



Single Nucleotide Polymorphisms: Discovery, Detection and Analysis

Mostafa Ronaghi & Taimour Langaee

To cite this article: Mostafa Ronaghi & Taimour Langaee (2005) Single Nucleotide Polymorphisms: Discovery, Detection and Analysis, Personalized Medicine, 2:2, 111-125, DOI: [10.1517/17410541.2.2.111](https://doi.org/10.1517/17410541.2.2.111)

To link to this article: <https://doi.org/10.1517/17410541.2.2.111>



Published online: 24 Nov 2005.



Submit your article to this journal [↗](#)



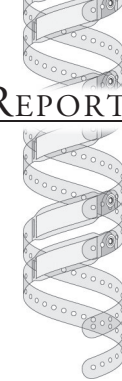
Article views: 375



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)



Single nucleotide polymorphisms: discovery, detection and analysis

Mostafa Ronaghi^{1,2} &
Taimour Langaee^{3†}

[†]Author for correspondence

¹Stanford University, Stanford
Genome Technology Center,
Palo Alto, CA, USA

²Tehran University,
Institute for Biochemistry and
Biophysics,
Tehran, Iran

³University of Florida,
Department of Pharmacy
Practice,
Center for Pharmacogenomics,
Gainesville, FL,
USA

E-mail: langaee@cop.ufl.edu

Single nucleotide polymorphisms (SNPs) are the most abundant and simple form of DNA variation. Analyses of SNPs in the human population have the potential to greatly improve human health, both by predicting susceptibility to disease and guiding choice of therapy. This review describes new tools for SNP discovery, and current and emerging technologies for large-scale SNP analysis, as well as providing a guide to choosing the best approach for SNP analysis.

There are several types of variations at the genome level in individuals. The ultimate goal is to drive the cost of genome sequencing down to a US\$1000 level in order to map all types of variations in a genome. However, this might take more than a decade and several strides need to be taken before reaching this goal. Analyses of the most common type of genetic variation, single nucleotide polymorphisms (SNPs), provide a significant amount of information about an individual. There are approximately 10–20 million SNPs occurring commonly in the human genome. The short-term objective is to discover the most common SNPs and analyze the most informative ones for comprehensive genetic analyses. A series of performance metrics are now commonly applied to measure the quality of genotyping data generated by different techniques. ‘Conversion rate’ is a measure of the SNPs in the genome that can be assayed, and is a function of both the quality of the SNPs chosen and the technology used to score them. In addition, some SNPs can be rejected *in silico*, prior to probe synthesis. The ‘call rate’ for a given marker is the percentage of DNA samples in a study whose genotype is successfully measured. For a set of converted SNPs, the percentage of total genotypes returned across all markers is defined as the ‘completeness’ of the study. Finally, ‘accuracy’ is the percentage of the genotypes that are correct. This review focuses on the major technologies currently in use or under development for comprehensive genomic analyses (it is not the paper’s purpose to cite all the available technologies).

SNP discovery

A key area of investigation in genetics is discovery of novel SNPs. The nature of previously discovered SNPs has thus far been determined

by DNA sequencing. Comparative genomic and genetic [1–3] tools have been used for large-scale extraction of SNPs. Currently, more than 10 million SNPs have been identified covering the most common polymorphisms. The first step toward understanding the nature of a variation is DNA sequencing. Here, the current and future methodologies that are impacting the way we discover and analyze these single variations shall be reviewed.

DNA sequencing approaches

Since the invention of robust DNA sequencing less than three decades ago, the technique has revolutionized the way we study biology. In the first two decades, DNA sequencing cost was reduced by three orders of magnitude, and that enabled the sequencing of the human genome. Currently, the cost of sequencing a mammalian genome is approximately US\$5–10 million (according to RFA-HG-04-002 released by the US National Institutes of Health [NIH]). The cost should be further reduced to find population-specific polymorphisms and, subsequently, to enable the profiling of individual genomes. Integration and miniaturization are the most common efforts in cost reduction of current sequencing technologies. Several different technologies have the potential to make a major impact in the realization of individual genome profiling.

Miniaturization of conventional DNA sequencing

Conventional DNA sequencing consists of a multistep process to produce DNA fragments, each differing by a single nucleotide, which are normally then analyzed by capillary electrophoresis [4]. Although more than 200 genomes have thus far been sequenced, this technique

Keywords: analysis, assay,
DNA sequencing, genome,
genotyping, multiplex, SNP
discovery, throughput

future
medicine

faces limitations in terms of cost, time and throughput for future applications. Retaining the advantages of current DNA sequencing methods (i.e., well-developed community infrastructure, commercial availability of reagents and existing analysis software) Microchip Biotechnologies, Inc. [201] is developing a completely automated, bench-top system based on the Sanger method for sequencing. The heart of the system will be a microchip-based device that can label and process DNA fragments from individual microbeads in low-volume reactions, followed by ultra-fast separation and analysis on microfabricated capillary electrophoresis channels. This technique may reduce the cost of Sanger DNA sequencing to approximately US\$100,000 for individual genomes.

Emerging DNA sequencing techniques

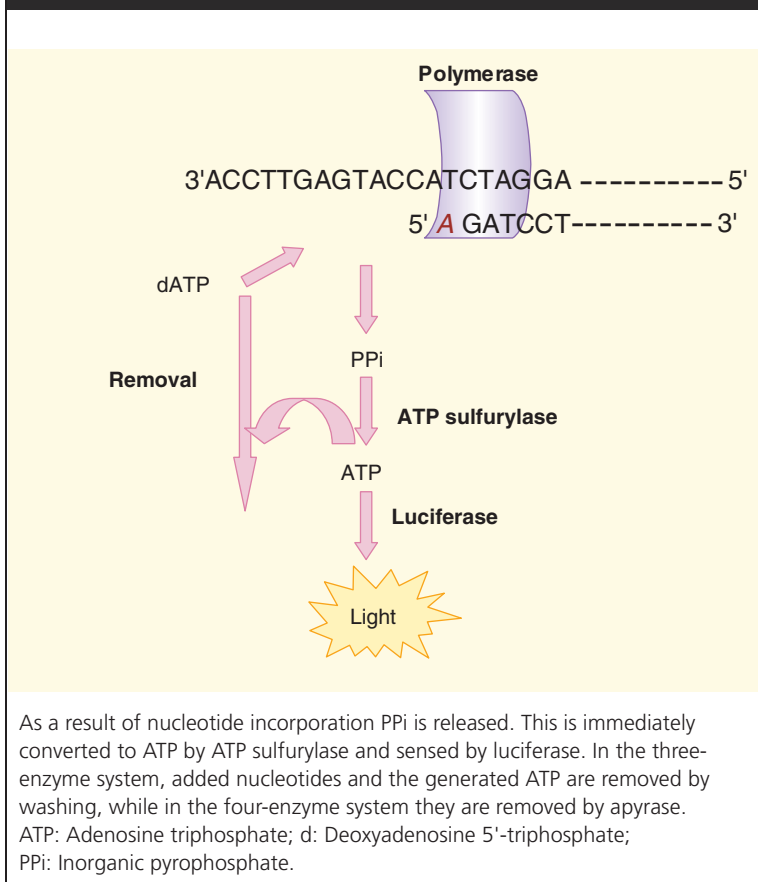
Hybridization-based technique

The hybridization-based technique, sequencing by hybridization (SBH), was proposed by several groups more than 15 years ago [5–7,101]. This technology is still under development for producing robust *de novo* data for genome-scale sequencing. The basic principle of SBH involves the use of differential hybridization of oligonucleotide probes to decode a target DNA sequence. The most common approach involves immobilizing the DNA to be sequenced on a solid support and then performing serial hybridizations with short probe oligonucleotides (5–6 nucleotides long). The extent to which specific probes bind can be used to infer the unknown sequence. The strategy has been applied to both genome resequencing and *de novo* sequencing [6,8]. Affymetrix and its spin-off Perlegen have pioneered a different approach to SBH by hybridizing sample DNA to high-density oligonucleotide arrays [9]. The current maximum density of Affymetrix arrays is about one oligonucleotide feature per 5 micron square, where each feature consists of roughly 100,000 copies of a defined 25 base pair (bp) oligonucleotide. For each bp of a reference genome to be resequenced, there are four features on the chip. By hybridizing labeled sample DNA to the chip and determining which of the four features yields the strongest signal for each bp in the reference sequence, a DNA sample can be rapidly resequenced. Miniaturization, bioinformatics and the availability of a reference human genome sequence has permitted Perlegen to greatly extend this approach and develop an oligonucleotide array for resequencing the human chromosome 21 [10], and more recently the whole genome [11]. SBH technology

possesses a unique set of advantages and challenges. It can be used to obtain an impressive amount of sequence (> 10⁹ bases) from many distinct chromosomes. The primary challenge that SBH will face is the cross-hybridization of probes to incorrect targets due to repetitive elements or chance similarities. These factors render a substantial fraction of chromosome 21 (> 50%) inaccessible [10], and might also contribute to the 3% false-positive SNP detection rate observed in that study. It is also worth noting that SBH still requires sample preparation steps, as the relevant fraction of the genome must be PCR-amplified prior to hybridization. Despite the low conversion rate achieved by this technique as a result of it being able to detect most of the polymorphisms, this technology provides the most inexpensive means of polymorphism discovery. As a part of the HapMap project, Perlegen is going to release more than four million SNPs (using its resequencing technology) in the coming year, in addition to the original one million released by the Hapmap community [202]. These SNPs will provide a comprehensive means of genome scanning with the Affymetrix 500,000 chip for SNP genotyping, which is being released.

Pyrosequencing

Pyrosequencing relies on the real-time detection of inorganic pyrophosphate (PPi) released on successful incorporation of nucleotides during DNA synthesis. PPi is immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of ATP generated is detected via luciferase-producing photons. Unused ATP and deoxynucleotides are washed in the three-enzyme system of pyrosequencing [12] and degraded in the four-enzyme system of pyrosequencing [13] by the nucleotide-degrading enzyme apyrase (Figure 1). The presence or absence of PPi, and therefore the incorporation or nonincorporation of each nucleotide added, is ultimately assessed on the basis of whether or not photons are detected. 454 Life Sciences has miniaturized the three-enzyme system of pyrosequencing, enabling more than hundreds of thousands sequencing to be performed on a single chip. This revolutionary technology is 100 times faster than the conventional capillary-based sequencing and is able to deliver more than 20 million bases of high-quality DNA sequence in only 4 h, with close to 20 times less the cost of terminator-based capillary sequencing [203]. An integrated platform for inexpensive genome sequencing using pyrosequencing is

Figure 1. Schematic presentation of the enzyme reaction in pyrosequencing.

under development using immobilized DNA on beads that have been isolated in microwells equipped with a complementary metal oxide semiconductor (CMOS) image sensor [14]. This technique may reduce the cost of genome sequencing to a few thousand dollars.

Proposed DNA sequencing techniques

Other sequencing-by-synthesis techniques

Sequencing-by-synthesis (SBS) methods generally involve enzymatic extension by polymerase through the iterative addition of nucleotides. Each nucleotide addition, called a cycle, only queries one or a few bases, but thousands or millions of features can be processed in parallel when the assay is performed in array format. These array features may be ordered or randomly dispersed. The progression of the primer strand can be monitored either by fluorescence using labeled nucleotides, or enzymatically using natural nucleotides such as in pyrosequencing [15]. Several groups reported on SBS using fluorescent nucleotides more than a decade ago [16,17,102]. Other groups have made advances in microfabri-

cation and integration of SBS using fluorescent nucleotides. Most notable is the work performed by Church and coworkers [18,19]. All of the above-mentioned SBS methods require an amplification step prior to sequencing, in order to achieve the required signal fidelity. Several companies have worked on developing single-molecule DNA sequencing techniques based on SBS, most notably Solexa [204], Genovox [205], Nanofluidics [20,206], and Helicos [21,207]. A method for sequencing single molecules of DNA directly would eliminate the need for costly and often cumbersome procedures, such as cloning and PCR amplification, and also avoid the problem of asynchrony during extension. Although SBS on single molecules may potentially decrease the costs to US\$1000 per genome, so far there has been limited published data by these companies to support it (Solexa, Genovox, Nanofluidics, and Helicos were started in 1998, 2000, 2003, 2003, respectively). There are two other companies, Visigen [208] and Li-Cor [103,209], attempting to engineer DNA polymerases or fluorescent nucleotides to provide real-time, base-specific signals, while synthesizing DNA at its natural pace.

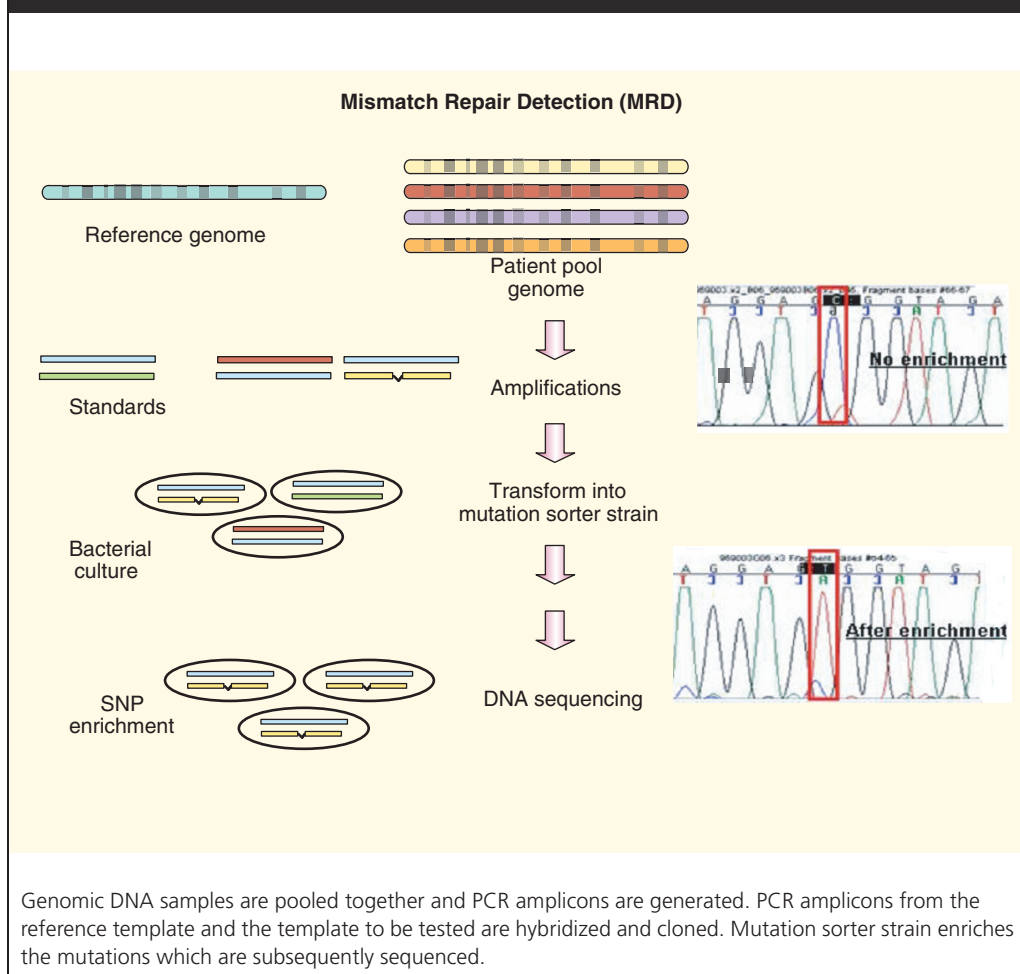
Non-enzymatic single-molecule DNA sequencing

Another single-molecule approach is nanopore sequencing [22–24]. As DNA passes through a nanopore, different bps obstruct the pore to varying degrees, resulting in fluctuations in the electrical conductance of the pore. The pore conductance can therefore be measured and used to infer the DNA sequence. The accuracy of base calling ranges from 60% for single events to 99.9% for 15 events [25]. However, the method has thus far been limited to the terminal bps of a specific type of hairpin. This method has a great deal of long-term potential for extraordinarily rapid sequencing with little to no sample preparation. However, it is likely that significant pore engineering will be necessary before a single-base resolution can be achieved.

Other SNP discovery techniques coupled with DNA sequencing

In vivo mismatch repair detection for variation scanning

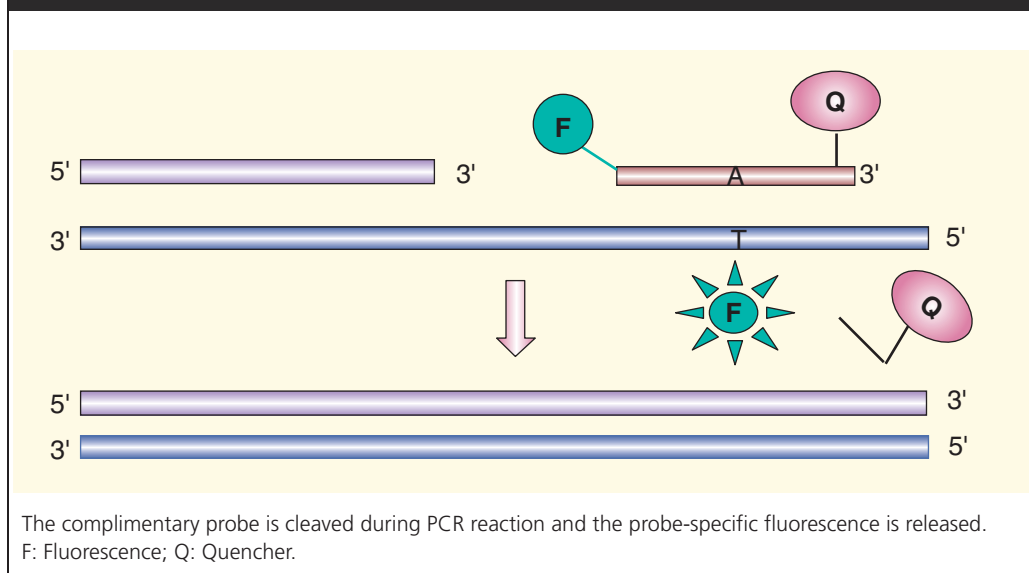
Recently, a new technique called *in vivo* mismatch repair detection (MRD) [26] was developed to enrich for variant alleles (Figure 2). MRD detects variants utilizing the mismatch repair system of *Escherichia coli* [210]. A specific strain is engineered to sort a mixture of transformed fragments into two pools: those carrying a variation

Figure 2. A schematic of the MRD SNP discovery process.

and those that do not. MRD has been described before as a method for multiplex variation scanning [27]. This technique has proven to be an excellent enrichment method for dideoxy terminator sequencing, in order to discover rare and common variant alleles in normal and disease populations. The cost of deep sequencing has been reduced 70-fold. In addition, rare mutations and SNPs can be identified, as each bacterial colony represents only one allele. In this way, homozygous sequencing traces are obtained. One of the main advantages of using MRD for finding genes associated with disease is the ability to enrich the genes containing rare mutations, rare SNPs, common mutations and common SNPs, as they will appear in individual colonies on the MRD variant plate. This is the unique feature of this technique, which is not possible with standard sequencing, as these mutations and SNPs are often missed in single-pass DNA sequencing without enriching the DNA fragment with MRD.

High-throughput mass spectrometry for SNP discovery

Recently, a comparative sequencing strategy based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was described that enables analyses of complete base-specific cleavage reactions of a target sequence [28,29]. In this technique, the target is converted to a DNA/RNA mosaic structure after PCR amplification using *in vitro* transcription. Cleavage with defined specificity is achieved by ribonucleases. The set of cleavage products is subjected to mass spectrometry without prior fractionation. The presented resequencing assay is particularly useful for SNP discovery. Using this technique in combination with mass spectra from four complementary cleavage reactions, 98% of all possible homozygous and heterozygous SNPs in target sequences, with a length of up to 500 bases, have been detected. The throughput of this technique is comparable with conventional

Figure 3. Schematic representation of a Taqman® assay.

DNA sequencing, while the read-length is much shorter and the accuracy of this technique in large-scale studies has not yet been determined.

Denaturing high-performance liquid chromatography

Denaturing high-performance liquid chromatography (DHPLC) is based on the formation of homoduplex and heteroduplex products in the absence and presence of differences in DNA sequence and analysis by ion-pair reversed-phase liquid chromatography, under partially denaturing conditions [30]. A mixture of two or three different DNA segments from unpurified PCR reactions is heated to cause them to denature, and then cooled gradually to reanneal to form homo- and heteroduplexes. The PCR mixture is injected into a DNA column and then eluted by a linear acetonitrile gradient. Based on the column retention times, the heteroduplexes (variant DNA fragments), which are less stable, elute faster than homoduplexes (identical DNA fragments). Both hetero- and homoduplexes are detected by an ultraviolet absorbance detector at 254 nm, and the resulting chromatograms are compared for mutation detection [31,32]. Despite the fact that this technology enables more accurate polymorphism discovery, it is laborious and not applicable for large-scale projects. Including labor costs, it will be cheaper and faster to perform brute-force sequencing.

SNP genotyping techniques

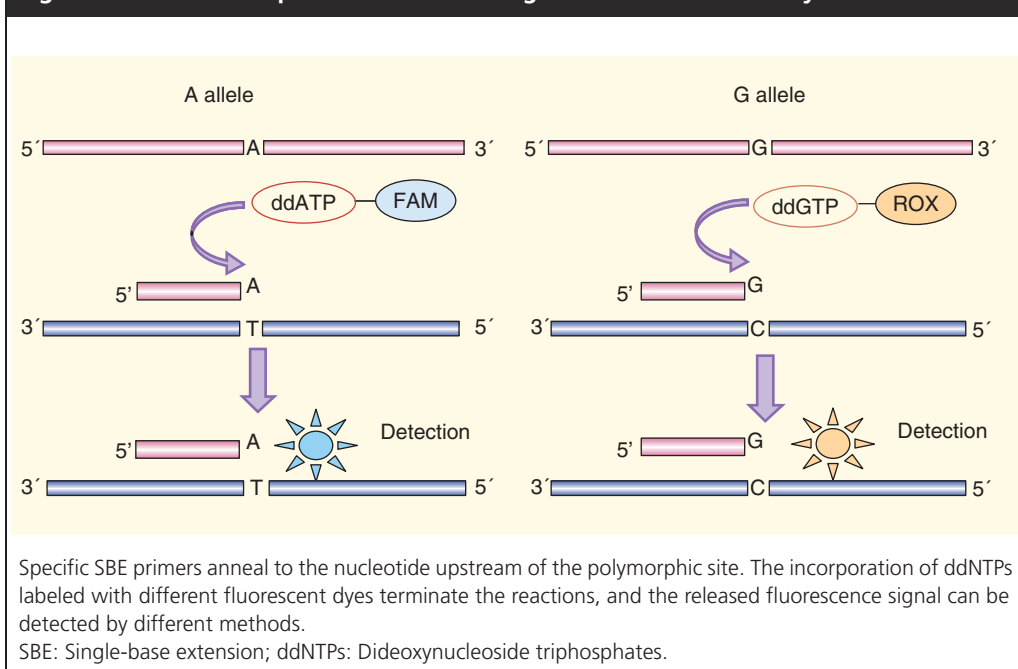
Overall, SNP genotyping techniques can be divided into two major categories: singleplex to

lowplex, and highly multiplexed technologies. Major singleplex-to-lowplex technologies include: pyrosequencing [33], single-base extension techniques [34], mass spectrometry-based techniques [35,36], Invader assay [37,38], and Taqman assay [39–41]. All these techniques, except Invader technology, require amplification prior to genotyping. More advanced technologies, which have allowed highly multiplex assay include: molecular inversion probe (MIP) assay [42,43], and oligonucleotide ligation assay (OLA) [44]. These techniques circumvent the amplification step prior to genotyping, enabling a high level of multiplexing. Whole-genome sampling assay (WGSa) was also recently developed, allowing highly multiplex SNP analysis [45,46]. A modification of the latter technique was recently introduced that enables custom genotyping, called multiplexed anchored runoff amplification (MARA) [47].

Low-to-medium throughput genotyping techniques

TaqMan assay

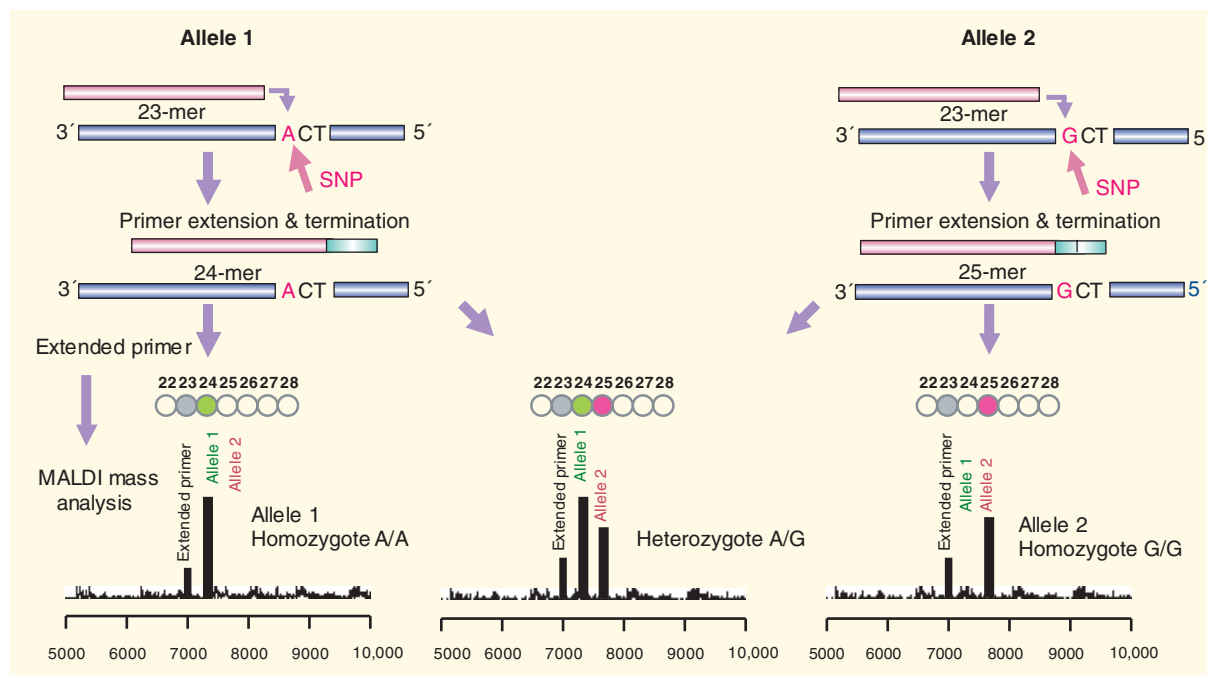
The TaqMan® assay is a microtiter plate-based fluorescent read-out system that uses an allele-specific hybridization method to discriminate between alleles (Figure 3). This assay was developed based on the 5' nuclease activity of Taq DNA polymerase, which cleaves the oligonucleotides hybridized to the DNA template during replication [39–41]. Taqman is considered as a simple, homogenous, and single-step genotyping assay. For SNP genotyping, one pair of TaqMan probes and one pair of PCR primers

Figure 4. Schematic representation of a single-base extension assay.

are used. Two TaqMan probes differ at the polymorphic site, with one probe complementary to the wild-type allele and the other to the variant allele. A 5' reporter dye and a 3' quencher dye are covalently linked to the wild-type or variant allele probes. The probe that perfectly matches the DNA target is cleaved by Taq DNA polymerase, resulting in the release of the reporter dye from quencher dye and an increase in reporter fluorescence intensity. Specific genotyping is determined by measuring the signal intensity of the two different reporter dyes after the PCR reaction. In addition to detecting SNPs, small gene deletions and insertions can also be identified by this method [48]. The TaqMan probes that were modified with minor groove-binding molecules increased the affinity for the target and enhanced allele discrimination rate [49]. Since the target amplification and genotype determination are both carried out into a single step, this alleviates the need to process and purify PCR product and reduces the time for genotyping assay. The main limitations of this method are the high cost of the fluorescence-labeled probes, the requirement to design and optimize probes for each specific polymorphism, and limited throughput. This technology is mostly suitable when large number of samples need to be genotyped for a limited number of polymorphisms.

Single-base extension assay

Single-base extension assay (SBE) is a simple method that can be adapted to different detection formats (Figure 4). SBE assay, also known as mini-sequencing, is based on the DNA synthesis activity where a DNA strand or primer is extended one bp at the 3' end. If the 3' base matches and pairs with DNA, the DNA polymerase will add a new base to the strand. When the extending base does not carry the 3'-hydroxyl group (3'-OH) no more base can be added and the DNA synthesis terminates [4]. Similar to regular DNA sequencing, mini-sequencing is an efficient way to detect SNPs through the addition of specific nucleotides to a single primer [34]. Several SNP markers can be analyzed in parallel by the use of locus-specific primers and analyzing the allele-specific incorporation of labeled nucleotides. To increase the throughput, new sets of reaction products can be loaded onto the same gel. Recently a modified SBE assay, called arrayed SBE, was developed at Stanford Genome Center, which has been optimized to do parallel SNP genotyping of many samples at much lower cost [50]. Overall, gel-based genotyping assays are relatively straightforward and are useful when dealing with a small number of samples. The methods are labor-intensive and require experienced and skilled technical staff for final analysis.

Figure 5. Schematic representation of a Sequenom MassExtend assay.

The hME assay from SEQUENOM® [211] uses a beadless, label-free primer extension chemistry to generate allele-specific products. In a series of reactions the amplified DNA template is dephosphorylated, then hME primer extension reaction generates allele-specific primer extension products 1–4 bases longer than the original ME primer. The extended primers are then cleaned by resin and transferred into the MALDI-TOF for mass analysis, and the genotypes are determined by a MassARRAY™ RT software.

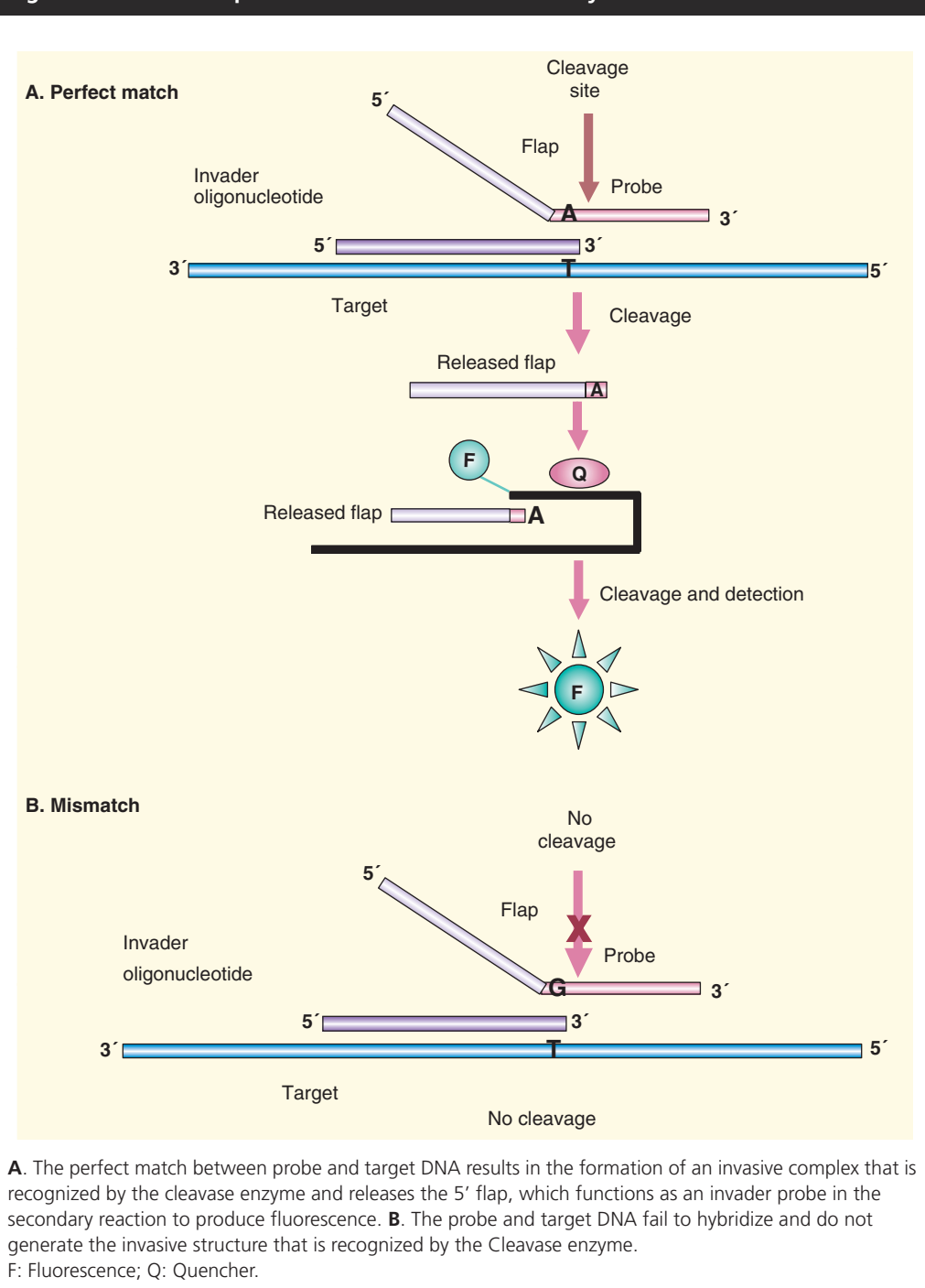
hME: Homogenous MassEXTEND™; MALDI-TOF: Matrix-assisted laser desorption ionization with time-of-flight detectors; SNP: Single nucleotide polymorphism.

One popular alternative detection system for SBE is MS (Figure 5). The most common form of MS for genotyping is MALDI-TOF. MS yields precise information on the molecular mass of the DNA fragments. The procedure can be fully automated, and both DNA strands can be analyzed in parallel [35,36]. Unlike the fluorescent genotyping methods described above, MS offers specificity and accuracy without requiring specially labeled probes or primers. One disadvantage of this technique is that it requires purified samples free of ions and other impurities, thus increasing the technical time and sample processing costs. PCR can be performed in 1 µl reactions directly in the chip wells in parallel. The PCR product can be detected *in situ* using MALDI MS. This miniaturization technique has the potential for medium throughput, low-cost genotyping. A chip-based genotyping method using MS was recently published providing data on more than 200,000 SNPs across the genome [51]. The advantages of MS are: its speed, high accuracy, low background, and internal standard for calibration. The disadvantages of this

method are the high cost of the instrument and the entire analytical process. There are several vendors for mass spectrometers including; Agilent, Applied Biosystems, Brukers, GenTech, Hitachi, JEOL, Keratos, LECO, Micromass, Shimadzu, ThermoFinnigan, Varian and Waters.

Pyrosequencing for genotyping

Pyrosequencing for genotyping has been widely used for polymorphism analyses. Recent advances include multiplex genotyping [52], allele frequency studies [53] and introduction of pyrosequencing Gold chemistry which uses single-stranded DNA-binding proteins in the enzyme mix [54]. The advantage of this chemistry is that it detects SNPs in the context of the surrounding sequence. This guarantees that each individual analysis is correct, without the need for other controls. Current machines enable genotyping of up to 13,000 SNPs per day, covering low to medium throughput demands. A disadvantage is that it requires a specialized machine to perform the analysis, which is relatively expensive.

Figure 6. Schematic representation of an Invader® assay.*Invader assay*

The Invader® assay (Third Wave Technologies, Inc., WI, USA) is an attractive fluorescence resonance energy transfer (FRET)-based genotyping method with the potential to genotype SNPs without PCR amplification. A fluorescent probe and upstream (Invader) oligonucleotide hybridize to the target DNA and create a 1 bp overlap

(flap). The Invader assay depends on cleavage enzymes (Flap endonucleases) that cleave the mismatched regions of DNA [37,38]. Invader assays are conducted isothermally, and a linear increase in signal over time will be produced (Figure 6). Each cleavage product then serves as an Invader oligonucleotide in a secondary reaction, where it directs the cleavage of a combined labeled FRET

probe-template construct [55]. This secondary oligonucleotide probe is 5' end labeled with the donor fluorophore, which is quenched by an internal acceptor dye. When the DNA is cleaved, the donor and acceptor dyes are no longer in close proximity, the quenching is abolished, and fluorescence is generated. Assays are read with a fluorescence plate reader, and genotypes are assigned after determination of the net wild type/variant signal ratio for each sample [56,57]. The Invader assay (invasive cleavage reaction) can be used for genotyping directly from human genomic DNA without any PCR amplification of the target DNA. In this method the probe oligonucleotides labeled with a fluorescent and a quencher are coated on a polystyrene microsphere surface. After the genomic DNA is applied to the microspheres, the invasive cleavage reaction generates a fluorescent signal on the surface of the microspheres that is analyzed by the flow cytometry [58,59]. The disadvantage of invasive cleavage reaction is that a high concentration of genomic DNA (25–35 µg) is needed for each reaction.

High-throughput genotyping techniques

Molecular inversion probe

In molecular inversion probe (MIP) a single oligonucleotide probe with recognition sequences at each terminus is hybridized with a genomic target sequence, such that it forms a circular structure, with the ends of the probe abutting, leaving a single base gap at the location of a SNP [42,43]. This gapped-duplex is then tested in four separate reactions, each with a single deoxynucleoside triphosphate (dNTP) species present, in which successful polymerization/ligation provides allelic differentiation. The probes are subsequently released from the genomic DNA and those that have been covalently circularized in the correct allele/nucleotide reaction combinations are amplified using a 'universal' PCR primer pair. Each amplified probe contains a unique tag sequence that is complementary to a sequence on the universal tag array. Tags have been selected to have a similar thermal stability (T_m) and base composition, and to be maximally orthogonal in sequence complementarity. Amplicons are fluorescently labeled and the tag sequences are released from the genome homology regions using a restriction endonuclease treatment. The tags are then detected using a complementary tag array (Figure 7). A level of multiplexing of 22,000 SNPs scored per assay has been performed with MIP, without any reduction in accuracy and conversion rate of this technology.

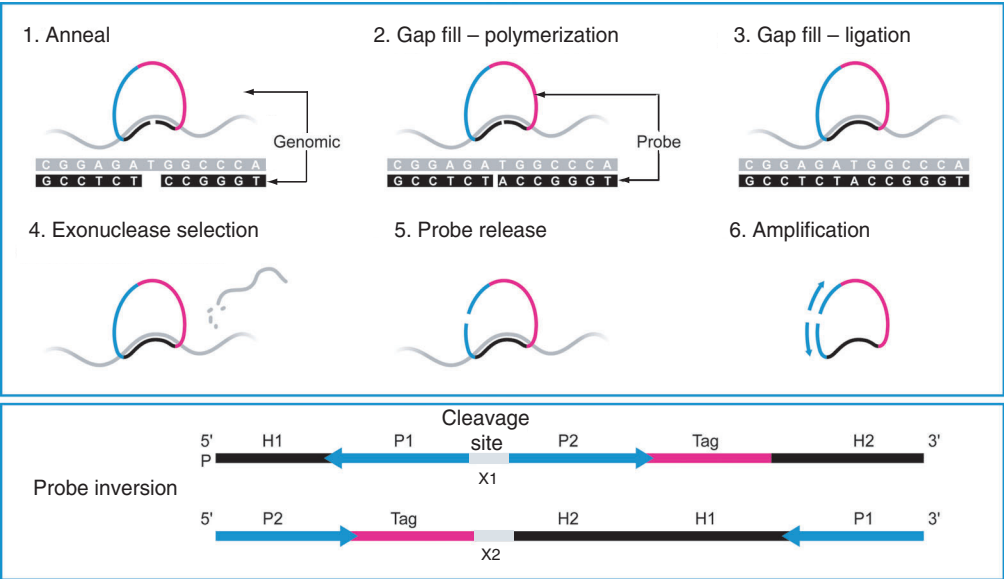
Oligonucleotide ligation assay

The oligonucleotide ligation assay (OLA) relies on hybridization with specific oligonucleotide probes that can effectively discriminate between the wild-type and variant sequences (Figure 8). Three oligonucleotides are used in OLA: two allele-specific oligonucleotide probes (one specific for the wild-type allele and the other specific for the mutant allele) plus a fluorescent common probe. The 3' ends of the allele-specific probes are immediately adjacent to the 5' end of the common probe. Two approaches have emerged enabling multiplexing of OLA: SNPlex™ and GoldenGate™ assay [60].

SNPlex: This technique has been developed by Applied Biosystems and is based on OLA. The ligase reaction is performed, followed by a multiplex PCR and capillary gel electrophoresis. The two OLA probes have 5'-tags complementary to a universal PCR primer pair. The SNPlex system's multiplexed OLA/PCR assay, along with the ZipChut™ Mobility Modifier probes, apply spatial separation and multi-color fluorescence technology to detect each allele and reduce the need to optimize the assay. This system allows for multiplexed genotyping of 48 SNPs simultaneously in a single biological sample, and can detect 4500 SNPs in parallel. Genotyping detection is performed with Applied Biosystems 3730 and 3730xl platforms. The technique is mostly suitable for medium to high throughput genotyping.

GoldenGate assay: This method is based on a highly multiplexed, allele discriminating, extension reaction (OLA-based platform) that uses three oligonucleotides (two allele-specific oligos [ASOs] and a locus-specific oligo [LSO]). The ASOs contain target-specific regions and universal PCR primer sites; and the LSOs contain a unique address sequence that targets the complementary oligonucleotide probe for a particular bead type on the array [61,212]. These three assay oligonucleotide sets, for up to 1152 different SNP loci, are pooled and hybridized to a small sample of genomic DNA. Hybridized allele-specific probes whose 3' ends are complementary to the dimorphic base at the SNP position are extended by a DNA polymerase, while allele-specific probes whose 3' ends mismatch the base at the SNP position will not be extended efficiently. Since all nucleotides are used for extension, strong strand-displacement activity of DNA polymerase limits the choice of DNA polymerases available because extended allele-specific assay

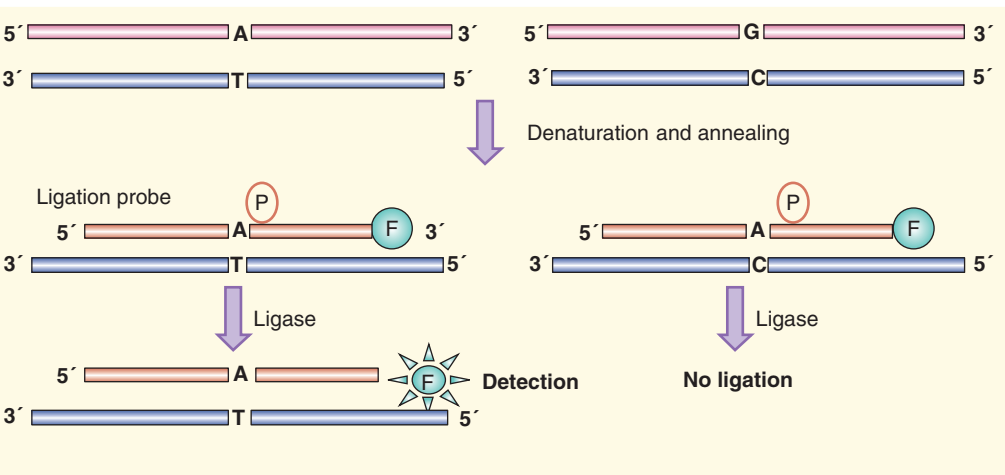
Figure 7. Schematic representation of a molecular inversion probe.



Upper panel describes the enzymatic steps in a MIP assay for genotyping, and the lower panel illustrates the different parts of the probe before and after inversion. In the lower panel, unreacted probe (top) and inverted probe (bottom) are demonstrated. MIP consists of two regions of homology to target genomic DNA (H1 and H2 are at the termini of the probe and are unique to each probe), two PCR primer regions which are common to all probes (P1 and P2), one barcode which is specific for each locus, and two common cleavage sites (X1 and X2). The cleavage sites X1 and X2 are used to release the circularized probe from genomic DNA and for postamplification processing, respectively.

MIP: Molecular Inversion Probe.

Figure 8. Schematic representation of an oligonucleotide ligation assay.



The high fidelity DNA ligase joins the nick junction when the bases are perfectly matched and the released fluorescent signal can be detected by different methods. There will be no ligation between mismatched base pairs. F: Fluorescence; P: Probe.

oligonucleotides should terminate adjacent to the 5' phosphate of the locus-specific probe which is ligated to the locus-specific oligonucleotide. These full-length copies provide the amplification template for PCR using Cy3- and Cy5-labeled primers (for the allele-specific assay oligonucleotides) and a biotin-labelled primer. After thermocycling, the PCR products are bound to streptavidin-coated paramagnetic particles, and the dye-labeled strands are isolated. The fluorescently labeled, single-stranded products are then hybridized to a predecoded random array. This platform allows multiplexing of up to 1536 SNPs in a single reaction tube. GoldenGate assay relies on the Sentrix™ array and Sherlock scanner commercialized by Illumina. The Sentrix 96 Array Matrix is capable of genotyping 96 samples at a time, or 150,000 in a single microtiter plate.

Whole-genome sampling analysis

Recently, a hybridization-based technique was introduced for SNP genotyping. The WGSa method allows more than 100,000 SNPs to be genotyped simultaneously by allele-specific hybridization probes synthesized on a single array [45,46]. In this technique genomic DNA is digested by restriction endonucleases into fragments. Followed by adapter ligation, *in vitro* amplification is performed to selectively amplify fragments between 400–800 nucleotides for 10,000 arrays and 250–2000 for 100,000 arrays, respectively (Figure 9). Amplification represents approximately 4×10^7 bps of genomic DNA for 10,000 arrays and 3×10^8 bps of genomic DNA for 100,000 arrays [45,46]. Successful genotyping of tens of thousands of SNPs has been reported [45,46]. The number of SNPs to be analyzed on this platform can potentially be increased since the chip manufacturing technique is photolithographic and smaller feature size can be achieved. Two platforms for 10,000 and 100,000 SNPs are available through Affymetrix, and a 500,000 SNP platform is going to be on the market this year. This technology demonstrates limited flexibility in the choice of SNPs to be genotyped and has a conversion rate approaching 15%. However, the products serve as an excellent whole-genome scan to achieve a relatively high-resolution linkage peak for further custom genotyping.

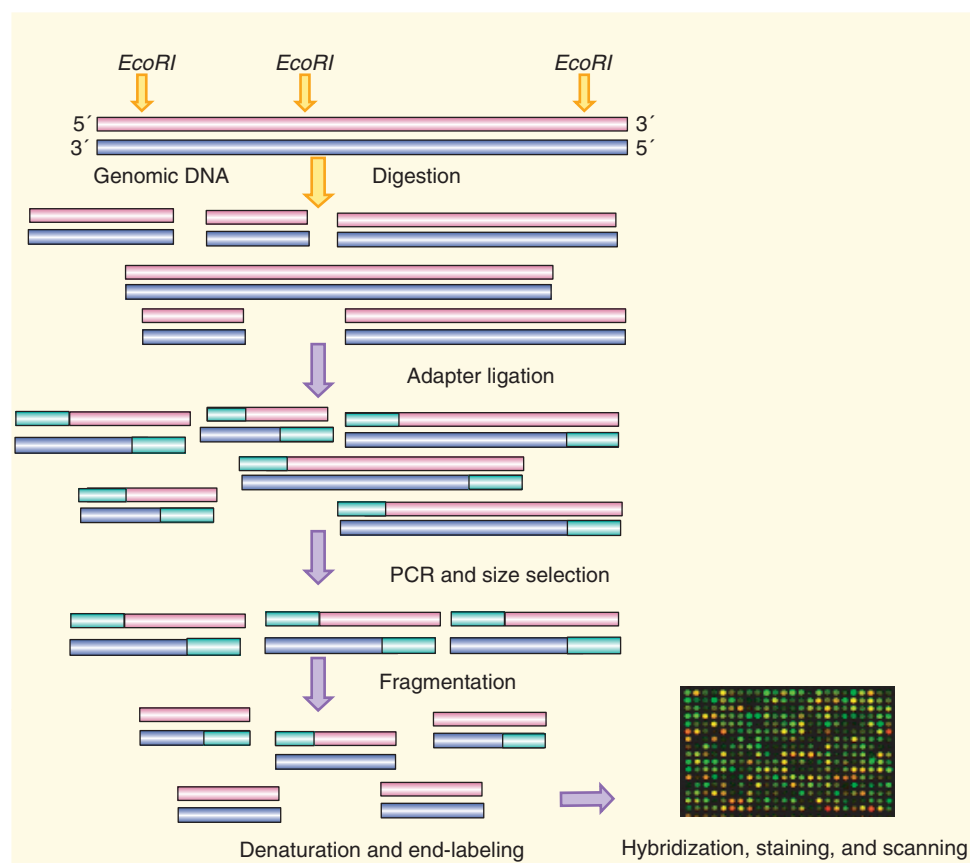
Multiplexed anchored runoff amplification assay

The multiplexed anchored runoff amplification (MARA) assay uses a single primer per SNP, in conjunction with restriction-enzyme-digested,

adapter-ligated genomic DNA. Each primer is composed of a common sequence at the 5' end followed by locus-specific sequence at the 3' end. Following a primary reaction in which locus-specific products are generated, a secondary universal amplification is carried out using a generic primer pair corresponding to the oligonucleotide and genomic DNA adapter sequences. Allele discrimination is achieved by hybridization to high-density DNA oligonucleotide arrays. Multiplexing reaction of more than 4000-plex was recently reported with 75% success rate [47]. This technique improves the metrics of WGSa since it involves an enzymatic step. However, the data presented is based on chromosome 21 and SNPs were chosen from the Perlegen database, therefore, a higher decrease in conversion rate is expected from chip hybridization on other chromosomes.

Hapmap and SNP analysis

The goal of the International HapMap Project [202] is to compare the genetic sequences of different individuals to identify chromosomal regions where genetic variants are shared. In many parts of chromosomes, just a handful of haplotypes are found in humans. In a given population, 55% of people may have one version of a haplotype, 30% may have another, 8% may have a third, and the rest may have a variety of less common haplotypes. The International HapMap Project is identifying these common haplotypes in four populations from different parts of the world. In addition, 'tag' SNPs that uniquely identify these haplotypes in an individual's DNA will be identified. These SNPs will be a part of the whole-genome SNP library. The number of tag SNPs that contain most of the information about the patterns of genetic variation is estimated to be about 300,000 to 600,000, which is far fewer than the 10–20 million common SNPs. Once the information on tag SNPs from the HapMap is available, researchers will be able to use them to locate genes involved in medically important traits by focusing on specific candidate genes that may be associated with a disease, or even by looking across the entire genome to find chromosomal regions that may be associated with a disease. In Phase I of the HapMap, 1,073,663 SNPs were genotyped in Centre d'Etude du Polymorphisme Humain individuals, 1,044,686 SNPs in Han Chinese individuals, 1,044,416 SNPs in Japanese individuals and 1,034,205 SNPs in Yoruba individuals (about one million common SNPs were genotyped in 270 individuals). In Phase II, an additional 4 million

Figure 9. Schematic representation of whole-genome sampling analysis.

Genomic DNA is digested by restriction enzyme (e.g., *EcoRI*), adaptors are ligated to different DNA fragments, PCR reaction is used to amplify selectively the 400–800 bp fragments. The amplified DNA targets are fragmented, denatured, labeled and hybridized to microarray, stained, scanned and analyzed.

SNPs in the same samples will be well underway. Five major technologies were used in HapMap projects: direct hybridization on high-density oligonucleotide arrays from Perlegen, Molecular Inversion Probe assay from ParAllele BioSciences, GoldenGate assay from Illumina, mass spectrometry assay from Sequenom, and Invader assay from Thirdwave Technologies.

Concluding remarks

SNP technologies will dominate the field of large-scale genome analysis for at least another decade. While the HapMap organization is trying to reduce the number of SNPs to be analyzed, the price has continuously fallen by almost twofold per year, from more than US\$1 per SNP, to a few cents for custom genotyping and below a cent for random genotyping (in whole-genome analysis). The new way to use SNPs will be another area of interest in this field. Some of the SNPs may have diagnostic

value, and some others may serve as potential drug targets. Currently, high-throughput technologies play an important role in determining the value of the SNPs. However, low-throughput technologies used in highly integrated devices may be of interest for future diagnostic tests. The selection of an appropriate genotyping technology for current applications can be a challenging task. The ideal genotyping technology would identify any chosen nucleotide in the genome in all DNA samples with perfect accuracy at low cost. Ultimately, the quality of the genotype data are measured by the power that it provides to elucidate genetic associations. A high-quality genotyping technology will have a power to find associations that closely approximates the power of the idealized technology. Conversion rate is defined to be the fraction of probes per batch that yielded strong signals with discernable clusters to indicate the different genotypes. As mentioned previously, *in silico* design rates are relevant in

Highlights

- Technologies related to sequencing and genotyping will continue to evolve and play a major role in large-scale genome analysis for at least another decade.
- The completion of the International HapMap project will provide a tool that allows researchers to screen the entire genome and find associations between genes, genetic variations and diseases more efficiently and at a lower cost.
- An ideal technology for genotyping would identify any selected nucleotide in the genome, in all DNA samples, with high accuracy at a low cost.
- Advances in sequencing and genotyping technologies will further decrease the cost of single nucleotide polymorphism discovery and whole-genome analysis in the near future.
- The production of universal and whole-genome arrays from several vendors will further decrease the cost of DNA array.
- Large investments by pharmaceutical and biotechnology companies for developing genomic strategies will result in discovering more biomarkers and novel therapeutic targets.
- In the next decade, genome projects, along with DNA array technologies, high-throughput screening systems and advanced bioinformatics, will facilitate the elucidation of complex multigenic diseases.
- As the cost of genomic screening and analysis continues to dramatically decrease, personalized medicine may not be far from becoming a reality.
- Although powerful high-throughput technologies look promising in producing large amounts of data at a low cost, the lack of an efficient and comprehensive tool for analyzing these data remains a limiting factor.

assessing the overall SNP yield. Another measure of accuracy is trio concordance. The use of mother–father–child trios allows accuracy to be monitored by looking at the Mendelian inheritance patterns across markers. This test is able to capture some forms of systematic errors that cannot be estimated by repeatability. It is important to note that all these parameters – conversion rate, completeness, and concordance – are related through parameters that can be changed in the clustering algorithm used to assign genotypes. A permissive algorithm can result in high levels of completeness and conversion at the expense of accuracy. In addition to the factors mentioned above, cost of a technology for genome-based analysis is also important. Standard panels provide lowest genotyping cost. However, custom genotyping techniques are emerging as it provides flexibility in use and increases the power of a genetic study. High-throughput custom genotyping techniques, involve three dominating costs: the target-specific probe, DNA array and enzyme cost. The cost of the DNA probe is mainly determined by the number of nucleotides in a probe and the amount of probe being used in an assay. The cost of the enzyme proportionally decreases with the increase in the number of multiplexing. The cost of the DNA array is also decreasing as whole genome arrays and universal arrays will be available

from several vendors. In addition, the set up cost for instruments that are used to perform these assays varies in different technologies.

Outlook

Substantial investments are being made within the pharmaceutical and biotechnology industries to develop and use genomic strategies for the discovery of biomarkers and novel therapeutic targets. Over the next decade it is anticipated that genome projects, coupled with DNA array technologies, high-throughput screening systems, and advanced bioinformatics, will permit rapid elucidation of complex genetic components of human health and disease. Personalized medicine will emerge, as the cost of analysis will decrease by orders of magnitude. Such analysis, which is only performed once for each individual, can then become the blueprint for personalized therapy. The technology roadmap is relatively clear and several parallel efforts are likely to lead us to the goal. There are three sweet spots for genotyping markets: whole-genome random SNP analysis useful for mapping, large-scale custom genotyping to find the causing haplotype and diagnostic genotyping. The first two markets are likely to be dominated by array-based technique and careful study design, and the use of various technologies along with different discovery steps will have a major impact on comprehensive genome analyses in complex disorders. The diagnostic market will probably favor the use of label-free technologies that provide easier and less inexpensive automation. These technologies will likely help us to sort out the large number of false positive results that will be generated by powerful high-throughput technologies. In summary, although the advanced technologies enable us to produce large amounts of data, the lack of a sufficient number of well-characterized samples from different populations and efficient analyses remains a limiting factor in pharmacogenomics. The ongoing efforts by several companies and institutions to collect well-defined samples from different diseased populations, and develop more advanced and comprehensive analysis tools for large data processing and interpretation, are expected to overcome these limitations.

Acknowledgments

Dr Langaee is supported by NIH/NHLBI RO1 HL747430 grant and Dr Ronaghi is supported by NIH grant # 2PO1 HG00205.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Marth GT, Korf I, Yandell MD *et al.*: A general approach to single-nucleotide polymorphism discovery. *Nature Genet.* 23(4), 452–456 (1999).
2. Buetow KH, Edmonson MN, Cassidy AB: Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genet.* 21(3), 323–325 (1999).
3. Weckx S, Del-Favero J, Rademakers R *et al.*: novoSNP, a novel computational tool for sequence variation discovery. *Genome Res.* 15, 436–442 (2005).
4. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* 74(12), 5463–5467 (1977).
5. Bains W, Smith GC: A novel method for nucleic acid sequence determination. *J. Theor. Biol.* 135, 303–307 (1988).
6. Drmanac S: Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nature Biotechnol.* 16, 54–58 (1998).
7. Fodor SP, Read JL, Pirrung MC *et al.*: Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767–773 (1991).
8. Drmanac R, Drmanac S, Baier J *et al.*: DNA sequencing by hybridization with arrays of samples or probes. *Methods Mol. Biol.* 170, 173–179 (2001).
9. Lipshutz RJ, Morris D, Chee M *et al.*: Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* 19(3), 442–447 (1995).
10. Patil N, Berno AJ, Hinds DA *et al.*: Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294, 1719–1723 (2001).
11. Hinds DA, Stuve LL, Nilsen GB *et al.*: Whole-genome patterns of common DNA variation in three human populations. *Science* 307(5712), 1072–1079 (2005).
12. Ronaghi M, Karamohamed S, Pettersson B *et al.*: Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242, 84–89 (1996).
13. Ronaghi M, Uhlen M, Nyren P: A sequencing method based on real-time pyrophosphate. *Science* 281, 363 (1998).
- Outlines the principles of pyrosequencing, and modification of this technique by the addition of a nucleotide-degrading enzyme to obtain a four-enzyme mixture.
14. Salama K, Eltoukhy H, Hassibi A *et al.*: Modeling and simulation of integrated bioluminescence detection platforms. *Biosens. Bioelectron. J. Special issue: Micro and Nano Bioengineering* 19(11), 1377–1386 (2004).
15. Ronaghi M: Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 11, 3–11 (2001).
16. Metzker ML, Raghavachari R, Richards S *et al.*: Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.* 22, 4259–4267 (1994).
17. Hultman T, Stahl S, Hornes E *et al.*: Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res.* 17(13), 4937–4946 (1989).
18. Mitra RD, Shendure J, Olejnik J *et al.*: Fluorescent *in situ* sequencing on polymerase colonies. *Anal. Biochem.* 320(1), 55–65 (2003).
19. Mitra RD, Church GM: *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* 27, 1–6 (1999).
20. Levene MJ, Korlach J, Turner SW *et al.*: Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299(5607), 682–686 (2003).
21. Braslavsky I, Hebert B, Kartalov E *et al.*: Sequence information can be obtained from single DNA molecules. *Proc. Natl Acad. Sci. USA* 100, 3960–3964 (2003).
22. Deamer DW, Branton D: Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* 35, 817–825 (2002).
23. Deamer DW, Akeson M: Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotech.* 18, 147–151 (2000).
24. Li J, Gershow M, Stein D *et al.*: DNA molecules and configurations in a solid-state nanopore microscope. *Nature Mater.* 2, 611–615 (2003).
25. Winters-Hilt S, Vercoutere W, DeGuzman VS *et al.*: Highly accurate classification of Watson-Crick basepairs on termini of single DNA molecules. *Biophys. J.* 84(2, Pt 1), 967–976 (2003).
26. Faham M, Baharloo S, Tomitaka S *et al.*: Mismatch repair detection (MRD): high-throughput scanning for DNA variations. *Hum. Mol. Genet.* 10(16), 1657–1664 (2001).
- Modifications of MRD for high-throughput variant scanning that is capable of detecting rare alleles.
27. Fakhrai-Rad H, Zheng J, Willis TD *et al.*: SNP discovery in pooled samples with mismatch repair detection. *Genome Res.* 14(7), 1404–1412 (2004).
28. Stanssens P, Zabeau M, Meersseman G *et al.*: High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res.* 14(1), 126–133 (2004).
29. Honisch C, Raghunathan A, Cantor CR *et al.*: High-throughput mutation detection underlying adaptive evolution of *Escherichia coli*-K12. *Genome Res.* 14(12), 2495–2502 (2004).
30. Oefner PJ: Allelic discrimination by denaturing high-performance liquid chromatography. *J. Chromatogr. B. Biomed. Sci.* 739(2), 345–355 (2000).
31. Wolford JK, Blunt D, Balcer C *et al.*: High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Hum. Genet.* 107(5), 483–487 (2000).
32. Tammara A, Bracco ACS, Esposito M *et al.*: Scanning for mutations of the ryanodine receptor (*RYR1*) gene by denaturing HPLC: detection of three novel malignant hyperthermia alleles. *Clin. Chem.* 49(5), 761–768 (2003).
33. Ronaghi M: Pyrosequencing for SNP genotyping. *Methods Mol. Biol.* 212, 189–195 (2003).
34. Syvanen AC, Aalto-Setälä K, Harju L *et al.*: A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* 8(4), 684–692 (1990).
- Introduction of a new technique (single base extension or mini-sequencing) that identifies the immobilized DNA by a one-step primer extension reaction.
35. Ross P, Hall L, Smirnov I *et al.*: High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol.* 16(13), 1347–1351 (1998).
36. Pusch W, Wurmbach JH, Thiele H, Kostrzewa M: MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 3(4), 537–548 (2002).
37. Lyamichev V, Mast AL, Hall JG *et al.*: Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnol.* 17(3), 292–296 (1999).
38. Hall JG, Eis PS, Law SM *et al.*: Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl Acad. Sci. USA* 97(15), 8272–8277 (2000).
39. Holland PM, Abramson RD, Watson R *et al.*: 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(16), 7276–7280 (1991).
40. Livak KJ: Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal.* 14(5–6), 143–149 (1999).
 41. Livak KJ: SNP genotyping by the 5'-nuclease reaction. *Methods Mol. Biol.* 212, 129–147 (2003).
 - **References 39–41 describe how the 5'--3' exonuclease activity of the enzyme Taq DNA polymerase can be used to discriminate alleles that differ by a single base substitution. A simple and single-step assay used for SNP genotyping.**
 42. Hardenbol P, Baner J, Jain M *et al.*: Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnol.* 21(6), 673–678 (2003).
 43. Hardenbol P, Yu F, Belmont J *et al.*: Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15(2), 269–275 (2005).
 - **References 42 and 43 give a description of the development of molecular inversion probe (MIP) and highly multiplexed MIP genotyping, an efficient technology used for large-scale SNP analysis**
 44. Barany F: Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl Acad. Sci. USA* 88(1), 189–193 (1991).
 45. Matsuzaki H, Loi H, Dong S *et al.*: Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* 14(3), 414–425 (2004).
 46. Matsuzaki H, Dong S, Loi H *et al.*: Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* 1(2), 109–111 (2004).
 47. Shapero MH, Zhang J, Loraine A *et al.*: A novel approach for highly multiplexed locus-specific SNP genotyping using high-density DNA oligonucleotide arrays. *Nucleic Acids Res.* 32(22), E181 (2004).
 48. Kuimelis RG, Livak KJ, Mullah B *et al.*: Structural analogues of TaqMan probes for real-time quantitative PCR. *Nucleic Acids Symp. Ser.* 37, 255–256 (1997).
 49. Van Hoeyveld E, Houtmeyers F, Massonet C *et al.*: Detection of single nucleotide polymorphisms in the mannose-binding lectin gene using minor groove binder-DNA probes. *J. Immunol. Methods* 287(1–2), 227–230 (2004).
 50. Jain M, Thorstenson YR, Faulkner DM *et al.*: Genotyping African haplotypes in ATM using a co-spotted single-base extension assay. *Hum. Mutat.* 22(3), 214–221 (2003).
 51. Nelson MR, Marnellos G, Kammerer S *et al.*: Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res.* 14(8), 1664–1668 (2004).
 52. Pourmand N, Elahi, E., Davis RW *et al.*: Multiplex Pyrosequencing. *Nucleic Acids Res.* 30, E31 (2002).
 53. Lavebratt C, Sengul S, Jansson M *et al.*: Pyrosequencing-based SNP allele frequency estimation in DNA pools. *Hum. Mutat.* 23, 92–97 (2004).
 - **Describes how pyrosequencing can be used for allele frequency estimation in DNA pools of SNPs with accuracy and precision.**
 54. Ronaghi M: Improved performance of Pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286(2), 282–288 (2000).
 55. Nazarenko IA, Bhatnagar SK, Hohman RJ: A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res.* 25(12), 2516–2521 (1997).
 - **Describes the direct detection method of PCR-amplified DNA in a closed system using FRET as a detection method, which is used for medium to high-throughput SNP genotyping.**
 56. Ledford M, Friedman KD, Hessner MJ *et al.*: A multi-site study for detection of the factor V (Leiden) mutation from genomic DNA using a homogeneous invader microtiter plate fluorescence resonance energy transfer (FRET) assay. *J. Mol. Diagn.* 2(2), 97–104 (2000).
 57. Hessner MJ, Budish MA, Friedman KD: Genotyping of factor V G1691A (Leiden) without the use of PCR by invasive cleavage of oligonucleotide probes. *Clin. Chem.* 46(8, Pt 1), 1051–1056 (2000).
 58. Wilkins SP, Hall JG, Lyamichev V *et al.*: Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. *Nucleic Acids Res.* 29(16), E77 (2001).
 59. Rao KV, Stevens PW, Hall JG *et al.*: Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. *Nucleic Acids Res.* 31(11), E66 (2003).
 60. Landegren U, Kaiser R, Sanders J *et al.*: A ligase-mediated gene detection technique. *Science* 241(4869), 1077–1080 (1998).
 - **This paper shows, for the first time, that the enzyme DNA ligase can be used as a tool for SNP genotyping, in oligonucleotide ligation assay.**
 61. Oliphant A, Barker DL, Stuelpnagel JR *et al.*: BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques* (Suppl.), 56–58, 60–61 (2002).
 - **Describes the development of a very high-throughput genotyping system that uses the BeadArray platform with a high level of multiplexing, accuracy and low cost for SNP analysis.**

Patents

101. Southern EM: WO10977 (1989). Analysing polynucleotide sequences.
102. Cheesman PC: US5302509 (1994). Method for sequencing polynucleotides.
103. Williams J: US6306607 (2001). Heterogenous assay for pyrophosphate.

Websites

201. www.microchipbiotech.com
Microchip Biotechnologies, Inc.
202. www.hapmap.org
HapMap project
203. www.454.com
454 Life Sciences
204. www.lynxgen.com
Solexa
205. www.genovoxx.de
Genovoxx
206. www.nanofluidics.com
Nanofluidics
207. www.helicosbio.com
Helicos
208. www.visigenbio.com
Visigen
209. www.licor.com
Li-Cor
210. www.parallelebio.com
ParAllele Bioscience
211. www.sequenom.com
Sequenom
212. www.illumina.com
Illumina