Technology evolution for genomic revolution

✉ Hanlet Camacho, Alberto Cintado, Marta G Dueñas

Biomedical Research, Center for Genetic Engineering and Biotechnology (CIGB)
Ave 31 / 158 and 190, Cubanacán, Playa, PO Box 6162, Havana 10 600, Cuba
Fax: (53 7) 271 4764; E-mail: hamlet.camacho@cigb.edu.cu

ABSTRACT

The Human Genome Project has provided near-total sequence information about our genes and their surroundings. This monumental accomplishment has been accompanied by equally remarkable progress in biomedical research and computer science. Biologists and clinical investigators are now well positioned to perform global analysis of genetic polymorphisms-mutations, gene expression, protein identity and function (proteomics), and the 3-dimentional structure of proteins (structural genomics). The resulting information should accelerate the development of novel and targeted therapeutics that are tailored to individual genomic profiles (pharmacogenomics). The massive amount of data generated by genome-wide experiments requires the development of comprehensive genomic-proteomic databases and software that enable megadata access, transfer, and manipulation (bioinformatics). Advances in molecular biology techniques have allowed accurate and rapid characterization of DNA sequences as well as the identification and quantification of cellular RNA and protein. Global analytic methods and human genetic mapping are expected to accelerate the process of identification and localization of disease genes. In this review article in genomics, selected principles and methods in molecular biology are recapped, with the intent to prepare the reader for forthcoming articles with a more direct focus on aspects of the subject matter.

Key words: Genomics, genome techniques, functional genomics, population genomics

RESUMEN

Evolución tecnológica para revolución genómica. El Proyecto del Genoma Humano nos ha provisto de la secuencia casi completa de nuestros genes y sus alrededores. Este logro monumental ha llegado acompañado de un progreso similar en las investigaciones biomédicas y las ciencias computarizadas. Los investigadores en biología y medicina clínica tienen ahora una privilegiada posición para realizar análisis globales de los polimorfismos y las mutaciones, de la expresión de genes, de la identidad y función de las proteínas (proteómica) y de la estructura tridimensional de las proteínas (proteómica estructural). La información resultante deberá acelerar el desarrollo de terapias novedosas dirigidas a blancos moleculares que sean lo ideal de acuerdo con el perfil genómico de cada individuo (farmacogenómica). La masiva cantidad de datos generados mediante la realización de experimentos con el genoma completo, requieren del desarrollo de bases de datos genómicos y proteómicos y programas que posibiliten el acceso, la transferencia y la manipulación de "megadatos" (Bioinformática). El avance en las técnicas de biología molecular ha permitido una rápida y precisa caracterización de las secuencias de ADN, así como la caracterización de ARN celular y la identificación de proteínas. Se espera que el empleo de métodos analíticos globales y estrategias de mapeos genómicos aceleren el proceso de identificación y localización de genes asociados a las enfermedades. En este artículo de revisión de genómica se ha recapitulado en algunos principios y métodos de la biología molecular, con la intención de preparar al lector para artículos venideros con un enfoque más directo en aspectos relacionados con esta temática.

Palabras clave: técnicas en genómica, genómica funcional, genómica poblacional

Introduction

The completion of the Human Genome Project represents the first step in a process that will revolutionize medicine in the 21st century. The "race to the finish line" between the "public" and "private" Human Genome Project depicted in the popular press was really a race to the starting line with regard to the application of genomics to medicine.

Research activities downstream from the early, descriptive portions of the Human Genome Project promise to fundamentally alter the way in which future generations of health care professionals will diagnose, treat, and -ultimately- prevent disease. Unlike the early experiences with genetic medicine, in which gross aberrations of genetic material and their influence on relatively small populations of patients were typically studied, the new genomic revolution has the promise to better explain the pathomechanisms underlying highly prevalent illnesses such as coronary artery disease, cancer, and neurodegenerative disorders. Improved understanding of the multiple, interacting genetic components of disease pathophysiology should make it possible to design specific, "targeted" therapies. Furthermore, it should be possible to "individualize" drug and other treatments to accommodate both inherited differences in the pathophysiology of a single patient's disease process and genetic variation in that patient's ability to metabolize drugs. These issues are addressed by the field of "pharmacogenomics". Achieving the promise of the genomic revolution for medicine will also require the development of new clinical laboratory tests, including those that use microarrays (chips) that in turn can be used for DNA sequence analysis and for evaluation of gene expression in normal and diseased tissue.

Because genes encode proteins, the genomic revolution has also led to the development of "proteomics".

✉ Corresponding author

Proteomics involves the measurement not of single proteins, but rather of all the proteins expressed in a cell or tissue. The bounty of information resulting from the development of genomic and proteomic science has required the parallel development of computer-based techniques for managing vast collections of data and, even more important, providing ready access to those data by medical scientists and practitioners. In response, the discipline of bioinformatics has developed new techniques for organizing, storing, and analyzing the acquired information and, in many instances, disseminating the information via the World Wide Web. Although the promise of the genomic revolution for medicine is truly unprecedented, it is important that enthusiasm for the science be tempered by sensitivity to public concerns regarding the potential misuse of genomic information. The possibility of loss of confidentiality, issues of group stigmatization, and the difficult issue of genomic predestination must all be addressed if humankind is to reap the full benefits of the potential of this exciting new science.

As is readily apparent, evolving genomic science has associated concepts and a vocabulary that may challenge the study of medicine and biology. The scope of genomic knowledge and the applications are so broad that it is difficult to envision a disease process, patient, or clinician who will not be affected in the future. It is important that graduate, postgraduate, and continuing medical education keep abreast of the new information so that new discoveries can be translated into clinical practice rapidly and appropriately.

The development of a structural and functional genomic infrastructure will undoubtedly accelerate our understanding of diseases and help refine methods of disease prediction, diagnosis, staging, and treatment. Clinicians, investigators, and students of medicine and biological sciences must be prepared to participate effectively in the "new" genomics revolution and tackle the economical and ethical issues that are integral to all aspects of medical genomics. The intent of the current review is to begin the process of education in this regard, with the basic principles and methods in molecular biology that are relevant to genomics. This review is not intended to be comprehensive but to serve as an introduction to this subject.

Polymerase chain reaction

Polymerase chain reaction is an in vitro method of replicating relatively small DNA sequences into millions of copies over a short period [1]. A typical PCR reaction requires (A) two oligonucleotide primers (PCR primers), (B) a thermally stable DNA polymerase, (C) an ample amount of free deoxynucleotides, and (D) a small amount of DNA (the sample) that contains the sequence of interest (ie, the region to be amplified) [2]. The PCR primers (short DNA sequences of approximately 20 bp in length) are synthesized in the laboratory, and their nucleotide sequences are designed to be complementary to DNA sequences that flank (lie on either side of) the DNA fragment of interest (the target) [3, 4]. Therefore, one has to know the nucleotide sequence of the target DNA fragment to use PCR technology. Because DNA synthesis can proceed only in the 5´ to 3´ direction, a PCR primer allows the extension of DNA in only 1 direction (the

other side is a dead end). Therefore, to replicate both strands of the DNA molecule (the complementary strands), 2 PCR primers (forward and reverse primers) are needed, and they are placed on either side of the DNA sequence of interest on complementary strands (the forward primer on 1 strand of DNA and the reverse primer on the second strand of DNA that is complementary to the first strand). This approach allows the amplification of the double-stranded target sequence that is located between the 2 primers [5-8].

In general, PCR technology works best when the DNA fragment of interest is relatively small (<1 000 bp) [5]. The technique is easily automated to allow the detection and amplification of nanomolar quantities of DNA with a rapid turnaround time (a few hours). Because of the speed of the assay and the small amount of test sample needed (nanogram quantities), PCR is an extremely useful and efficient method of genetic testing in disease diagnosis, monitoring of treatment response, detection of minimal residual disease, and tissue typing [9-14]. Furthermore, the ability to extract DNA from various biological samples combined with PCR technology has substantially affected both forensic medicine and archeological studies [15-18]. However, the extreme sensitivity of PCR makes it vulnerable to the possibility of false-positive test results due to trace amounts of contaminants [19].

Nearly all techniques start with PCR amplified DNA containing the SNP of interest. Although PCR is an enormously powerful technique, it does have certain disadvantages. First, a single PCR reaction generally amplifies a single or limited number of SNP regions. Efforts have been made to use multiplex PCR techniques where several SNP locations are amplified at once, but these multiplex techniques are still not generally used and can only be applied to certain combinations of SNPs. Thus, future improvements in genotyping technology will likely involve better multiplex PCR protocols.

Real-time PCR

Highly sensitive detection of specific RNA species in a complex RNA sample can be achieved if the reverse transcription (RT) reaction is followed by PCR. Real-time detection of the ongoing PCR via the TaqMan assay [20, 21] exploits the nuclease activity of the *Taq*-polymerase [22]. A probe complementary to the amplified target sequence, labeled with a reporter fluorophore at the 5'-end and a quencher fluorophore at the 3'-end [23, 24], is added to the PCR. During the extension step *Taq*-polymerase degrades hybridized probes, releasing the 5' fluorophore. The fluorescence from the released fluorophores is proportional to the amount of degraded probes, and hence the amplification cycle at which the fluorescence exceeds a threshold value can be taken as a measure of the amount of the target present in the reaction. Alternative probe designs have been developed for real-time detection, such as molecular beacons [25] and scorpion probes [26]. These probes contain reporter fluorophores and quencher molecules as in the TaqMan probes previously described, but are constructed so that degradation of the probes is not necessary. Furthermore, they can be used to distinguish transcript variants that differ in a single-nucleotide position [20, 21]. The PCR step renders the

1. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp Quant Biol 1986;51(1):263- 73.

2. Baumforth KR, Nelson PN, Digby JE, O'Neil JD, Murray PG. Demystified the polymerase chain reaction. Mol Pathol 1999;52:1-10.

3. Dieffenbach CW, Lowe TM, Dveksler GS. General concepts for PCR primer design. PCR Methods Appl 1993;3:30-7.

4. Rychlik W. Selection of primers for polymerase chain reaction. Mol. Biotechnol 1995;3:129-34.

5. Eisenstein BI. The polymerase chain reaction: a new method of using molecular genetics for medical diagnosis. N Engl J Med 1990;322:178-83.

6. Saiki RK, Scharf S, Faloona F, *et al*. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985;230:1350-4.

7. Erlich HA. Principles and applications of the polymerase chain reaction. Rev Immunogenet 1999;1:127-34.

8. Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. Science 1991;252:1643-51.

9. Aurelius E, Johansson B, Skoldenberg B, Staland A, Forsgren M. Rapid diagnosis of herpes simplex encephalitis by nested polymerase chain reaction assay of cerebrospinal fluid. Lancet 1991;337:189-92.

10. Na BK, Kim JH, Shin GC, *et al*. Detection and typing of respiratory adenoviruses in a single-tube multiplex polymerase chain reaction. J Med Virol 2002;66:512-7.

11. Bustin SA, Dorudi S. Molecular assessment of tumour stage and disease recurrence using PCR-based assays. Mol Med Today 1998;4:389-96.

12. Biondi A, Rambaldi A. Polymerase chain reaction (PCR) approach for the evaluation of minimal residual disease in acute leukemia. Stem Cells 1994;12:394-401.

13. Raj GV, Moreno JG, Gomella LG. Utilization of polymerase chain reaction technology in the detection of solid tumors. Cancer 1998;82:1419-42.

14. Jackson JB. The polymerase chain reaction in transfusion medicine. Transfusion 1990;30:51-7.

15. Westwood SA, Werrett DJ. An evaluation of the polymerase chain reaction method for forensic applications. Forensic Sci Int 1990;45:201-15.

16. Decorte R, Cassiman JJ. Forensic medicine and the polymerase chain reaction technique. J Med Genet 1993;30:625-33.

17. Hagelberg E, Bell LS, Allen T, Boyde A, Jones SJ, Clegg JB. Analysis of ancient bone DNA: techniques and applications. Philos Trans R Soc Lond B Biol Sci 1991; 333:399-407.

18. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. Neandertal DNA sequences and the origin of modern humans. Cell 1997;90:19-30.

19. Lo YM, Mehal WZ, Fleming KA. False-positive results and the polymerase chain reaction [letter]. Lancet 1988;2:679.

Taqman method sensitive to false positive results due to the contamination of the samples, and multiplexing is difficult, demanding labor-intensive optimizations.

Polymorphism

The human genome is extensively affected by sequence variation [27]. When the rate of variation at a specific point in the DNA yields a variant sequence that is found in more than 1% of a given population, it is referred to as a *polymorphism*. When the incidence of a variant sequence is less than 1%, it is referred to as a *mutation*. However, these definitions are imprecise because what may be considered a mutation in a given population may qualify as a polymorphism in another population (e.g., sickle cell mutation].

Genetic polymorphism may be revealed by restriction enzyme digestion that results in variably sized restriction fragments, a phenomenon known as *restriction fragment length polymorphism(RFLP)*. Usually occurring in intergenic DNA regions, RFLPs generally have no phenotypic effect. Certain regions of DNA are highly polymorphic because of the presence of variable numbers of tandem repeats (VNTRs) [28]. The latter are short DNA sequences that are repeated in sequence multiple times. Variations among individuals in the number and size of such repeat sequences may be revealed by using RFLPs induced by restriction enzymes that cut near the boundaries that encompass the repeat sequences. The degree of polymorphism in VNTRs is sufficiently high to allow DNA fingerprinting [29].

Single nucleotide polymorphisms occur as often as every 200 bp in the human genome [30]. Single nucleotide polymorphisms may be detected by microarrays and have the potential to be used in studies of genetic susceptibility to diseases [31]. Of note, however, is the fact that only about 1% of single nucleotide polymorphisms result in an alteration in protein coding at the amino acid level.

DNA Sequencing

DNA sequencing entails spelling out the nucleotide chain of a DNA region of interest [32-34]. Sufficient quantities of the target DNA region of interest are obtained from either a bacterial-cloned fragment or a PCR-amplified sequence.

A short DNA primer is used to start the synthesis of a complementary strand to the original template to be sequenced. With the use of specific chemical analogues of nucleotides (dideoxynucleotides), which are capable of terminating chain extension at specific nucleotide bases, the synthesis of the complementary strand is halted (nested) at specific nucleotide bases, resulting in the synthesis of DNA fragments with varying lengths that differ by 1 nucleotide . The set of nested fragments that are radiolabeled or tagged with fluorescent dyes can then be separated by gel electrophoresis, resulting in a pattern that allows the determination of the specific nucleotide at the end of each successive fragment. Currently, an automated computerized laser detection system is used for rapid accrual of sequence information.

Gene mapping

Human genetic mapping entails the localization or mapping of genes to each of the 23 chromosome pairs

and the determination of the order and spacing of the gene on the particular chromosome [35]. This is done in 1 of 2 ways: physical mapping (based on estimates of physical distance measured in bp) or genetic linkage mapping (based on the frequency of meiotic crossing-over that is observed between 2 loci of a chromosome) [36] Linkage distance is measured in centiMorgans (cM), named after Thomas H. Morgan, the father of *Drosophila* (fruit fly) genetics, who defined the concept of genetic linkage [37].

Locating a disease gene is the first step toward cloning the gene itself.

Physical mapping

Physical mapping of genes may be achieved by different methods with differing levels of resolution that range from an entire chromosome to a single base pair (the base sequence). Methods of low-resolution physical mapping include karyotyping, fluorescence *in situ* hybridization (FISH), and somatic cell hybridization. Human chromosomes may be identified by size and by distinctive banding patterns seen under light microscopy after staining with special DNA-binding dyes (Giemsa) [38]. This method of chromosome visualization is called *karyotyping* and provides a metaphase picture of the 22 homologous pairs of autosomal chromosomes and the 1 additional pair of sex chromosomes lined up by size. Such a chromosome picture allows the identification of gross abnormalities (chromosomal deletions, duplications, translocations) that may be associated with a particular disease phenotype (e.g., chronic myeloid leukemia is associated with a gross translocation between chromosomes 9 and 22) [39]. As a result, the disease gene is physically mapped to the vicinity of the abnormal karyotypic lesion.

Other physical maps include cDNA maps that localize the chromosomal positions of expressed genes [40, 41], RFLP maps of restriction cleavage sites [42], and contig maps of overlapping sequences of YACs, BACs, or cosmids. The resolution grade of karyotypic identification of a genetic lesion is estimated at 10 million bases (Mb), that of FISH at about 2 Mb, that of restriction map at 0.5 to 1 Mb, and contig maps at an even lower scale.

Genetic linkage mapping

Genes that are located on different chromosomes segregate independently during meiosis (the law of independent assortment). In other words, 2 genes on 2 different chromosomes have a 50% chance of being inherited together in a daughter cell. In contrast, genes on the same chromosome have a tendency to be inherited together (i.e., are genetically linked) [43]. However, the chance of segregating together is not 100% because of the "phenomenon" (process) of recombination (crossing-over) [44-48]. During recombination, the naturally occurring exchange of chromosome regions between homologous pairs of chromosomes occurs during meiosis [49]. The chance for a recombination event to occur depends on the distance between the 2 genes on the same chromosome [50]. The closer the genes are to each other, the less chance for a recombination event to occur and vice versa.

Genetic linkage is defined as the tendency for genes on the same chromosome to stick together during

20. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome Res 1996;6(10):986-94.

21. Gibson UE, Heid CA, Williams PM. A novel method for real time quantitative RT-PCR. Genome Res 1996;6(10): 995-1001.

22. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5' 3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci USA 1991;88(16):7276-80.

23. Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Res 1993;21(16):3761-6.

24. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 1995;4(6): 357-62.

25. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 1996;14(3):303-8.

26. Whitcombe D, Theaker J, Guy SP, Brown T. and Little, S. Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotechnol 1999; 17(8):804-7.

27. Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. Science 2001;291:1304-51.

28. Ramisse V, Houssu P, Hernandez E, Denoeud F, Hilaire V, Lisanti O, Ramisse F, Cavallo JD, Vergnaud G, Variable number of tandem repeats in Salmonella enterica subsp. enterica for typing purposes. J Clin Microbiol 2004;42:5722-30.

29. Jeffreys AJ, Wilson V, Wong Z, *et al*. Highly variable minisatellites and DNA fingerprints. Biochem Soc Symp 1987; 53:165-80.

30. Zhao GR, Yang F, Yuan YJ, Gao XM, Zhang JP. Progress in detection methods of single nucleotide polymorphisms. Yi Chuan 2005;27:123-9.

31. Khan AS. Genomics and microarray for detection and diagnostics. Acta Microbiol Immunol Hung 2004;51:463-7.

32. Graham CA, Hill AJ. Introduction to DNA sequencing. Methods. Mol Biol 2001; 167:1-12.

33. Toyoda A. Genome sequencing. Tanpakushitsu Kakusan Koso. 2004; 49:2620-6.

34. Marziali A, Akeson M. New DNA sequencing methods. Annu Rev Biomed Eng 2001;3:195-223.

35. Hafler DA, De Jager PL. Applying a new generation of genetic maps to understand human inflammatory disease. Nat Rev Immunol 2005;5:83-91.

36. Braendle C, Caillaud MC, Stern DL. Genetic mapping of aphicarus -a sex-linked locus controlling a wing polymorphism in the pea aphid (Acyrthosiphon pisum). Heredity 2005.Jan 26.

37. Morgan TH. Chromosomes and heredity. Am Nat 1910;44:449-6.

38. Korenberg JR, Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 1988;53:391-400.

meiosis. The strength of this linkage is used as a unit of measurement (Morgan) of the distance between the 2 linked genes [51]. One centiMorgan denotes a recombination event of 1% (i.e., the genes stick together 99% of the time during meiosis). A genetic distance of 1 cM therefore denotes 2 genes that are relatively close to each other.

Accordingly, a genetic linkage map may be perceived as a map of a chromosome with genes or DNA markers that are aligned in a linear fashion with intervening distances that reflect the frequency of recombination between the adjacent genes or markers [44]. The total genetic length of the human genome is estimated at 3 000 cM. As such, 1 cM corresponds to about 1 million bp. Note that chromosomal regions (DNA sequences) that are evaluated for linkage analysis may not be associated necessarily with a phenotypically apparent trait. For example, hair color is a phenotypically apparent trait, and variation in the DNA sequence responsible for hair color is reflected by variation in hair color. In contrast, DNA sequence variation of a phenotypically silent genetic marker is revealed by molecular techniques that exploit the presence or absence of specific restriction enzyme cleavage sites (RFLPs) and variations in the number of tandem repeats at specific chromosomal regions (VNTRs) [52]. Regardless of whether the study markers are phenotypically apparent (genes) or not (RFLPs, VNTRs), linkage analysis requires the existance of heterozygosity at the locus of both the gene and the genetic marker (i.e., alleles of the specific gene or DNA sequence must be heterozygous) [53].

Alleles are variants of the same gene occupying the same locus on chromosomes. Therefore, a homologous pair of chromosomes may either have the same allele on the 2 chromosomes forming the pair (i.e., are homozygous for the specific allele) or have 2 different alleles at a specific locus (i.e., are heterozygous). Genetic mapping methods are based on the observed segregation of alleles on a homologous pair of chromosomes. For example, if 2 separate traits are inherited together 99% of the time, then the alleles responsible for the traits are probably on the same chromosome at a distance between them of 1 cM.

Precise localization and sequencing of the gene of interest

In most diseases, a causal genetic aberration is not yet known. As described previously, one can use linkage analysis to associate a putative disease gene with a chromosomal locale that is detectable either molecularly or phenotypically [51]. Once the gross location of the putative disease gene is determined by linkage analysis, one can apply positional cloning for more precise identification and sequencing of the gene of interest [54, 55].

Positional cloning starts with the selection, by using the DNA sequence of the linked marker as a probe, of vector clones from a genomic library that carry DNA sequences that overlap with those from the linked genetic marker. As mentioned previously, a genomic library is constructed by partially digesting the entire genome with restriction enzymes, followed by cloning the resultant fragments into appropriate vectors [56]. Because the DNA products of restriction

digestion are relatively large, the vectors usually used are YAC and BAC. Using the DNA sequence of the linked marker, one can pick YAC or BAC clones with sequences that overlap with those of the linked marker [57]. Alternatively, sequence tagged sites, which are unique, relatively short DNA sequences with known chromosomal location, may be used to identify clones with overlapping sequences. Subsequently, the overlapping sequences themselves may be used as probes to select additional YAC or BAC clones with sequences that are partially overlapping with the probes used [58, 59].

The stepwise process of identifying YAC or BAC clones based on overlapping DNA sequences is called *chromosome walking* [60]. This process of chromosome walking continues until the disease gene is identified. The linear alignment of YACs or BACs with overlapping sequences results in a continuous DNA segment called a *contig map*. Such contig maps have been essential for the analysis of human and non human genomes as well as for specific chromosomal regions [58, 61-63].

Probe-identified clones can also be individually mapped for sites complementary to ESTs. An EST that cosegregates with the gene of interest is determined and then sequenced. Alternatively, the clone containing the linkage markers may have already been sequenced, and resident genes are sought by looking for ORFs. Finally, the candidate gene is identified by mutational analysis that compares ORFs from a patient-derived DNA library to those obtained from non affected individuals.

*P*rinciples of genotyping methods

The various approaches for allele discrimination are formally divided systematically into enzymatic approaches, in which the properties of different enzymes to discriminate between nucleotides are used (restriction enzymes type II, Cleavase and Resolvase, DNA polymerase, and ligase) [64]; electrophoretic methods, in which the allele discrimination is based on the difference in mobility in polymeric gels or capillaries (single -and double- stranded conformation assays, heteroduplex analysis, and DNA sequencing) [65]; solid-phase determination of allelic variants, including high-density oligonucleotide arrays for hybridization analysis, mini-sequencing primer extension analysis, and fiberoptic DNA sensor array [66]; chromatographic methods such as denaturing high-performance liquid chromatography (DHPLC) [67]; other physical methods of discrimination of allelic variants such as mass spectrometry (mass and charge) or fluorescence exchange-based techniques [68]; and in silico methods such as high-throughput analysis of expressed sequence tag data [69]. Some of the most frequently used techniques and instrumental settings applied in different combinations are described, and other methods that are less broadly used, but have interesting potentials are discussed in this review (Table 1).

Enzymatic approaches for discrimination of allelic variants

The first widely used method for the detection of polymorphisms exploited alterations in restriction

39. Stegmaier K, Ross KN, Colavito SA, O'malley S, Stockwell BR, Golub TR. Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. Nat Genet 2005; 37(3):328.

40. Reymond A, Friedli M, Henrichsen CN, *et al*. From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. Genomics 2001;78:46-54.

41. Hai Fang Yin, Zhi Hui Zhao, Bao Liang Fan, Zhao Liang Liu, Wei Lu, Yu Fang Liu and Ning Li. cDNA cloning, genomic structure, chromosomal mapping, and expression analysis of parotid secretory protein in pig . Genomics 2004;83: 9-18.

42. Vissing H, Grosveld F, Solomon E, *et al*. Progress towards construction of a total restriction fragment map of a human chromosome. Nucleic Acids Res 1987; 15:1363-75.

43. Kelly JK, Noor MA. Speciation by reinforcement: a model derived from studies of Drosophila. Genetics 1996;143: 1485-97.

44. Sturtevant AH. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Biol 1993;14:43-59.

45. Putrament A. Recombination and chromosome structure in eukaryotes. Genet Res 1971;18:85-95.

46. Bowcock AM, Hebert JM, Wijsman E, Gadi I, Cavalli-Sforza LL, Boyd CD. High recombination between two physically close human basement membrane collagen genes at the distal end of chromosome 13q. Proc Natl Acad Sci USA 1988; 85:2701-5.

47. Smith KN, Nicolas A. Recombination at work for meiosis. Curr Opin Genet Dev 1998;8:200-11.

48. Lynn A, Kashuk C, Petersen MB, *et al*. Patterns of meiotic recombination on the long arm of human chromosome 21. Genome Res 2000;10:1319-32.

49. Tease C. Cytological detection of crossing-over in BUdR substituted meiotic chromosomes using the fluorescent plus Giemsa technique. Nature 1978;272: 823-4.

50. Morton NE. Sequential tests for the detection of linkage. Hum Genet 1955; 7:277-318.

51. Robson EB. The human gene map. Philos Trans R Soc Lond B Biol Sci 1988; 319:229-37.

52. Pulst SM. Genetic linkage analysis. Arch Neurol 1999;56:667-72.

53. White R, Lalouel JM, Leppert M, Lathrop M, Nakamura Y, O'Connell P. Linkage maps of human chromosomes. Genome 1989;31:1066-72.

54. Xu L, Li Y. Positional cloning. Methods Mol Biol 2000;136:285-96.

55. Bashiardes S, Veile R, Wise CA, Szappanos L, Lovett M. Positional cloning strategies for idiopathic scoliosis. Stud Health Technol Inform 2002;91:86-9.

56. Song A, Zhou Y. Comparative genomics- an application for positional cloning of the weissherbst mutant. Methods Cell Biol 2004;77:459-73.

enzymes sites. A variety of enzymatic approaches for the discrimination of allelic variants has since been described. However not all can be considered high throughput in their present format; data suggest that they have the potential to be developed into fully automated assays.

Restriction fragment length polymorphism and amplified fragment length polymorphism

Restriction fragment length polymorphism (RFLP) is comprised of PCR amplification of a fragment of interest and the subsequent digestion with a restriction enzyme [70]. Because of its simplicity and straightforwardness, the method has been extensively used and is still popular, although it confers certain limitations. Only a subset of polymorphisms that happen to reside in an endonuclease restriction site can be studied with the conventional method. The approach has recently been generalized by swapping the amplification and restriction events -the DNA sample is first digested by a combination of restriction enzymes and then amplified. Amplification of the fragments is achieved by ligating adapter sequences on both sides of the fragments, which, together with the restriction target sequence, can be amplified using a specific end-labeled primer [amplified fragment length polymorphism (AFLP)]. A unique triplet of nucleotides following the adapter and restriction enzymes sequences on the primers enables specific amplification of fragments beginning or ending with the complementary triplet. This fingerprinting method allows co-amplification of a high number of restriction fragments without the knowledge of the nucleotide sequence [71]. A complicating factor may be that the successful implementation is dependent on the completeness of digestion in both studied and reference samples.

Cleavase fragment length polymorphism analysis

The invasive cleavage assay [72] is a probe cycling, signal amplification reaction used for detection of single nucleotide polymorphisms (SNPs) [73] and quantitative determination of gene expression and viral load. The reaction requires two synthetic oligonucleotides, called the 'upstream oligonucleotide' and 'probe', that anneal to the target sequence with an overlap of 1 nt. This creates a bifurcated overlapping complex that resembles a structure generated during strand displacement DNA synthesis. Structure-specific 5'-nucleases, whose primary cell function is believed to be processing of Okazaki fragments [74, 75], cleave the bifurcated substrate at the site of the overlap, releasing the 5'-arm and one base paired nucleotide of the probe [76, 77]. The cleaved 5'-arm serves as a signal indicating the presence of the target in an analyzed sample. By performing the invasive reaction at the probe melting temperature (T_m), multiple cleavage events can be achieved for each target. Typically, an invasive signal amplification reaction generates 30-50 cleaved probes/target/min, resulting in 10^3-10^4-fold signal amplification in a 1-3 h reaction [78]. By combining two invasive reactions into a serial assay, the signal amplification can be increased to 10^7-fold, which is sufficient to detect 600 copies of unique sequence in samples of

human genomic DNA in 2-4 h using a standard fluorescence plate reader [79].

The number of different formats that can be applied for signal detection emphasizes the versatility of the invasive cleavage assay. These include electrophoresis [80, 81], microplate enzyme-linked immunosorbent assay (ELISA) [65] and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry methods [82]. More recently, a fluorescence resonance energy transfer (FRET) methodology enabled homogeneous detection of SNPs by the invasive cleavage reaction using zeptomole (10^{-21} mol) amounts of target DNA [79].

Adapting the invasive cleavage assay to a solid phase format presents the possibility of analyzing multiple SNPs in parallel. At the present time, SNP detection using the invasive cleavage reaction is performed in 96-well microplates with nanogram amounts of human genomic DNA per SNP [79].

Electrophoretic discrimination of allelic variants

High-Performance DNA sequencing

DNA sequencing has been proclaimed the "gold standard" used to establish the identity of both known and unknown sequence specific nucleotide variations [83], which in turn precipitated a number of publications questioning this statement [84, 85]. Four different dyes are used to label each of the DNA bases (A, C, G, and T) that can be electrophoresed together in the ABI sequencing system. The DNA polymerase

Table 1. Different approaches to identify and visualize genetic alterations

Approaches to identify and visualize genetic alterations	Methods and techniques
1. Enzymatic approach for discrimination of allelic variants	• Restriction fragment length polymorphism (RFLP) • Amplified fragment length polymorphism (AFLP) • Cleavase fragment length polymorphism (CFLP) • Ligation-based assays • Random amplified polymorphic DNA • [RAPD (AP-PCR)] • Direct termination PCR (DT-PCR)
2. Electrophoretic discrimination of allelic variants	• Single strand conformation polymorphism (SSCP) • Heteroduplex analysis • Fragment analysis • High-performance DNA sequencing • Doublex sequencing • Single sequencing reaction (SSR) • Clipper sequencer
3. Solid-Phase determination of allelic variants	• Oligonucleotide arrays • High-density oligonucleotide arrays for hybridization analysis • Minisequencing primer extension analysis • Fiber optic DNA sensor array
4. Chromatographic methods for discrimination of allelic variants	• Denaturing high performance liquid chromatography (DHPLC)
5. Physical methods for discrimination of allelic variants	• Differential sequencing with mass spectrometry • Fluorescence exchange-based methods

57. Hasegawa Y. Systematic genome-wide approach to positional candidate cloning for identification of novel human disease genes. Tanpakushitsu Kakusan Koso 2004;49:2894-901.

58. Moir DT, Dorman TE, Smyth AP, Smith DR. A human genome YAC library in a selectable high-copy-number vector. Gene 1993;125:229-32.

59. Korenberg JR, Chen XN, Sun Z, et al. Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. Genome Res 1999;9:994-1001.

60. Spoerel NA, Kafatos FC. Isolation of full-length genes: walking the chromosome. Methods Enzymol 1987;152:598-603.

61. Kim UJ, Shizuya H, Kang HL, et al. A bacterial artificial chromosome- based framework contig map of human chromosome 22q. Proc Natl Acad Sci USA. 1996;93:6297-301.

62. Rubio JP, Triglia T, Kemp DJ, de Bruin D, Ravetch JV, Cowman AF. A YAC contig map of Plasmodium falciparum chromosome 4: characterization of a DNA amplification between two recently separated isolates. Genomics 1995;26:192-8.

63. Lu-Kuo JM, Le Paslier D, Weissenbach J, Chumakov I, Cohen D, Ward DC. Construction of a YAC contig and a STS map spanning at least seven megabasepairs in chromosome 5q34-35. Hum Mol Genet 1994;3:99-106.

may discriminate between the differently labeled ddNTPs, resulting in different intensities of terminating signals [i.e., uneven peaks, which may undermine the detection efficiency of heterozygotes, a problem practically solved with the use of stronger fluorescent signals -the BigDyes'(Applied Biosystems)]. This increased precision is combined with high throughput in the currently available systems (ABI 310 and ABI 3700). On the other hand, the Pharmacia-ALF sequencing technology (Amersham Pharmacia Biotech, Piscataway, NJ, USA) employs a single-dye chemistry format in which Cy5 is used to label DNA fragments for the detection of laser-induced fluorescence. This contributes to obtaining easily interpretable raw data, high accuracy, and no mobility differences between fragments caused by different-size fluorophore labels. However, the disadvantages are lower throughput and higher labor cost. Several attempts to increase the throughput of ALF DNA sequencing have been described. Doublex sequencing involves the combination of an ALF DNA sequencer and an ALFexpress DNA sequencer, which allows the detection of two different labels, fluorescein isothiocyanate (FITC) and Cy5. With the use of two differently labeled primers, one can sequence both strands in a single reaction and gel run [86]. Another approach for the detection of known SNPs is to use single sequencing reactions (SSR) in which termination reactions for only one of the four ddNTPs are generated and analyzed [87]. New platforms for rapid automated DNA sequencing have been introduced using ultra-thin gels and thin capillaries (the Clipper sequencer) [81], and protocols for simultaneous PCR amplification and cycle sequencing in a single tube (CAS, DEXAS, or CLIP) have been described [88-90].

Solid-Phase determination of allelic variants

Oligonucleotide arrays

Short oligonucleotides (8-80 bases) or longer cDNA (more than 100 bases) attached to a solid support (microscope slide or microplate) are referred to as microarrays. High-density oligonucleotide arrays on miniature glass supports are also designated DNA chips. The attachment of one of the components of a reaction to a solid support allows fast separation of the different components of the reaction by simple washing. The immobilized nucleic acid may be either the known sequence, referred to as the probe, or the unknown (questioned or interrogated) sequence, also referred to as the target [91].

DNA microarrays are simple and versatile tools for experimental explorations of genome structure, gene expression programs, gene function, and cell and organismal biology.

There are many types of DNA microarray, but the two main platforms are GeneChip oligonucleotide microarray, in which a known oligonucleotide (probe) is synthetisized directly on glass arrays *in situ*, and DNA-spotted microarrays, in which the probe, generally a PCR product, is spotted onto a substrate by precision robotic [6, 9]. In both cases, an experiment consists of hybridization of a fluorescently labeled cDNA against the probe. The raw data are obtained by measuring the resulting intensity values of the spots by fluorescent microscopy with laser excitation (Fluorescent scanners).

Of great interest recently has been the potential application of microarray technology to the identification of SNPs.

Oligonucleotide arrays can be used to determine the nucleotide sequence of DNA. Genomic DNA of many individuals in a population can be sequenced with microarrays in order to identify commonly occuring polymorphisms. Since the human genome sequence has been determined, the "reference sequence" for this region is known and can be used to design an "oligonucleotide array" chip to determine, within a DNA region, the frequency and distribution of nucleotide variation in a population. Genomic DNA is isolated from individuals and then PCR is used to amplify as well as fluorescent-tag it. This tester DNA is hybridized to the "oligonucleotide microarray", which can distinguish single-base mismatches from perfect hybrids [31].

The ability of hybridization with allele-specific oligonucleotides (ASO) to detect a single base mismatch was first shown in 1979 [92], and then used to detect the sickle-cell mutation in the β-globin gene by Southern blot hybridization to human genomic DNA in 1983 [93]. Identification of a single base change in the 6 x 109 bp of the diploid human genome is, however, a demanding task. Not until the PCR technique was invented [91, 94], did it become possible to design useful assays for genotyping SNPs in complex genomes.

Some of the early SNP-genotyping assays used for PCR products were based on ASO hybridization in dot blot [95] or reverse dot blot [96] formats. The reverse dot blot format can be viewed as the precursor of the high-density microarray-based methods for multiplex genotyping of SNPs by ASO hybridization [97-99]. The use of ASOs as hybridization probes or as PCR primers is the basis for the SNP genotyping assays that are referred to as 'homogeneous', because they contain no separation steps and are monitored in real time during PCR. These assays are frequently used for large-scale genotyping of SNPs today [100-102]. The solid-phase assays for enzyme-assisted genotyping, using a DNA ligase [103] or a DNA polymerase [104], were also introduced more than a decade ago. Because the enzyme-assisted methods have proven to be more robust and to provide more specific allele distinction than ASO hybridization [105], these methods have been multiplexed, automated and adapted to various detection strategies, and they provide most of the current high-throughput SNP-genotyping platforms.

In methods based on single nucleotide primer extension -minisequencing- the distinction between genotypes of the SNPs is based on the high accuracy of nucleotide incorporation by the DNA polymerase [104]. The primer extension reaction is robust, allowing specific genotyping of most SNPs at similar reaction conditions. These features are advantageous for high throughput applications because the effort required for assay design and optimization are minimized. Consequently, single nucleotide primer extension is gaining acceptance as the reaction principle of choice for high-throughput genotyping of SNPs, and

64. Andersen JS, Svensson B, Roepstorff P. Electrospray ionization and matrix assisted lazer desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. Nat Biotechnol 1996;14:449-57.

65. Argüello JR, Little AM, Pay AL, Gallardo D, Rojas I, Marsh SGE, Goldman JM, Madrigal JA. Mutation detection and typing of polymorphic loci through double strand conformation analysis. Nat Genet 1998; 18:192-4.

66. Arnold N, Gross E, Schwarz-Boeger U, Pfisterer J, Jonat W, and Kiechle M. A highly sensitive, fast, and economical technique for mutation analysis in hereditary breast and ovarian cancers. Hum Mutat 1999;14:333-9.

67. Braun A, Little DP, Koester H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 1997;43:1151-8.

68. Buetow KH, Edmonson ME, Cassidy AB. Reliable identification of large number of candidate SNPs from public EST data. Nat Genet 1999;21:323-5.

69. Chee M, Yang R, Hubbel E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS. Accessing genetic information with high density DNA arrays. Science 1996;274:610-4.

70. Mirhendi H, Makimura K, Zomorodian K, Yamada T, Sugita T, Yamaguchi H. A simple PCR-RFLP method for identification and differentiation of 11 Malassezia species. J Microbiol Methods 2005; 61:2814.

71. Zhou J, Xue YL, Guan YX, Yang YD, Fu SB, Zhang JC. Association study of apolipoprotein e gene polymorphism and cerebral infarction in type 2 diabetic patients. Yi Chuan 2005;27:35-8.

72. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Sander TJ, Arco DA, Neri BP, and Brow MAD. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotides probes. Nat Biotechnol 1999;17:292-6.

73. Brookes AJ. The essence of SNPs. Gene 1999;234:177-86.

74. Rao KV, Stevens PW, Hall JG, Lyamichev V, Neri BP, Kelso DM. Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. Nucleic Acids Res 2003;31(11):66.

75. Rao KV, Stevens PW, Hall JG, Lyamichev V, Neri BP, Kelso DM Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. Nucleic Acids Res 2003;1:31(11):66.

76. Chen Y, Shortreed MR, Peelen D, Lu M, Smith LM. Surface amplification of invasive cleavage products. RNA 2003;9: 1552-61.

77. Lyamichev V, Brow MAD, Varvel VE, and Dahlberg JE. Comparison of the 5'nuclease activities of Taq DNA Polymerase and its isolate nuclease domain. Proc Natl Acad Sci USA 1999;96:6143-8.

78. Lyamichev V, Kaiser MW, Lyamicheva N, Hall JG, Ma W, and Neri BP. Experimental and theoretical analysis of the invasive signal amplification reactions. Biochemistry 2000;39:9523-32.

has been adapted to various assay formats, detection strategies and technology platforms.

Physical methods for discrimination of allelic variants

Fluorescence exchange-based method

Fluorescence polarization (FP) single base extension is a fairly robust and specific assay for genotyping SNPs. A tremendous advantage of this assay is that it is homogeneous so that all reactions can be carried out stepwise in a single tube (or well). This assay utilizes the physical property of light polarization in order to discriminate between alleles. In this strategy, the nucleotides added in the primer extension reaction are labeled with two different fluorescent tags (TAMRA or ROX) corresponding to each of the two alleles. By measuring which fluorescent marker is attached, one can determine which allele is present. When polarized light falls upon a small fluorescent molecule, the molecule will become excited and emit the light in all directions because the emission is accompanied by kinetic motion and rotation. However, when the fluorescent molecule is tethered to a larger molecule, this rotation will be hindered and a greater proportion of fluorescence emission will be in the polarized plane. The measurement of the relative amount of polarized emission versus non-polarized emission will differ between bound and unbound fluorescent molecules. Thus, when a fluorescent dye labeled nucleotide is attached during a primer extension reaction the increased measure of polarization will indicate whether a particular nucleotide had been attached to the extension primer.

The FP genotyping method is generally robust but it is limited by the number of PCR reactions one can do in a single well. This eliminates the possibility of multiplexing which is essential to cost-reduction and an increase in throughput necessary for affordable candidate gene and linkage mapping studies.

A second method of detecting primer extension reactions is analogous to the detection method used for the DNA sequencing reaction. Applied Biosystems (formerly, ABI/PE whose parent company is Applera) manufactures instrumentation for DNA sequencing and the same instruments can be used to discriminate between allele-specific primer extension products by the size and fluorescence of the product. Indeed, PE Biosystems and Orchid Biosciences have announced a collaborative agreement that enables PE Biosystems to develop a product based on Orchid's SNP-IT reaction chemistry and a detection platform using PE Biosystems' instrumentation and detection technology.

Affymetrix, Inc, the leading provider of DNA oligonucleotide chips (used primarily for mRNA expression analysis) has also entered into collaboration with Orchid. Orchid's primer extension technology will be coupled to Affymetrix's solid chip format and fluorescence detection system. This is an example of a technique that uses both a homogeneous liquid reaction and a solid support assay environment. In this system a bipartite oligonucleotide is designed for each SNP to be interrogated (see below). One part of the oligonucleotide is a traditional primer for primer extension and corresponds to the region adjacent to the

SNP. The other segment of the oligonucleotide consists of a nucleotide "tag" which serves as a capture probe and is unique to each SNP being assessed. Thus, if there are 100 SNPs being assessed there are 100 different capture probes. Following primer extension and the incorporation of a fluorescently labeled nucleotide numerous reactions are captured on a DNA chip by way of the capture probes. In this format, numerous primer extension reactions can be performed simultaneously and each individual SNP reaction can be captured and localized at a particular position on the solid oligonucleotide chip. The genotype of each SNP is determined by measuring fluorescence.

A third method is the Mini-sequencing technique related to primer extension. Unlike primer extension where a primer is extended by a single nucleotide, mini-sequencing extends a primer by several nucleotides. Pyrosequencing, Inc. has developed a format for mini-sequencing. In traditional DNA sequencing, the reaction products are separated according to size on a polyacrylamide gel and both size and fluorescence is used to identify the nucleotide at a given position.

To obviate the need for gel electrophoresis, Pyrosequencing technology is performed in a microtiter plate format. In brief, the identity of the nucleotides surrounding and including the SNP position is determined by adding known nucleotides sequentially to a DNA sequencing reaction [106, 107]. If the nucleotide is incorporated into the primer, a cascade of reactions involving the enzymes sulfuryase and luciferase as well as the pyrophosphate released by nucleotide incorporation occur and ultimately produce detectable light, which indicates nucleotide identity [108]. The enzyme apyrase degrades the excess nucleotides (which could interfere with the subsequent reaction) and enables another cycle of Pyrosequencing.

In contrast to most other techniques for SNP genotyping, Pyrosequencing technology not only provides information about the variant base but also takes into account the sequence adjacent to the SNP. The surrounding sequence verifies the SNP scores, reveals sequence artifacts and thereby eliminates false results. With the help of dedicated software, SNP genotypes are scored automatically to further speed up analysis and improve reliability. An additional benefit of Pyrosequencing is that the assay can be performed in a dedicated 96-well-plate or 384 well microtitre-plate [109] format to facilitate automation and provide a suitable platform for large-scale SNP genotyping efforts. Due to its flexibility one or a few samples can be analyzed simultaneously for several different SNPs, or a single SNP scored in 96 different samples. A limitation of the method is that the sequential identification of bases prevents genotyping of several SNPs per reaction in diploid genomes.

Differential sequencing with mass spectrometry

Sequencing with mass spectrometry is an attractive alternative for high-throughput detection of polymorphism because it allows the generation of accurate data in seconds per sample without any probes or target labeling [110]. This method can be easily automated. Design of assays based only on local sequence allows automated assay design with uniform assay

79. Hall JG, Prudent JR, Mast AL, Dahlberg JE, Elis PS, Law SM, Allawi HT, Reinaldo LP and Lyamichev V. Sesitive etection of DNA polymorphisms by the serial invasive signal amplification reaction. Proc Natl Acad Sci USA 2000;97:8272-7.

80. O'Connell CD, Atha DH, Oldenburg MC, Tian J, Siebert M, Handrow R, Heisler L, and the Arruda M. Detection of P53 gene mutation: analysis by single- strand conformation polymorphism and cleavase fragment length polymorphism. Electrophoresis 1999;20:1211-23.

81. Sander T, Olson S, Hall JG, Siebert M, Neri B, Heisler L, and Arruda M. Comparison of detection platform and post-polymerase chain reaction DNA purification methods for use in conjunction with clavase fragment length polymorphism analysis. Electrophoresis 1999;20:1131-40.

82. Oldenburg MC, and Siebert M. New cleavase fragment length polymorphism method improves the mutation detection assay. Biotechniques 2000;28:351-7.

83. Dianzani I, Camaschella A, Ponzone A, and Cotton RGH. Dilemmas and progress in mutation detection. Trends Genet 1993;9:403-5.

84. Cotton RGH. Slowly but surely towards better scanning for mutations. Trends Genet 1997;13:43-6.

85. Eng, C, and Vjig J. Genetic testing: the problems and the promise. Nat Biotechnol 1997;15:422-6.

86. Plaschke J, Voss H, Hahn M, Ansorge W, and Schackert HK. Doublex sequencing in molecular diagnosis of hereditary diseases. BioTechniques 1998;24:838-41.

87. Kristensen T, Nedelcheva KV, Borresen-Dale A-L. High throughput screening for known mutations by automated analysis of single sequencing reactions (SSR). BioTechniques 1998;24:832-5.

88. Yager TD, Baron L, Batra R, Bouevitch A, Chan D, Chan K, Darach S, Gilchrist R, Ismailov A, Lacroix J.-M, *et al*. High performance DNA sequencing and the detection of mutations and polymorphisms on the Clipper sequencer. Electrophoresis 1999;20:1280-300.

89. Kilger C, Paabo S. Direct DNA sequence determination from total genomic DNA. Nucleic Acids Res 1997;25:2032-4.

90. Ruano G, Kidd KK. Coupled amplification and sequencing of genomic DNA. Proc Natl Acad Sci USA 1991;88:2815-9.

91. Mullis KB, Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol 1987;155:335-50.

92. Wallace RB, *et al*. Hybridization of synthetic oligodeoxyribonucleotides to p-? 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 1979;6:3543-57.

93. Conner BJ, *et al*. Detection of sickle cell S-globin allele by hybridization with synthetic oligonucleotides. Proc Natl Acad Sci USA 1983;80:278-82.

94. Saiki RK, *et al*. Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985;230:1350-4.

conditions. This similarity of assay conditions permits the extensive use of robotics, which limits human error. Mass spectrometry data collection is fast and automated, based on the size of extended products.

The ionization of the analyte is a prerequisite for all ms techniques. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) [111] and electrospray ionization mass spectrometry (ESI-MS) [112] are techniques for creating ionized gas-phase DNA whose ionized molecules are accelerated in an electric field, followed by a flight through a vacuum chamber to a detector. This detection method, is based on the time taken by each particle to fly to the detector. It is inversely related to the velocity, which in turn is proportional to the mass/charge ratio of a flying particle, a specific characteristic for the samples under examination [110], since the mass/charge ratio is an intrinsic property of the molecule (fragment DNA), dependent only on base composition.

Mass spectrometry analysis has been successfully applied as a detection step in the cleavage assay or in the primer extension mini-sequencing techniques as PROBE or Pin-Point. Mutation analysis in several different types of cancer, diseases such as cystic fibrosis [113], and hypertension has been performed by PROBE. The method has been particularly useful in analyzing "difficult" sequences [e.g., tri-nucleotide repeats [114]. MALDI-MS has been used to visualize direct reads from sequencing TP53 (exons 5-8) and was shown to be able to reliably detect all mutations,

heterozygous and homozygous and false stops of the sequencing reaction [115], as well as a number of SNPs [111, 116].

Conclusions

The drug industry has and will continue to benefit from the relatively new disciplines of genomics and proteomics. Advances made in the multidisciplinary field of genomics have lead to the near completion of the once unimaginable goal of sequencing the entire human genome. Once complete, the conversion of vast amounts of DNA sequencing data into the identification of novel genes and gene products will lead to a better understanding of normal physiology and disease states. Further, proteomics holds great potential as it represents a depiction of an organism at the phenotypic rather than the genotypic level. Information gained through proteomic studies includes the characterization of novel proteins and the identification of new protein-protein interactions and pathways. Eventually, genomic and proteomic studies will lead to a better understanding of the underlying mechanisms of disease. Subsequently, a better understanding of diseases will lead to superior therapeutics with higher efficacies and reduced adverse effects. The emerging discipline of pharmacogenomics attempts to apply the innovative techniques of genomics and proteomics in order to better understand disease, drug response and eventually to produce more effective drugs with fewer side effects.

95. Saiki RK, *et al*. Diagnosis of sickle cell anemia and ß-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. N Engl J Med 1988;319:537-41.

96. Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc Natl Acad Sci USA 1989; 86:6230-4.

97. Southern EM, Maskos U, Elder JK. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics 1992; 13:1008-17.

98. Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, Yakhini Z, Bruhn L, Laderman S. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. Proc Natl Acad Sci USA 2004; 101:17765-70.

99. Zekri AR, Bassuoni MA, Hammad OM, Sakr MA, Ibrahim AA. Application of refractory fragment amplification system for detection of Egyptian variant of Familial Mediterranean Fever. Egypt J Immunol 2004;11:103-10.

100. Martin ER, *et al*. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 2000; 67:383-94.

101. Ranade K, *et al*. High-throughput genotyping with single nucleotide polymorphisms. Genome Res 2001;11:1262-8.

102. Grupe A, *et al*. In silico mapping of complex diseaserelated traits in mice. Science 2001;292:1915-8.

103. Dai SM, Chen HH, Chang C, Riggs AD, Flanagan SD. Ligation-mediated PCR for quantitative in vivo footprinting. Nat Biotechnol 2000;18:1108-11.

104. Brouwer DA, van Doormaal JJ, Muskiet FA. Clinical chemistry of common apo-lipoprotein E isoforms. J Chromatogr B Biomed Appl 1996;678:23-41.

105. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 1997; 7:606-14.

106. Olsson C, Liljedahl U, Syvanen AC. Quantitative analysis of SNPs in pooled DNA samples by solid-phase minisequen-cing. Methods Mol Biol 2003;212:167-76.

107. Ronaghi M, Karamohamed S, Pe-ttersson B, Uhlen M, and Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 1996; 242:84-9.

108. Tully G, Sullivan KM, Nixon P, Stones RE, Gill P. Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. Genomics 1996;34:107-13.

109. Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by realtime pyrophosphate

DNA sequencing. Genome Res 2000;10: 1249-58.

110. Rowley PT, Loader S, Kaplan RM. Prenatal screening for cystic fibrosis carriers: an economic evaluation. Am J Hum Genet 1998;63:1160-74.

111. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Biotechnology 1992;24:104-8.

112. Kahraman S, Kiykim AA, Altun B, Genctoy G, Arici M, Gulsun M, Erdem Y, Yasavul U, Turgan C, Caglar S. Apolipo-protein E gene polymorphism in renal transplant recipients: effects on lipid metabolism, atherosclerosis and allograft function. Clin Transplant 2004; 18:288-94.

113. Davignon J, Gregg RE, Sing CF. Apoli-poprotein E polymorphism and atherosclerosis. Arteriosclerosis 1988;8:1-21.

114. Pastinen T, Partanen J, Syvanen AC. Multiplex, fluorescent, solid-phase mini-sequencing for efficient screening of DNA sequence variation. Clin Chem 1996; 42:1391-7.

115. Collins A, Lonjou C, and Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 1999;96:15173-7.

116. Hacia JG, *et al*. Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. Genome Res 1998;8:1245-58.