Use of DNA chromatography for sequence analysis

SP Yip,[1] KH Leung,[1] PM Ngan,[1] WC Tang,[1,2] MKH Yap[2]

[1]Department of Health Technology and Informatics, and
[2]School of Optometry, The Hong Kong Polytechnic University

Correspondence to: Professor S.P. Yip
E-mail: shea.ping.yip@polyu.edu.hk

Abstract

Analysis of DNA sequence variations, either polymorphisms or disease-causing mutations, is crucial to biomedical science research and to molecular testing for the diagnosis of diseases. DNA chromatography differentially separates DNA molecules, double- or single-stranded, on the basis of their interaction through amphiphilic ions with the non-polar stationary phase and the polar mobile phase. Under non-denaturing conditions, DNA chromatography separates DNA molecules according to their size. Under partially denaturing conditions, DNA chromatography can differentially separate homoduplexes and heteroduplexes carrying mismatches under suitable column temperature, which varies according to the sequence of the DNA fragments being analysed. As such, it has now become a very powerful tool for screening and identifying unknown sequence variations. Under completely denaturing conditions, DNA chromatography separates short single-stranded DNA fragments according to both the length and base composition of the fragments. Thus, coupled with primer extension reaction, it allows rapid and reliable genotyping of known sequence variations. Multiplexing of primer extension reactions can increase the throughput and is particularly useful in molecular diagnostics. On the other hand, DNA chromatography can also be used in conjunction with DNA pooling. Under partially denaturing conditions, DNA pooling can speed up the process of screening sequence variations. Under completely denaturing conditions, DNA pooling allows accurate estimation of relative allele frequencies of single nucleotide polymorphisms analysed by primer extension coupled with DHPLC. This approach is extremely useful in mapping genes involved in complex diseases using genetic association studies.

Key words

DNA chromatography, denaturing high performance liquid chromatography, polymorphisms, mutations.

This article serves to give a brief account of the principle underlying DNA chromatography and the use of this technology for DNA analysis in both research and diagnostic settings. References have been purposely kept to a minimum in the main text. Interested readers can refer to two recent excellent reviews, one by Xiao and Oefner[1] and another by Gjerde and co-workers[2].

An overview of DNA sequence variations

The Human Genome Project was completed a few years ago. The human genome is estimated to carry about 30,000 genes which, however, occupy less than 5% of the genome. It also becomes very clear that there are a very large number of sequence variations in our genome[3]. On average, there is a difference of 0.1% in the DNA sequence between any two persons randomly chosen from a population. In other words, there are about 3 million sequence differences between any two unrelated persons.

From a clinical viewpoint, there are two broad types of DNA sequence variations: polymorphisms and disease-causing mutations. Polymorphisms refer to those sequence variations that are found in normal individuals and do not result in diseased phenotypes. Examples of polymorphisms include single nucleotide polymorphisms (SNPs), microsatellites and minisatellites. A SNP (pronounced as *snip*) is a sequence variation due to change in a single nucleotide (the building block of DNA). Microsatellites and minisatellites are due to variation in the number of repeat units that are themselves a short stretch of DNA sequence. They are very useful in research for locating the position of genes in our chromosomes, a process known as gene mapping, and in individual identification in forensics. On the other hand, disease-causing mutations are those sequence variations that result in diseased phenotypes because they adversely affect the functions of the proteins of interest in either a quantitative or a qualitative manner. For example, a single base mutation in the human globin gene can give rise to sickle cell disease. In clinical medicine, identification of mutations is important for the diagnosis of genetic diseases.

However, the distinction between polymorphisms and disease-causing mutations becomes blurred when one understands that they just represent the two extreme situations in a spectrum of sequence variations with varying degree of effects on the phenotypes of the individuals concerned. There are sequence variations that produce subtle functional changes in the proteins encoded or give rise to disease phenotypes only under certain environmental conditions.

From a technical viewpoint, DNA sequence variations can be classified into two groups: unknown and known variations[4]. Unknown sequence variations refer to those that are not known to exist previously and have to be detected by a group of methods called screening or scanning methods. Known sequence variations are known to exist and their genotypes can be determined by methods known as diagnostic methods. Many methods are available for detecting sequence variations, and vary from each other in terms of the principle of the method, cost, ease of optimization and use, requirement of special instruments and turnaround time. Examples include DNA sequencing, single strand conformation analysis, DNA chromatography and many others. DNA chromatography can be used to screen unknown sequence variations, and genotype known sequence variations.

DNA chromatography

The hardware for DNA chromatography

DNA chromatography is the use of modern high performance liquid chromatography (HPLC) for separating DNA molecules with different length and/or base composition. The hardware of DNA chromatography consists of components similar to those in conventional HPLC (Figure 1). Together with a gradient system, a pressure pump delivers buffers or solvents from buffer reservoirs in appropriate proportions to a column for the separation of DNA molecules. The buffer is the mobile phase while the column is the stationary phase. DNA samples are placed in an autosampler plate and injected into the column through an injection unit. The column is housed in a temperature-controlled oven. Under appropriate conditions, DNA molecules are separated in the column into individual components and then eluted from the column. The eluted components are monitored by a detector and the data collected in a computer system. An ultraviolet detector is the most common option for measuring DNA molecules at a wavelength of 260 nm although a fluorescence detector can also be installed. The results are displayed as chromatograms or elution profiles showing the amounts and elution times for various components separated by the column. An optional fragment collector can be connected to the detector to collect the separated components into vials for further analysis and processing. The heart of the system lies in the column-the stationary phase. The most widely used column is the *DNASep* marketed by the company Transgenomic. DNA separating columns from other manufacturers use a different type of stationary phase, and are less popular when compared with DNASep.

The principle of chromatographic separation of DNA molecules

DNA chromatography is based on a reversed phase system in which the stationary phase is non-polar and the mobile phase polar. For the DNASep column, the hydrophobic stationary phase is made up of alkylated non-porous poly(styrene-divinylbenzene) particles 2-3 microns in diameter. The polar mobile phase is acetonitrile (CH_3-CN). However, DNA molecules are large anions because of the negative charges on the phosphate groups in the phosphate-sugar backbones of the DNA strands. Organic cations are required to allow interaction between DNA anions and the non-polar stationary phase. The organic cation carries a positively charged portion to interact with the negative charge of DNA molecules on the one hand, and also a hydrophobic portion to interact with the non-polar stationary phase on the other hand. The most commonly used organic cation is triethylammonium, $(CH_3CH_2)_3N^+$, in the form of triethylammonium acetate (TEAA). Thus, TEAA is used as an ion pairing reagent. The triethylammonium cations bind to the phosphate groups of DNA molecules and hence effectively coat the DNA molecules with a hydrophobic layer (the triethyl portion). The

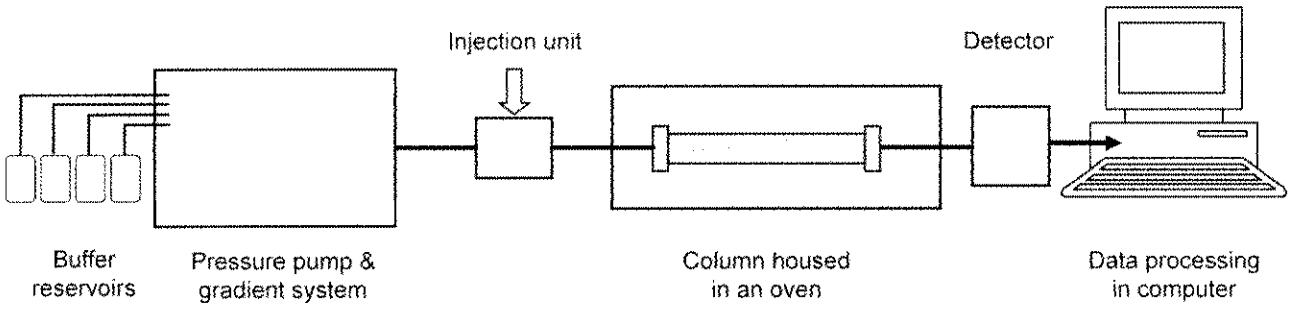

Figure 1. Hardware components of DNA chromatography.

number of TEAA molecules coating the DNA molecules is proportional to the length of the DNA molecules and in turn determines the degree of interaction between the DNA molecules and the stationary phase. DNA molecules are eluted from the column in an increasing gradient of acetonitrile which weakens the interaction between coated DNA molecules and the stationary phase. In other words, coated DNA molecules bind onto the stationary phase and will be released from the stationary phase when acetonitrile in the mobile phase reaches a specific concentration. Thus, shorter DNA molecules are eluted earlier from the column than and hence separated from longer DNA molecules under the same buffer condition. In summary, the separation of DNA molecules is based on the principle of ion-pair reversed phase liquid chromatography.

Three modes of operation are available for DNA chromatography, depending on the temperature of the column. They are non-denaturing, partially denaturing and completely denaturing modes. Each mode of operation serves a different purpose in DNA analysis.

Non-denaturing HPLC

The mode of operation is non-denaturing when the column temperature is maintained at 50°C and DNA molecules remain double-stranded (ds). Under this condition, the separation of dsDNA molecules depends on the length of the molecules, but not the base composition. Thus, non-denaturing HPLC can replace conventional gel electrophoresis for accurate sizing of fragments amplified by polymerase chain reaction (PCR) (Figure 2). One recent application is the high-throughput typing of *Mycobacterium tuberculosis* strains based on twelve loci of variable number of tandem repeat present in mycobacteria[5]. Typing results based on non-denaturing HPLC showed 100% concordance with those generated by agarose gel electrophoresis. However, such applications do not fully utilize the benefits of non-denaturing HPLC because the amounts of PCR products are not measured.

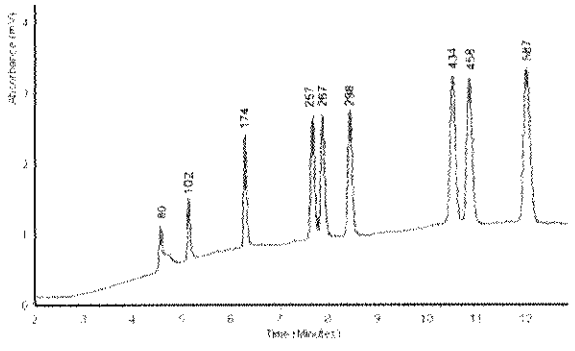

Figure 2. Use of non-denaturing HPLC for separating double-stranded DNA molecules. The chromatogram shows the elution profile of DNA molecules of known size (size standards) with the size in basepairs indicated above the peaks.

Non-denaturing HPLC can be used to size and quantify PCR products. To allow for quantification of PCR products, the number of PCR cycles has to be around 25 so that the amount of product amplified is proportional to the initial gene copy number. This type of analysis is very useful for detecting gene rearrangements such as deletions and insertions. Unlabelled products can be used for analysis even with an ultraviolet detector if the injection volumes are large. Use of unlabelled products reduces the cost of analysis. With small injection volumes, reliable quantification can also be achieved by adding dsDNA intercalation dye such as SYBR Green I and measuring the green fluorescence with a fluorescence detector. The dye is mixed with the DNA samples after elution from the column (post-column addition). The throughput of such analyses can further be increased by multiplex PCR in which several different fragments are amplified in the same tube. Applications are exemplified by the detection of large deletions of the X-linked dystrophin gene in Duchenne muscular dystrophy[6] and the demonstration of exon deletions and duplications in the *RB1* tumour suppressor gene[7]. With a wildtype or normal chromatogram for comparison, homozygous deletion is indicated by a complete absence of a particular elution peak, and heterozygous deletion (or a carrier) by a peak of half height (Figure 3).

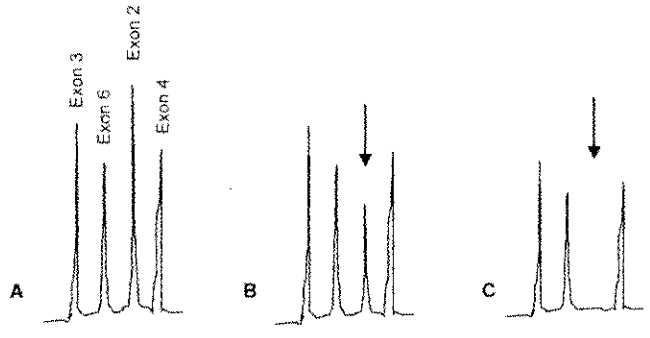

Figure 3. Use of non-denaturing HPLC to detect exon deletions. The chromatograms show the elution profiles of multiplex PCR products amplified from a hypothetical X-linked gene. Panel A shows the profile for a normal female, panel B for a carrier female with a heterozygous deletion of exon 2, and panel C for an affected hemizygous male with deletion of exon 2.

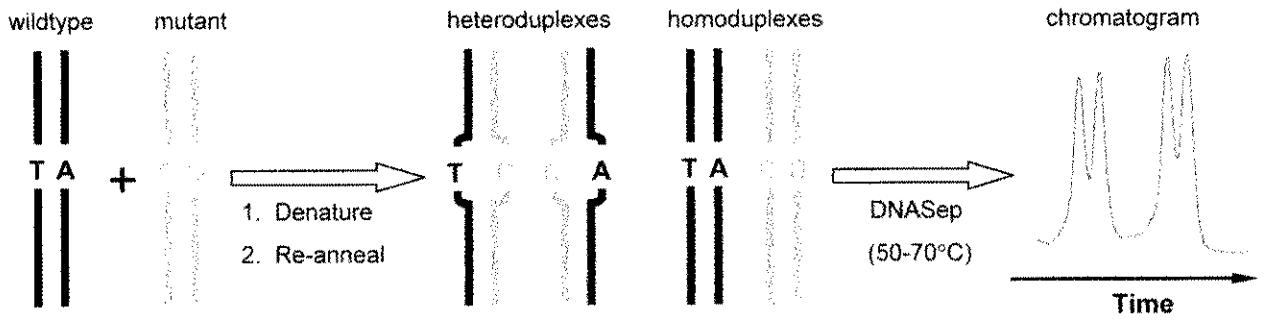

Figure 4. Ue of partially denaturing HPLC to separate homoduplexes and heteroduplexes. This mode of operation allows detection of sequence variation.

Partially denaturing HPLC

The partially denaturing mode operates at an optimized column temperature maintained above 50°C and below 70°C. When two dsDNA molecules differing by a single basepair (e.g. T-A changed into C-G) are mixed, denatured and allowed to re-anneal, two homoduplexes and two heteroduplexes are formed (Figure 4). Stability of the dsDNA duplexes determines the order of elution from the column: the more stable the duplexes, the longer the elution time. Heteroduplexes with mismatches are less stable than and are thus eluted before homoduplexes. In other words, partially denaturing HPLC allows separation of homoduplexes and heteroduplexes produced as a result of even a single base difference between two otherwise identical dsDNA molecules. With reference to a homozygous wildtype control, any difference in elution profile is indicative of the presence of a sequence variation. In fact, a four-peak pattern is not frequently seen.

Partially denaturing HPLC can be used as a screening method to detect unknown sequence variations. Heteroduplexes can be generated before analysis by mixing the PCR product from a test sample with a homozygous reference PCR product in equal volumes. This step is essential for the detection of X-linked mutations in males and homozygous mutations. The screening throughput can be increased by mixing several test samples with one reference sample. The ideal size of PCR products is 150-450 bp for detection of unknown sequence variations although mutations have been detected in fragments as large as 1500 bp. Long DNA fragments tend to have more than one melting domain and hence require several column temperatures for complete screening of the fragment.

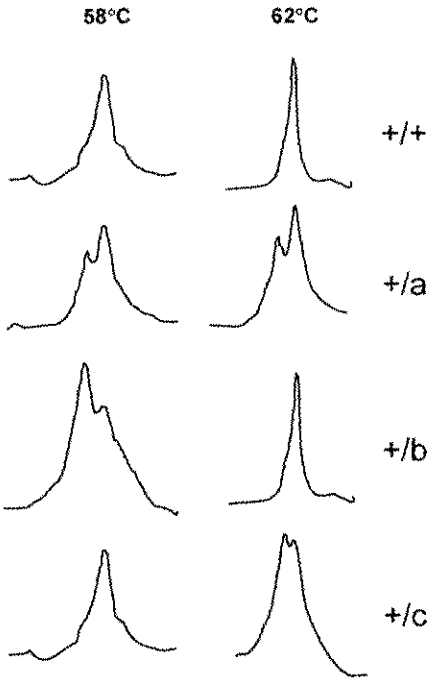

Figure 5. The effect of column temperature on the detection of sequence variation. The diagram shows the elution profiles at two column temperatures (58°C and 62°C) for three different hypothetical heterozygous mutations (+/a, +/b and +/c) in comparison with the homozygous normal (+/+). The mutations are found within the same DNA fragment in different individuals. Mutation a is detected at both temperatures, mutation b at 58°C only, and mutation c at 62°C only.

The column temperature is the most important parameter for the detection of sequence variations. Some sequence variations can be detected only at a particular temperature while others can be detected at several temperatures[8] (Figure 5). The software package accompanying the WAVE DNA Fragment Analysis system (Transgenomic) recommends only one column temperature for the analysis of a particular PCR fragment while the freeware MELT (http://insertion.stanford.edu/melt.html) may recommend more than one column temperature for a PCR fragment. Experience has shown that optimized analysis temperature may differ from the one recommended by a software package[9].

Partially denaturing HPLC is very sensitive in detecting heteroduplexes even present in small amounts only. Being not error-proof, DNA polymerase can mis-incorporate nucleotides during PCR and hence generate in the test mixture a small proportion of heteroduplexes which may produce a shoulder peak or background heteroduplex peaks. Such false positive results have to be excluded by follow-up study with DNA sequencing, and are thus wastage of time and money. Use of DNA polymerase with proofreading activity or a mixture of such enzyme and *Taq* DNA polymerase can help reduce such errors.

There are occasions in which several sequence variations are found within a small DNA region in different chromosomes (i.e., in different individuals). These are well illustrated by the diverse mutations in the *HBB* and *CFTR* genes[8,10]. Mutations in *HBB* give rise to β-thalassaemia or sickle cell disease while mutations in *CFTR* result in cystic fibrosis. Distinct mutations in a PCR product usually give consistent distinct chromatograms. Therefore, partially denaturing HPLC can also be used to genotype known sequence variations, particularly known mutations, once their corresponding distinct chromatograms are established. However, it is still possible that different mutations may share indistinguishable chromatograms.

Partially denaturing HPLC is the most widely used mode of operation for DNA chromatography. To date, over 300 genes have been analysed by partially denaturing HPLC (http://insertion.stanford.edu/dhplc_genes1.html). For screening mutations, its sensitivity and specificity approach 100% in many published studies and compare very favourably with direct DNA sequencing, which is widely regarded as the gold standard for comparison. Partially denaturing HPLC can also detect mosaic mutations that account for only a small proportion of alleles in a PCR product and that may remain undetected even by DNA sequencing. Mosaic mutations are found in tumour samples where tumour cells with mutations are surrounded by normal cells without mutations, and in minimal residual disease where malignant cells are not completely eradicated.

Completely denaturing HPLC

The mode of operation is completely denaturing if the column temperature is maintained between 70°C and 80°C. Under such a high temperature, DNA molecules are completely denatured and become single-stranded. Completely denaturing HPLC can differentiate single-stranded DNA (and RNA) molecules with the separation depending on both the length and the base composition of the single-stranded nucleic acid molecules. It can be used to analyse and isolate synthetic oligonucleotides and RNA molecules. More commonly, it is used to analyse the products from primer extension reactions (also known as minisequencing).

In primer extension, a modified DNA polymerase adds one dideoxyribonucleotide (ddNTP) in a template-dependent manner to a primer annealing immediately upstream of a sequence variation (Figure 6). The primer extension products are then analysed by

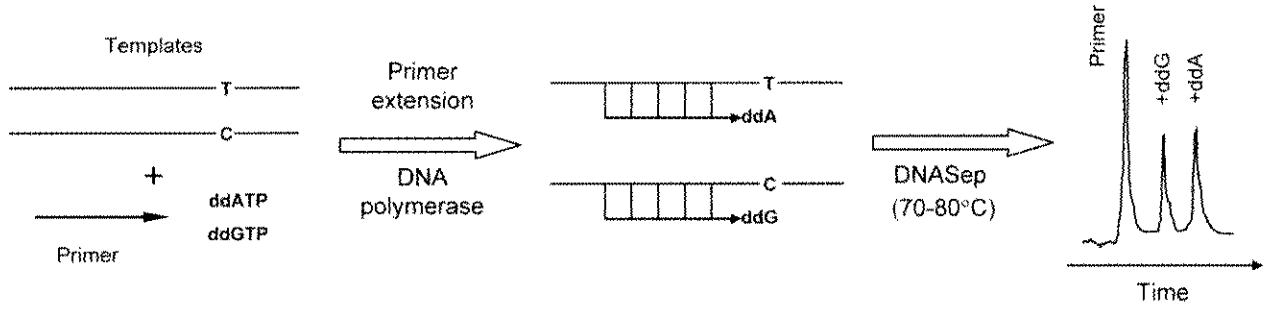

Figure 6. Use of completely denaturing HPLC for analysing primer extension products. Under this conditions, genotypes can be determined for known sequence variations for individual DNA samples or DNA pools.

completely denaturing HPLC. As such, completely denaturing HPLC provides a robust platform of medium throughput for genotyping known mutations or SNPs. The throughput can be increased by multiplexing several primer extensions in a single reaction[11].

The primer extension products can be quantified by their absorbance at 260 nm. If the starting test sample is a mixture prepared by pooling many DNA samples in equal amounts (a process known as DNA pooling), the relative allele frequencies of a SNP in the DNA pool can be estimated by measuring the relative signal intensities of the two extension products of the DNA pool with reference to those of a heterozygote sample[12]. Note that a SNP only has two alleles and their relative allele frequencies sum up to one. This provides a very cost-effective method for estimating the relative allele frequencies of a large number of different samples. Conventionally, estimation of allele frequencies is achieved by genotyping all samples individually, and this approach is of course very expensive and time-consuming if the number of samples is very large (in the range of several hundreds, or more preferably over 1000).

In other words, completely denaturing HPLC coupled with primer extension provides a convenient method for estimating relative allele frequencies in DNA pools, and is very useful in mapping genes involved in complex diseases. Many human diseases are complex in nature in that they are caused by genetic factors, environmental factors such as lifestyle and diet, and possibly the interaction between genetic and environmental factors. Examples of complex disease include diabetes, hypertension, infectious diseases and many others. Many genes are expected to be involved in a complex disease and the effect of each gene on the disease is usually small. Genetic association studies are very powerful in identifying genes of small effects in complex diseases[13]. One approach of genetic association studies is called case-control study, in which the allele frequencies of a SNP in a candidate gene are compared between a group of patients with the same disease under study (the "cases") and a group of control individuals without the disease (the "controls"). Instead of genotyping all samples one by one, estimation of relative allele frequencies in DNA pools is a very attractive alternative[14]. Two DNA pools are usually constructed: one *case pool* prepared from all patients' samples and one *control pool* from all control samples in equal amounts. Once estimated, the allele frequencies of DNA pools can be compared by statistical procedure. If initial comparison of DNA pools shows statistically significant difference in allele frequencies, confirmatory study is carried out by genotyping individual samples. If initial comparison does not show any significant difference, then the SNP will not be further investigated. Therefore, this approach allows more time and effort being spent on sequence variations that are worthy of further investigation, and unpromising sequence variations are abandoned after initial testing. In brief, completely denaturing HPLC plays an important role in mapping genes involved in complex diseases.

Conclusions

The heart of DNA chromatography lies in the column that is capable of separating DNA molecules under different conditions (non-denaturing, partially denaturing or completely denaturing). Under appropriate conditions, DNA chromatography can be used to size and quantify PCR products, detect unknown (and known) sequence variations, and genotype individual samples or DNA pools by analysis of primer extension products. Such applications can be used for both research and diagnostic purposes. In addition, HPLC has already established itself as one of the standard technologies used in clinical laboratories, It is envisioned that DNA chromatography will easily find its way into clinical laboratories and establishes itself as a standard platform in molecular diagnostics.

Acknowledgements

The WAVE DNA Fragment Analysis System was purchased with a PolyU Big Equipment Grant (G.53.27.9027). Research work using DNA chromatography as one of the analytic methods was funded by PolyU's Internal Competitive Research Funds and funds from School of Nursing, Department of Health Technology and Informatics, and the Centre for Myopia Research under the School of Optometry, The Hong Kong Polytechnic University.

References

1. Xiao WH, Oefner PJ. Denaturing high-performance liquid chromatography: A review. Hum Mutat 2001; 17: 439-474.

2. Gjerde DT, Hanna CP, Hornby D. DNA chromatography. Weinheim: Wiley-VCH, 2002.

3. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 49: 928-933.

4. Cotton RGH. Mutation detection. Oxford: OUP, 1997.

5. Evans JT, Hawkey PM, Smith EG et al. Automated high-throughput mycobacterial interspersed repetitive unit typing of Mycobacterium tuberculosis strains by a combination of PCR and nondenaturing high-performance liquid chromatography. J Clin Microbiol 2004; 42: 4175-4180.

6. Hung CC, Su YN, Lin CY et al. Denaturing HPLC coupled with multiplex PCR for rapid detection of large deletions in Duchenne muscular dystrophy carriers. Clin Chem 2005; 51:1252-1256.

7. Dehainault C, Lauge A, Caux-Moncoutier V et al. Multiplex PCR/liquid chromatography assay for detection of gene rearrangements: application to RB1 gene. Nucleic Acids Res 2004; 32:e139

8. Ravnik-Glavac M, Atkinson A, Glavac D, Dean M. DHPLC screening of cystic fibrosis gene mutations. Hum Mutat 2002; 19: 374-383.

9. Han W, Yip SP, Wang J, Yap MKH. Using denaturing HPLC for SNP discovery and genotyping, and establishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. J Hum Genet 2004; 49: 16-23.

10. Colosimo A, Guida V, De Luca A et al. Reliability of DHPLC in mutational screening of beta-globin (HBB) alleles. Hum Mutat 2002; 19:287-295

11. Yip SP, Pun SF, Leung KH, Lee SY. Rapid simultaneous genotyping of five common Southeast Asian β-thalassemia mutations by multiplex minisequencing and denaturing HPLC. Clin Chem 2003; 49: 1656-1659.

12. Hoogendoorn B, Norton N, Kirov G et al. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. Hum Genet 2000; 107: 488-493.

13. Lewis CM. Genetic association studies: design, analysis and interpretation. Brief Bioinform 2002; 3: 146-153.

14. Norton N, Williams NM, O'Donovan MC, Owen MJ. DNA pooling as a tool for large-scale association studies in complex traits. Ann Med 2004; 36: 146-52.