High throughput genotyping technologies

Andrew M. Dearlove

Abstract
A comprehensive genetic map containing several hundred microsatellite markers resulted from a large microsatellite mapping project. This was the first real study that introduced high throughput methods to the genetic community. This map and the concurrent technological advances, which will briefly be reviewed, led to further numerous mapping investigations of simple and complex diseases.

The annotated draft sequence of approximately three billion base pairs (bp) of the human genome has been completed much sooner than many imagined, due to considerable technological advancements and the international enterprise that resulted. This was a major development for the genetics community, but is only the precursor to the next phase of studying and understanding the variation within the human genome. The awareness of the differences may help us understand the effects on the genetics of the variation between individuals and disease. It is these variations at the nucleotide level that determine the physiological differences, or phenotypes of each individual, including all biological functions at the cellular and body level.

Single nucleotide polymorphisms (SNPs) will provide the next high density map, and be the genetic tool to study these genetic variations. There are many sources of SNPs and exhaustive numbers of methods of SNP detection to be considered. The focus in this paper will be on the merits of selected, varied SNP typing methodologies that are emerging to genotype many individuals with the required huge number of SNPs to make the study of complex diseases and pharmacogenomics a practical and economically viable option.

INTRODUCTION
The first simple map using just restriction fragment length polymorphisms (RFLPs)\textsuperscript{1} was followed by a second generation comprehensive microsatellite genetic map\textsuperscript{2,3} containing several thousand microsatellites covering the human genome. Microsatellites, also known as short tandem repeat polymorphisms (STRPs), are useful genetic markers\textsuperscript{4,5} as they are polymorphic regions of tandem repeats of a very simple DNA sequence. The map resulted from the immense undertaking at Genethon using the best available technology at the time.

Genetic mapping studies, completed using radiolabelling techniques, were challenging and time consuming. The introduction of new fluorescent technologies utilising automated sequencers and hundreds of evenly spaced markers from these maps made major genome screen studies for complex diseases, in which many loci contribute to the risk of developing disease, more feasible.\textsuperscript{6} The projects still took a relatively large amount of time to complete, requiring major funding in specialist laboratories. These studies are now more feasible for every scientist with a good DNA collection on their disease of interest, due to the advances in the labelling and detection technologies, microsatellite design, thermal cycling enzymes and data analysis. Indeed, these types of investigations with significant results are now commonplace for all types of single gene disorders,\textsuperscript{7} and there have been publications on studies screening for linkage for most common complex diseases,\textsuperscript{8-10} with some significant results being replicated.\textsuperscript{11,12} These technologies...
Microsatellites have been the primary tool for genetic mapping during the last ten years. They are relatively high throughput, although not the most efficient for the large numbers of extensive populations that need to be typed for the study of complex diseases. Microsatellites will most likely still have a place in many studies, perhaps starting with a 10 centiMorgan (cM) genome screen to suggest areas for further, higher density investigations.

A single nucleotide polymorphism (SNP) is a single base biallelic substitution sequence polymorphism that occurs in more than one per cent of the population. They are mutationally more stable than microsatellites, and more common, probably occurring about every 500–1,000 base pairs (bp) in the human genome. So, functional genes, several mega bp in length, may contain many SNPs. Indeed, SNPs are significant contributors to genetic variation, directly influencing protein structure or expression levels, and anonymous SNPs will be near to disease gene loci. They comprise 80 per cent of all polymorphisms in the human genome, so although they are only biallelic, and therefore less heterozygous, giving less information than microsatellites, they have many genetic applications. They provide conserved hereditary markers near or in the locus of interest and are extremely useful genetic mapping markers. They are especially used in association analysis (often referred to as linkage disequilibrium), with enough statistical power to identify subtle genetic risk factors. Association analysis uses controls to test the relationship between alleles and traits across families, not just patterns of linkage inherited within families. Indeed, the right statistical approach for the analysis of data is critical to success. Automation of the assays for typing interpretation, including automatic scoring, is often more feasible to set up due to the biallelic nature of the data produced, so there is a potential for huge volumes of data to be produced as soon as it is logistically plausible and economically not prohibitive.

Following the enormous amount of work that led to the publication of the sequence of the human genome, the SNP Consortium (TSC) was set up in April 1999 as the second phase of the human genome mapping project. The plan was to create a dense map of up to 300,000 SNPs distributed evenly throughout the human genome by the end of 2001. It was funded with cooperation by the Wellcome Trust, APTiotech, AstraZeneca PLC, Aventis, Bayer AG, Bristol–Meyers Squibb Company, F. Hoffmann-LaRoche, Glaxo Wellcome PLC, IBM, Motorola, Novartis, Pfizer Inc, Searle and SmithKline Beecham PLC. The work supported by the consortium is performed at major centres for molecular genetics: The Whitehead Institute for Biomedical Research, Washington University School of Medicine (St. Louis), The Wellcome Trust’s Sanger Centre and Stanford Human Genome Center. The current release consists of well over one million SNPs, which have been mapped by 'in silico' mapping, and, in some cases, further validated by further external sources. The information related to these SNPs is available to the public without intellectual property restrictions.

METHODS

Microsatellites

As detailed, microsatellites have become the primary tool for genetic mapping during the last ten years, as high density maps became publicly available. Many public human genome databases, including those detailed, can be reached via the UK HGMP Resource Centre main menu. Whole genome screens only became feasible both time wise and economically with the application of the fluorescent sequencing technology to microsatellite genotyping. Several commercial maps — for example, Applied Biosystems and Research Genetics — became available, comprising of 200–800 di-, tri- or tetra-nucleotide repeat markers evenly spaced...
Microsatellites have a high heterozygosity

throughout the human genome. Primer pairs, with one of the primers 5’ labelled, are utilised to amplify the microsatellite region. The single-stranded labelled products can be detected and sized accurately using sensitive automated sequencers. The Applied Biosystems technology utilising the 400 markers in the 10 cM Linkage Mapping Set 2 is probably the most frequently used system, although other technology is available. The microsatellites used have significant, different frequencies of various alleles, with a high heterozygosity, so data contain a relatively high amount of information when studying family-based transmission of alleles. The first polyacrylamide gel-based automated fluorescent sequencers (ABI 373A) were more efficient than autoradiography due to the sensitive multicomponent real time data collection, with much improved throughputs and quality of data assisted by internal size standards. Over the past few years, primer sets have improved in quality, with common reaction conditions, improved ‘PIG tailed’ reverse primers to remove one base stutter,31 and more spectrally separate dyes. There are presently five dyes available, and these are under continual review and optimisation, for better product synthesis and detection. The sequencers have also advanced so that gels have a longer well to read distance for much better resolution of products, which is possible in a shorter run time as gels are temperature controlled and thinner. They now also have virtual optical filters that collect and disperse the fluorescent light across the charge-coupled device (CCD) detector for more sensitive, simultaneous all-colour fluorescence detection. This provides the versatility to use new primer chemistries as they are developed without optical hardware changes. The throughput on these gels was increased as the lane capacity increased from 24 to 96 samples, but it was the advent of capillary machines that has automated the loading and running process completely. Analysis is also semi-automated, with software such as GeneScan for tracking data, and Genotyper for allele assignment, binning the allele data peaks while filtering out the background, and software such as GeneMapper can make this more efficient and less subjective with score quality values. Data can be exported electronically in tabular format to add to a pedigree file for subsequent checking and linkage analysis. Primer multiplexing has been utilised to reduce the numbers of reactions that need to be set up, although this requires much optimisation. DNA pooling methods32 have been utilised by some laboratories to reduce the total amount of separate genotyping samples that need to be run separately. The current electrophoresis equipment is the 16 capillary (ABI 3100), 48 capillary (Amersham Biosciences MagaBACE 500), 96 capillary (ABI 3700 or Amersham Biosciences MagaBACE 1000) and, more recently, 384 capillary (Amersham Biosciences MagaBACE 4000). These are 96 and 384 microtitre plate, compatible for automated efficiency. A capability for setting up lots of reactions, using small volume liquid handling robots, and a large thermal cycling capacity for polymerase chain reaction (PCR) is required, using a large water immersion system (KBiosystems) or numerous heating blocks. Utilising these types of machine together means large genotyping projects can successfully be undertaken when it is appropriate and the strategy of choice, with a realistic time frame and cost.

SNP discovery and validation

There are many sources of freely available data, including information of known SNPs, on searchable databases.33-36 As the number of SNPs increases well over one million,37 it will be increasingly unnecessary for the researcher to search for SNPs biologically in their own region of interest. SNPs are commonly identified ‘in silico’ when there are a number of sequences available for the region of interest. Indeed, this huge scale sequencing is a significant approach by those groups already doing such
Sequencing machines with 96 capillaries have increased the rate of sequencing and automation dramatically over the last few years. Software such as BLAST (Basic Local Alignment Search Tool), FASTA and CLUSTALW are utilised to find the sequences, to align them and subsequently to identify potential polymorphic sites. Success is obviously dependent on the quality and accuracy of sequencing data.

There are many traditional techniques in addition to sequencing that can be applied to SNP discovery by searching for areas of variation. Some have been developed further by biotechnology companies or, when applied appropriately, can provide a reasonable throughput, but they are still relatively inefficient, labour intensive and costly. Perhaps some of these are better suited for diagnostic use, where a large amount of time can be given to the development of each assay, or for finding and validating SNPs rather than typing them.

It is estimated that one-half of all SNPs can be determined by the difference in the size of allelic restriction fragments as a result of the polymorphic presence or absence of a particular restriction recognition site. After PCR, and cutting with a restriction endonuclease, the resulting different length products are visualised. One electrophoresis system that has been developed to assist with the efficient analysis of these products is the ‘microtitre array diagonal gel electrophoresis’ technique (MADGE). Essentially, it enables quick agarose electrophoresis of many products from amplification refractory mutation system (ARMS) or RFLP, which run at an angle, using only minimal reagents and equipment.

The ARMS method relies on allele-specific PCR. Two different forward primers are designed with differing 3’ nucleotides to complement the two expected alleles. Synthesis occurs during PCR, only if correct base pairing is present throughout the whole primer site. With proper primer design, the two different SNP-specific primers can be made to produce distinguishable products, possibly using different 5’ fluorescent labels or different 5’ extensions to produce product length differences. Multiplexing, using multiple sets of SNP-specific primers to give a range of identifiable products, perhaps by size and colour fluorescence, can lead to efficient screening of sufficient numbers.

For the oligonucleotide ligation assay (OLA), two primers are designed that are directly next to each other when hybridised to the complementary target DNA sequence in question. The two adjacent primers must be directly next to each other with no interval, or mismatch, for them to be covalently joined by ligation. This discriminates whether there is a SNP present. There are many different labelling and detection methods, including ELISA, or electrophoresis and detection on a fluorescence sequencer.

Another classic popular method for DNA mutation (including insertion and deletion) screening, detection and analysis is single stranded conformational polymorphism (SSCP) analysis. Primers are designed to produce fragments up to 300 bp in length. These products are heat denatured and snap cooled, and the resulting single stranded products are loaded on a non-denaturing acrylamide gel. Products may be radio-labelled, fluorescently labelled or if unlabelled could be detected by silver staining. The differing conformations brought about by the mutation lead to a different DNA mobility due to the intra-strand secondary structure, so the banding patterns are visualised and compared with controls of known variants. Again, with the latest electrophoresis labelling and electrophoresis methods, SSCP methods can achieve a relatively high throughput, but they are complex to set up and not as robust as would be liked.

Most mutations occur in heterozygous form, so heteroduplexes can be formed by denaturing the PCR products by heating, before cooling them slowly. If the complementary strands of two different
heterozygous alleles re-anneal, there is a bulge in the complex as the non-complementary bases are aligned. The mobility of these denatured heteroduplex fragments changes and can be studied under suitable conditions using denaturing gradient gel electrophoresis (DGGE) and denaturing high performance liquid chromatography (dHPLC). This latter method has been made more efficient by Transgenomic in their WAVE Nucleic Acid Analysis System and has become a laboratory standard for SNP detection.

**SNP genotyping**

SNP typing methods need to be amenable to high throughput analysis to be used for large studies of complex diseases. The majority of protocols involve amplification of the area of sequence to be considered, before specific determination of the variable sequence by further interrogation, whether it be enzymatic or by the addition of further nucleotides or hybridisation, for example. A selection of methods will be considered briefly below, with full details in the original publications.

**Primer extension**

Several stages are required for any protocols designed around the principle of primer extension. Products covering the SNP region are amplified by PCR, and any remaining amplification primers and dinucleotide triphosphates (dNTPs) are inactivated or removed prior to minisequencing. A dye-labelled primer is specifically annealed next to the SNP site of interest, which may then be extended by the addition of just one specific base (A, C, G or T) by adding only dideoxynucleotide triphosphates (Figure 1A), before interrogation by a suitable detection system. The fact that all these methods require multiple steps means that it is much more difficult to automate the whole process, with a significant amount of user intervention even if each part of the process is automated. This also makes the technology less robust in the first instance and then harder to implement initially in a laboratory, allowing for variation at each stage. The following variety of methods are based on primer extension with different ways of determining which base(s) have been added.

If dideoxynucleotide triphosphates labelled with different dyes are used, single base extension (SBE) products can be run on a fluorescent sequencer, either gel or capillary based, for electrophoresis, where data can be visualised and collected for analysis. By multiplexing the PCR and extension reactions using primers of different lengths, perhaps by adding 5' ddGTP ddATP ddCTP ddTTP 5'3'5'3' (A) 5'3'5'3' (B) T G NFQ MGB V F & NFQ MGB V F

**Various detection systems**

**SNPs are biallelic — more amenable to high throughput analysis**

**Multiple steps in primer extension based methods**

![Figure 1: SNP assays. (A) Primer extension: the specific extension product is then determined by a suitable method. (B) 5' nuclease assay: the enzyme activity cleaves the probe if hybridisation occurs, releasing the reporter from the quencher. F is FAM label; V is VIC label; NFQ is non-fluorescent quencher; MGB is minor groove binder.](https://example.com/figure1.png)
Primer extension — lower set up cost, potential for multiplexing

Hybridisation based 5’ nuclease assay is homogeneous with standardised reaction conditions

tails, interrogation of up to ten or so different SNPs can be achieved in one lane or capillary for each DNA. Using the latest chemistries and automated mult-capillary fluorescent sequencers, quite a large volume of data can be collected: over several thousand typings per day with good multiplexing. Throughputs will increase with a greater number of capillaries per machine, shorter capillaries to reduce run time and more efficient automated analysis software.

Conventional detection methods can be used to detect the SBE — an immunochemical assay, for example. Two different antibodies recognise allele-specific epitopes in the primer extension products, and the subsequent secondary enzyme-catalysed chemical reactions generate a coloured signal, which is an indirect measure for each allele.

Matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) can be used to separate and detect products simultaneously from the underlying primer extension chemistry to determine the genetic variants of a SNP site directly without the need for any labelled tags. The products are separated to a high degree of precision by their respective small, different molecular masses that are dependent on which base(s) has been added. Each assay again requires multiple steps before being spotted on a target plate that is inserted into the machine for reading. A high throughput mass spectrometer is capable of reading thousands of readings a day from an array, desorbing them individually for detection. Reactions can be multiplexed to a small degree with careful consideration to the expected masses of products, which are quickly automatically analysed. There are several different strategies available that vary by the exact extension and sample preparation chemistries and MALDI-TOF machine used; that may be part of a system like the widely used Sequenom MassARRAY or stand alone in an efficient production line. A high throughput of genotyping (~50,000/day) can be achieved with increased multiplexing, density of spotting and automation of multiple target plate loading. An advantage with primer extension-based SNP methods is that assays can be designed and tested relatively inexpensively, which is useful for SNP validation, before progressing to large scale genotyping. Consumable cost savings can be made by multiplexing reactions, but a degree of optimisation for suitable groups of assays may be involved for good results, leading to inflexible groups of SNPs to be typed. There is a subsequent loss of multiplexing with cost implications if some assays fail or are no longer required to be typed for the same samples.

In pyrosequencing, single stranded DNA is produced before the complementary strand synthesis is performed in the absence of dideoxynucleotides. Each dNTP is added individually with a nucleotidase enzyme, and incorporation of each nucleotide is then monitored by the release of inorganic pyrophosphate from the dNTP which is converted to ATP, fuelling a luciferase reaction. If the dNTP is not incorporated, it is degraded with no light emitted. The sequence of events is followed, and is specific to the sequence of the variant: the surrounding sequence giving confidence to genotype data. The primer can afford to be a little further away from the variable site if optimal for extension, as a small length of sequence can be produced in this manner — which means this is the preferred choice if two SNPs close together are to be typed. Reagent costs are high for standard volumes, but can be minimised with the latest machines and automated so that many reactions can be performed in parallel. Savings should be possible using pyrosequencing for allele quantification using pooled DNA samples; but work is required to determine the sensitivity to small allele frequencies (about five per cent) for different numbers of pooled samples negating any preferential amplification.
Hybridisation

A common assay is the 5’ nuclease assay (also known as the TaqMan assay): the allelic discrimination is based on the characteristic 5’ to 3’ exonuclease activity of Taq DNA polymerase.\textsuperscript{51-53} PCR using flanking primers is performed, including fluorescent oligonucleotide probes in a homogeneous assay. The probes consist of a 5’ reporter dye and a 3’ quencher dye, and are specific to the region containing the base change in the region to be amplified. The 5’ nuclease activity cleaves the probe if hybridisation occurs, releasing the reporter from the quencher. Two different probes with different fluorogenic reporters are put in the reaction for allele discrimination, one specific to complement each of the variant alleles to be typed. If there is a mismatch between the probe and target DNA sequence, the hybridisation is significantly reduced, therefore stopping the cleavage of reporter from quencher and release of fluorescent signal (Figure 1B). The amounts of each signal released indicate which allele(s) of the target region is present (Figure 2). The probes have been improved further by the introduction of non-fluorescent quenchers that contain a minor groove binder in the probe, thereby increasing the specificity of the probe to the SNP region and reaction standardisation and sensitivity. A complete system of synthesising probes and automated detection equipment (7900HT Sequence Detection System) has been commercially developed by Applied Biosystems, with automated 384-well plate loading for end-point detection. This has the potential for \textasciitilde 100,000 genotypes a day, although specialised probes are required to be synthesised after design at present, until a pre-developed robust genome-wide set of assays is released shortly. The system is, therefore, presently inefficient for typing small sample sizes (<1,000) due to the set-up charges, but is competitively priced for larger sample sizes using small reaction volumes. The homogeneous nature of the assay with standard reaction conditions.

![Figure 2: Scatter plot of the results from 5’ nuclease assay reactions analysed on an ABI PRISM 7900HT sequence detection system. The clusters of output of the fluorescent data are seen: only FAM signal, homozygous allele 1; only VIC signal, homozygous allele 2; and increase in both FAM and VIC signal, heterozygous, both alleles present. The negative control is also shown.](https://academic.oup.com/bfg/article-abstract/1/2/139/207874)
means implementation and automation is quite simple, with flexibility as no multiplexing is involved. Scoring of data can be automated with suitable clustering software, although this is still in development, and some assays may be difficult to score, or worth redesigning. The same microarray hybridisation chip technology utilised for gene expression studies can be applied to SNP typing. Although a technological challenge, DNA samples or oligonucleotides can be arrayed on the surface of a chip at high density. There are differences in the allele-specific oligonucleotide hybridisations (ASO) of short oligonucleotides to matched or mismatched target sequence. After stringent washing, hybridised mismatch probes fall off, and only perfectly matched probe oligonucleotides remain attached. The resulting detailed fluorescent signals on the arrayed chip are compared with reference samples to identify the SNP. A problem due to the common hybridisation conditions used for multiplexing means they may be sub-optimal for specificity for all of the SNPs, with subsequent implications for data quality. By generating lots of short PCR products for each SNP in multiplex PCR, and screening lots of SNPs in parallel, a huge amount of information can be obtained. Affymetrix can provide such a complete system to automate the hybridisation, staining, data scanning and analysis with minimal hands-on time. At present, there are about 1,500 SNPs on each chip that can be interrogated in a single experiment, although this number is sure to increase significantly.

One assay type which utilises the structure-specific endonuclease activity of cleavage enzymes is called ‘cleavage fragment length polymorphism’, or, more commonly, the Invader assay (Third Wave Technologies). If the invader probe and labelled signal probe hybridise to the target DNA, and overlap, there is subsequent invasion by signal—probe target duplex, displacing the single stranded flap containing label. This juncture between flap and duplex is recognised and cleaved, releasing the unhybridised fragment of probe, that can be detected by various methods — eg fluorescence resonance energy transfer (FRET). The reaction is performed near the probes’ melting temperature, and no cycling is required because new signal is released as new probe hybridises, with cleaved probe accumulating. The probes are designed specifically by the technology provider to cause cleavage on the site of the altered base. Clusters are scored manually, although this may be quite subjective in some assays.

Dynamic allele-specific hybridisation (DASH) is another method for scoring SNPs, and can be used for scoring insertion/deletion polymorphisms and complex sequence rearrangements. It involves steadily increasing the temperature of a duplex formed between PCR amplified target DNA and an allele-specific oligonucleotide probe. The hybridisation state of this duplex is followed dynamically via fluorescence output from an intercalating double-strand specific dye, in order to determine the complete denaturation profile. Distinct target sequence alleles are then readily distinguished by their melting temperature differences.

A solution-based, microtitre plate method for SNP genotyping is based on allele discrimination by ligation of open circle probes, followed by rolling circle amplification of the signal using fluorescent primers. Only the probe with a 3’ base complementary to the SNP is circularised by ligation.

Another type of probe, called a molecular beacon, allows the simultaneous amplification and detection of nucleic acids within a closed reaction vessel. The method is based on the incorporation of energy transfer-labelled hairpin primers into the amplification product. The hairpin primers are designed so that a fluorescent signal is generated only when the primer is unfolded during its incorporation into an amplification product. The fluorescence signal
produced directly correlates with the accumulation of PCR product at each cycle. Unincorporated primers have an extremely low fluorescence signal, eliminating the need to purify the PCR reaction prior to quantitation; thus, PCR and fluorescent signal detection can occur in a single reaction vessel. Signal is measured either during the reaction in real-time or at the end-point after the reaction.

Electronically active microfabricated arrays or electronic chips have been developed by Nanogen, so the hybridisation can be carefully controlled by the charge on array. DNA is immobilised, and matches and mismatches of allele-specific oligonucleotides are measured by analysing the fluorophore signals.

Novel tagging technology is used in the Masscode system (Qiagen) technology that has a resultant high throughput. The system takes advantage of the powerful and rapid data acquisition capabilities of a standard mass spectrometer to read DNA by developing a system of specially designed cleavable Masscode tags. The system has been designed to be more specific and to provide hundreds of easily detectable tags, allowing massive parallel measurements, although there are several steps in the process.

Bead-based hybridisation assays are in development by several companies; these include Illumina, using fibre-optic bundles, Luminex using microspheres labelled with fluorescent dyes and Lynx Therapeutics using microbeads. Signal intensity is measured for thousands of reactions simultaneously and genotype readouts are obtained.

**CONCLUSIONS**
The technology of genotype microsatellites has been developed and can now be accessible to all, producing several thousands of genotypes a day. Genome screens can be performed routinely for large numbers of DNA samples with the application of this technology supported by reliable, reproducibly accurate, high capacity liquid handling robotics and sufficient thermal cycling capacity. It is commonly believed that screens will always have a role in genetic analysis, but their application in the huge numbers needed to allow genotyping for future studies will always be limited by the requirement for electrophoresis detection and allele assignment.

In contrast, SNPs are biallelic by their nature, so it is a realistic challenge to genotype them automatically, even though more need to be typed to gain the same amount of information. The exact density or number of markers required to cover the human genome is an area of debate: perhaps about 2,000 for linkage purposes, and 200,000 for an association map. The development and use of a huge range of different technological advances in various methodologies and platforms for SNP genotyping have been illustrated: all with high pass rates and an accuracy capability of >99 per cent. The concepts behind these potential SNP applications are well studied, as it is realised that the quality of data produced is only as good as the underlying principle behind the assay utilised. Indeed, it also has to be remembered that the results from the assays rely on using enough good quality source DNA. The methods vary, sometimes only in subtle details, but all of the technologies and platforms discussed have their own merits and disadvantages. They complement each other and are suitable for different laboratories as preferred, depending on project type and circumstances, with current throughputs between 5,000 and 100,000 typings each day. This number of SNP genotypes that can successfully be produced is continually rising, so it is difficult to judge the longevity of current methods. The basis of the assays may remain the same, relying on fundamental biological genetic principles, but the platforms and manner in which they are performed will evolve dramatically, perhaps analogous to microsatellite genotyping. In common with microsatellite genotyping, the set-up...
costs of equipment and consumables such as probes and chips are sure to decrease and become more accessible to all. Currently, capital costs may be in excess of £250,000 for complete set up, including pipetting work stations and suitable detection equipment. Reagent costs also vary considerably — between £0.20 and £2.00 per assay — for each type of technology. These figures are very specific to the exact numbers of assays and samples to be typed with economics of scale, sample size reduction and multiplexing to be taken into consideration. There may of course be a shift to a whole new developing technology that proves successful — such as the promising DNA chips, microbeads or fibre-optics, with a magnitude drop in costs to a couple of pence per reaction. A technology that can use less DNA or pooled samples would have major advantages, or perhaps whole genome amplification (WGA) will become common prior to analysis. What is certain is that miniaturisation, parallel processing and automatic scoring, all supported by an appropriate laboratory information management system (LIMS), will be vital. Due to the not inconsiderable capital outlay to set up for SNP analysis, it may be more sensible for large centres to provide SNP genotyping resources for other groups that have access to the large DNA collections to be typed. Access to high throughput, small volume liquid dispensing, and a capacity for a large amount of thermal cycling, is also normally required, as well as assay detection equipment. This might not be realistic for all laboratories more interested in studying the results for the genetic or proteomic implications. While it is not certain which technology will be the most utilised in the future, SNPs are certainly in vogue and the time to start typing SNPs in the genetics community is now.

The value of this large amount of variation data remains to be proven62, with the correct analysis strategy vital for success. The eventual goal must be to help understand the way these differences in sequence affect the function and disease in all genes, leading to results of pharmacogenetic importance to, and resulting in improvement in, our future health.

Acknowledgements
The UK HGMP Resource Centre is funded by the Medical Research Council (MRC). I thank colleagues for valuable comments and discussions on this paper.

References


23. URL: http://snp.cshl.org/

24. The UK HGMP Resource Centre, URL: http://www.hgmp.mrc.ac.uk/


27. CEPH database, URL: http://www.cephb.fr/cephpd/

28. The Marshfield Center for Medical Genetics, URL: http://research.marshfieldclinic.org/genetics/

29. Whitehead Institute/MIT Genome Center, URL: http://www.genome.wi.mit.edu/

30. The Genome Database, URL: http://www.gdb.org/


34. URL: http://hgvbase.cgb.ki.se/

35. URL: http://snp.cshl.org/data/

36. URL: http://www.genome.wi.mit.edu/SNP/human/index.html


44. Spiegelman, J. I., Mindros, M. N. and Oefner, P. J. (2000), ‘High-accuracy DNA sequence variation screening by