HUMAN MUTATION 17:475–492 (2001)

REVIEW ARTICLE

Automation in Genotyping of Single Nucleotide Polymorphisms

Ivo G. Gut
Centre National de Génotypage, Evry, France

Communicated by Stylianos Antonarakis

Automation for genotyping of single nucleotide polymorphisms (SNPs) can be split into the automation of the sample preparation and the automation of the analysis technology. SNP genotyping methods are reviewed and solutions for their automation discussed. A panacea for SNP genotyping does not exist. Different scientific questions require adapted solutions. The choice of a technology for SNP genotyping depends on whether few different SNPs are to be genotyped in many individuals, or many different SNPs are to be genotyped in few individuals. The requirements of throughput and the ease of establishing an SNP genotyping operation are important, as well as the degree of integration. The potential and state-of-the-art of different solutions are outlined. Hum Mutat 17:475–492, 2001. © 2001 Wiley-Liss, Inc.

KEY WORDS: SNP; automation; genotyping; DNA analysis; mutation detection; variation; robotics; arrays; microarrays; microspheres; mass spectrometry

INTRODUCTION

An annotated draft of the DNA sequence of the human genome was recently announced [Genome International Sequencing Consortium, 2001; Venter et al., 2001]. Rather than this being the conclusion of over a decade of work on the human genome project, it provides a starting point and a baseline for future projects, like the systematic characterization of the estimated 30,000–140,000 genes encoded in the human genome [Roest Crollius et al., 2000]. This will be followed by the systematic attribution of gene function and association with diseases of genetic origin.

The identification of disease-causing variants of genes relies on the association with disease phenotypes. Genotyping, the assignment of different variants in an otherwise conserved DNA region, is used for the identification of susceptibility genes in case/control populations. It can be the first step in the identification of gene function. Ever faster and more efficient methods for genotyping are required to support gene characterization. Due to the vast numbers of genotypes that are required for this, powerful analysis tools with substantial automation and organization are essential [Kruglyak, 1999]. In the future these methods might find their application in diagnostics; pharmacogenomics [Shi et al., 1999; McCarthy and Hilfiker, 2000], where tailored medicines will be developed to suit the needs of the individual patient; in animal- and plant-breeding; and animal- and plant-identification programs. Clearly, due to the vast numbers of genotypes that are being projected, automation of genotyping will play a key role in this field. Recently, the application of automation in genomics was extensively described in two reviews [Meldrum, 2000a, b].

Received 24 October 2000; accepted revised manuscript 7 March 2001.

*Correspondence to: Ivo G. Gut, Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France. E-mail: ivogut@cng.fr

Contract grant sponsor: French Ministry of Education, Research and Technology (MENRT).

GENOTYPING OF SINGLE NUCLEOTIDE POLYMORPHISMS

Single nucleotide polymorphisms (SNPs) are changes in a single base at a specific position in the genome, in most cases with two alleles. SNPs are found at a frequency of about one every 1,000 bases in humans [Kruglyak, 1997]. By definition the rarer allele should be more abundant than 1% in the general population. The relative simplicity of SNP genotyping technologies and the abundance of SNPs in human genome has made them very popular in recent years. First projects using SNPs as markers for genotyping have been shown [Wang et al., 1998]. Yet, there still is some debate about the usefulness of SNP markers compared to microsatellite markers for linkage studies and how many SNP markers will have to be analyzed for meaningful association studies [Kruglyak, 1999; see also Brookes et al., 2001].

Several SNP genotyping technologies have reached maturity in the last few years and are being integrated into large scale genotyping operations supported by automation. SNP genotyping methods are very diverse, as are automation solutions for them. Appropriate automation entirely depends on the method. On the other hand, the choice of method depends on the scale and the scientific question a project is trying to answer. A project might require genotyping of a limited number of SNP markers in a large population or the analysis of a large number of SNP markers in one individual. Flexibility in choice of SNP markers and DNA to be genotyped or the possibility to precisely quantify an allele frequency in pooled DNA samples might be issues. Currently there are few one-stop-shops for a high-throughput SNP genotyping process. All systems are combinations of products from different suppliers and are the results of alliances. Still most systems are custom products that are built around commercial elements to specification of individual projects.

SNP genotyping methods are very diverse [Shi et al., 1999]. Broadly, each method can be separated into two elements. The first element is a method for interrogating an SNP. This is a sequence of molecular biological, physical, and chemical procedures for the distinction of the alleles of an SNP (Fig. 1). The second element is the actual analysis or measurement of the allele-specific products. Often, very different methods share elements, like reading out a fluorescent tag in a plate reader (SNP genotyping methods with fluorescent detection were recently reviewed by Landegren et al. [1998]), or the method of generating allele-specific products (for example by primer extension or oligonucleotide ligation) which can be analyzed in different analysis formats.

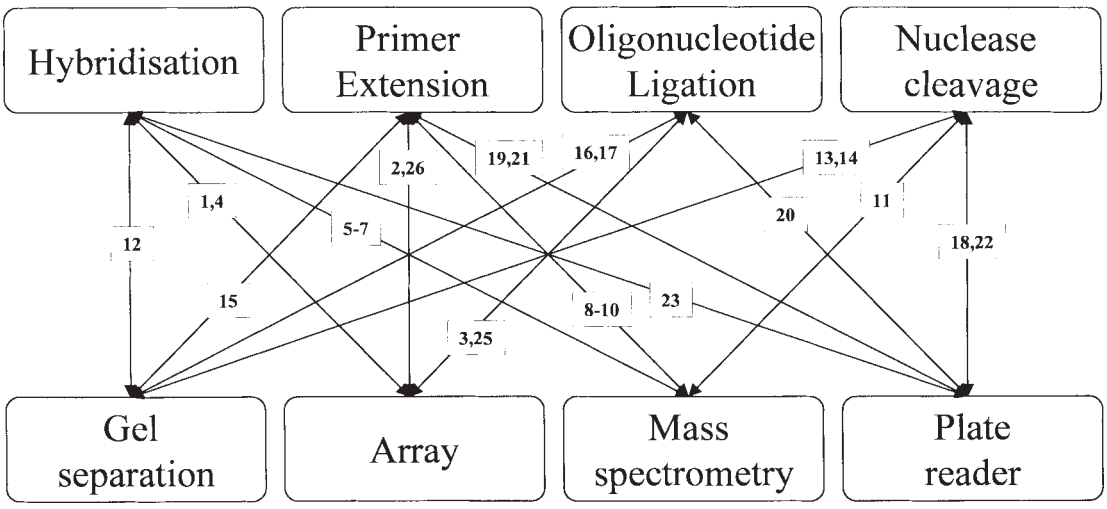

FIGURE 1. **SNP genotyping methods separated into sample preparation and analysis. Most SNP genotyping methods are a combination of one of these sample preparation techniques and one analysis technique. The numbering of methods is in Table 1.**

METHODS FOR INTERROGATING SNPS

Please see Figure 2 and Table 1 in the context of this discussion.

Hybridization

Alleles can be distinguished by hybridizing complementary oligonucleotide sequences (Fig. 2a). The stringency of hybridization is a physically controlled process. As the two alleles of an SNP are very similar in sequence significant cross-talk occurs. The stringency can be increased by the choice of appropriate buffers and temperatures, or by using modified oligonucleotides like PNAs [Egholm et al., 1993]. Direct hybridization is used in conjunction with stringent washing to remove not fully complementary probes. Molecular beacons, allele-specific oligonucleotide probes, use a fluorescent dye/quencher system for visualization [Tyagi and Kramer, 1996; Tyagi et al., 1998]. Molecular beacons are added to a PCR. The center of the molecular beacon is complementary to one allele of an SNP while the two ends are complementary to each other, allowing the formation of a stem-loop structure. A fluorescent dye is attached to one end of the molecular beacon while a quencher is attached to the other end. If the PCR product is fully complementary to the central sequence of the molecular beacon, it hybridizes and the two ends of the beacon are separated. Fluorescence is detected.

Enzymatic Methods

Adding an enzymatic step to distinguish alleles increases the fidelity. A hybridization event is followed by the intervention of an enzyme. Many different enzymes, like DNA polymerases, DNA ligases, or nucleases, can be applied. Most enzymatic methods allow the generation of both allele products distinguishable in a single reaction.

ARMS and kinetic PCR. The formation of a product in a PCR is dependent on complementary primers. For amplification refractory mutation systems (ARMS) [Newton et al., 1989] and kinetic PCR [Germer et al., 2000], one of the primers ends with correct complementary base on a polymorphism (Fig. 2b). Thus the formation or the rate of formation of a PCR product depends on the allele present in the template. One reaction for each allele is run in parallel. Successful generation of a PCR product is either analyzed on a gel (ARMS) or in real-time by the increased fluorescence of dyes intercalating into the forming PCR products.

5′nuclease assay. This assay is also called the TaqMan® assay (Fig. 2c). It actually is based on plain hybridization similar to molecular beacons, except that the probe is degraded enzymatically. A labeled oligonucleotide probe complementary to an internal sequence of a target DNA is added to a PCR [Holland et al., 1991; Livak et al., 1995; Kalinina et al., 1997]. The nucleotide probe carries a fluorescent dye and a fluorescence quencher molecule. Successful hybridization of the oligonucleotide probe due to matching with one allele of the SNP results in its degradation from the 5′end whereby the fluorescent dye and quencher are separated. This promotes fluorescence.

Primer extension. Primer extension is one way of distinguishing alleles of an SNP (Fig. 2d). For primer extension, an oligonucleotide is hybridized next to an SNP. Nucleotides are added by a DNA polymerase generating allele-specific products. Primer extension was recently extensively reviewed [Syvänen, 1999].

Oligonucleotide ligation (OLA). For OLA, two oligonucleotides adjacent to each other are ligated enzymatically by a DNA ligase when the bases next to the ligation position are fully complementary to the template strand [Barany, 1991; Samiotaki et al., 1994] (Fig. 2e). Padlock is a variant of OLA in which one oligonucleotide is circularized by ligation [Nilsson et al., 1994, 1997; Landegren et al., 1996]. Circularized DNA can subsequently be used in rolling circle replication to generate a substantial, easily detectable amount of single stranded DNA.

Restriction fragment length polymorphism (RFLP). RFLP is one of the most commonly used formats for SNP genotyping in a standard laboratory [Parsons and Heflich, 1997] (Fig. 2f). PCR products are digested with restriction endonucleases that are specifically chosen for the base change at the position of the SNP, resulting in a restriction cut for one allele but not the other. Fragment patterns are used for allele assignment after gel separation. Due to the limited number of restriction enzymes, the complex patterns that

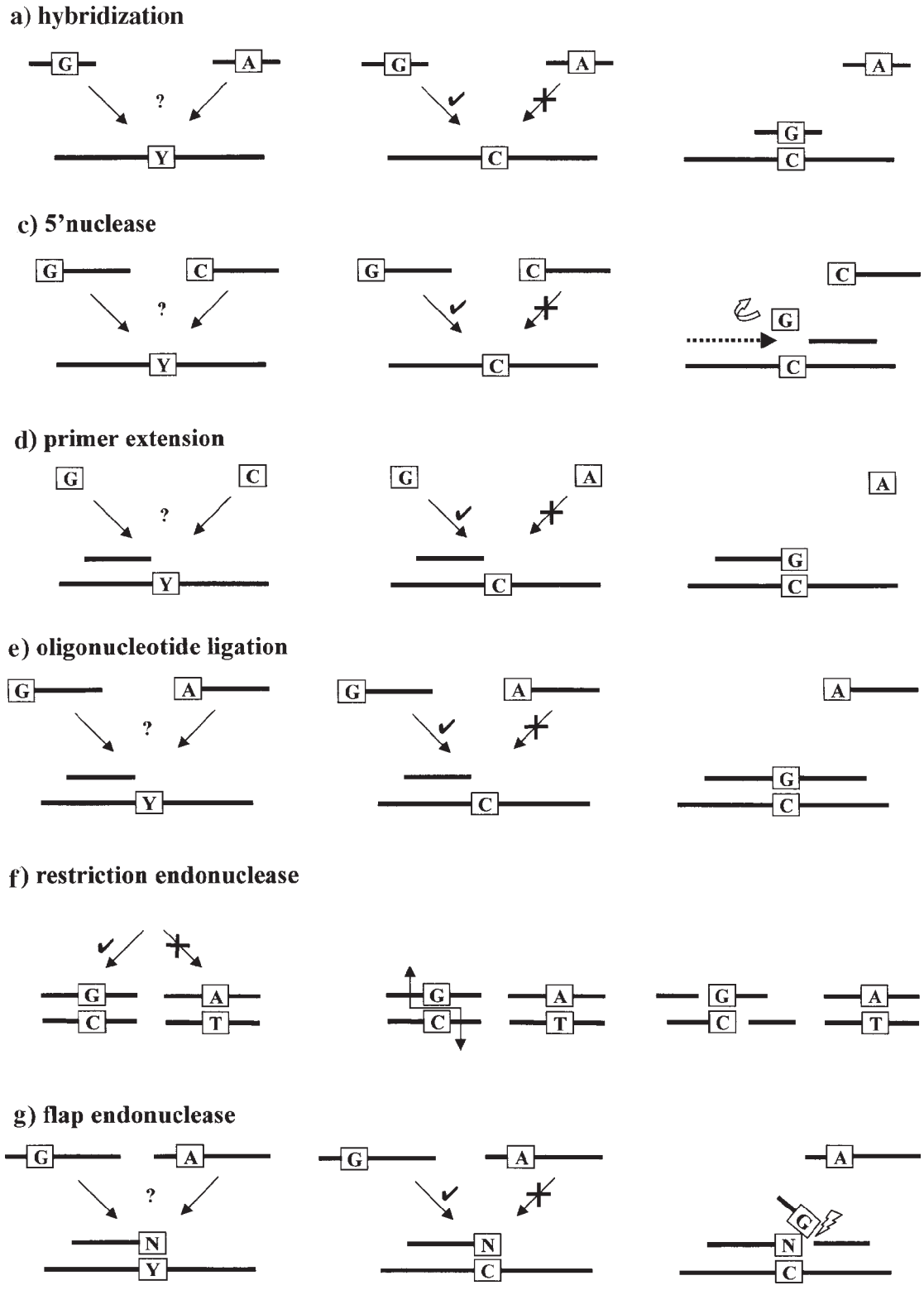

FIGURE 2. Schematic representation of different methods for the preparation of allele distinguishable products. Techniques described in detail within the text: **a:** hybridization; **b:** ARMS and kinetic PCR; **c:** 5′nuclease assay; **d:** primer extension; **e:** oligonucleotide ligation (OLA); **f:** restriction fragment length polymorphism (RFLP); **g:** flap endonuclease; **h:** DNA sequencing. (**b** and **h** are not depicted.)

may result, and gel separation it is a very labor-intensive method and does not lend itself to automation. The cost of RFLP depends on the required restriction endonuclease.

Flap endonuclease. The Invader® assay makes use of flap endonucleases (FEN) for the discrimination between alleles of SNPs [Harrington and Lieber, 1994; DeFrancesco, 1998] (Fig. 2g). An invader oligonucleotide and a signal oligonucleotide with a 5′overhang (flap) over the invader oligonucleotide are hybridized to a target sequence. Only in the case of a perfect match of the signal oligonucleotide with the target sequence is the flap cleaved off. For the Invader® squared assay, the cleaved off flap is used to drive a secondary, universal cleavage reaction thereby increasing the rate of signal generation [Ryan et al., 1999]. Cleavage of the flap can be linked to a change in fluorescence, for example with a fluorescence resonance energy transfer system. Alternatively, the cleaved off flap can also be analyzed by mass spectrometry [Griffin et al., 1999]. Processing for the Invader® assay is isothermal. It is heralded as a PCR-free genotyping method, which from an automation point of view is very enticing as there is no need for thermocycling. Unfortunately, rather large quantities of genomic DNA (50 ng/reaction) are required for each SNP. In a recent study, an automated SNP genotyping production run with the Invader® assay was shown [Mein et al., 2000]. For this, PCR was used to generate sufficient amounts of template DNA. The Invader® assay is marketed by Third Wave Technologies (Madison, WI, www.twt.com).

DNA sequencing. DNA sequencing is also a viable method for assigning alleles of polymorphisms (Fig. 2h). Currently, it is not competitive for the kind of throughput that is being targeted. A very imaginative protocol for sequencing a few tens of bases from a primer is Pyrosequencing™ [Ronaghi et al., 1996, 1999; Pyrosequencing AB, Uppsala, Sweden, www.pyrosequencing.com]. This protocol makes use of a DNA polymerase, an ATP sulfurylase, a luciferase, and an apyrase to visualize the successful incorporation of a nucleotide. The DNA polymerase in the presence of one of the four nucleotide-triphosphates makes a template directed extension of a primer, if the next base on the template DNA strand is complementary to the nucleotide-triphosphate. Pyrophosphate is released by a successful extension, which in turn is then used to drive a chemiluminescent reaction with the luciferase which is monitored in real-time. Nucleotide-triphosphates are added in turn to establish sequence ladders. The dynamic range of luminometric detection is very large. In addition, this system uses piezo pipettes liquid dispensing which is very accurate. For these reasons quantification of results is excellent.

ANALYSIS FORMATS

Over the past years diverse analysis formats have been devised. They include gels, microtiter plate fluorescent readers with integrated thermocyclers, oligonucleotide microarrays (DNA chips), coded spheres, and mass spectrometers.

Gel-Based Analysis

DNA sequencing using gel analysis can be used to genotype SNPs. This is considered to be "the gold standard" to which all other SNP genotyping technologies are compared. Slab gel formats have, in fact, been applied to SNP genotyping. However, the major drawback with this kind of gel is that despite all the efforts in the past a significant manual labor goes into pouring and loading gels. MASDA is a hybridization based SNP genotyping method with gel analysis [Shuber et al., 1997]. Primer extension using slab gel fluorescent analysis has been shown [Pastinen et al., 1996]. The oligonucleotide ligation assay with analysis on gels has also been shown [Grossman et al., 1994; Day et al., 1995]. With the emergence of 96-capillary electrophoresis sequencers, protocols are becoming available for their use with SNP genotyping. Advantages of capillary systems over slab gel systems include the potential for 24-hour unsupervised operation, the elimination of cumbersome gel pouring and loading, and that no lane tracking is required. These systems provide significant throughput. SNapShot™ is a commercially available system suited to this format (Applied Biosystems, Foster City, CA, www.appliedbiosystems.com). A facile and quite efficient method using gel separation is microtiter array diagonal gel electrophoresis (MADGE) [Day and Humphries, 1994]. SNP genotyping using it is usually based on the gen-

TABLE 1. SNP Genotyping Methods and Their Features

Method	Allele distinction	Detection	TC	Main advantage	Automation		Reference
					Allele distinction	Analysis	
Microarray							
1 GeneChip®	Hybridization	Fluorescence	Yes	Sophisticated software	—	✓	Wang et al. [1998]; www.affymetrix.com
2 APEX	Primer extension	Fluorescence	Yes	Data quality	—	✓	Shumaker et al. [1996]; Pastinen et al. [1997]
3 OLA	Oligonucleotide ligation	Fluorescence	Yes	High multiplexing	—	✓	Khanna et al. [1999]
4 EF microarray	Hybridization	Fluorescence	Yes	Flexibility, stringency	—	✓	Edman et al. [1997]; Sosnowski et al. [1997]
Mass spectrometry							
5 PNA	Hybridization	Direct	Yes	Simple principle, rapid data accumulaton	—	—	Butler et al. [1996]; Griffin et al. [1997]
6 Masscode™	Hybridization	Mass tag	Yes	Very high-throughput	NIA	NIA	www.qiagen.com
7 Mass tags	Hybridization	Mass tag	—	High-throughput	—	NIA	Shchepinov et al. [1999]
8 PROBE™	Primer extension	Direct	Yes	Accurate, rapid data accumulation	✓	✓	Little et al. [1997]
9 PinPoint™	Primer extension	Direct	Yes	High quality data, rapid data accumulation	—	—	Haff and Smirnov [1997]; Ross et al. [1998]
10 GOOD	Primer extension	Direct	Yes	Accurate, easy handling, no purification	✓	✓	Sauer et al. [2000 a,b]
11 Invader®	Endonuclease cleavage	Direct	No	No PCR, isothermal	—	—	Griffin et al. [1999]
Gel							
12 MASDA	Hybridization	Fluorescence	Yes	Easy access	—	—	Shuber et al. [1997]
13 RFLP	Endonuclease cleavage	Fluorescence	Yes	Easy access	—	—	Parsons and Heflich [1997]
14 MADGE	PCR	Fluorescence	Yes	Easy access	—	—	Day and Humphries [1994]
15 SNapShot™	Primer extension	Fluroescence	Yes	Multiplexing	—	✓	www.appliedbiosystems.com
16 OLA	Oligonucleotide ligation	Fluorescence	Yes	High plex factor for allele generation	—	—	Baron et al. [1996]
17 Padlock	Ligation	Fluorescence	No	No PCR	NIA	NIA	Nilsson et al. [1994]; Landegren et al.

							[1996]
Plate reader							
18 Invader®	Endonuclease cleavage	Fluorescence, FRET	No	End point, no PCR, isothermal	(✓)	✓	DeFrancesco [1998]; Mein et al. [2000]
19 SNP-IT™	Primer extension	Fluorescence	Yes	End point, integrated system	✓	✓	www.orchidbio.com; Syvänen [1994]; Syvänen et al. [1993]
20 OLA	Oligonucleotide ligation	Fluorescence	Yes	End point, easy format	—	✓	Barany [1991]; Samiotaki et al. [1994]
21 FP-TDI	Primer extension	Fluorescence polarization	Yes	End point, homogeneous assay	—	✓	Chen et al. [1997, 1999]
22 5′nuclease	Exonuclease cleavage	Fluorescence, FRET	Yes	Real-time, homogeneous assay	(✓)	✓	Holland et al. [1991]; Livak et al. [1995]
23 Mol. beacons	Hybridization	Fluorescence, FRET	Yes	Real-time, homogeneous assay	(✓)	✓	Tyagi et al. [1996, 1998]
24 Kinetic PCR	PCR	Fluorescence	Yes	Real-time, homogeneous assay	(✓)	✓	Germer et al. [2000]
Flow cytometer							
25 Coded spheres	Oligonucleotide ligation	Fluorescence	Yes	Flexibility, high-throughput	—	✓	Iannone et al. [2000]
26 Coded spheres	Primer extension	Fluorescence	Yes	Flexibility, high-throughput	—	✓	Cai et al. [2000]; Chen et al. [2000]
Others							
27 Pyro-sequencing	Sequencing	Chemilumi-nescence	Yes	Quantitation	(✓)	✓	Ronaghi et al. [1996, 1999]
28 Coded spheres	—	Fluorescence	—	Optical fibre reader, high-throughput	—	✓	www.illumina.com

Many protocols are only very vaguely described in the literature. They have been inferred as high-throughput SNP genotyping methods, though their integration into a high-throughput SNP genotyping products is still being awaited. The transition of a protocol into a product and scale-up usually leaves the realms of academia, which in most cases also means that information is significantly harder to come by. The numbering of methods is maintained for Figure 1.
✓, has been shown; (✓), has not been shown yet but would be reasonably easy to implement with standard robots; NIA, indicates that no information was available; TC, thermocycling.

eration of allele-specific products by ARMS or RFLP. Three hundred eighty-four reactions can be analyzed simultaneously on a horizontal gel. In terms of set-up cost this is by far the cheapest SNP genotyping technology.

Fluorescent Reader-Based Analysis

Several genotyping technologies use microtiter plate fluorescent readout systems for detection. Oligonucleotide ligation with oligonucleotides covalently linked to fluorescent chromophores is a sensitive analysis method [Chen et al., 1998]. Template-directed dye-terminator incorporation (TDI) is a primer extension procedure adapted to a fluorescent reader system [Chen and Kwok, 1997; Chen et al., 1997]. Drawbacks of these systems are that unreacted fluorescent reagents have to be removed prior to reading, PCR products used as templates for the reactions have to be prepared beforehand, and the probing oligonucleotides have to be immobilized on a solid surface. Fluorescence resonance energy transfer (FRET) allows non-immobilized processing with real-time analysis without the separation of reacted and unreacted fluorescent dyes (FP-TDI) due to the change in polarization of the fluorescence after primer extension [Chen et al., 1999]. The Invader® assay is also monitored by endpoint fluorescence. A fluorescence polarization reader is manufactured by LJL BioSystems that recently merged with Molecular Devices (Sunnyvale, CA, www.ljlbio.com). The TaqMan® assay, molecular beacons, and kinetic PCR can be executed with similar detection devices using integrated thermocycling systems (these are homogeneous assay formats, wells are stocked with template DNA and reagents, they are processed and simultaneously analyzed by the reader). In these methods the appearance of an allele of an SNP is monitored in real-time. This is accomplished by a system like the GeneAmp 5700 of AppliedBiosystems (Foster City, CA, www.appliedbiosystems.com).

Array-Based Analysis

For oligonucleotide microarrays (DNA-chips) a series of oligonucleotides (features) is chemically attached to a solid surface, usually a glass slide [Shalon et al., 1996; Ramsay, 1998]. The position of a specific oligonucleotide on the solid surface is used as the identifier. In excess of 10,000 different oligonculeotides can be arrayed on 1 cm^2 of the solid surface. The preparation of microarrays on a solid surface is done by one of two methods, either by arraying oligonucleotides or DNA [Cheung et al., 1999], or by in-situ synthesis [Pease et al., 1994; Lipshutz et al., 1999]. Hybridizing an unknown sequence to a known sequence on the DNA chip identifies a complementary sequence element in the unknown sequence. Microarrays have found their application mainly in expression analysis [Brown and Botstein, 1999; Debouck and Goodfellow, 1999]. In principle, there are three different formats that are used for SNP genotyping in an oligonucleotide microarray format:

Hybridization arrays. This format uses allele-specific hybridization. Oligonucleotides covering the complementary sequence of the two alleles of an SNP are on specific positions of the array. Fluorescently labeled PCR products containing the SNP sequences to be queried are hybridized to the array. In most cases it is not sufficient to represent only one sequence for each allele, rather a series of oligonucleotides that walk over each variant of the SNP are used (Affymetrix, Santa Clara, CA, www.affymetrix.com). This is referred to as tiling [Wang et al., 1998; Mei et al., 2000]. The hybridization pattern of all oligonucleotides spanning the SNP are used to evaluate positive and negative signals with sophisticated computer algorithms. The choice of hybridization buffers, hybridization times, and washing conditions are crucial. Discrimination of the hybridization tends to give very low signal/background as only a small difference in fluorescence upon hybridization of labeled target to matched or mismatched sequence is achieved. Using longer oligonucleotides does not improve the situation too much, as synthesis errors start playing into the results and discrimination decreases.

Arrays with electrostringent hybridization electric field microarrays. Rather than chemically binding an oligonucleotide to a solid support, this format uses electric fields (EFs) to direct the oligonucleotides to a specific address on the support [Edman et al., 1997; Sosnowski et al., 1997]. The support is a modified agarose matrix. The hybridization of templates is also

supported by applying electric fields, which results in a dramatic decrease of hybridization and wash times, while accuracy is increased. Electrostringent arrays are an interesting format that have a flexibility advantage over other array methods in terms of choice of SNPs. Currently available commercial devices are suitable for the analysis of 100 SNPs (Nanogen, San Diego, CA, www.nanogen.com).

Arrays with enzymatic processing. Arrayed primer extension (APEX) is a format in which oligonucleotides corresponding to sequences up to an SNP are arrayed on a solid surface [Shumaker et al., 1996; Pastinen et al., 1997]. PCR products containing the SNP sequences are hybridized to the arrayed oligonucleotides. Each oligonucleotide on the array acts as a primer for a primer extension reaction with a DNA polymerase and four differently fluorescently labeled dideoxynucleotides. The fluorescence emission of the incorporated nucleotide identifies the next base on the hybridized template. In contrast to hybridization arrays, only one sequence per SNP is required on the array. The specificity of the genotyping is markedly increased due to the use of enzymatic discrimination rather than differentiation by hybridization, yet enzymatic processing in the proximity of surfaces is not unproblematic. An interesting new variant of this has just been presented [Pastinen et al., 2000]. In this approach, the arrayed primers are allele-specific (the 3′ base of the primer matches one of the alleles of the SNP; two oligonucleotides are required for each SNP), and are extended with multiple fluorescent deoxynucleotides. More than one fluorescent chromophore is incorporated for each matching allele, increasing the detection efficiency. An array based detection of allele-specific products generated by OLA was used by Khanna et al. [1999].

The immediate use of genomic DNA on a microarray is not possible due to the high complexity and low concentration of target sequence. Thus all oligonucleotide array protocols require the template DNA to be amplified by PCR. To match the highly parallel analysis qualities of the microarrays, either highly multiplex PCRs have to be established or multiple PCR products have to be pooled. Multiplex PCR is difficult to control, and the risk of cross-talk increases. PCR has a bias towards smaller products, which means that only very small PCR products can be generated in highly multiplex PCR. Multiplexes beyond 8 with small fragment size PCR products (<200 bases) require a significant optimization effort.

Another analysis system that recently has been applied is fluorescence-labeled (coded) microspheres. In principle, these are arrays that have been released from the two-dimensional format. Each sphere contains a fluorescent color code. Currently, two unique fluorescent dyes at 10 different concentrations provide a set of 100 distinguishable entities. A third dye is used to identify the allele information of an SNP, for example by hybridizing a fluorescently labeled sequence complementary to a sequence bound to the bead. Photon counting is used to identify the "address" of this coded sphere and the genotyping information. In one case, a flow cytometer is used to identify the address of the sphere that is being interrogated and the allele information (Luminex, Austin, TX, www.luminexcorp.com). The readout is interpreted automatically. Both oligonucleotide ligation and primer extension have been demonstrated with this technology [Iannone et al., 2000; Chen et al., 2000; Cai et al., 2000]. The analysis of a pool of spheres can be carried out in a few seconds. This system provides more flexibility than oligonucleotide microarrays as the pool of beads can be rearranged. A similar carrier system with yet a different detection system has been presented [Healey et al., 1997]. Here the coded spheres are captured in solid phase wells to which a fiber optic detection system is coupled. This system was developed by Illumina (La Jolla, CA, www.illumina.com).

Mass Spectrometry-Based Analysis

Mass spectrometry, specifically matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, has been demonstrated as an analysis tool for SNP genotyping [Little et al., 1997; Haff and Smirnov, 1997; Kwok, 1998; Li et al., 1999; Griffin et al., 1999; Sauer et al., 2000a, b]. Allele-specific products are deposited with a matrix on the metal surface of a target plate. The matrix and

the analyte is desorbed into the gas phase with a short laser pulse. Analytes are ionized by collision and accelerated toward the detector. An accurately recorded time-of-flight of a product to the detector is directly related to its mass. The analysis is done serially (one sample after another) but at very high speed. Mass spectrometers are capable of recording a single trace in under a second. Apart from the speed and accuracy, a major strength of mass spectrometric analysis is the number of available detection channels. Resolution of the current generation of mass spectrometers allows the distinction of base substitutions in the range of 1,000–6,000 Da (this corresponds to product sizes of 3–20 bases, the smallest mass difference for a base change thymine to adenine is 9 Da). In principle, this would allow the analysis of hundreds of products in a single trace. However, for reasons of difficulty in multiplexing the sample preparation, only SNP genotyping multiplexes on the order of up to 12 SNPs have been shown [Ross et al., 1998]. Sample consumption is very low, the data accumulation can be automatic, and absolute mass information allows for automatic allele calling with high confidence. The current generation of mass spectrometers is capable of recording close to 40,000 spectra per day. This puts the theoretical capacity for SNP genotyping of a single mass spectrometer in excess of 250,000 genotypes per day. A production setting with this sort of throughput awaits to be demonstrated.

One of the major problems in the analysis of DNA by MALDI is the purity of the sample required for analysis. The purity has to be significantly higher than for other genotyping methods due to the sensitivity of the analysis to impurities. For the PROBE™ assay, the first published protocol, magnetic bead purification with a biotin/streptavidin binding system was applied [Little et al., 1997]. The PROBE™ assay with an integrated automation is commercially available as the MassArray™ system (Sequenom, San Diego, CA, www.sequenom.com). This was shortly followed by the PinPoint™ assay for which purification with ZipTips (Millipore, Bedford, MA, www.millipore.com) was applied [Haff and Smirnov, 1997]. These are solid phase purification procedures that are not easy to automate. For magnetic beads batch to batch variability has to be dealt with, while ZipTips require frequent changing as they get blocked. These purification procedures can be quite costly. The GOOD Assay was recently presented [Sauer et al., 2000a, b]. Its main feature is that although the products are analyzed by mass spectrometry, no purification is required due to a chemical modification strategy. This strategy does require additional reaction steps and modified oligonucleotides, but all reaction steps are simple additions of solutions into one reaction vial. For this reason, this method is well suited for automated sample preparation. Another advantage of this preparation format is that the analyzed products are in the size range of 4–5 bases, where the performance of MALDI mass spectrometers is optimal. For a protocol presented by Li et al. [1999], products also are in this favorable mass range, but a solid phase separation is applied.

PNAs are a suitable substrate for mass spectrometric detection. SNP genotyping using MALDI mass spectrometry and hybridized PNAs has been demonstrated [Butler et al., 1996; Griffin et al., 1997], yet this approach has so far not been integrated into a high-throughput system.

Other mass spectrometric detection methods offer further possibilities for SNP genotyping. Mass tags that can be analyzed on a number of different mass spectrometers have been shown [Shchepinov et al., 1999]. A commercially available system applying this principle is MassCode™ (Qiagen Genomics, Bothell, WA, www.qiagen.com). A reporter entity (mass tag), similar to a fluorescent dye molecule, is attached as the identifier to an oligonucleotide complementary to one allele of an SNP. For analysis, the tags are cleaved from the oligonucleotides and analyzed by mass spectrometry. In contrast to fluorescent dyes and their detection, mass tags and mass spectrometers offer a dramatically larger number of distinguishable entities. Using isotopically pure materials for the mass tags makes it possible to place one mass tag every Da. This means that a fairly simple mass spectrometer can be used to distinguish thousands of mass tags at once from oligonucleotides that have been hybridized allele-specifically.

AUTOMATION

The same way SNP genotyping can be separated into two parts (preparation of allele-specific products and analysis), so also the automation for SNP genotyping can be split into the same two parts, the automation of sample preparation and the automation of analysis.

Automation of Sample Preparation for SNP Genotyping

Preparation of allele-specific products is a significant part of any SNP genotyping procedure. Liquid and plate handling are the basic elements of automation of this preparation. Many suppliers of laboratory robotics offer liquid handling robots. Over the years, most suppliers have developed their systems from a single needle or tip to 4-, 8-, 12- and 96-fold systems. Liquid handling in macroscopic systems works by displacement, virtually identical to a standard handheld pipet. The tip and the syringe (displacement system) are connected directly or by tubing. In some cases the tubing is rather long, which has the disadvantage that drag, capillary forces, and compressibility within the tube influences the pipetting process. Air bubbles can accumulate inside water filled tubing, decreasing the pipetting accuracy. Displacement systems with short distances and low dead volumes are preferable because they provide more accuracy without a need for frequent flushing. Low dead volume liquid handling systems are used in robots of Robbins (Sunnyvale, CA, www.robsci.com), the Automation Partnership (Royston, UK, www.autoprt.co.uk), Packard (Meriden CT, www.packardinst.com), Zymark (Hopkinton, MA, www.zymark. com), and CRS (Burlington, Ontario, Canada, www.crsrobotics. com). These systems are capable of accurate liquid handling down to the low microliter range, also in a 96-tip format. At lower volumes, piezopipetting has frequently been suggested as a replacement for displacement systems. The accuracy of piezopipetting is very good as a single droplet has a volume of only a few hundred picoliters. Piezopipetting still has major drawbacks. In most cases, only a single piezopipet is operated at a time. Systems with multiple piezopipets in parallel are still at a very experimental stage. Loading, discharging, and rinsing piezopipets is very slow. They do not perform well with inhomogeneous solutions and are very expensive. They are, therefore, not yet well adapted for high-throughput applications. The only systems where they are integrated in SNP genotyping is for sample deposition in the MassArray™ system (Sequenom, San Diego, CA, www.sequenom.com) and for reagent dispensing in Pyrosequencing™ (Pyrosequencing AB, Uppsala, Sweden, www.pyrosequencing.com).

Plate handling is the second aspect of automation for sample preparation in SNP genotyping. Most SNP genotyping protocols require samples to be incubated and/or thermocycled. Therefore, a standalone sample preparation robot has to be capable of accurately handling liquids and then transferring tubes or microtitre plates into an incubator or a thermocycler. Many state-of-the-art laboratory robots are capable of doing this. Very popular robots for this are the BioMek from Beckman (Fullerton, CA, www.beckman.com), the Genesis Workstation from Tecan (Männedorf, Switzerland, www. tecan-us.com), the Multiprobe II from Packard (Meriden, CT, www.packardinst.com), the Plato from Rosys (Hombrechtikon, Switzerland, www.rosys.ch), and the Roboamp from MWG (Ebersberg, Germany, www.mwgdna.com). The Roboamp features an integrated PCR machine. CRS (Burlington, Ontario, Canada, www.crsrobotics.com) is very strong on plate handling solutions, and many systems integrating hardware from different manufacturers use their plate handling systems. The Allegro from Zymark (Hopkinton, MA, www.zymark.com) applies an assembly line modular approach that hands plates from one position to the next. Their liquid handling allows individual addressing of tips in a 96-tip format. While an excellent development, this is typically not needed in a high-throughput sample preparation system. Due to the repetitive nature of high-throughput procedures, ruggedness is preferred over flexibility. Highly sophisticated systems that support industrial scale high-throughput liquid and plate handling are built by the Automation Partnership (Royston, UK, www.autoprt.com). These systems clearly leave the realm of laboratory automation.

Many SNP genotyping methods require PCR. Most robot manufacturers will custom integrate PCR machines. Conventional PCR machines are very bulky and have large footprints, which means they occupy a lot of space on a robot bed. This limits the number of PCR machines that can be placed on a robot bed and so limits the throughput. PCR machines tend to be rather high due to the ventilation systems used to carry heat away from the heat sinks. To place PCR machines onto the robot bed in some cases requires substantial modification. A solution is to have the PCR machines outside the liquid handling area of the robot and transport plates with a robotic arm. This requires extensive effort and the price scales with transport distance. Many integrated systems like MassArray™ (Sequenom, San Diego, CA, www.sequenom.com) and SNPstream (www.orchidbio.com) take this approach.

PCR is frequently quoted as being "expensive" due to reagent cost. Reagent consumption of PCR can dramatically be reduced by optimization of reactions. A marked drawback of PCR is that thermocycling is required. From an engineering point of view, thermocycling at high-throughput is laborious, demanding, expensive, and currently there is not much hope that state-of-the-art PCR technology will be replaced by something more efficient. A high-throughput PCR device based on submersion in waterbaths has been described [Maier and Lehrach, 1997]. A commercial version is produced by KBiosystems (Basildon, UK, www.kbiosystems.com), and is based on submersion of sealed microtiter plates in three waterbaths at three different temperatures. This system is difficult to load with a plate handling robot.

Likewise the manipulation of solid particles, such as magnetic beads by a robot, is problematic. Batch-to-batch variation and adhesion of solid particles to surfaces makes them difficult to handle reliably.

Reversed-phase purification (ZipTip, Millipore, Bedford, MA, www.millipore.com) and gel-filtration, required for the PinPoint assay or many gel-based genotyping methods, are not that amenable to automation. The reproducible preparation of gel-filtration systems with a robot is difficult. Tips charged with reversed-phase material vary in loading pressures, which influences the delivery accuracy of a liquid handling robot.

In general, the ease of automation is directly correlated to the complexity of an SNP genotyping protocol. A protocol that requires only standard liquid handling is easier to implement on a standard liquid handling robot than one with "special" protocol steps. Very few robotic systems offer turnkey solutions. Usually, substantial work has to be invested into taking a robotic system into production. Sample preparation for SNP genotyping methods that process serially (for example the MALDI mass spectrometer) were the first to benefit from full-scale automation. Sample preparation for microarray SNP genotyping with robotic preparation systems have, so far, not been demonstrated. For this sample, preparation still is a manual process. Clearly, all high-throughput systems benefit from automation. Robots do not get bored with doing the same task day in and day out, and if set up correctly they do not make mistakes.

Automatic Data Accumulation and Analysis

Several providers of SNP analysis instrumentation also offer automatic data accumulation and interpretation with a large number of features. Bruker Daltonic has recently introduced a MALDI mass spectrometer for SNP genotyping (Autoflex™) backed up with the "Genotools SNP manager" software that will automatically record and analyze spectra [Pusch et al., 2001] (Billerica, MA, www.daltonics.bruker.com). A Twister (Zymark, www.zymark.com) can be coupled to the front end of the Autoflex™ for automatic target loading. This allows 24-hour stand alone operation. Due to the in-format analysis (the target plate has the same dimensions as a standard microtiter plate) robotic sample transfer is easy, and the data association is straight forward.

There are many suppliers of microarray scanners. Axon Instruments (Union City, CA, www.axon.com) offers a slide scanner called the GenePix that is supported by a sophisticated software package (GenePix Pro). Applied Precision (Attleboro, MA, www.api. com) offers the arrayWoRx. This system supports scanning 40 slides unattended. This is an interesting feature

in a microarray scanner that should be considered for high-throughput operation.

ORGANIZATION

One of the biggest challenges in running SNP genotyping at high-throughput is keeping track of samples and streamlining a production line. Repetitive and facile procedures have a marked advantage over ones where samples have to be transferred or formats have to be condensed and then deconvoluted. For example, if hundreds of PCRs have to be pooled for hybridization to one oligonucleotide microarray, a failure in one or several PCRs results in missing data points that can only be regenerated by repeating the entire experiment. As the cost of the individual experiments is very high, this becomes nearly prohibitive.

A laboratory information management system (LIMS) is an essential tool for tracking samples in high-throughput systems. With a low cost per analysis, inevitable failures or errors in some analysis can be corrected while the LIMS can seamlessly integrate the corrected data. Methods with software driven troubleshooting are of great interest. Most commercial systems are very inconclusive about the degree of implementation of quality control measures. Adequate review of LIMS is called for.

INTEGRATED SYSTEMS

Clearly a high-throughput SNP genotyping platform benefits from full integration, starting from DNA sample administration, to sample preparation, genotyping, data accumulation, data analysis, and LIMS tracking. A commercial system that covers all of this is currently not available.

Solutions are available that take care of sample preparation, and data accumulation and analysis. An example is the previously mentioned MassArray™ system of Sequenom, which is based on mass spectrometric analysis (San Diego, CA, www.sequenom. com). It consists of a robotically controlled automated process line that links instrument modules for streamlined processing of DNA samples including PCR amplification, enzymatic primer extension, conditioning for SNPs analysis, transfer of samples onto a SpectroChip™, and analysis by mass spectrometry backed up with appropriate software. It is designed for a throughput of 10,000 sample preparations per day. Orchid Bioscience has developed the SNPstream™ product line capable of generating 25,000 SNP genotypes per day (Princeton, NJ, www.orchidbio.com). The system is based on the proprietary SNP-IT primer extension SNP genotyping method. The system was developed using instruments of original equipment manufacturers. It is claimed that many different SNP genotyping protocols can be adapted to their system. LJL Biosystems/Molecular Devices (Sunnyvale, CA, www.ljlbio.uk.co or www.ljlbio.com) offers an automated fluorescent plate reader system for FP-TDI with an automatic plate loading system for 24-hour operation with complete up-stream robotic integration. The Allele Sorter software package provides allele calling. This system allows screening of several hundred thousand wells per day. Affymetrix offers the GeneChip® system which consists of high-density oligonucleotide arrays; a hybridization system; an array scanner; and software to analyze, manage, and mine the data (Santa Clara, CA, www.affymetrix. com). These microarrays are quite sophisticated in that they have built-in quality controls, but offer very little flexibility. A new microarray system with more flexibility was recently shown [Fan et al., 2000]. Unfortunately, automatic sample preparation and automatic microarray loading into the scanner that would allow SNP genotyping in thousands of individuals is not yet provided for.

Several biotech companies are starting to offer SNP genotyping as a service in addition to selling the instrumentation. This is partly precipitated by the high initial investments for many technologies.

COST OF SNP GENOTYPING

Often the cost of SNP genotyping by the different technologies is discussed, though comparison is very difficult if not impossible. The initial set-up cost of a method is of primary consideration. Automation is expensive, but results in staff cost savings. Repetitive tasks are more reliably executed by robots. Clearly, reagent cost is another consideration, but there are far more

factors that have to be drawn into the calculation. Even a "fully automated" system incurs staff costs, for example, for operation. How much effort is required for setting up a new SNP for genotyping? Flexibility, the failure rate (on first pass and ease of finishing), and accuracy are important. The usage of instrumentation and reagents dramatically affects costs. For example, with mass spectrometric techniques the initial investment is high. The set-up for genotyping a specific SNP involves manual labor. Thereafter, the cost per typed SNP dramatically sinks with the number of times this SNP is genotyped. Multiplexing can significantly increase throughput without significantly increasing the costs. Pooling approaches decrease the cost per SNP/DNA though individual results are lost. Should the cost of DNA extraction be factored into the cost of a genotype? Until now no comprehensive comparison of the cost of different SNP genotyping methods has been undertaken. It would be a difficult task to uncover hidden cost factors in each implementation, but would be of great merit.

Finally, the choice of an appropriate technology for the scientific question asked dramatically contributes to the cost effectiveness.

CAVEAT

There are several crucial considerations in SNP genotyping. First, is the quality of the DNA. The result obtained by any genotyping method is directly related to the DNA quality. The better the quality of the DNA, the better the result. However, it is possible to make protocols amenable to low quality DNA. SNP genotyping for agricultural applications does not allow elaborate DNA extraction procedure for cost reasons. Good results were obtained with Proteinase K digested tissue samples genotyped with the GOOD assay [Sauer et al., 2000b]. The lower the quality the more DNA generally is required.

A second problem with the assessment of SNP genotyping technologies is the first pass success rate. All technologies generate failures. It is important to be able to establish the identity and cause of these failures. Mass spectrometric techniques are very useful in this respect as the quality of reagents (primers, enzymatic activities, etc.) can be analyzed on the same platform as the final results.

A third consideration is the effort required for establishing an assay for an SNP. Oligonucleotide arrays are very unfavorable, as very early on in a project a decision has to be taken as to which SNPs should be genotyped. Due to the high set-up cost of a study, it is difficult to add a new SNP to a study. An assay that fails on the array is difficult to replace or substitute. Formats like the electrostringent arrays and coded spheres do not suffer from these limitations, yet still retain the advantage of parallel processing of oligonucleotide arrays.

A fourth consideration is the cost of failure of an SNP assay during optimization. Modifications of oligonucleotides, like biotinylation that is required in pyrosequencing or the PROBE assay, or complex fluorescent dyes that are used for the TaqMan assay or molecular beacons, make redesign of assays very costly.

A fifth consideration is the stringency of an interrogation method. Intrinsic methods using an enzymatic preparation of allele-specific products (like primer extension, ligation, or cleavage) make use of nature's ability of selectively generating or cleaving specific DNA structures. All of these methods use hybridization as the initial event and then let a specific enzyme process the allele. Therefore, an enzymatic processing step provides more fidelity then pure hybridization, though with a concomitant increase in cost due to the enzymes.

One approach to achieve high-throughput is genotyping of pooled DNA samples [Collins, 2000]. The identification of disease causing variants of SNPs can be achieved this way by analyzing pools of cases and pools of controls separately. The concentrations and quality of the used DNAs have to be very precisely quantified and the analysis technique has to have good means of quantification. A method using PCR and ligation with fluorescence detection has shown to resolve pools of 500 individuals [Khanna et al., 1999]. This provides the possibility of identifying rare alleles or small changes in allele frequencies. Kinetic quantitative PCR has been demonstrated for the determination of SNP allele frequencies [Germer et al., 2000].

Pyrosequencing provides excellent signal/noise, which makes it well suited to quantification. Quantification of SNP genotyping with MALDI detection gives less resolution, but coupled with the throughput of this technology could be very efficient [Ross et al., 2000]. These approaches, while still in experimental stages, are very promising, and applications remain to be shown.

MINIATURIZATION

Miniaturization is widely considered the future of SNP genotyping in terms of decreasing cost and facilitating automation. There are efforts to place genotyping methods onto a microfluidic system [Kalinina et al., 1997; Anderson et al., 2000; Pang and Yeung, 2000; Schmalzing et al., 2000]. There are limitations to this. The delivery of reagents to a microfluidic system is significantly more difficult then to a macroscopic system. Genomic DNA is the starting material for SNP genotyping. In the macroscopic world (in low microliter reaction volumes) SNP genotyping reactions can be done with as little as 0.5–1.0 ng of genomic DNA. This equates to 150–300 copies of the human genome. In heterozygous samples, an allele might thus be represented by as few as 75 copies. Reducing this too far obviously results in allele loss due to statistical factors. A starting template ratio of alleles of 2:1 can result in a difference of 10 after amplification with an "exponential" process such as PCR, which leads to errors. Also, enzymes can be lost in miniaturized processes due to dilution. Another problem with miniaturized systems is shearing on the system walls. Reagents (especially DNA and enzymes) can be destroyed by it. This is less critical for the DNA than for enzymes as the latter lose their activity. Caliper Technologies (Mountain View, CA, www.calipertech.com), together with Agilent Technologies (Palo Alto, CA, www.agilent.com), has developed a lab-on-a-chip system. This can be applied to size DNA products. It has not yet been shown for genotyping SNPs.

SNP genotyping using coded beads beautifully avoids drawbacks of miniaturization [Iannone et al., 2000; Cai et al., 2000; Chen et al., 2000]. The reaction volume of a single SNP in a 10 μl reaction with 100 different spheres is only 200 nl. Liquid handling in the 10 μl range is reliable in most robotic systems.

OUTLOOK

Today, a fully functional and truly integrated SNP genotyping system is not commercially available. Manufacturers for the different SNP genotyping methods and analyzers are starting to provide automated solutions for sample preparation together with an SNP genotyping method. High-throughput SNP genotyping still has to be tailored to the needs of the individual operators. In most cases, end-users themselves have made significant efforts toward the integration of their systems.

The future demands ever cheaper and higher throughput SNP technologies to make use of the hundreds of thousands of SNPs that are currently being discovered for whole genome association studies. Clearly, the thrust of SNP genotyping technologies is to deliver a large number of genotypes for a large number of individuals. Methods providing a limited number of genotypes per preparation, like the mass spectrometric techniques, will be developed toward an increase in genotypes per preparation. Technologies that are suited for the analysis of a large number of SNPs in one individual, like the microarrays, will need to be directed toward the analysis of large numbers of individuals, which could be achieved by suitable automation of the sample preparation. One should see new innovative approaches that are capable of providing large numbers of genotypes in large numbers of individuals emerge.

In many respects, solutions to high-throughput SNP genotyping are not as well developed as one might wish, considering the status of the genome sequencing projects.

Reviewing SNP genotyping and automation thereof can not, unfortunately, be done without mentioning dozens of companies. Their products are usually not well described in accessible scientific terms. An effort was made to provide up-to-date information and show problems encountered in high-throughput SNP genotyping. Of course, this is in no way complete and several providers may have not been reviewed. An effort was made to highlight crucial points of this rapidly evolving field.

ACKNOWLEDGMENTS

I thank Jacqui Beckmann, Sascha Sauer, and Noah Christian for reading the manuscript and making valuable suggestions.

REFERENCES

Anderson RC, Su X, Bogdan GJ, Fenton J. 2000. A miniature integrated device for automated multistep genetic assays. Nucleic Acids Res (methods online) 28:e60.

Barany F. 1991. Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc Natl Acad Sci USA 88:189–193.

Baron H, Fung S, Aydin A, Bahring S, Luft FC, Schuster H. 1996. Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. Nat Biotechnol 14:1279–1282.

Brookes AJ, Kwok P-Y, White PS, Oefner PJ. 2001. SNiPpets from the Third International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis, September 8–11, 2000, Taos, New Mexico, USA. Hum Mutat 17:241–242.

Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. Nat Genet 21(suppl): 33–37.

Butler JM, Jiang-Baucom P, Huang M, Belgrader P, Girard J. 1996. Peptide nucleic acid characterization by MALDI-TOF mass spectrometry. Anal Chem 68:3283–3287.

Cai H, White PS, Torney D, Deshpande A, Wang Z, Marrone B, Nolan JP. 2000. Flow cytometry-base minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. Genomics 66:135–143.

Chen J, Iannone MA, Li M-S, Taylor D, Rivers P, Nelsen AJ, Slentz-Kesler KA, Roses A, Weiner MP. 2000. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. Genome Res 10:549–557.

Chen X, Kwok P-Y. 1997. Template-directed dye-terminator incorporation (TDI) assay: a homogeneous DNA diagnostic method based on fluorescence resonance energy transfer. Nucleic Acids Res 25:347–353.

Chen X, Livak KJ, Kwok PY. 1998. A homogeneous ligase-mediated DNA diagnostic test. Genome Res 8:549–556.

Chen X, Zehnbauer B, Girnke A, Kwok P-Y. 1997. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. Proc Natl Acad Sci USA 94:10756–10761.

Chen X, Levine L, Kwok PY. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9:492–498.

Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. 1999. Making and reading microarrays. Nat Genet 21(suppl):15–19.

Collins F. 2000. Keynote lecture at the Cold Spring Harbor meeting of genome sequencing and biology. Cold Spring Harbor, NY, USA, May 10–14.

Day DJ, Speiser PW, White PC, Barany F. 1995. Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. Genomics 29:152–162.

Day INM, Humphries SE. 1994. Electrophoresis for genotyping: microtitre array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. Anal Biochem 222:389–395.

Debouck C, Goodfellow PN. 1999. DNA microarrays in drug discovery and development. Nat Genet 21(suppl):48–50.

DeFrancesco L. 1998. The next new wave in genome analysis: Invader™ assays developed by Third Wave Technologies Inc. The Scientist 12:16.

Edman CF, Raymond DE, Wu DJ, Tu E, Sosnowki RG, Butler WF, Nerenberg M, Heller MJ. 1997. Electric field directed nucleic acid hybridization on microchips. Nucleic Acids Res 25:4907–4914.

Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Norden B, Nielsen PE. 1993. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen bonding rules. Nature 365:566–568.

Fan J-B, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. 2000. Parallel genotying of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res 10:853–860.

Genome International Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Germer S, Holland MJ, Higuchi R. 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. Genome Res 10:258–266.

Griffin TJ, Tang W, Smith LM. 1997. Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. Nat Biotech 15:1368–1372.

Griffin TJ, Hall JG, Prudent JR, Smith LM. 1999. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. Proc Natl Acad Sci USA 96:6301.

Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S, Iovannisci DA, Woo S, Winn-Deen ES. 1994. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res 22:4527–4534.

Haff L, Smirnov IP. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 7:378–388.

Harrington JJ, Lieber MR. 1994. Functional domains within FEN-1 and RAD2 define a family of structure specific endonucleases: implications for nucleotide excision repair. Genes Dev 8:1344–1355.

Healey BG, Matson RS, Walt DR. 1997. Fiberoptic DNA sensor array capable of detecting point mutations. Anal Chem 251:270–279.

Holland PM, Abramson RD, Watson R, Gelfand DH. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci USA 88:7276–7280.

Iannone MA, Taylor JD, Chen J, Li M-S, Rivers P, Slentz KA, Weiner MP. 2000. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. Cytometry 39:131–140.

Kalinina O, Lebedeva I, Brown J, Silver J. 1997. Nanoliter scale PCR with TaqMan detection. Nucleic Acids Res 25:1999–2004.

Khanna M, Park P, Zirvi M, Cao W, Picon A, Day J, Paty P, Barany F. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. Oncogene 18:27–38.

Kruglyak L. 1997. The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17:21–24.

Kruglyak L. 1999. Prospect for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144.

Kwok P-Y. 1998. Genotyping by mass spectrometry takes flight. Nat Biotechnol 16:1314–1315.

Landegren U, Samiotaki M, Nilsson M, Malmgren H, Kwiatkowski M. 1996. Detecting genes with ligases. Methods (a companion to Methods Enzymol) 9:84–90.

Landegren U, Nilson M, Kwok P-Y. 1998. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res 8:769–776.

Li J, Butler JM, Tan Y, Lin H, Royer S, Ohler L, Shaler TA, Hunter JM, Pollart DJ, Montforte JA, Becker CA. 1999. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. Electrophoresis 20:1258–1265.

Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. Nat Genet 21(suppl):20–24.

Little DP, Braun A, Darnhofer-Demar B, Frilling A, Li Y, McIver RT, Köster H. 1997 Detection of RET proto-oncogene codon 634 mutations using mass spectrometry. J Mol Med 75:745–750.

Livak K, Marmaro J, Todd JA. 1995. Towards fully automated genome-wide polymorphism screening. Nat Genet 9:341–342.

Maier E, Lehrach H. 1997. Neue Techniken zur Automatisierung in der DNA-Sequenzierung. Chemie in unserer Zeit 31:66–75.

McCarthy JJ, Hilfiker R. 2000. The use of single nucleotide polymorphism maps in pharmacogenomics. Nat Biotechnol 18:505–508.

Mei R, Galipeau PC, Prass C, Berno A, Ghandour G, Patil N, Wolff RK, Chee MS, Reid BJ, Lockhart DJ. 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res 10: 1126–1137.

Mein CA, Barrat BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Phillips MS, Jarvis LS, deArruda M, Todd JA. 2000. Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. Genome Res 10:330–343.

Meldrum D. 2000a. Automation for genomics, part one: preparation for sequencing. Genome Res 10:1081–1092.

Meldrum D. 2000b. Automation for genomics, part two: sequencers, microarrays, and future trends. Genome Res 10:1288–1303.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503–2516.

Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 265:2085–2088.

Nilsson M, Krejci K, Koch J, Kwiatkowski M, Gustavsson P, Landegren U. 1997. Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. Nat Genet 16:252–255.

Pang H-M, Yeung ES. 2000. Automated one-step DNA sequencing based on nanoliter reaction volumes and capillary electrophoresis. Nucleic Acids Res (methods online) 28:e73.

Parsons BL, Heflich RH. 1997. Genotypic selection methods for the direct analysis of point mutations. Mutat Res 387:97–121.

Pastinen T, Partanen J, Syvänen A-C. 1996. Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. Clin Chem 42:1391–1397.

Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvänen A-C. 1997. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 7:606–614.

Pastinen T, Raito M, Lindroos K, Tainola P, Peltonen L, Syvänen A-C. 2000. A system for specific, high-through-

put genotyping by allele-specific primer extension on microarrays. Genome Res 10:1031–1042.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes C, Fodor SPA. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 91:5022–5026.

Pusch W, Kraeuter K-O, Froehlich T, Staalgies Y, Kostrzewa M. 2001. The GenoTool SNP manager: a new software for automated high-throughput MALDI-TOF MS SNP genotyping. Biotechniques 30:210–215.

Ramsay G. 1998. DNA chips: state-of-the-art. Nat Biotech 16:40–44.

Roest Crollius H, Jaillon O, Bernot A, Dasilva C, Bouneau L, Fischer C, Fizames C, Winckler P, Brottier P, Quetier F, Saurin W, Weissenbach J. 2000. Estimate of human gene number provided by genome-wide analysis using Tetraodon negroviridis DNA sequence. Nat Genet 25: 235–238.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242:84–89.

Ronaghi M, Nygren M, Lundeberg J, Nyren P. 1999. Analyses of secondary structures in DNA by pyrosequencing. Anal Biochem 267:65–71.

Ross P, Hall L, Smirnov I, Haff L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16:1347–1351.

Ross P, Hall L, Haff LA. 2000. Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. BioTechniques 29:620–629.

Ryan D, Nuccie B, Arvan D. 1999. Non-PCR dependent detection of factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. Mol Diag 4:135–144.

Samiotaki M, Kwiatkowski M, Parik J, Landegren U. 1994. Dual-color detection of DNA sequence variants by ligase-mediated analysis. Genomics 20:238–242.

Sauer S, Lechner D, Berlin K, Lehrach H, Escary J-L, Fox N, Gut IG. 2000a. A novel procedure for efficient genotyping of single nucleotide polymorphisms. Nucleic Acids Res (methods online) 28:e13.

Sauer S, Lechner D, Berlin K, Plançon C, Heuermann A, Lehrach H, Gut IG. 2000b. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. Nucleic Acids Res (methods online) 28:e100.

Schmalzing D, Belenky A, Novotny MA, Koutny L, Salas-Solano O, El-Difrawy S, Adourian P, Ehrlich D. 2000. Microchip electrophoresis: a method for high-speed SNP detection. Nucleic Acids Res (methods online) 28:e43.

Shalon D, Smith SJ, Brown PO. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6:639–645.

Shchepinov MS, Chalk R, Southern EM. 1999. Trityl mass-tags for encoding in combinatorial oligonucleotide synthesis. Nucleic Acids Symp Ser 42:107–108.

Shi MM, Bleavins MR, de la Iglesia FA. 1999. Technologies for detecting genetic polymorphisms in pharmacogenomics. Mol Diagn 4:343–351.

Shuber AP, Michalowsky LA, Nass GS, Skoletsky J, Hire LM, Kotsopoulos SK, Phipps MF, Barberio DM, Klinger KW. 1997. High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. Hum Mol Genet 6:337–347.

Shumaker JM, Metspalu A, Caskey CT. 1996. Mutation detection by solid phase primer extension. Hum Mutat 7:346–354.

Sosnowski RG, Tu E, Butler WF, O'Connell JP, Heller MJ. 1997. Rapid determination of single base mismatch mutations in DNA hybrids by electric field control. Proc Natl Acad Sci USA 94:1119–1123.

Syvänen A-C, Sajantila A, Lukka M. 1993. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. Am J Hum Genet 52:46–59.

Syvänen A-C. 1994. Detection of point mutations in human genes by the solid-phase minisequencing method. Clin Chim Acta 226:225–236.

Syvänen A-C. 1999. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 13:1–10.

Tyagi S, Kramer FR. 1996. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14:303–308.

Tyagi S, Bratu DP, Kramer FR. 1998. Multicolor molecular beacons for allele discrimination. Nat Biotechnol 16:49–53.

Venter JC, et al. 2001. The sequence of the human genome. Science 291:1304–1351.

Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082.