



Review

Typing of single nucleotide polymorphisms by MALDI mass spectrometry: Principles and diagnostic applications

Sascha Sauer*

Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics (Prof. H. Lehrach), Ihnestr. 73, D-14195 Berlin, Germany

Received 3 April 2005; received in revised form 17 May 2005; accepted 18 May 2005

Available online 31 August 2005

Abstract

Background: After the completion of the human genome sequencing project human genetics has now shifted its focus to DNA variation. DNA variation analysis is considered to be a key in partly understanding the mechanisms of complex diseases or varying patient responses in drug treatment. One of the major goals in genetics is finding the DNA variants that can act as diagnostic markers for predisposition to specific diseases. Moreover, in microbiology DNA variation has long been known to help discriminate and identify bacterial strains and viruses. Diagnostics based on DNA or RNA detection might be advantageous as an early-stage indication can be provided.

Methods: Many simple and efficient methods for the analysis of nucleic acids are already available. Consequently, the last few years have seen an increased in the use of large-scale analysis of nucleic acids, in basic DNA variation studies along with diagnostics. Mass spectrometry techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) can be of great use for genome variation analysis. In particular high-throughput SNP analysis by MALDI can be performed using fully integrated platforms.

Conclusions: Mass spectrometry-based procedures have promise for SNPs analysis especially for clinical diagnostics.

© 2005 Elsevier B.V. All rights reserved.

Keywords: SNP; Genotyping; DNA; Clinical diagnostics

Contents

1. Introduction	95
2. Mass spectrometry	96
2.1. Why use MALDI mass spectrometry for analysis of nucleic acids?	97
2.2. Principles of MALDI mass spectrometry	97
3. SNP genotyping by MALDI	98
3.1. Quantification	99
3.2. Applications	100
4. Molecular haplotyping	100
5. Re-sequencing	102
6. Conclusions	103
Acknowledgements	103
References	103

1. Introduction

With the availability of the human genome sequence along with sequences of other genomes including bacteria,

* Tel.: +49 30 84131565; fax: +49 30 84131365.

E-mail address: sauer@molgen.mpg.de.

DNA variation analysis enters a new stage of research. Many projects in current human genetics aim to dissect complex traits by making use of DNA markers, mainly by single nucleotide polymorphisms (SNPs) [1]. Currently, around 10 million SNPs are stored in public databases. In general, in diploid organisms as humans SNPs are biallelic polymorphisms meaning base substitutions, occurring at least once per 300–1000 bp, with the minor allele occurring at least for 1% in a population — otherwise they are termed point mutation. Although single base deletions and insertions are not strictly SNPs, they are commonly included in this definition. SNPs are currently being used as genetic markers, primarily for whole-genome scanning of genomes to gain first indication of interesting regions that contribute to the traits under investigation or in more focused candidate gene association studies [2,3]. Many SNP-based studies are still at their early stages and it is not yet clear how far SNP genotyping can eventually contribute to the genetic dissection of (complex, polygenic) traits [4]. Once an SNP or a group of SNPs has been identified as a disease marker, it can be used for diagnostics. Thus, researchers hope that SNP genotyping might in general allow for systematic screening of genetic risk variants in the future and change medicine from case-driven intervention to predictive medicine. Currently, however, only few examples such as for Crohn's disease have been shown [5]. Furthermore, SNPs are currently being used as genetic markers to distinguish related organisms, for example, in forensics, or to specifically detect microorganisms in food or in patient sera.

High throughput methods have been developed for *in vitro* SNP genotyping in different fields of applications. The (cost-) efficiency of the different methods can in principle be measured by the number of SNPs and DNAs under investigation. Technologies using array-based read-out combined with high-multiplex generation of SNP products such as the Affymetrix 10 or 100 K chips [6], the Illumina bead arrays [7] or the ParAllele chips [8] are efficient when several thousand SNPs are analyzed in parallel. Solution-based one-step procedures such as the homogenous 5'-nuclease TaqMan™ assay [9] using real-time fluorescent detection or MALDI mass spectrometry-based procedures [10] are particularly economical in studies where up to 50 SNPs have to be typed in cohorts of several thousand DNAs.

Many of the SNP typing methods require several steps. Short operational time is desirable because it reduces labor costs. Short reaction sequences become even more important in clinical diagnostics, where rapid analysis can be crucial to save the life of a diseased (e.g., infected) patient. Many of the SNP genotyping techniques are currently being optimized in terms of higher throughput and cost efficiency; however, it seems that the focus is still rather on making SNP genotyping procedures more cost-efficient for fundamental genome research by parallelization and miniaturization of assays [11] rather than simplifying assays for diagnostic routine. While current technology developments

thus mainly concentrate on the requirements of research laboratories, in clinics some specific criteria are important that are sketched here: Firstly, a simple and short protocol, which can be easily applied for medium throughput applications, is required. It is less likely in diagnostic laboratories that thousands of different SNPs have to be typed in one individual patient DNA. Rather the repetitive day-to-day analysis of a limited number of SNPs in several thousand DNA samples per day seems more realistic. In a clinical environment, in many cases hospitalization costs become an important issue, which lie in the range of several hundred, if not thousand Euro or US-Dollar per day. Therefore in clinics the raw genotype price is not as important as in a scientific environment. Although the price per genotype is high due to the use of expensive modified oligonucleotide probes, the TaqMan™ assay [9] is one of the most frequently used techniques in clinical laboratories. This can be explained by its short reaction sequence (basically a one-step procedure requiring about 2 h), and, provided that suitable standards are used for clustering of fluorescence signals, TaqMan™ results are, in general, sufficiently accurate. Furthermore, fully integrated platforms including assay design and analysis software are available facilitating easy application.

2. Mass spectrometry

In principle, mass spectrometry-based technologies provide attractive solutions for biomolecule analysis because they can be used for direct and fast detection of proteins, peptides and nucleic acids [12]. No labeling and clustering of signals are required. With mass spectrometry not only one detection platform can be used for many applications but data obtained can be easily scored and handled by automated data management systems. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [13] and electrospray ionization (ESI) mass spectrometry [14] have revolutionized the analysis of nucleic acids and other biomolecules. It has been shown that ESI can be routinely coupled to (liquid) chromatography to provide separation of DNA molecules as well as sample clean-up [15]. ESI can be used to detect large molecules (>20 kDa) and was applied to analyse small PCR products [16,17]. However, only a few principles using ESI or related procedures based on atmospheric pressure chemical ionization (APCI) have been developed for SNP detection. The reason for this might be the ionization process that requires very pure samples and does presently not allow for as high-throughput testing as MALDI (described below). Moreover, very sophisticated and expensive instruments such as ESI-FT-ICR (Fourier transform ion cyclotron resonance) mass spectrometers [18] would be the best choice to perform the procedures accurately and with high resolution. Additionally, sequential analysis of many samples per day seems to remain more difficult with these equipments than with robust

MALDI–TOF mass spectrometers. Other approaches using ion trap, quadrupole or time-of-flight ESI mass spectrometers focused on generating and detecting small products of SNPs [19], indirect monitoring of substrates of allele-specific reactions such as primer extension [20], or analyzing tags such as photolabile groups attached to allele-specific products [21]. Particularly, the method of Kokoris et al. using photolinker chemistry [21] could be powerful in terms of throughput; however, it seems that this approach has been commercialized by Qiagen (Hilden, Germany) with marginal success. The requirement of several molecular biological and purification steps might represent important disadvantages of these procedures. However, since ESI and related methods do presently not allow for as high-throughput and are hence not as widely used as MALDI this review concentrates on MALDI-based technologies for SNP analysis.

2.1. Why use MALDI mass spectrometry for analysis of nucleic acids?

A characteristic feature of mass spectrometry is that the detection relies on a physical property (the mass to charge ratio) of target sequences, in contrast to many other methods based on indirect detection making use of fluorescent or more old-fashioned radioactive reporter tags. Mass analysis can contribute to highly accurate data [10]. The sensitivity limitation of available MALDI mass spectrometers is in the range of ~ 1 fmol per analyte. Current MALDI instruments are capable of recording about 1 analysable mass spectrum and due to the fact that MALDI provides many detection channels it can be used to record as many data points per experiment. In this way multiple DNA products can be analyzed in a single experiment (multiplex analysis). Moreover, due to the chemical nature of nucleic acids, which, in contrast to peptides are rather monotonous molecules virtually displaying no differences in desorption efficiency, quantification is routinely feasible. Based on

these features of MALDI – accuracy, high-throughput capability, multichannel detection, and quantification – several procedures and software for SNP analysis were successfully introduced; some of them suitable for efficient SNP diagnostics [10,22].

2.2. Principles of MALDI mass spectrometry

The principle of MALDI is shown in Fig. 1. Irradiation of co-crystals formed by suitable small and acidic organic molecules, termed the “matrix”, and biomolecules with a pulsed laser (usually an UV laser) at a wavelength close to a resonant absorption band of the organic (aromatic) molecules induces an energy transfer and desorption process, evaporating matrix and analyte ions into the gas-phase [10,12,23]. A common matrix used for nucleic acids is 3-hydroxypicolinic acid [24]. By MALDI predominantly either positively or negatively single charged molecules are observed [25]. The ions are generated by a proton-transfer reaction of matrix and analyte molecules in the gas-phase. After ionization, matrix and analyte molecules are accelerated by an electric field. Usually, MALDI–MS is performed with time-of-flight separation (MALDI–TOF–MS). In this way molecules are guided – with virtually the same kinetic energy – by ion optics into a field-free flight-tube where they are separated by their mass to charge ratios before they finally reach the detector.

The equipment of MALDI mass spectrometers with delayed extraction [26] has greatly improved the resolution of mass signals allowing the distinction of nucleobase substitutions in the range of 1000–8000 Da, which approximately corresponds to DNA lengths of 3–24 nucleobases. One contributing factor to this limitation is probably the size-dependent tendency of the phosphodiester backbone of the DNA to fragment during the MALDI process, which results in a loss of signal intensity for intact DNA [23]. An additional contributing factor to this restriction could be a bias of

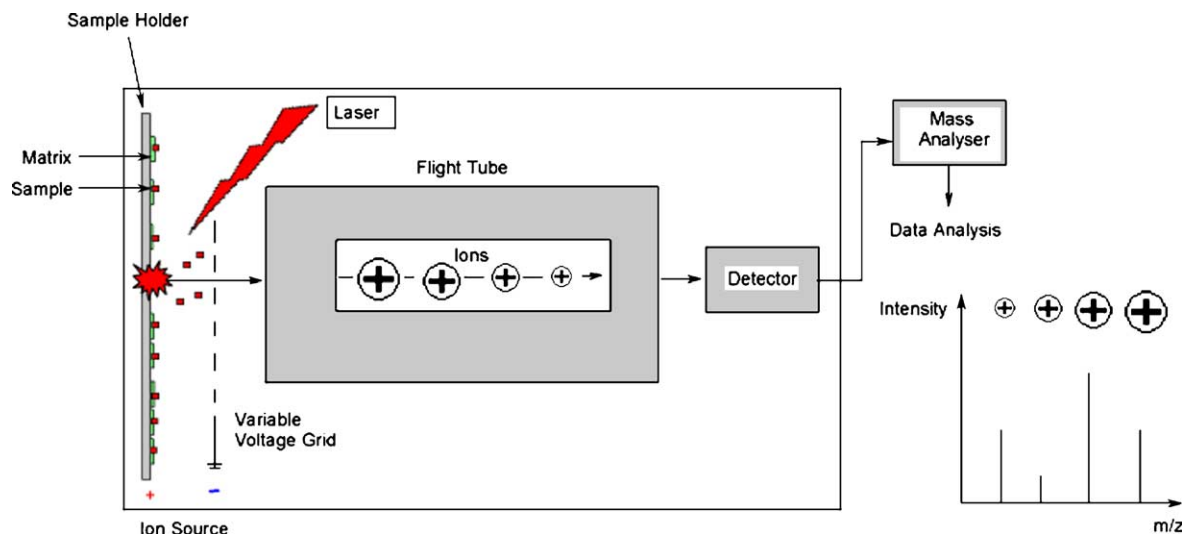


Fig. 1. The principle of MALDI is shown. The figure is simplified and the scale is not proportional. For more details see text.

MALDI towards smaller DNA oligomers. Interestingly, RNA molecules are more stable in MALDI due to the additional 2'-hydroxyl group that stabilizes the glycosidic bond leading to significantly reduced depurination and fragmentation of the whole oligomer [27].

The main problem in detecting nucleic acids by mass spectrometry is related to the negatively charged sugar-phosphate backbones [10]. Each phosphate linkage provides a site of negative charge in solution so that nucleic acids carry as many negative charges as phosphate bridges. The affinity of the phosphate residues for alkali counterions commonly used in PCR buffers, such as sodium and potassium, is not high enough to result in complete saturation. Salt ions interfere in the ionization process by inducing adducts with nucleic acids, which causes significantly reduced signal intensities. Problems with alkali ions can be alleviated by adding ammonium ions during matrix preparation [28]. However, to satisfactorily solve this problem, stringent purification procedures were introduced for removal of salt ions and detergents that can disturb matrix crystallization. Protocols include use of magnetic bead separation, reversed-

phase column purification, gel filtration, ion exchange chromatography and ethanol precipitation [10]. Using DNA modification chemistry termed “charge-tagging” [29,30] in the molecular biological procedure for generating products of SNPs, purification could be circumvented by a simplified, one-pot reaction method called the GOOD assay [31–34]. Applying “charge-tag” chemistry, the detection sensitivity is increased ~100-fold compared to unmodified DNA and DNA products as well as peptide nucleic acids [35] are rendered insensitive to impurities in MALDI by conditioning products to carry either a single fixed positive or a single fixed negative charge.

3. SNP genotyping by MALDI

In the mid-90s, several approaches combined enzymatic Sanger-sequencing or related procedures with MALDI read-out to provide an efficient alternative to relatively slow slab-gel or capillary-based technology [36,37]. However, it became apparent that the highlighted problems with MALDI

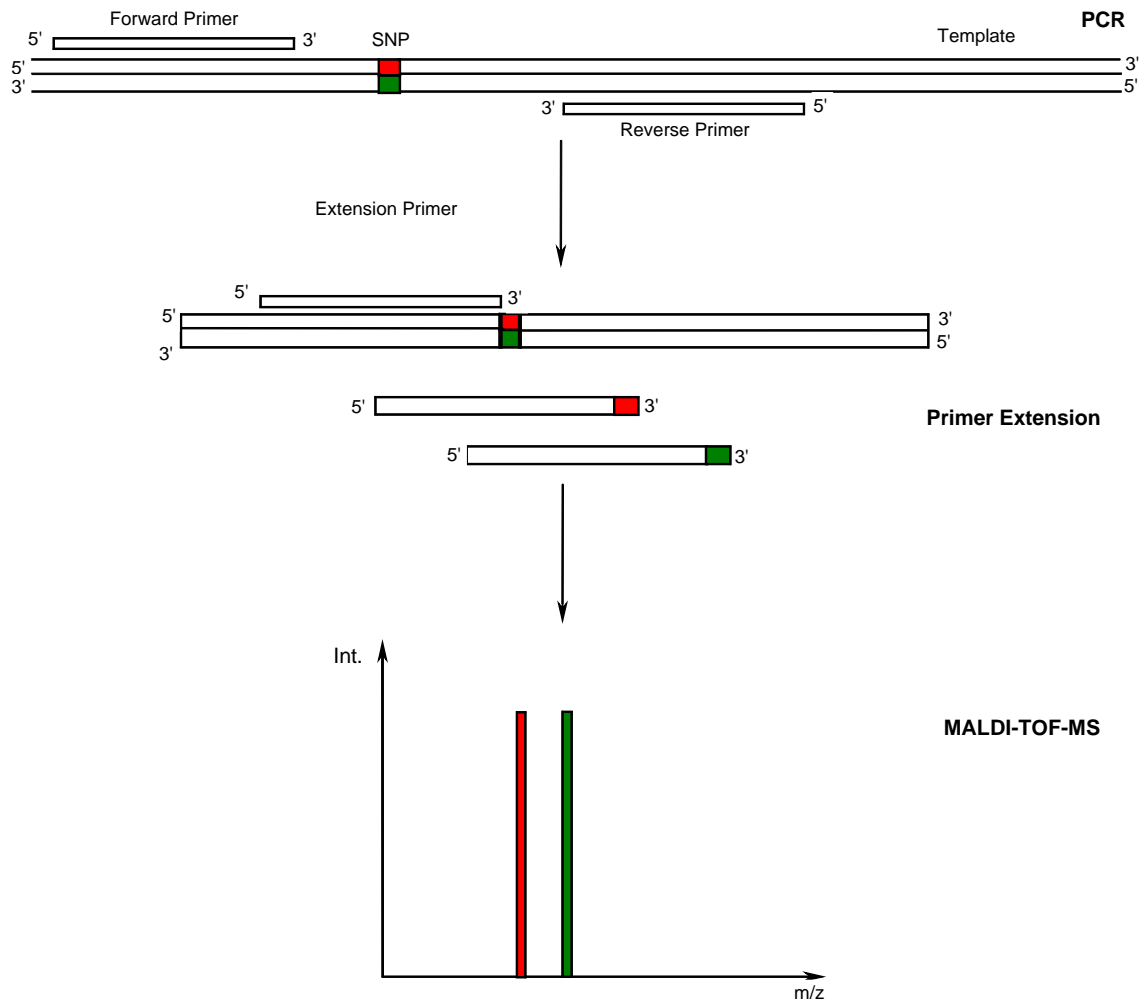


Fig. 2. The principle of the primer extension reaction approach is shown. A DNA polymerase specifically extends an oligonucleotide upstream of a SNP with a set of ddNTPs (and optionally dNTPs) on a PCR product resulting in allele-specific products for MALDI analysis.

such as limited size range of detection, fragmentation, adduct formation, and resolution severely impeded to accurately and efficiently analyse DNA sequence ladders. This was the reason why molecular biological sample preparation approaches for MALDI have shifted from sequencing to related primer extension reactions; coinciding with the increasing popularity of SNPs as markers in human genetics.

Primer extension has – not only in mass spectrometry – become one of the most widely used procedures for the generation of products of SNP because it is robust, flexible, easy to design and, what is crucial for MALDI, it generates products with masses of around 5000–8000 Da that can be easily detected [10,38]. The principle of the procedure is shown in Fig. 2. A reaction of a PCR product with an extension primer that is chosen upstream of the SNP under investigation, ddNTPs (and optionally dNTPs), and a thermostable DNA polymerase yields allele specific products. The polymerase extends the 3'-end of the primer by specifically incorporating nucleotides that are complementary to the DNA template. The extension reaction terminates at the first nucleobase in the template where a nucleotide occurs that is complementary to one of the ddNTPs in the reaction mix. A thermostable DNA polymerase employed in a temperature-cycled reaction leads to linear amplification of the extended primers. As several steps are required for all of these procedures, operational time becomes an important issue. It can be estimated that all these assays would require about 6 h for sample preparation prior to analysis.

Primer extension reaction approaches enable a limited degree of multiplexing. In general, 3–5 SNPs can be routinely typed in parallel. However the quality of enzymatic reactions in multiplex assays strongly depends on the SNPs that have to be combined. Surrounding DNA sequences can significantly influence the quality of the PCR and primer extension. In contrast to basic epidemiological studies, it is likely that the same SNPs have to be genotyped many times in diagnostics. In this case investing some time for optimisation of multiplexed assays seems to be worthwhile.

Sequenom (San Diego, CA; <http://www.sequenom.com>) offers a complete commercially available solution for high-throughput MALDI-based SNP genotyping by primer extension termed MassARRAY™. The Sequenom package includes assay design software, reagents, liquid handling, detection, and data analysis. A unique feature is the use of piezoelectric pipetting of nanoliter quantities of SNP products onto SpectroCHIPS™ covered with matrix spots. These chips, which are not recyclable, do significantly contribute to the cost. Current prices for a chip covered with 384 matrix spots lie in the range of about 200 € or 260 US\$. Therefore, the multiplexing of genotyping assays at the level of PCR and primer extension is proposed to lower costs deriving from the chips. Another commercially available method is the recent GenoSNIP™ assay [39] from Bruker Daltonics (Leipzig, Germany; <http://www.bdal.com>).

Many alternative methods based on PCR, primer extension and MALDI have been developed but are not available as complete integrated systems. Significantly more efforts are required to implement such procedures in a laboratory. However, the implementation of MALDI-based SNP assays, including the PinPoint assay [40,41], the VSET (very short extension) assay [42], the purification-free GOOD assay [31–34] or related assays [43], is feasible and could lead to more cost-effective analysis of SNPs than the commercially available methods. Recently, MALDI assays were introduced using biotinylated ddNTPs in the extension reaction and solid-phase capturing to provide an enrichment of primer extension products during the purification step [44,45]. This approach results in a higher degree of multiplexing. However, all mentioned assays require several steps. Therefore much more operational time is needed than in competing assays such as TaqMan™. Future developments of mass spectrometric SNP typing methods should focus on shortening and simplifying reaction sequences [46]. Such improvements can be crucial for diagnostic applications in clinical practice.

3.1. Quantification

It was postulated that several hundred thousand SNPs are required to detect subtle effects of DNA variations in complex diseases [2]. That would represent an enormous task for a SNP typing laboratory including unreasonably high costs. Therefore pooling of DNA samples has been suggested as a strategy to decrease the number of genotypes necessary for large-scale association studies. The principle of the procedure is shown in Fig. 3. A crucial point of this approach is the preparation of well-defined mixtures of different individual DNAs.

The MassARRAY™ system from Sequenom has been used in a study to analyse allele frequencies of about 9000 gene-based SNPs in DNA pools of 95 individuals [47]. In this study high reproducibility of $\pm 1.6\%$ for alleles with a frequency $>10\%$ was achieved. The deviation of the detected allele frequencies (about 0.7%) was checked by comparison of genotypes measured in pooled DNA and for individual DNAs.

A pooling approach, combined with Sequenom technology to determine allele frequencies in case and control populations, was also shown by Werner et al. [48]. The accuracy and reliability of the procedure was determined in pools of eight previously genotyped individuals. The allele frequencies were in the range of 0.1 to 0.9. The accuracy of detected SNP allele frequencies was also demonstrated using DNA pools of 142 to 186 individuals and typing for additional markers. Allele frequencies observed in the pooled samples deviated from the real frequencies by about 3%.

Shifman et al. compared MALDI-based technologies with restriction fragment length polymorphism, real-time pyrophosphate DNA sequencing, single base extension with fluorescently labeled ddNTPs, and the TaqMan™ assay to evaluate their applicability to perform accurate SNP allele

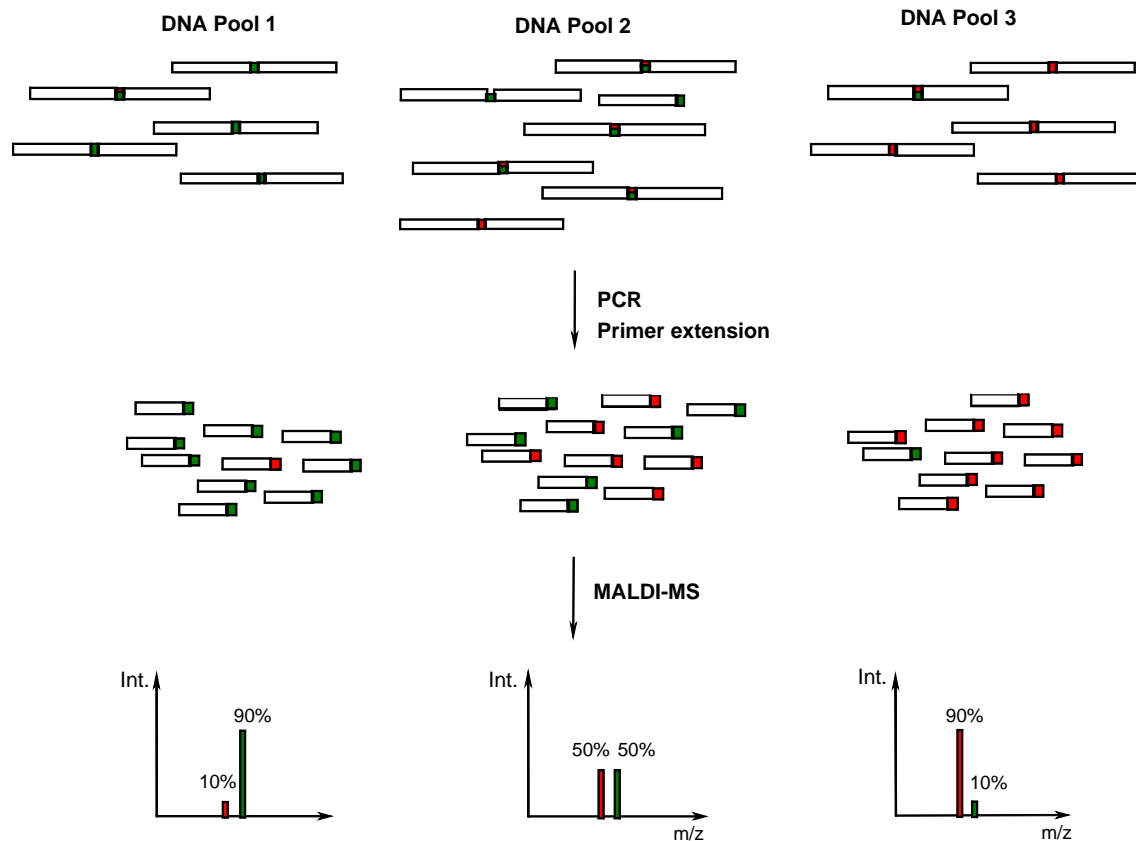


Fig. 3. Quantification of SNP alleles in pooled DNA samples. Mass spectra are obtained that reflect the corresponding proportion of SNP alleles (illustrated by red and green caskets) in the three DNA examples shown.

frequency determination in samples of pooled DNA [49]. The authors concluded that although the procedures slightly differ in accuracy, they all could be used for quantification of SNP allele frequencies in DNA pools.

In a nutshell, the combination of pooled DNA and primer extension-based MALDI assays provides a cost- and time-efficient method if common alleles are in the focus of the researcher's interest.

3.2. Applications

Primer extension and MALDI mass spectrometry can be most efficiently applied to a limited number of SNPs (<50) in large DNA cohorts. Numerous candidate gene association studies and even whole genome scans, diagnostic assays including bacterial and viral genotyping, and SNP typing in (breeding and model) animals and plants have been published in the last years, which used MALDI as detection tool. Technically, most of such studies relied on the Sequenom system. However, any other mentioned alternative MALDI procedure is principally suitable for such studies. In this section, I present two examples in which MALDI SNP typing has been used in a candidate gene and in a regional genome scan approach.

SNPs in ADAM33 were found to be associated with asthma and related traits in two German populations [50]. The

transmission disequilibrium test (TDT) was applied for association analysis in a family-based study. Additionally, association tests of SNPs were carried out in case-control DNA cohorts. Both approaches revealed SNPs that were significantly associated with asthma and related traits. This study replicated a more recently published association between asthma and ADAM33 gene variants. Unfortunately, most of the associated SNPs were found at non-identical positions in different population samples impeding their use in diagnostics.

The same group investigated allele frequencies using pooled DNA from individuals suffering from asthma, Crohn's disease (CD), schizophrenia, type 1 diabetes (T1D), and controls of healthy patients [51]. In this study 534 SNPs were selected spanning a 25-Mb region on chromosome 6p21. One of the key conclusions of the authors was that the analysis of SNP allele frequencies in pooled DNA is a valuable procedure to assign genotyping priorities in large regions of linkage.

4. Molecular haplotyping

In diploid organisms such as humans, looking at the haplotype of SNPs – the pattern of alleles on a single chromosome – can be more informative than just looking at

single SNPs [52]. Interacting SNPs can represent more powerful markers; moreover, haplotypes reflect the sequence and eventually the structure of a transcribed protein. This is one of the reasons why haplotypes have become interesting in fields such as pharmacogenetics. For n heterozygous SNPs, 2^n haplotypes are possible. Although the theoretically maximal number is rarely found, it can often be too large to accurately infer a haplotype from SNP alleles. Haplotypes can be deduced using SNP genotype data and applying computer programs that are build on a number of different algorithms. However, these programs have a limited degree of accuracy; statistical uncertainty particularly arises when many heterozygous SNPs occur in close proximity. The uncertainty associated with computational haplotyping is unacceptable in diagnostics. Molecular (direct) procedures that rely on the separation of parts or whole chromosomes are more accurate alternatives to determine the haplotype. Many methods have been introduced to directly detect the haplotype by wet-lab methods including mass spectrometry. As previously described, mass spectrometry has the advantage that several SNPs can be analyzed at the same time

(multiplex detection). Mass spectrometry approaches, such as the Sequenom technology or the GOOD assay, can be converted to a haplotyping procedure by using either single molecule PCR amplification, allele-specific PCR, or cloning procedures in the first step. The three different procedures are illustrated in Fig. 4.

Single molecule PCR amplification of diploid genomic DNA is a feasible strategy to generate haploid DNA for subsequent primer extension reactions. Such a procedure necessitates DNA samples that are diluted to a single copy. This haplotyping method is principally limited by the size of the DNA fragments. In many methods for isolation of genomic DNA, fragments of about 25 kb are frequently seen. This can impede efficient analysis of haplotypes over tens of kilobases. In practice, contamination problems can represent an additional significant drawback, particularly in clinical practice. A proof-of-principle study of the approach described here has been shown with the Sequenom technology [53].

Allele-specific PCR amplifies one allele of a heterozygous SNP and thereby a haploid piece of human DNA.

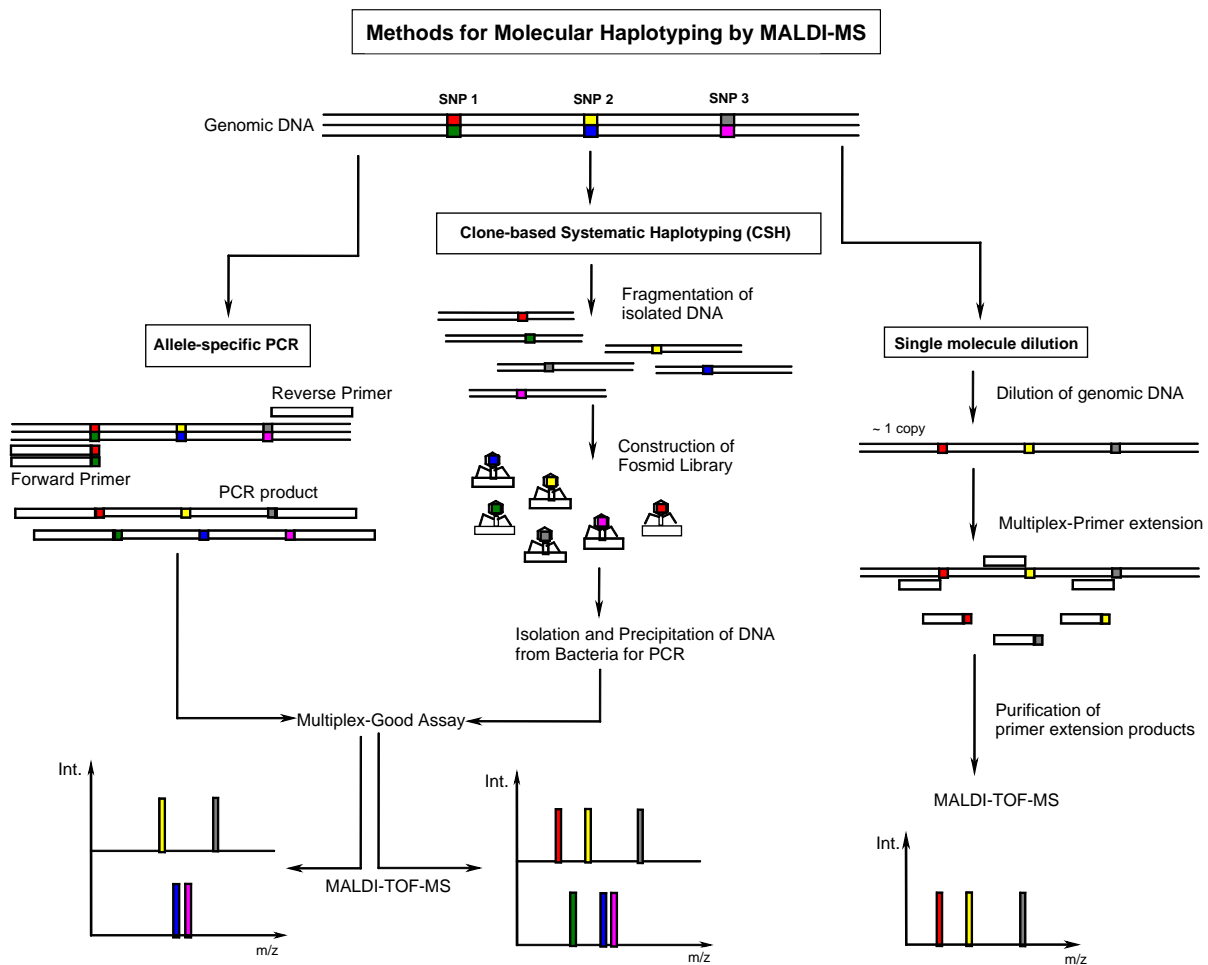


Fig. 4. Three different strategies for molecular haplotyping of SNPs by MALDI. On the left side, the combination of allele specific PCR and the GOOD assay is shown, while in the middle, the combination of the cloning approach and the GOOD assay is depicted. On the right side, the procedure using single molecule PCR and subsequent SNP typing is shown. For more details see text.

The following steps for SNP allele scoring can be similar to mentioned primer extension-based genotyping assays. A proof-of-principle study of this approach including multiplex analysis of SNPs was shown using the GOOD assay [54]. In this procedure, molecular haplotypes are monitored by looking at SNP alleles that are correlated with the allele-specific PCR. However, this method could be limited by the size of PCR products. Haplotyping using this approach has been shown on PCR products of about 1000 bp. Long-range haplotyping over several kilobases would require cumbersome walking from one fragment to the adjacent fragment.

Recently, we have developed a molecular haplotyping method in which the alleles of SNPs are statistically separated by a cloning approach using a pooling strategy and subsequently determining the phase of SNPs by multiplexed GOOD assays [55]. We have termed the procedure “Clone-based Systematic Haplotyping” (CSH). This method is particularly valuable when studying several genes in parallel or large genomic regions (>30 kb). Therefore it is more likely that this method will rather be used in basic studies than in diagnostics.

5. Re-sequencing

Recently, some related procedures for re-sequencing DNA using MALDI have been reported [56,57]. All of these methods make use of the accurate multichannel detection capability of MALDI. As is shown in Fig. 5, the procedures rely on the analysis of base-specific fragments. The fragments are produced by initially performing PCR with a primer tagged with a 5'-tail containing for example a T7 or SP6 promoter site for RNA polymerase. In the next step the PCR product serves as template for transcription into RNA. To increase information content, usually sense and antisense transcription is performed. However, there is one considerable drawback: RNA polymerases used for transcription do not have 3'–5' proofreading activity, which can increase false nucleotide incorporation. Several related RNA transcription protocols using natural or modified pyrimidine nucleotides that resist subsequent hydrolysis by RNases have been introduced. After transcription the RNA is specifically fragmented by base-specific hydrolysis using RNases such as RNase T1 (G-specific), RNase U2 (A-specific) or RNase A (C- and U-specific). Differences between the masses

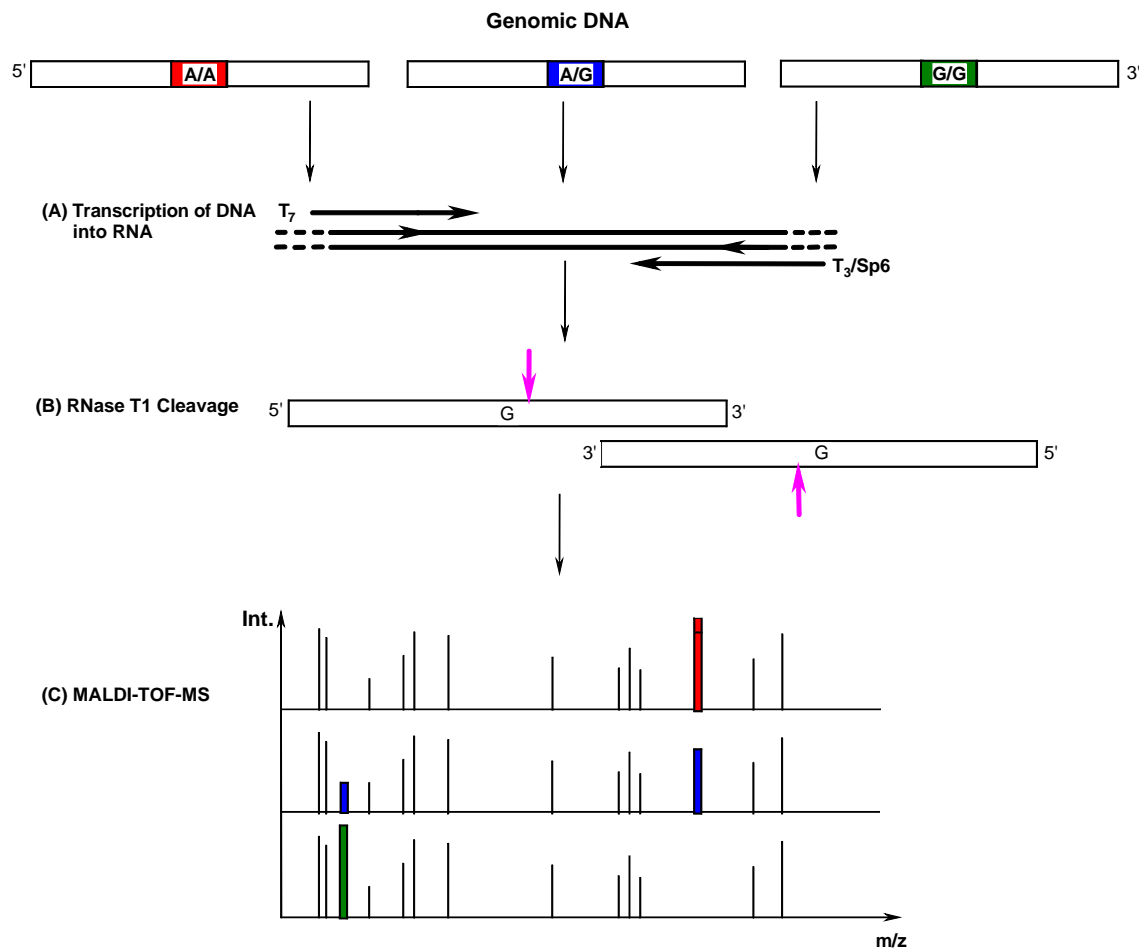


Fig. 5. The principle of re-sequencing by sequence-specific RNase digestion of transcribed PCR products and MALDI analysis. The mass signals obtained are compared with masses predicted by *in silico* digests of the reference sequence. Sequence deviations can be assigned to the reference sequence by specific software.

obtained for the sample under investigation and an *in silico* digested reference sequence allow finding sequence variations such as SNPs. The technology described here is available from Sequenom.

A similar approach was applied to bacterial identification using the 16S rRNA gene (16S rDNA) [58]. Base-specific fragmentation of PCR products using uracil-DNA-glycosylase was combined with MALDI detection. 16S rDNA signature sequences were amplified by PCR from bacteria in the presence of dUTP instead of dTTP. The PCR products were subsequently immobilised onto a solid surface coated with streptavidin to selectively produce either sense or antisense templates. Single-stranded amplicons were then treated with uracil-DNA-glycosylase to generate T-specific abasic sites and fragmented by alkaline treatment. The resulting fragment patterns were analyzed applying the same principles as described above. This procedure enabled the characterization of cultured and as-yet-uncultured bacteria.

Chemical cleavage is an alternative to enzymatic approaches. In an interesting non-enzymatic base-specific cleavage approach with primer extension products, acid-labile phosphoramidite bonds replace the 5' phosphodiester bonds of newly incorporated pyrimidine nucleotides [59]. Sequence variations as SNPs are observed by specific hydrolysis of the acid-labile bonds of the incorporated nucleotides and MALDI analysis of the resulting fragments. The analysis of genomic targets ranging from 100 to 450 bp in length is possible.

The methods described for re-sequencing might be particularly useful for discovery of new SNPs and in bacterial fingerprinting. Recently, researchers at Epigenomics (Berlin, Germany; <http://www.epigenomics.com>) have shown that these procedures can also be used to quantitatively analyse DNA methylation in the human genome [60]. Software is available for the interpretation of complex mass spectra obtained by described RNase-based approaches [61]. However, it seems that there are some caveats in these re-sequencing approaches. In the low mass range, mass signals usually do not differ well from each other, no matter which sequence is under investigation. Additional information can be gained from larger fragments (>10 bases) but in this mass range signals have lower quality in terms of sensitivity and resolution. A further problem could be that not all products of the RNase digestion can be detected by MALDI due to interference in the desorption/ionization process, or it might occur that the RNase digestion is not as efficient as expected.

6. Conclusions

Mass spectrometers, particularly those based on MALDI, are mature and stable instruments that can be efficiently applied for SNP analysis. Jurinke et al. have recently highlighted the wide application range of MALDI in genetics; as it can be used for SNP discovery and genotyping,

molecular haplotyping, methylation analysis, and RNA and allele-specific expression [62].

Many different molecular biological procedures for the generation of SNP products have been coupled to MALDI mass spectrometry. Currently, those based on primer extension reactions are mostly used for candidate gene association studies and in pharmacogenetics. More recent approaches developed for re-sequencing exploit the multi-channel detection capability of MALDI mass spectrometers. These approaches are presently transferred, for example, to enable fingerprinting of bacteria.

While MALDI technologies are cost-effective when applied in multiplex applications, these approaches have one drawback in clinical environment: they require multiple steps and hence generally too long operational time. Future developments focusing on detection of PCR products with novel MALDI mass spectrometers are sought-after to overcome current limitations discussed in this article. One promising approach in this regard consists of the development of novel sample preparation and MALDI ionization techniques as well as improved detection devices to improve mass signal resolution and to enable detection of analyte molecules being significantly larger than 100 kDa. Research efforts in the mass spectrometry community should focus on this issue.

In the past, technology development for the analysis of SNPs mainly focused on *in vitro* procedures. Future efforts will also include the specific analysis of nucleic acids at the low to sub-zeptomole level in cells or tissues, either *in situ* or *in vivo* [11].

A remaining issue, which has not fully been addressed in this review, concerns possibilities of quality control and certification of mass spectrometry-based procedures for clinical diagnostics. To my knowledge, so far no DNA mass spectrometry assay has been approved by the FDA. However, since MALDI provides accurate detection of the mass of SNP products and does not rely on clustering methods, it might become an interesting diagnostic tool in the future.

Acknowledgements

I would like to acknowledge the European Union (grant LSHG-CT-2004-503155), the German Ministry for Research and Education (NGFN2, grant 01GR0414), and the Max-Planck Society for support. I would like to thank Hans Lehrach and Richard Reinhardt for fruitful discussions. I would also like to thank Stephen Gelling for the critical reading of the manuscript and Anett Smyra for great help with the figures.

References

- [1] Brookes AJ. The essence of SNPs. *Gene* 1999;234:177–86.
- [2] Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999;22:139–44.

- [3] Matisse TC, Sachidanandam R, Clark AG, et al. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *Am J Hum Genet* 2003;73:271–84.
- [4] Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? *Nat Genet* 2000;26:151–7.
- [5] Lesage S, Zouali H, Cezard JP, et al. CARD15/NOD2 mutational analysis and genotype–phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;70:845–57.
- [6] 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* 2004;14:414–25.
- [7] Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002;56–61 [Suppl.].
- [8] Hardenbol P, Baner J, Jain M, et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol* 2003;5:1–5.
- [9] 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 U S A* 1991;88:7276–80.
- [10] Sauer S, Gut IG. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;782:73–87.
- [11] Sauer S, Lange BMH, Gobom J, et al. Miniaturization in functional genomics and proteomics. *Nat Rev Genet* 2005;6:465–76.
- [12] Bonk T, Humeny A. MALDI–TOF–MS analysis of protein and DNA. *Neuroscientist* 2001;7:6–12.
- [13] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. *Anal Chem* 1988;60:2299–303.
- [14] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989;246:64–71.
- [15] Huber CG, Oberacher H. Analysis of nucleic acids by on-line liquid chromatography–mass spectrometry. *Mass Spectrom Rev* 2001;20:310–43.
- [16] Krahmer MT, Johnson YA, Walters JJ, Fox KF, Fox A, Nagpal M. Electrospray quadrupole mass spectrometry analysis of model oligonucleotides and polymerase chain reaction products: determination of base substitutions, nucleotide additions/deletions, and chemical modifications. *Anal Chem* 1999;71:2893–900.
- [17] Oberacher H, Oefner PJ, Holz G, Premstaller A, Davis K, Huber CG. Re-sequencing of multiple single nucleotide polymorphisms by liquid chromatography–electrospray ionization mass spectrometry. *Nucleic Acids Res* 2002;30:e67.
- [18] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998;17:1–35.
- [19] Laken SJ, Jackson PE, Kinzler KW, et al. Genotyping by mass spectrometric analysis of short DNA fragments. *Nat Biotechnol* 1998;16:1352–6.
- [20] Zhang S, Van Pelt CK, Schultz GA. Electrospray ionization mass spectrometry-based genotyping: an approach for identification of single nucleotide polymorphisms. *Anal Chem* 2001;73:2117–25.
- [21] Kokoris M, Dix K, Moynihan K, et al. High-throughput SNP genotyping with the Masscode system. *Mol Diagn* 2000;5:329–40.
- [22] Pusch W, Kraeuter KO, Froehlich T, Stalgies Y, Kostrzewa M. Genotools SNP manager: a new software for automated high-throughput MALDI–TOF mass spectrometry SNP genotyping. *Biotechniques* 2001;30:210–5.
- [23] Griffin TJ, Smith LM. Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. *Trends Biotechnol* 2000;18:77–84.
- [24] Wu KJ, Steding A, Becker CH. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun Mass Spectrom* 1993;7:142–6.
- [25] Karas M, Gluckmann M, Schafer J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom* 2000;35:1–12.
- [26] Juhasz P, Roskey MT, Smirnov IP, Haff LA, Vestal ML, Martin SA. Applications of delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry to oligonucleotide analysis. *Anal Chem* 1996;68:941–6.
- [27] Kirpekar F, Nordhoff E, Kristiansen K, et al. Matrix assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa. *Nucleic Acids Res* 1994;22:3866–70.
- [28] Pielele U, Zurcher W, Schar M, Moser HE. Matrix-assisted laser desorption time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res* 1993;14:3191–6.
- [29] Gut IG, Jeffery WA, Pappin DJC, Beck S. Analysis of DNA by ‘charge tagging’ and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 1997;11:43–50.
- [30] Berlin K, Gut IG. Analysis of negatively ‘charge tagged’ DNA by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 1999;13:1739–43.
- [31] Sauer S, Lechner D, Berlin K, et al. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 2000;28:e13.
- [32] Sauer S, Lechner D, Berlin K, et al. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res* 2000;28:e100.
- [33] Sauer S, Gelfand DH, Boussicault F, et al. Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Res* 2002;30:e22.
- [34] Sauer S, Gut IG. Extension of the GOOD assay for genotyping single nucleotide polymorphisms by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:1265–72.
- [35] Bauer O, Guerasimova A, Sauer S, et al. Multiplexed hybridizations of positively charge-tagged peptide nucleic acids detected by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2004;18:1821–9.
- [36] Fu DJ, Tang K, Braun A, et al. Sequencing exons 5 to 8 of the p53 gene by MALDI–TOF mass spectrometry. *Nat Biotechnol* 1998;16:381–4.
- [37] Nordhoff E, Luebbert C, Thiele G, Heiser V, Lehrach H. Rapid determination of short DNA sequences by the use of MALDI–MS. *Nucleic Acids Res* 2000;28:e86.
- [38] Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001;2:930–42.
- [39] Wenzel T, Ellsner T, Fahr K, et al. GenoSNIP: SNP genotyping by MALDI–TOF MS using photocleavable oligonucleotides. *Nucleosides Nucleotides Nucleic Acids* 2003;22:1579–81.
- [40] Haff LA, Smirnov IP. Single nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI–TOF mass spectrometry. *Genome Res* 1997;7:378–88.
- [41] Haff LA, Smirnov IP. Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res* 1997;25:3749–50.
- [42] Sun X, Ding H, Hung K, Guo B. A new MALDI–TOF based mini-sequencing assay for genotyping of SNPs. *Nucleic Acids Res* 2000;28:e68.
- [43] Sauer S, Lehrach H, Reinhardt R. MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging. *Nucleic Acids Res* 2003;31:e63.
- [44] Kim S, Edwards JR, Deng L, Chung W, Ju J. Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Res* 2002;30:e85.

- [45] Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. MALDI–TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 Y-chromosome single-nucleotide polymorphisms. *Anal Chem* 2004;76:6039–45.
- [46] Sauer S, Kepper P, Smyra A, et al. Automated solid-phase extraction for purification of single nucleotide polymorphism genotyping products prior to matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *J Chromatogr A* 2004;1049:9–16.
- [47] Buetow KH, Edmonson M, MacDonald R, et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A* 2001;98:581–4.
- [48] Werner M, Sych M, Herbon N, Illig T, Konig IR, Wjst M. Large-scale determination of SNP allele frequencies in DNA pools using MALDI–TOF mass spectrometry. *Hum Mutat* 2002;20:57–64.
- [49] Shifman S, Pisante-Shalom A, Yakir B, Darvasi A. Quantitative technologies for allele frequency estimation of SNPs in DNA pools. *Mol Cell Probes* 2002;16:429–34.
- [50] Werner M, Herbon N, Gohlke H, et al. Asthma is associated with single-nucleotide polymorphisms in ADAM33. *Clin Exp Allergy* 2004;34:26–31.
- [51] Herbon N, Werner M, Braig C, et al. High-resolution SNP scan of chromosome 6p21 in pooled samples from patients with complex diseases. *Genomics* 2003;81:510–8.
- [52] Hoehe MR, Timmermann B, Lehrach H. Human inter-individual DNA sequence variation in candidate genes, drug targets, the importance of haplotypes and pharmacogenomics. *Curr Pharm Biotechnol* 2003;4:351–78.
- [53] Ding C, Cantor CR. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. *Proc Natl Acad Sci U S A* 2003;100:7449–53.
- [54] Tost J, Brandt O, Boussicault F, et al. Molecular haplotyping at high throughput. *Nucleic Acids Res* 2002;30:e96.
- [55] Burgtorf C, Kepper P, Hoehe M, et al. Clone-based systematic haplotyping (CSH): a procedure for physical haplotyping of whole genomes. *Genome Res* 2003;13:2717–24.
- [56] Krebs S, Medugorac I, Seichter D, Forster M. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res* 2003;31:e37.
- [57] Stanssens P, Zabeau M, Meersseman G, et al. High-throughput MALDI–TOF discovery of genomic sequence polymorphisms. *Genome Res* 2004;14:126–33.
- [58] von Wintzingerode F, Bocker S, Schlotelburg C, et al. Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. *Proc Natl Acad Sci U S A* 2002;99:7039–44.
- [59] Smylie KJ, Cantor CR, Denissenko MF. Analysis of sequence variations in several human genes using phosphoramidite bond DNA fragmentation and chip-based MALDI–TOF. *Genome Res* 2004;14:134–41.
- [60] Schatz P, Dietrich D, Schuster M. Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI–TOF. *Nucleic Acids Res* 2004;32:e167.
- [61] Bocker S. SNP and mutation discovery using base-specific cleavage and MALDI–TOF mass spectrometry. *Bioinformatics* 2003;19:i44–53.
- [62] Jurinke C, Denissenko MF, Oeth P, et al. A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY. *Mutat Res* 2005;573:83–95.