http://www.ashley-pub.com

***Review***


1. Introduction
2. Review
2.1 Hybridisation approaches
2.2 Primer extension approaches
2.3 Oligonucleotide ligation approaches
2.4 Flap probe cleavage approach
3. Conclusion and expert opinion
Bibliography


**Genotyping**

# High-throughput genotyping assay approaches


Pui-Yan Kwok

*Washington University School of Medicine, 660 South Euclid Avenue, Box 8123, St. Louis, MI 63110, USA*



High-throughput genotyping approaches are being developed to meet the demands of pharmacogenomics, where numerous individuals are studied with thousands of single nucleotide polymorphism (SNP) markers. All non-gel-based genotyping approaches achieve allelic discrimination by one of four mechanisms: allele-specific hybridisation, allele-specific primer extension, allele-specific oligonucleotide ligation and allele-specific cleavage of a flap probe. By combining one of these allelic discrimination mechanisms with either a homogeneous or solid-phase reaction format and a detection method such as fluorescence intensity, fluorescence polarisation or mass spectrometry, a number of viable high-throughput genotyping methods have been developed and are being readied for routine use. With the biochemistry for robust genotyping in place, good engineering solutions are needed to make high-throughput genotyping a reality.



***Keywords:***

*Pharmacogenomics (2000) **1**(1):*


## 1. Introduction

There is a concerted effort to develop a large set of single nucleotide polymorphism (SNP) markers for the human genome, funded by both public and commercial sources [1]. The impetus for this enormous undertaking is the belief that the genetic factors associated with common diseases can be identified by studying the patient and control populations, if enough genetic markers are used [2]. The theoretical models predict that a large number of individuals must be studied with an even larger number of markers for association study approaches to have adequate power to find genetic factors with moderate effects on a particular condition [3,4]. With the many SNP discovery projects currently underway, there is no doubt that > 500,000 human SNPs will become available within the next two years [5]. With a super-dense set of markers, it is possible to initiate large-scale association studies designed to elucidate the genetic risk factors for common diseases such as infection, cancer, cardiovascular diseases, autoimmune disorders and psychiatric disorders [3]. In addition, one can also conduct association studies to identify the genetic factors that determine if a patient is likely to benefit from, or be adversely affected by, a particular medication [6]. Regardless of the application, the use of SNP markers in population studies will demad the ability to perform a large number of genotyping tests. In this review, the high-throughput genotyping approaches currently under development will be described.




2000 © Ashley Publications Ltd. ISSN 1462-2416

## 2. Review

Determining which base-pair is present at a particular location in the genome is not a trivial task. In the early days of SNP genotyping, one relied on gel electrophoresis to differentiate between two alleles [7]. Examples of these approaches include those exploiting the mobility difference between fully annealed and partially melted DNA fragments, single stranded DNA molecules with conformational difference between the alleles, and homoduplex and heteroduplex DNA fragments. Alternatively, one could take advantage of the difference in the cleavage or binding efficiency between an unpaired nucleotide in a heteroduplex DNA fragment *versus* a paired nucleotide in a homoduplex DNA fragment. In addition, one could also differentiate between the alleles if one of them destroyed or created a new restriction site. Even with the use of fluorescent labels and automated gel and capillary electrophoretic instruments, the throughput of these methods is still substantially below that required to sustain large scale studies.

As the field turns to methods not based on electrophoresis, a number of promising approaches have emerged [8]. Besides the requirements that the genotyping methods be robust and cheap, high-throughput methods must also be amenable to automation. In general, three aspects of the genotyping method must be considered.

- First, an allele discriminating method must be chosen. For non-gel-based methods, four molecular mechanisms can be used; namely, allele-specific hybridisation, allele-specific primer extension, allele-specific oligonucleotide ligation and allele-specific cleavage of a flap probe.
- Second, a reaction format must be selected. This can be done in solution, on solid support, or as a combination of the two.
- Third, a choice must be made of detection modality with the accompanying analytical algorithm to determine the genotype of the test sample. The detection methods available range from colorimetric approaches based on the ELISA to changes in fluorescence intensity and polarisation to mass spectrometry.

All the high-throughput genotyping methods currently in use or being developed are based on a combination of the three requirements outlined above. Each combination has its strengths and weaknesses. Depending on the specifics of a particular project, one method may be more appropriate than another.

### 2.1 Hybridisation approaches

Although distinguishing between two DNA molecules differing by one base by hybridisation is a demanding task, several strategies have been used successfully. In the solid-phase reaction format, one strategy is to use fluorescently-labelled PCR products in a hybridisation reaction with sets of oligonucleotides complementary to the allelic sequences placed on a microarray [9]. After the hybridisation reaction, unbound PCR products are washed away and the fluorescence pattern is analysed to determine which allelic set of probes give positive hybridisation signals. With robust multiplex PCR, numerous SNPs can be typed in one experiment. The major determinant in this approach is the multiplex PCR step. A major drawback of this approach is the common observation that 10-15% of the SNPs on the microarray fail to yield a definitive genotype and it is difficult to find a way to recover the data, since repeating the experiment is not a viable option. In addition, the fluorescence pattern of the microarrays requires specialised imaging, and image-analysis algorithms must be used to compute the genotypes.

When the hybridisation is performed in solution, there are two promising strategies. One is based on the 5'-nuclease activity of Taq DNA polymerase; the other relies on the opening of a stem-loop structure upon hybridisation. In the 5'-nuclease reaction, a oligonucleotide probe hybridised to a target being amplified by the polymerase chain reaction is cleaved by the Taq DNA polymerase. If a set of probes can be built such that they bind to the target DNA only when they are perfectly complementary to the target, one can infer the genotype of the target DNA by determining which allele-specific probe is cleaved during the PCR reaction [10,11]. With fluorescence energy transfer probes, cleavage of the probes leads to an increase in fluorescence of the reporter dye when it is separated from the quencher found on the other end of the probe. In the second approach, a fluorescent dye and a quencher are brought together when a probe forms a stem-loop structure in solution. When a target DNA molecule containing an SNP is perfectly complementary to the loop sequence, hybridisation occurs and the stem is opened up and the fluorescent reporter dye is separated from its quencher [12,13]. Once again, an increase in

© Ashley Publications Ltd. All rights reserved.

fluorescence is observed when allele-specific hybridisation occurs. Both of these reactions can be observed in solution during PCR without any purification or separation, making them the simplest reaction to set up. The major drawback of these two approaches is the cost of the labelled probes.

## 2.2 Primer extension approaches

Several strategies utilising the ability of DNA polymerase to incorporate the nucleotide complementary to the template are contenders in the high-throughput genotyping arena. These strategies are based on either allele-specific PCR or single base extension with dideoxynucleotide terminators or direct sequencing of PCR products. In the allele-specific PCR approach, two allele-specific PCR primers are constructed with the allelic bases at their 3'-end. When the target DNA is incubated with the allele-specific primers and a common reverse primer, only the allele-specific primer complementary to the allele(s) found in the target will produce a PCR product [14]. By determining which product is amplified, one can infer the allele(s) present in the test sample. The most promising allele-specific PCR genotyping method distinguishes the two allele-specific PCR products by attaching a GC-rich tail to one of the allele-specific PCR primers. The resultant PCR products therefore exhibit very different melting curves when analysed in real time [15]. The advantage of this approach is the simplicity of assay set up and the inexpensive primers required in the assay. The only drawback of the method is the effort needed in optimising each assay.

In the single base extension approach, the target region is amplified by PCR followed by a single base sequencing reaction using a primer that anneals one base shy of the polymorphic site. Because it is a very simple and robust assay, there are at least four reaction formats with high-throughput potential.

- First, one can immobilise the sequencing primers of a large number of SNPs on a solid support and perform a large number of single base extension reactions in parallel. Since the only reagents needed are the four dye-labelled dideoxynucleotides and the DNA polymerase, as long as the many SNP targets can be amplified efficiently in multiplex PCRs, the reaction is quite simple [16,17]. The drawbacks are similar to those of any microarray-based systems where inflexibility of the probe set is always an issue.
- Second, one can circumvent the inflexible probe array design by constructing a set of probes (called 'tags' by one group) with very robust and uniform hybridisation characteristics and placing them on a generic microarray. One can then design SNP-specific sequencing primers tailed with a sequence complementary to one of the generic probes found on the array. The multiplex PCR and single base extension reaction are performed in solution and the mixture of products is incubated with the 'tag-array' [18]. The products are segregated by their tail sequence and the hybridisation pattern is analysed to determine the alleles found in each test sample.
- Third, the single base extension product can be analysed by mass spectrometry [19,20,21]. Here, no dye-label is needed since each dideoxynucleotide has a unique mass. The advantage of mass spectrometry detection is that the genotypes are not inferred, but determined accurately as a physical property of the single base extension product. The drawback of mass spectrometry detection is the intolerance of the instrument to ionic and other contaminants, therefore requiring meticulous sample preparation and handling.
- Fourth, a single base extension assay based on fluorescence polarisation detection has been developed as a homogeneous assay. The detection method takes advantage of the change in molecular weight when a dye-terminator is incorporated in the single base extension reaction. By monitoring the change in fluorescence polarisation, one can infer which dye-terminator is incorporated and determine the genotype present in the test sample [22]. The advantage of this method is its simplicity and low reagent cost. One drawback of all the single base extension reaction methods is that several steps are involved.

A non-gel-based sequencing method, called pyrosequencing, has been developed for SNP genotyping [23,24]. In pyrosequencing, one monitors the release of the by-product of DNA polymerisation, pyrophosphate, when a natural deoxynucleotide is added to the sequencing reaction. In a series of enzymatic reactions, pyrophosphate is converted to ATP, the energy source for a luciferase reaction that produces light. Therefore, by monitoring which deoxynucleotide's addition is associated with light emission, one can infer the based being incorporated onto the primer. One advantage of pyrosequencing is that one

© Ashley Publications Ltd. All rights reserved.

can determine the sequence of up to 50 bases, making it the method of choice in cases where multiple SNPs are found in one locus. The drawbacks include the need for multiple enzymes and problems inherent in reactions that require sequential addition of reagents.

### 2.3 Oligonucleotide ligation approaches

Oligonucleotide ligation is probably the genotyping method with the highest level of specificity. By designing oligonucleotides complementary to the target sequence, with the allele-specific base at its 3'-end or 5'-end, one can determine the genotype of the PCR amplified target sequence by determining whether an oligonucleotide complementary to the DNA sequencing adjoining the polymorphic site is ligated to the allele-specific oligonucleotide or not. Just like the other allele discrimination approaches, oligonucleotide ligation assays can be performed with oligonucleotide probes immobilised on solid support, with the ligation reaction performed in solution followed by sorting of the products on 'tag-arrays', or with the assay performed entirely in solution.

The homogeneous oligonucleotide ligation assay based on fluorescence resonance energy transfer (FRET) detection is based on the fact that certain dye-pairs exhibit FRET when they are in close proximity. In the ligation reaction, the donor dye is placed on the oligonucleotide common to both alleles while the acceptor dye is found on the allele-specific oligonucleotide probe. When the donor dye is excited, FRET occurs and the acceptor fluorescence is observed only when the two oligonucleotides are ligated together in the presence of the appropriate target [25]. Because ligation and PCR are different molecular reactions, reagents for both reactions can be assembled in the same reaction tube at the beginning of the assay. The advantage of this assay is its simplicity and specificity. The major drawback of the ligation approach is the cost of the three oligonucleotides needed for each SNP. Since DNA ligation requires a 5'-phosphorylated oligonucleotide at the junction, at least one of the oligonucleotides must be modified. With FRET detection, all three oligonucleotides must be labelled with an additional dye.

### 2.4 Flap probe cleavage approach

A new molecular mechanism for single base detection has been described recently. This approach, called the 'invader assay', is based on the observation that flap endonucleases (also called cleavases) isolated from archaea recognise and cleave a structure formed when two overlapping oligonucleotides hybridise to a complementary DNA target. When the downstream oligonucleotide is designed with a 'flap' consisting of non-complementary sequence, this stretch of DNA is cleaved in the presence of an upstream 'invader' oligonucleotide and target DNA. If the cleaved 'flap' is used as the 'invader' oligonucleotide in a secondary assay with a second 'flap' probe and a synthetic template, one can amplify the signal without amplifying the genomic DNA target. The final detection can be based on fluorescence intensity changes when the FRET probe is cleaved [26]. The main advantages of this approach are the possibility of eliminating PCR amplification, the low cost of unlabelled allele-specific probes and the relative simplicity of the reaction protocol. The only drawback is the need for relatively large amounts of genomic DNA in the invader assay when PCR is not used.

## 3. Conclusion and expert opinion

There are at least twelve serious contenders in the high-throughput genotyping arena. Some require very high start-up cost in terms of automation and detection equipment, some need expensive probes that must be prepared before the assay can be optimised, some assays involve many steps, and some utilise expensive reagents. When a set of SNP markers is used to genotype thousands of samples, the cost of probes and the effort expended on optimisation are amortised over a large number of assays. In this scenario, the personnel cost and the variable cost of each reaction become the most important factors. In cases where one must quickly move from one set of markers to the next and develop the assays along the way, methods that require little or no optimisation are more appropriate. While all of the approaches described in this review have high-throughput genotyping potential, none has gone through extensive 'field testing' and several of them await engineering solutions to realise their full potential.

The speed with which the goals of pharmacogenomics is met largely depends on the development of a large set of SNP markers, a suite of high-throughput genotyping methods, and analytical tools that can handle the deluge of data and identify the genetic factors important in drug action. Given the many genotyping approaches developed over the last few years, there is great optimism that when a large set of characterised SNP markers become available for pharmacogenomics applications at the end of 2001, a

© Ashley Publications Ltd. All rights reserved.

number of robust, inexpensive, high-throughput genotype methods will be ready for routine use.

## Bibliography

Papers of special note have been highlighted as:
- • of interest
- •• of considerable interest


1. COLLINS FS, GUYER MS, CHAKRAVARTI A: **Variations on a theme: Cataloging human DNA sequence variation.** *Science* (1997) **278**:1580-1581.
2. LANDER ES, SCHORK NJ: **Genetic dissection of complex traits.** *Science* (1994) **265**:2037-2048.
3. RISCH N, MERIKANGAS K: **The future of genetic studies of complex human diseases.** *Science* (1996) **273**:1516-1517.
4. KRUGLYAK L: **Prospects for whole-genome genome linkage disequilibrium mapping of common disease genes.** *Nat. Genet.* (1999) **22**:139-144.
5. MARSHALL E: **Drug firms to create public database of genetic mutations.** *Science* (1999) **284**:406-407.
6. HOUSEMAN D, LEDLEY FD: **Why pharmacogenomics? Why now?** *Nature Biotech.* (1998) **16**(Suppl.):2-3.
7. KWOK PY, CHEN X: **Detection of single nucleotide polymorphisms.** In: *Genetic Engineering, Principles and Methods.* Setlow JK (Ed.), Plenum Press, New York, USA (1998) **20**:125-134.
8. LANDEGREN U, NILSSON M, KWOK PY: **Reading Bits of Genetic Information: Methods for Single-nucleotide Polymorphism Analysis.** *Genome Res.* (1998) **8**:769-776.
9. WANG DG xxxxxxxxxxx *et al.*: **Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.** *Science* (1998) **280**:1077-1082.
10. LIVAK KJ: **Allelic discrimination using fluorogenic probes and the 5' nuclease assay.** *Genet. Anal.* (1999) **14**:143-149.
11. LEE LG, LIVAK KJ, MULLAH B *et al.*: **Seven-color, homogeneous detection of six PCR products.** *Biotechniques* (1999) **27**:342-349.
12. TYAGI S, BRATU DP, KRAMER FR: **Multicolor molecular beacons for allele discrimination.** *Nat. Biotechnol.* (1998) **16**:49-53.
13. KOSTRIKIS LG, TYAGI S, MHLANGA MM *et al.*: **Spectral genotyping of human alleles.** *Science* (1998) **279**:1228-1229.
14. SOMMER SS, GROSZBACH AR, BOTTEMA CD: **PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes.** *Biotechniques* (1992) **12**:82-87.
15. GERMER S, HIGUCHI R: **Single-tube genotyping without oligonucleotide probes.** *Genome Res.* (1999) **9**:72-78.
16. SHUMAKER JM, METSPALU A, CASKEY CT: **Mutation detection by solid phase primer extension.** *Hum. Mutat.* (1996) **7**:346-354.
17. PASTINEN T, KURG A, METSPALU L *et al.*: **Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays.** *Genome Res.* (1997) **7**:606-614.
18. FAN JB: Personal communication.
19. HAFF LA, SMIRNOV IP: **Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry.** *Genome Res.* (1997) **7**:378-388.
20. TANG K, FU DJ, JULIEN D *et al.*: **Chip-based genotyping by mass spectrometry.** *Proc. Natl. Acad. Sci. USA* (1999) **96**:10016-10020.
21. LI J, BUTLER JM, TAN Y *et al.*: **Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry.** *Electrophoresis* (1999) **20**:1258-1265.
22. CHEN X, LEVINE L, KWOK PY: **Fluorescence polarization in homogeneous nucleic acid analysis.** *Genome Res.* (1999) **9**:492-498.
23. NYREN P, KARAMOHAMED S, RONAGHI M: **Detection of single-base changes using a bioluminometric primer extension assay.** *Anal. Biochem.* (1997) **244**:367-373.
24. RONAGHI M, UHLEN M, NYREN P: **A sequencing method based on real-time pyrophosphate.** *Science* (1998) **281**:363-365.
25. CHEN X, LIVAK K, KWOK PY: **A homogeneous, ligase-mediated DNA diagnostic test.** *Genome Res.* (1998) **8**:549-556.
26. LYAMICHEV V, MAST AL, HALL JG *et al.*: **Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes.** *Nat. Biotechnol.* (1999) **17**:292-296.


Pui-Yan Kwok
Washington University School of Medicine, 660 South Euclid Avenue, Box 8123, St. Louis, MI 63110, USA
Tel.:.+1 314 362 8236; Fax: +1 314 362 8159
Email: kwok@genetics.wustl.edu

© Ashley Publications Ltd. All rights reserved. 