently thermostable endonuclease V has been described that may have potential. Any enzymatic system will need to be competitive against increasing facile physico-chemical methods and direct sequencing itself.

13.3.3 Ribonuclease mismatch cleavage

Ribonuclease mismatch cleavage was the first of the mismatch cleavage techniques to be developed. It relies on the ability of RNase A and other RNases to cleave RNA:RNA and RNA:DNA duplexes at or near single base mismatch. Different mismatches are cleaved with differing efficiency with sequence context perhaps accounting for at least part of this variability; small insertions and deletions are also detected. Detection rates are typically in the range of 60-90% and RNase cleavage is able to analyse fragments of up to 1 kb or more. The major disadvantage of RNase cleavage is the requirement to synthesise RNA in vitro. The non-isotopic (NIRCA) format devised by Goldrick has the advantage of requiring no specialized equipment, and is available in commercial kit form.

13.3.4 Cleavage fragment length polymorphism (CFLP)

Cleavase I is a proprietary structure-specific endonuclease which cleaves single stranded DNA at sites of secondary structure to produce a characteristic pattern of bands for any fragment. Mutations in the DNA fragment result in a change to the band pattern. Reported mutation detection rates are 92-100% in fragments of up to 550 bp, with indications that fragments of up to 1 kb can be analysed.

13.3.5 MutS

The *E.Coli* MutS protein binds to mismatched DNA. This property has been exploited in both a gel shift assay and an exonuclease protection assay. The latter method reports the position of the mutation, although the sensitivity of the assay has not been established over a large range of samples. Solid phase immobilised

MutS has also been used to detect mutations by binding to nitrocellulose filters or magnetic capture.

13.4 Sequencing methods

There are two basic sequencing formats in current use: sequencing using dideoxynucleotide chain terminators and the less widely used chemical cleavage method. Alternative methods do exist, but sequencing by hybridisation has yet to deliver large scale sequencing, whilst pyrosequencing is making some progress, and resequencing by mass spectroscopy requires further improvements.

Assuming perfect data quality, the Sanger method provides absolute information about the position and nature of a sequence change. It is universally applied in mutation detection for defining mutations identified by scanning techniques and is generally regarded as the "gold standard" to which other techniques are compared. Sequencing is also widely used as a primary mutation screening technique which probably reflect the easy commercial availability of the technology together with the familiarity of the technique.

The requirements of the human genome project have pushed technological development so that sequencing is now a high throughput, high accuracy technique. Finished human genome sequence has accuracy of 99.99%. However, to achieve this, each base has been sequenced on average at least 8-10 times, a depth of coverage not generally used for mutation screening.

Few objective analyses of the mutation detection sensitivity of sequencing have been carried out, not least because of the inherent difficulty in determining the false negative rate. Several studies have shown that mutation detection rates can be substantially less than 100% and that factors including sequencing chemistry, the nature of the samples analysed, the depth of coverage and the method of data analysis undoubtedly influence the sensitivity.

For sequencing, as for any method, failure to detect a mutation can occur because the mutation does not generate a difference between wild type and mutant data or because the method of data analysis fails to detect a difference which is present. DNA sequencing generates a more significant burden for data analysis than most other scanning methods, as sequencing with both forward and reverse primers, which would be regarded as the minimum acceptable standard for diagnostic work, generates two pieces of data per base pair analysed whereas most other techniques generate one or a few pieces of data per fragment analysed. There are two ways of analysing DNA sequence data: either by visual inspection, which is the only method available for manual gels, and often also used for fluorescent electropherograms. The alternative, which is to use software such as PolyPhred or TraceDiff is only available for automated fluorescent sequencing and is still dependant on good quality raw data.

Comparative sequence analysis (CSA) and its close relative, orphan peak analysis is an alternative method of analysing the products, making a direct comparison of mutant and wild type sequencing data without the use of base calling software. Although sensitivity is high and mutations are defined as well as identified, the limitations which apply to sequencing also apply to CSA.

Sequencing of heterozygotes by matrix assisted laser desorption/ ionisation time-of-flight mass spectrometry (MALDI TOF MS) has been developed. This technique, which is fast, accurate and fully automated, has tremendous potential for mutation scanning, although current technical limitations on read length need to be improved.

The use of high density oligonucleotide microarrays for mutation scanning is an application of sequencing by Hybridisation (SBH), which in principle can screen kilobase lengths of DNA for novel mutations with near 100% sensitivity. The principle has been tested for the BRCA1, p53, ATM, and APC genes, amongst others. Sensitivity is in the range of 91-99% and is greater for homozygous than for heterozygous changes. Detection of insertion or deletion mutations, especially at repeated sequences remains problematic.

13.5 Protein methods

A fourth group of methods are those which detect sequence variation at the protein level, either as functional assays or by examining the protein product directly. As a group, these methods are characterised by being highly labour intensive, with low throughput. However, these disadvantages are offset by being able to screen large fragments of DNA in a single reaction and obtaining information about the biological consequences of the mutation.

13.5.1 The protein truncation test (PTT)

The PTT also known as the *in vitro* protein synthesis assay detects mutations which result in premature truncation of translation. Labelled protein synthesised *in vitro* is analysed by SDS-PAGE, with the presence of a truncating mutation indicated by a change in size of the protein compared to a wild type control. Sensitivity for truncating mutations is high with most false negative results due to mutations at the ends of the fragment. Fragment size for PTT analysis is typically in the range of 1-1.5kb for the majority of genes. PTT analysis requires cDNA or large exons as a starting material. The biggest advantage of PTT is that only mutations with a functional consequence, i.e. truncating mutations, are identified. A yeast *in vivo* assay for truncating mutations, with the ability to screen fragments of up to 3.5 kb has also been described.

13.5.2 Functional assays

A small number of assays which directly test protein function from a cloned DNA sequence have been described. Successful applications of functional assays have been described, however, applications for functional assays are limited, not least because of the paucity of information about the molecular function of many disease-associated proteins. A functional assay can only exist where the function of the protein is known; functional protein can be expressed *in vitro* or *in vivo* and a quantifiable assay designed. Many proteins have multiple functional domains: an assay which tests one function does not necessarily test all the functions of the protein. Furthermore functional assays only test nucleotide function at the protein level: nucleotide changes may also have effects on function at the RNA level.

13.6 Future developments

No current mutation scanning method is entirely satisfactory or meets even current diagnostic demands. Recent trends have been to

adapt existing methods to automated processes using automated data collection and robotic sample handling.

Microarray sequencing, which now exist in a variety of formats, is potentially a tremendously powerful technique, capable of far higher throughput than any other and may be the only technique than can match the demands for sequence variation data generated as a consequence of the completion of the human genome sequence. However whether the arrays will be read by mass spectroscopy, fluorescence or some other technique remains to be established. These techniques will have to compete with micro-fabricated alternatives to established electric field separation technologies. Improvements to the sensitivity of mutation detection, will inevitably push the burden of genetic diagnostic work into data analysis, and also sample preparation. The likely increase in numbers and types of mutation identified is a potentially valuable resource not only for the clinical insights concerning genotype-phenotype relationships, but also as part of the ongoing process to document human genome sequence variation.

References

1. Krook et al (1992) Rapid and simultaneous detection of multiple mutations by pooled and multiple single nucleotide primer extension. Hum Mol Genet, 1,391-395.
2. Head SR et al (1997) Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumour suppressor gene. NAR, 25, 5065-5071.
3. Sokolov BP (1990) Primer extension technique for the detection of single nucleotide in genomic DNA. NAR, 18, 3671.
4. Parsons BL et al (1998) Detection of basepair substitution mutation at a frequency of 1x10(-7) by conmbining two genotypic selection methods, MutEx enrichment and allele-apecific competitive blocker PCR. Environmental and molecular mutagenesis, 32, 200-211.
5. Ronaghi et al (2000) Direct analysis of single-nucleotide polymorphism on double stranded DNA by pyrosequencing. Biotechnol Appl Biochem, 31, 107-112.
6. Garcia CA et al (2000) Mutation detection by pyrosequencing : sequencing of exons 5-8 of the p53 tumour suppressor gene. Gene, 253, 249-257.
7. Mein CA et al (2000) Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. Genome Res, 10, 330-343.
8. Ryan D et al (1999) Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. Molec Diagn, 4, 135-144.
9. Lipshutz RJ et al (1999) High density synthetic oligonucleotide arrays. Nature Genet, 21 20-24.
10. Cronin MT et al (1996) Cystic-fibrosis mutation detection by hybridization to light-generated DNA-probe assay. Hum Mutat, 7, 244-255.
11. Chee M et al (1996) Accessing genetic information with high density DNA arrays. Science, 274, 610-614.
12. Ahrendt SA et al (1999) Rapid p53 sequence analysis in primary lung cancer using an oligo probe array. PNAS, 96, 7382-7387.
13. Wen WH et al (2000) Comparison of TP53 mutations identified by oligo microarray and conventional DNA sequence analysis. Cancer Res, 60, 2716-2722.
14. Metspalu et al (1997) Arrayed primer extension (APEX) for mutation detection using gene specific DNA chips. Am J Hum Genet, 61, 1301.
15. Favis R et al (2000) Universal DANN array detection of small insertion and deletions in BRCA1 and BRCA2. Nature Biotech, 18, 561-564.
16. Syvanen AC (2001) Accessing genetic variation: genotyping SNPs. Nat Rev Genet, 2, 930-942.
17. Syvanen AC et al (2002) A primer guided nucleotide incorporation assay in the genotyping of ApoE. Genomics, 8, 684-692.
18. Pastinen T et al (2000) A system for specific, high troughpiut genotyping by allele-specific primer extension on microarrays. Genome Res, 10, 1031-1042.
19. Ferrari M. et al (2003) Molecular diagnostics by microelectronic microchips. Clin Chem Lab Med, 41, 462-467.
20. Aitman TJ (2001) DNA microarrays in medical practice. BMJ, 323, 611-615.
21. Cotton, RGH (1997) Slowly but surely towards better scanning for mutations. Trends in Genetics, 13, 43-46.
22. Jones, AC et al (2000) Application and evaluation of denaturing HPLC for molecular genetic analysis in tuberous sclerosis. Hum Genet, 106, 663-668.
23. Arnold N et al (1999) A highly sensitive, fast, and economical technique for mutation analysis in hereditary breast and ovarian cancers. Hum Mutat, 14, 333-339.
24. Dobson-Stone C et al (2000) Comparison of fluorescent SSCP analysis and denaturing HPLC for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses. Eur J Hum Genet, 8, 24-32.
25. Mir KU, Southern EM (1999) Determining the influence of structure on hybridization using oligonucleotide arrays. Nature Biotech, 17, 788-792.
26. Vidal-Puig A, Moller DE (1994) Comparative sensitivity of alternative SSCP methods. Biotechniques, 17, 490-2, 494, 496.
27. Larsen L et al, (1999) High-throughput SSCP analysis by automated capillary electrophoresis: Robust multiplex analysis and pattern based identification of allelic variants. Hum Mutat 13, 318-327.
28. Liu Q et al (1996) Bi-directional dideoxy fingerprinting (Bi-ddF): a rapid method for quantitative detection of mutations in genomic regions of 300-600bp. Hum Mol. Genet, 5, 107-114.
29. Edwards SM (2001) An improved high throughput heteroduplex mutation detection system for screening BRCA2 mutations-fluorescent mutation detection (F-MD). Hum Mutat, 17, 220-232.
30. Rozycka M et al (2000) Rapid detection of DNA sequence variants by conformation-sensitive capillary electrophoresis. Genomics, 70, 34-40.
31. Gejman PV et al (1998) The sensitivity of denaturing gradient gel electrophoresis: a blinded analysis. Mut Res, 382, 109-114.
32. Gao et al (2000) High throughput detection of unknown mutations by using multiplexed capillary electrophoresis with poly(vinylpyrrolidone) solution. Analytical Chem, 72, 2499-2506.
33. O'Donovan MC et al (1998) Blind analysis of DHPLC as a tool for mutation detection. Genomics, 52, 44-49.
34. Liu WG et al (1998) DHPLC used in the detection of germline and somatic mutations. NAR, 26, 1396-1400.
35. Hecker KH et al (1999) Mutation detection by denaturing DNA chromatography using fluorescently labelled PCR products. Analytical Biochemistry, 272, 156-164.
36. Hansen LL (1996) Sensitive and fast mutation detection by solid phase chemical cleavage. Human Mutat, 7, 256-263.
37. Mashal RD et al (1995) Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. Nature Genet, 9, 177-183.
38. Kulinski J et al (2000) CEL I enzymatic mutation detection assay. Biotechniques, 29, 44-6, 48.

39. Goldrick MM et al (1996) NIRCA ™ - A rapid robust method for screening for unknown point mutations. Biotechniques, 21, 106-112.

40. Wagner R et al (1995) Mutation detection using immobilized mismatch binding-protein (MutS). NAR, 23, 3944-3948.

41. Lander ES et al (2001) Initial sequencing analysis of the human genome. Nature, 409, 860-921.

42. Nickerson DA et al (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based reseqeuncing. NAR, 25, 2745-2751.

43. Bonfield JK Automated detection of point mutations using fluorescent sequence trace subtraction. NAR, 26, 3404-3409.

44. Mattocks C et al (2000) CSA: a new sequence-based method for the identification and characterization of mutations in DNA. Hum Mutat, 16, 437-443.

45. Fu DJ et al (1998) Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. Nat Biotechnol, 16, 381-384.

46. Wilkman FP et al (2000) Evaluation of the performance of a p53 sequencing microarray chip using 140 previously sequenced bladder tumour samples. Clin Chem, 46, 1555-1561.

47. Hacia JG et al (1996) Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nature Genet, 14, 441-447.

48. Den Dunnen JT, van Ommen GJ (1999) The PTT: a review. Hum Mutat, 14, 95-102.

49. Ishioka C et al (1997) Detection of heterozygous truncating mutations in the BRCA1 and APC genes by using a rapid screening assay in yeast. PNAS, 94, 2449-2453.

50. Flaman JM et al (1995) A simple p53 functional assay for screening cell lines, blood, and tumours. PNAS, 92, 3963-3967.

51. Schwartz H et al (1998) Mutation detection by a two-hybrid assay. Hum Mol Genet, 7, 1029-1032.

52. Duddy et al (2000) Improving the detection of p53 mutations in breast cancer by use of the FASAY, a functional assay. J Mol Diagn, 2, 139-144.

53. Bougeard G et al (2001) Detection of 11 germline inactivating TP53 mutations and absence of TP63 and HCHK2 mutations in 17 French families with Li-Fraumeni or Li-Fraumeni-like syndrome. J Med Genet, 38, 253-257.

54. Davies RC et al (2000) Did nucleotides or aminoacids drive evolutionary conservation of the WT1+/-KTS alternative splice? Hum Mol Genet, 9, 1177-1183.

55. Wue G et al (2001) On-line nanoliter sequencing reaction with capillary zone electrophoresis purification for DANN sequencing. J Cromatogr, 914, 245-256.