

33. Benner, S. A., Ellington, A. D. & Tauer, S. A. Modern metabolism as a palimpsest of an RNA world. *Proc. Natl Acad. Sci. USA* **86**, 7054–7058 (1989).
34. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
35. Rogers, J. & Joyce, G. F. The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7**, 395–404 (2001).
36. Reader, J. S. & Joyce, G. F. A ribozyme composed of only two different nucleotides. *Nature* **420**, 841–844 (2002).
37. Mac Dónaill, D. A. A parity code interpretation of nucleotide alphabet composition. *Chem. Commun.* **18**, 2062–2063 (2002).
38. Mac Dónaill, D. A. Why nature chose A, C, G and U/T: an error-coding perspective of nucleotide alphabet composition. *Orig. Life Evol. Biosphere* **33**, 433–455 (2003).
39. Mac Dónaill, D. A. & Brocklebank, D. An *ab initio* quantum chemical investigation of the error-coding model of nucleotide alphabet composition. *Mol. Phys.* **101**, 2755–2763 (2003).
40. McGinness, K. E. & Joyce, G. F. In search of an RNA replicase ribozyme. *Chem. Biol.* **10**, 5–14 (2003).
41. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54–63 (1998).
42. Szathmáry, E. The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**, 223–229 (1999).
43. Wong, J. T. A coevolution theory of the genetic code. *Proc. Natl Acad. Sci. USA* **72**, 1909–1912 (1975).
44. Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
45. Benner, S. A. Synthetic biology: act natural. *Nature* **421**, 118 (2003).
46. Hamming, R. W. Error detecting and error correcting codes. *Bell Syst. Techn. J.* **29**, 147–160 (1950).

Acknowledgements

I thank the biologists at the Wissenschaftskolleg zu Berlin for vivid discussions. Also, B. Papp and V. Müller who kindly read the manuscript before submission.

Competing interests statement

The author declares that he has no competing financial interests.

Online links

FURTHER INFORMATION

Eörs Szathmáry's web page:

<http://www.colbud.hu/fellows/szathmary.shtml>

Scripps Research Institute: <http://www.scripps.edu>

Steven Benner's web page:

<http://www.chem.ufl.edu/benner.html>

Access to this interactive links box is free online.

INNOVATION

Digital genotyping using molecular affinity and mass spectrometry

Sobin Kim, Hameer D. Ruparel, T. Conrad Gilliam and Jingyue Ju

The goal of DNA sequencing and genotyping is to efficiently generate accurate high-throughput digital genetic information that unambiguously identifies sources of genetic variation and clearly distinguishes heterozygous from homozygous variants. Recent advances in mass-spectrometry-based DNA sequencing and genotyping bode well for meeting these criteria. Pilot studies show that these recently developed approaches allow unambiguous multiplex detection of heterozygous variants and the identification of deletion and insertion variants.

The completion of the Human Genome Project has set the stage for screening genetic mutations to identify disease genes on a genome-wide scale¹. Accurate high-throughput methods for resequencing the intron/exon regions of candidate genes are needed to explore the complete human genome sequence for disease-gene discovery. State-of-the-art technology for high-throughput DNA sequencing, such as that used in the Human Genome Project, uses

capillary-array DNA sequencing with LASER-INDUCED FLUORESCENCE DETECTION^{2–5}. Although this technology meets the throughput and read-length requirements of large-scale DNA sequencing projects, the accuracy that is required for mutation detection needs to be improved for a wide range of applications, ranging from disease-gene discovery to personalized medicine. For example, the unambiguous detection of heterozygotes is difficult with electrophoresis-based DNA sequencing methods. Problems also arise in GC rich regions owing to COMPRESSION^{6,7}, which leads to poor resolution in sequencing DNA fragments. Also, the first few bases downstream of the priming site are often masked by high levels of fluorescence from the excess dye-labelled primers or dye-labelled terminators, and are therefore difficult to identify.

Many recent advances in DNA sequencing technology address these limitations. In particular, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has emerged as a rapid and efficient analytical tool in DNA sequencing and genotyping. Here, we review these advances and discuss new approaches that use

molecular affinity for the accurate and simultaneous detection of genetic variations, which have wide applications in PHARMACOGENOMICS and clinical diagnostics.

DNA sequencing by MALDI-TOF MS

MALDI-TOF MS has been widely used in DNA sequencing^{8–12} (FIG. 1). SANGER DNA SEQUENCING is generally used to produce DNA sequencing fragments in MALDI-TOF MS¹³. Compared with gel electrophoresis-based sequencing systems, MALDI-TOF MS produces high-resolution short DNA sequencing fragments of <100 base pairs (bp), rapid fragment separation on microsecond time-scales and the complete elimination of the compressions that are associated with gel electrophoresis.

An important challenge for DNA sequencing using mass spectrometry is the stringent purity requirement for the sequencing fragments that are introduced into the mass detector. Because DNA sequences are determined by accurately measuring the mass of the DNA fragments, DNA must be free from alkaline earth salts and other contaminants.

Approaches for purifying DNA samples that rely on the strong interaction of a small molecule (biotin) and a protein (streptavidin¹⁴) on solid surfaces (such as magnetic beads) are widely used^{15,16}. In DNA sequencing using MALDI-TOF MS, Monforte and Becker obtained read lengths of 100 bp by purifying DNA sequencing samples using a cleavable biotinylated primer¹¹. In this method, the primer-extension fragments are captured at their 5' end on streptavidin-coated magnetic beads, whereas the other components in the sequencing reaction are washed away. Fu *et al.* reported the sequencing of exons 5–8 of the human tumour suppressor *p53* gene (also known as *TP53*) by MALDI-TOF MS using a DNA template that was immobilized on a solid phase for one cycle of extension. In this study, extended DNA fragments were hybridized on the immobilized templates, whereas the other components in the sequencing reaction were removed. Neither method eliminates falsely terminated DNA sequencing fragments. Falsely terminated DNA fragments (false stops) are generated in Sanger DNA sequencing reactions when a DNA fragment is terminated by the incorporation of a deoxynucleotide (dNTP) rather than a dideoxynucleotide (ddNTP). False stops and dimerized primers can produce extra peaks in the mass spectra that prevent accurate base identification⁹. Also, four separate reactions are carried out in both methods, one for each dideoxynucleotide terminator, which is analogous to dye-labelled primer sequencing.

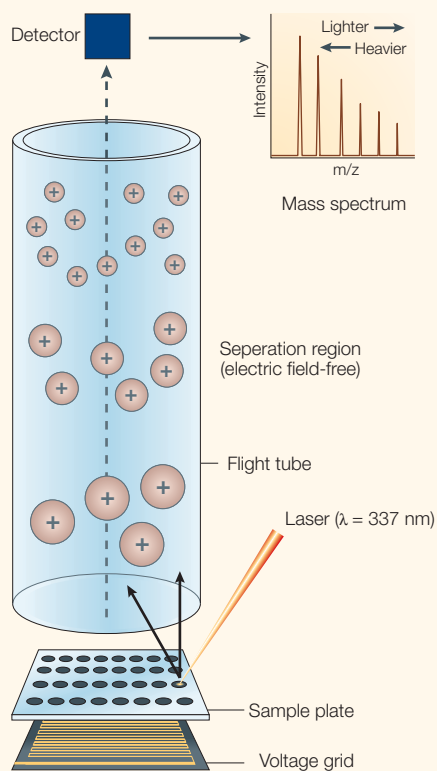


Figure 1 | Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This technique is widely used for detecting the masses of high-molecular weight molecules such as DNA and proteins. In this method, ANALYTE molecules (such as DNA sequencing fragments) and matrix molecules (typically ultraviolet (UV) or infrared (IR) light-absorbing small organic molecules) are mixed in solution. They are then co-crystallized on a flat sample plate, which is subsequently loaded into the vacuum chamber of the mass spectrometer. DNA molecules are gently desorbed and ionized along with the matrix molecules by UV laser irradiation and the resulting charged ions are accelerated under a constant electric voltage, which causes them to fly towards the ion detector. The charged molecules arrive at the detector at different times on the basis of their masses: lighter molecules fly faster and therefore arrive at the detector sooner than heavier molecules. The masses of the charged ions are determined from their time of flight to the detector. m/z , mass per charge ratio.

Ideally, for DNA sequencing with MALDI-TOF MS, a procedure would be established that allows sequencing reactions to be done in one tube to simplify sample preparation, uses cycle sequencing to increase the yield of the DNA sequencing fragments and isolates pure DNA sequencing fragments that are free of false stops. To this end, we developed a high fidelity DNA sequencing method using dye-labelled primers and solid-phase-capturable (SPC) biotinylated dideoxynucleotide terminators (biotin-ddNTPs)¹⁷. After

capture/release of the DNA sequencing fragments at the 3' end from the streptavidin-coated solid phase, only the pure DNA sequencing fragments are loaded and detected on sequencing gels. The method is effective in removing falsely stopped DNA fragments for unambiguous genetic-mutation detection, although GC compression still occurs because the method uses gel electrophoresis. We therefore subsequently used biotin-ddNTPs to develop an accurate DNA sequencing method (SPC sequencing) using mass spectrometry¹⁸. SPC sequencing takes advantage of the high affinity between biotin and streptavidin to isolate pure DNA sequencing fragments.

SPC sequencing

In this approach¹⁸, biotin is attached to position 5 of pyrimidines and position 7 of purines through a linker arm to generate a library of biotin-ddNTPs. It has been previously shown that when these nucleotide positions were modified with bulky organic dyes, DNA polymerase could still incorporate the modified nucleotides into the growing DNA strand¹⁹. The length of the linker arm was chosen to optimize the balance between capture efficiency by streptavidin and incorporation efficiency by the polymerase^{20,21}.

The scheme for isolating pure biotinylated DNA products in the SPC-sequencing approach is shown in FIG. 2a. DNA sequencing fragments that are terminated by biotin-ddNTPs at the 3' end in a cycle-sequencing reaction are bound to a streptavidin-coated solid phase. All excess primers, salts and falsely terminated DNA fragments are washed away to provide pure DNA samples for MALDI-TOF MS detection, which yields accurate sequencing data.

An example of a mass spectrum is shown in FIG. 2b. The first peak in the spectrum corresponds to the extension of the primer by the first nucleotide that is complementary to the corresponding nucleotide in the DNA template. The difference in mass between each successive peak can be measured to determine the identity of the nucleotide that corresponds to that peak, as each nucleotide has a unique molecular mass. No primer peak is seen in the mass spectrum, as the primers are not biotinylated and are removed after SPC, which eliminates the possibility of false peaks caused by primer dimers⁹. There are also no peaks owing to false stops in the spectrum. When biotin-ddNTPs with different molecular masses are used, the smallest mass difference between any two sequencing fragments is 16 daltons, compared with 9 daltons when standard terminators are used. It has been shown that as DNA fragment size

increases so does the mass-spectral peak width, which results in a diminished resolving capacity of the mass spectrometer for larger DNA fragments¹¹. Therefore, using biotinylated terminators with increased mass differences to generate DNA sequencing products greatly improves sequence identification by mass spectrometry. This increase in mass resolution facilitates heterozygote detection. DNA fragments with large masses are typically associated with larger peak-widths in their mass spectra, which makes heterozygote detection difficult if standard ddNTP terminators are used. FIGURE 3 shows that if biotinylated terminators — biotin-11-dd(A, C, G)TPs and biotin-16-ddUTP — are used, two distinct peaks of nearly equal intensity are generated in the mass-sequencing spectrum; in this case, the method was applied to a synthetic template that contains a polymorphic site.

In previously reported mass-spectrometry DNA sequencing methods, four reactions (one for each ddNTP) were carried out separately. Using biotinylated terminators and capturing the DNA sequencing fragments at the 3' end of the fragments on a solid surface, we developed a one-tube mass-spectrometry cycle-sequencing method. With improvement in mass-spectrometer detector sensitivity, this method will allow faster and more accurate detection of larger DNA fragments.

SPC sequencing in mutation detection

Although single-base extension (SBE) has been widely used to detect point mutations such as single nucleotide polymorphisms (SNPs; see below for further discussion) it might be less suitable to distinguish between point mutations and deletions or insertions that occur frequently throughout the genome. Direct DNA sequencing, in theory, is the most accurate technique for mutation detection. However, even the present state-of-the-art sequencing technology — fluorescent Sanger DNA sequencing — faces difficulties in detecting heterozygous nucleotide variation and accurately sequencing GC-rich regions owing to the compression of the DNA fragments^{6,7}. Pyrosequencing is an alternative sequencing approach, in which each nucleotide is identified as soon as it has been incorporated into a growing DNA strand by DNA polymerase²². Over the past few years, this approach has been successfully used to characterize short DNA sequences. Although pyrosequencing is now widely used to detect SNPs and other DNA variations, it faces difficulties in accurately sequencing single nucleotide-repeat regions²³. Furthermore, both of these sequencing methods have problems characterizing frameshift mutations, because a deletion or

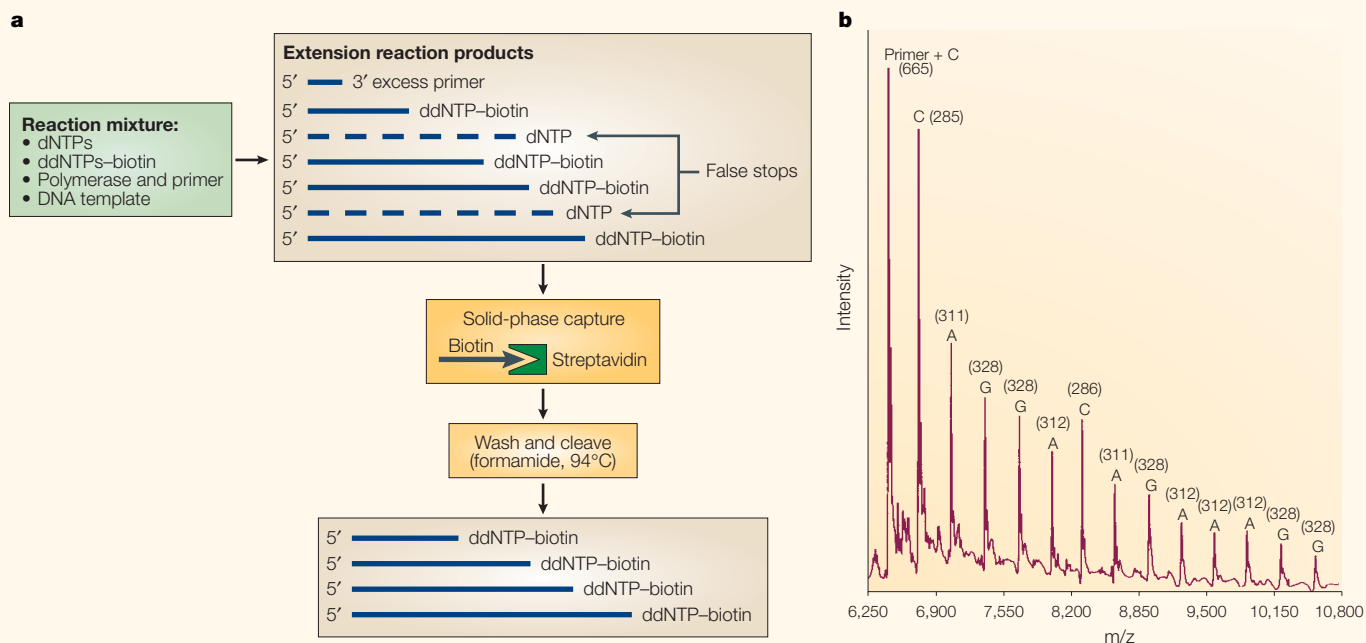


Figure 2 | Solid-phase capture (SPC) sequencing. a | The SPC-sequencing approach to isolate pure DNA fragments for analysis by mass spectrometry. False stops, excess primers and salts are eliminated from the reaction by capturing all correctly terminated DNA sequencing fragments with streptavidin-coated magnetic beads. The purified fragments are then cleaved from the beads with formamide for analysis by mass spectrometry. **b** | A DNA sequencing mass spectrum generated from biotinylated terminators: biotin-16-ddUTP and biotin-11-dd(A,G,C)TPs. The spectrum shows the sequence of a PCR template amplified from genomic DNA at the *5382insC* locus of the early-onset breast cancer 1 gene (*BRCA1*). The number assigned to each peak corresponds to the difference in mass between it and the preceding peak, which is used to identify the base that the peak represents. ddNTP, dideoxynucleotide; dNTP, deoxynucleotide; m/z, mass per charge ratio.

insertion in any one allele will cause the sequencing reactions for the two alleles to be out of phase, making it difficult to interpret the sequencing data.

SPC sequencing can be used to accurately detect and characterize frameshift mutations in PCR-amplified templates from genomic DNA while avoiding many of the limitations that are faced by the above mentioned mutation-detection techniques²⁴. The principal advantage of SPC sequencing in mutation detection is the highly accurate identification of the different bases that might coexist at a single locus along the DNA as a direct consequence of substitution or frameshift mutations. MALDI-TOF MS analysis displays the sequence data in the form of distinct mass peaks, which facilitates a more accurate characterization of frameshift mutations compared with other sequencing methods that measure fluorescent or radioactive signals that are emitted from the labelled DNA.

A schematic representation of the SPC-sequencing method for the detection of frameshift mutations is shown in FIG. 4. A mass spectrum that is produced by sequencing homozygous DNA, both copies of which have identical bases at every position, will yield a single peak at each position that identifies the base. However, in mutant DNA with a deletion or insertion in one of the copies, the two

alleles have a relative frameshift between them downstream of the mutation site. Therefore, each position in the DNA sequence following the mutation site might be occupied by two different bases. In SPC sequencing, every position upstream of the mutation will appear as a

single peak, whereas every position downstream of the mutation will have two distinct peaks that correspond to the masses of the two nucleotides that are present at that position. By calculating the differences in masses between subsequent peaks, the sequences of

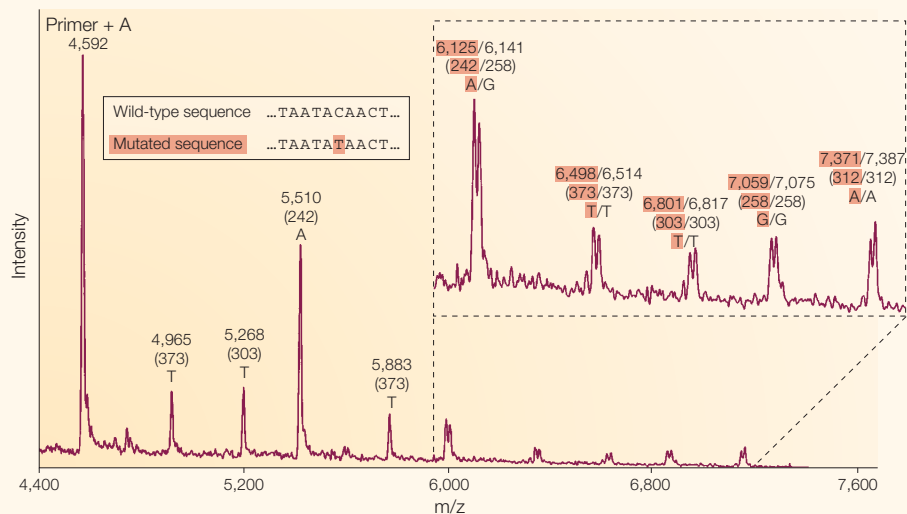


Figure 3 | Mass-sequencing spectrum from a DNA template that contains a polymorphic site. Each peak is labelled with the absolute mass value (m/z, mass per charge ratio). The mass difference (in parenthesis) between a peak and the previous one is used to determine the identity of the nucleotide. At the polymorphic site, double peaks are detected each corresponding to a mutant (A) or wild type (G) allele. The downstream peaks also appear as a doublet, owing to the mass shift (16 daltons for A/G) that occurs at the polymorphic site. The inset shows a magnified view of the heterozygote peaks. The signal of each DNA fragment is distinct and the heterozygote (A/G) can be easily recognized.

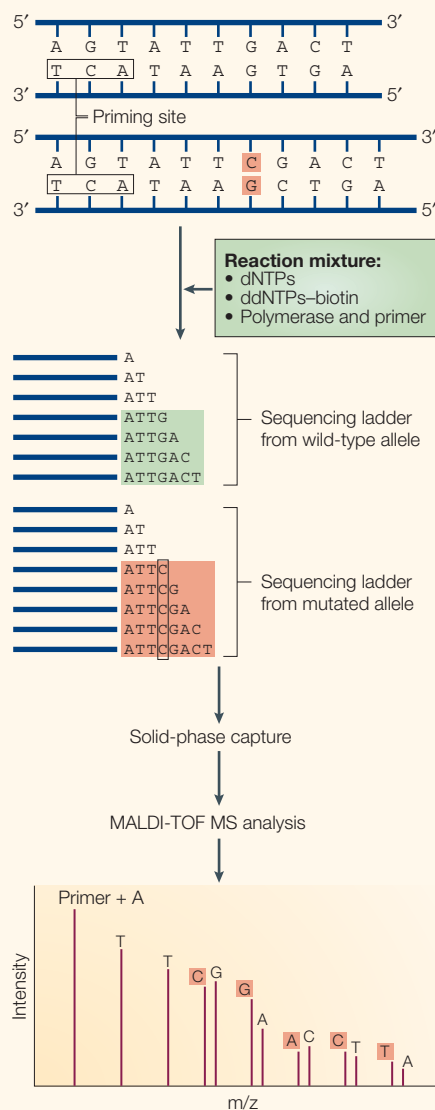


Figure 4 | Schematic representation of mutation detection using solid-phase capture (SPC) sequencing. A locus with an insertion or deletion (an insertion is shown here as an example) in one of its alleles can have two different bases coexisting at the same position. This leads to the formation of two distinct mass peaks at each position in the spectrum downstream of the mutation. By calculating the corresponding mass differences, the sequences of both alleles can be simultaneously read and the mutation site can be accurately characterized. ddNTP, dideoxynucleotide; dNTP, deoxynucleotide; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; m/z , mass per charge ratio.

the two alleles can be read simultaneously and the mutation site can be unambiguously identified.

SPC sequencing was used to characterize the *185delAG* and *5382insC* mutations in the early-onset breast cancer 1 gene (*BRCA1*) that

commonly occur among the Ashkenazi population with a strong family history of breast and ovarian cancer²⁵. A comparison of the results obtained from SPC sequencing and from electrophoresis-based DNA sequencing for *185delAG* is presented in BOX 1. These results show that SPC sequencing is highly accurate in characterizing frameshift mutations. The data acquisition is rapid and the sequencing results are clear and easily interpreted.

In electrophoresis-based sequencing that uses fluorescence detection, the first few bases are often masked by the high-intensity fluorescence signals from the labelled primers or terminators. Sequencing primers must therefore be designed a few bases away from the mutation site of interest to obtain coherent sequencing data for that region. SPC sequencing does not face this limitation because the data is obtained as a mass spectrum and false peaks are eliminated owing to molecular affinity-based purification of the sequencing fragments. Consequently, the sequencing primers can be designed close to the mutation site and few bases need to be sequenced to characterize the mutation. With electrophoresis-based fluorescent DNA sequencing, the identification of deletions and insertions requires manual analysis of the sequence data. But SPC-sequencing data analysis can be automated owing to the distinct peaks from the mass spectrum. The use of biotin-16-ddUTP, which has a longer linker between biotin and ddUTP than the other three biotinylated terminators (biotin-11-ddATP, biotin-11-ddCTP and biotin-11-ddGTP), ensures that all combinations of nucleotides produce mass differences that are significantly different from each other. Furthermore, as the data are in the form of distinct mass peaks, bases can be assigned to each peak by simply calculating the mass values. By contrast, the analysis of data from fluorescence-based DNA sequencing requires lengthy base-calling algorithms, because of the broad emission spectra from the fluorophores.

The Ashkenazi *BRCA1* mutations that are described in BOX 1 are just one example of many frameshift mutations that contribute to disease development. For example, mutations in the *p53* gene have an important role in the development of many types of cancer²⁶. With further developments in automation and sample processing, SPC sequencing can be an accurate tool for screening known mutations, with potential applications in clinical diagnostics. Mutations that are as yet uncharacterized can also be identified using the SPC sequencing in combination with the fluorescent Sanger DNA sequencing.

SPC-SBE multiplex SNP analysis

SNPs, which are the most common genetic variations in the human genome, are important markers for identifying disease genes and for pharmacogenetic studies^{27,28}. SNPs appear in the human genome with an average density of 1 per 1,000 bp²⁹. A rapid, precise and cost-effective method is required for large-scale SNP genotyping. MALDI-TOF MS³⁰ allows rapid and accurate sample measurements, and has been used in a range of SNP-detection methods including hybridization^{31,32}, INVASIVE CLEAVAGE^{33,34} and SBE^{35–39}. Recently, MALDI-TOF MS coupled with base extension has been successfully used to develop new approaches for measuring gene expression and single-copy DNA haplotyping^{40,41}.

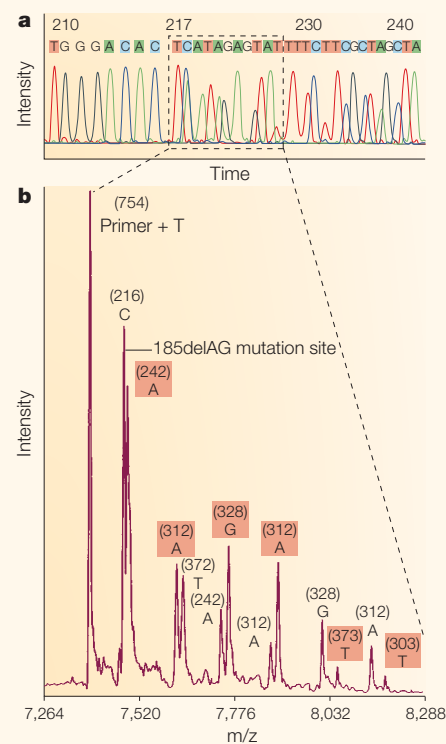
In SBE, which is widely used for multiplex SNP analysis, primers that are designed to anneal immediately next to a polymorphic site are extended by a single ddNTP that is complementary to the nucleotide at the variable site. A particular SNP can be identified by measuring the mass of the resulting extension product. Present SBE methods for multiplex SNP analysis using MALDI-TOF MS require the unambiguous simultaneous detection of a library of primers and their extension products. However, limitations in the resolution and sensitivity of MALDI-TOF MS for longer DNA molecules make it difficult to simultaneously measure DNA fragments over a large range of masses. The requirement to measure both primers and their extension products in this range limits the scope of multiplexing.

We recently used the SPC biotin-ddNTPs in SBE for multiplex genotyping by MALDI-TOF MS^{42,43}. In the SPC-SBE method, primers that have different molecular masses and are specific to the polymorphic sites in the DNA template are extended with biotin-ddNTPs to generate 3'-biotinylated DNA extension products. The 3'-biotinylated DNAs are then captured by streptavidin-coated magnetic beads, whereas the unextended primers and other components in the reaction are washed away. The pure DNA-extension products are subsequently released from the magnetic beads by denaturing the biotin-streptavidin interaction with formamide and are analysed with MALDI-TOF MS. The nucleotide at the polymorphic site is identified by measuring the mass of the primer-extension product. As the primer-extension products are isolated before MS analysis, the resulting mass spectrum is free of non-extended primer peaks and their associated dimers, which increases the accuracy and scope of multiplexing in SNP analysis. The solid phase-purification system facilitates the de-salting of the captured DNA products and therefore enhances

Box 1 | Solid-phase capture sequencing of *BRCA1* frameshift mutations

Solid-phase capture (SPC) sequencing was used to characterize the *185delAG* and *5382insC* mutations in the early-onset breast cancer 1 gene (*BRCA1*) that commonly occur among the Ashkenazi population with a strong family history of breast and ovarian cancer²⁴. The figure compares electrophoresis-based sequencing (a) with SPC sequencing (b) for characterizing the *185delAG* locus. The electrophoresis sequencing data are high quality and well resolved up to the T at position 217, after which the data become noisy and the sequence assignment is no longer accurate (a). This is caused by a frameshift mutation in one of the two alleles, which leads to misalignment between the sequences of the two alleles such that, beyond the mutation site, two different sequences are superimposed on each other, making the fluorescence signals unresolvable. It might be possible to confirm the presence of a mutation by recognizing this distinct pattern in the sequencing data in combination with sequencing data from the reverse direction, but an accurate characterization of the mutation from these data might still prove difficult and tedious.

Panel b shows the SPC sequencing results for the same region around the *185delAG* mutation site. The first position in the spectrum is occupied by a single large peak that corresponds to a T in both alleles. This is followed by a doublet peak at the second position in the spectrum, which identifies this position as the mutation site. The subtraction of the mass of the previous T-peak from the masses of these two peaks establishes their identities as C and A. Similarly, the identities of the two peaks at the next position are confirmed as T and A, and so on. Consequently, the two sequences identified from the spectrum are 3'-TCTAAGA...-5' and 3'-TAAGATT...-5'. The wild-type sequence in this region is 3'...TCTAAGATTT...-5'. Having compared the two parallel sequences obtained in the mass spectrum with the known wild-type sequence, the presence of a deletion of a C and a T after the first T in one of the alleles can be confirmed. Note that the spectrum shown in panel b actually indicates a deletion of 5'-CT-3', because a reverse primer was used to sequence this region and, therefore, the complementary sequence of the *185delAG* locus was obtained. SPC sequencing is equally efficient at characterizing insertion mutations. m/z, mass per charge ratio.



the accuracy and clarity of data. Also, four biotin-ddNTPs with distinct molecular masses were used to generate DNA extension products with significantly different masses to improve resolution and accuracy in detecting peaks from DNA fragments in the mass spectrum.

The method is schematically shown in FIG. 5. Genomic DNA is used to generate multiplex PCR products that serve as templates for SBE reactions in which a library of SNP primers with different masses, DNA polymerase and biotin-ddNTPs are used. After SPC of the 3'-biotinylated DNA extension fragments and elimination of all unextended primers and other components in the extension reaction, MALDI-TOF MS is used to analyse these DNA products to identify nucleotide variation.

Unextended primers occupy the mass range in the mass spectrum, which decreases the scope of multiplexing, and excess primers can dimerize to form false peaks in the mass spectrum⁹. The excess primers and their associated dimers also compete for the ion current, which reduces the detection sensitivity of MALDI-TOF MS for the desired DNA fragments. These complications are completely removed by using biotin-ddNTPs and SPC. Extension products for all four biotin-ddNTPs are clearly detected with well-resolved mass values. The masses of the primer-extension

products in comparison to the masses of the corresponding primers reveal the identity of each nucleotide at the polymorphic site. In the case of heterozygous genotypes, two peaks — one corresponding to each allele — are clearly distinguishable in the mass spectrum.

One advantage of MALDI-TOF MS compared with other detection techniques is its ability to simultaneously measure the mass of DNA fragments over a certain range. This is only possible if the masses of all primers and their extension products are sufficiently different to yield peaks that can be adequately resolved in the mass spectrum. Ross *et al.* simultaneously detected 12 SNPs by carefully tuning the masses of all primers and extension products so that they would lie in the range of 4.5–7.6 kDa without overlapping³⁷. As the unextended primers occupy the mass range in the mass spectrum, by eliminating them, the SPC-SBE approach should increase the scope of multiplexing in SNP analysis.

SPC-SBE in multiplex genotyping

Human hereditary haemochromatosis gene. Human hereditary haemochromatosis is a genetic condition that occurs in ~1 in every 400 Caucasians. The underlying *C282Y* mutation leads to iron overload and potentially to liver failure, diabetes and depression⁴⁴. A subset of individuals who are compound

heterozygotes for the *C282Y* and *H63D* mutations also accumulate excess iron in the body tissue. Because of the high prevalence of these mutations and the ability to prevent disease symptoms by PHLEBOTOMY, accurate methods for genotyping these two SNPs will foster genetic screening for this condition.

We used the SPC-SBE method to genotype the *C282Y* and *H63D* polymorphic sites from genomic DNA⁴². The result shows that both homozygous and heterozygous genotypes are clearly determined by the distinct mass difference in the mass spectra. Homozygotes are detected as single peaks with a distinct mass, whereas heterozygotes appear as two peaks that are separated by a specific mass difference (16 daltons for A/G and 39 daltons for C/G). All 14 genotypes from the 7 individuals tested, including heterozygotes, were accurately determined by this method. For this small-scale study, the genotypes were determined by eye from the distinct mass difference in the mass spectra with 100% accuracy.

Tumour-suppressor gene *p53*. *p53* is one of the most frequently mutated tumour-suppressor genes in human cancer^{45,46}. As most of the *p53* mutation hot spots are clustered in exons 5–8, we selected 30 polymorphic sites, including the most frequently

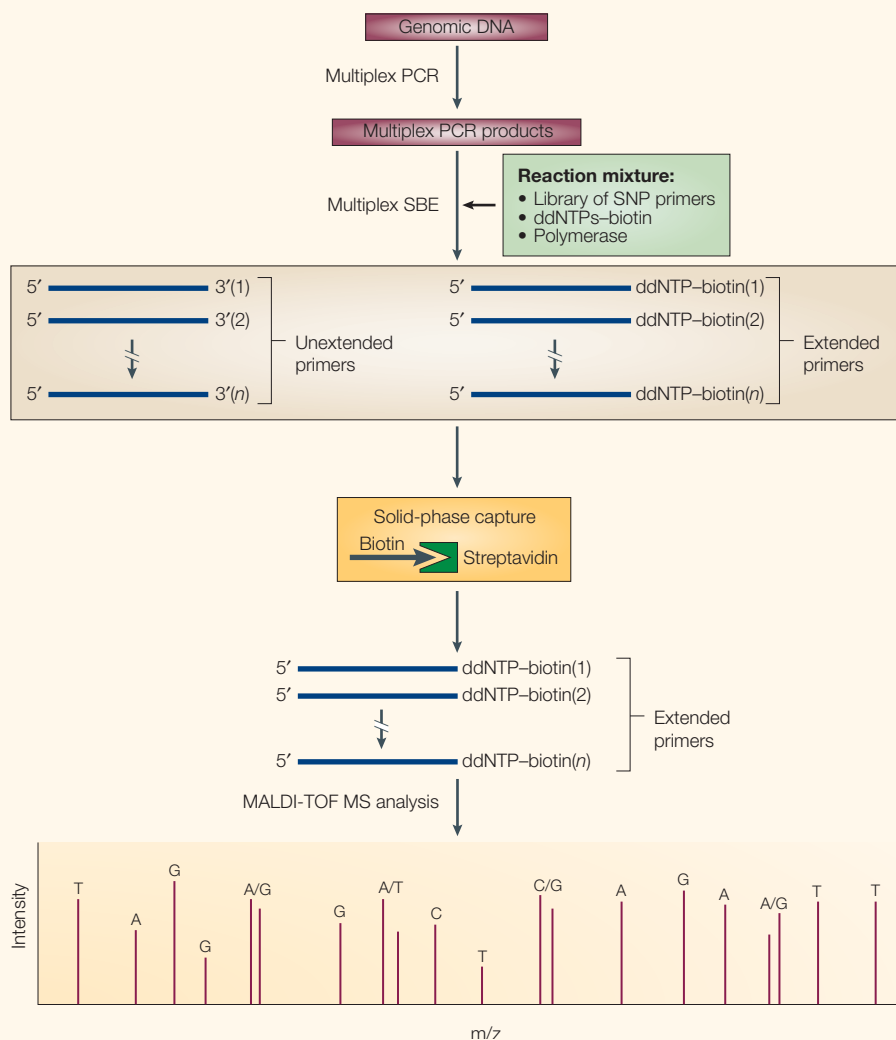


Figure 5 | The SPC-SBE approach for multiplex SNP analysis using biotinylated ddNTPs and MALDI-TOF MS. Genomic DNA is used to generate multiplex PCR products, which serve as templates for single-base extension (SBE) reactions using a library of single nucleotide polymorphism (SNP) primers with different masses, DNA polymerase and biotinylated dideoxynucleotides (ddNTPs). After solid-phase capture (SPC) of the 3'-biotinylated DNA extension fragments, and elimination of all unextended primers and other components in the extension reaction, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is used to analyse the DNA products, which results in identification of the nucleotide variations. *m/z*, mass per charge ratio.

mutated *p53* codons in exons 5, 7 and 8, to test the high multiplexing scope of the SPC-SBE method⁴⁷. Thirty primers specific to each polymorphic site were designed to yield SBE products with sufficient mass differences. To optimize mass differences between SBE products, we tuned the mass of some primers using methyl-dC and dU to replace dC and dT, respectively. So, without increasing the primer length to the point at which primer dimers or primer interaction can be a problem, a large library of primers with different molecular masses can be constructed for the SPC-SBE approach.

Multiplex PCR was used to amplify three *p53* exons from human genomic

DNA. The 30 primers were mixed with the PCR products and biotin-ddNTPs to generate 3'-biotinylated extension DNA products that were then captured by streptavidin-coated solid-phase magnetic beads. The pure DNA products were released from the solid phase and analysed by MALDI-TOF MS. The nucleotide at the polymorphic site was accurately identified by the mass of the DNA extension product in the mass spectrum.

We recently used the SPC-SBE genotyping approach to analyse nucleotide variations in *p53* from 30 Wilms tumours, 19 head and neck squamous carcinomas and 3 colorectal carcinomas — an experiment that represents the highest level of multiplexing

reported so far using mass spectrometry for SNP analysis⁴⁷. None of the 30 Wilms tumour samples had somatic mutations at the 30 polymorphic sites that were tested, yielding 30 distinct peaks that correspond to the wild-type *p53* sequences. By contrast, 2 of the 19 head and neck tumour samples contained a genetic variation: one at codon 157 (G/T heterozygous configuration; FIG. 6) and the other at codon 151 (C to T homozygous). In the 3 colorectal tumour cell lines that were evaluated, the first had 30 wild-type *p53* sequences for the 30 sites tested, whereas the others had a G to A homozygous mutation in codon 273. Both heterozygous and homozygous genotypes were clearly detected in the 30 codons with great accuracy. These data, which were confirmed by direct DNA sequencing, are consistent with the known paucity of the *p53* mutations in Wilms tumour^{48,49} and the known occurrence of such mutations in squamous⁵⁰ and colorectal carcinomas^{51,52}.

Other approaches to SNP analysis

Many alternative methods that use mass spectrometry and fluorescence detection have been developed for multiplex SNP analysis. For example, the MassARRAY assay⁵³ is widely used for SNP genotyping. In this automated high-throughput approach, the primer is extended in the presence of three ddNTPs and one dNTP that corresponds to one of the two alleles. So, unextended primers, SBE primers and primers extended with two or more bases are detected together in the mass spectrum. This method is simple and does not require labelling of any reaction components, although it does present limitations in the simultaneous analysis of many SNPs.

Other SNP genotyping methods involve fluorescence-based detection. The fluorescence polarization (FP)-template-directed dye-terminator-incorporation (TDI) assay⁵⁴ uses SBE with allele-specific dye terminators. The extension products are monitored for FP to determine genotypes. The FP-TDI approach is simple but might be limited with respect to multiplex analysis. In the molecular inversion probe (MIP) approach⁵⁵, locus-specific probes form a circular shape by the successive extension and ligation at polymorphic sites. After degrading linear probes, only circular probes that carry allelic information are amplified and detected using a microarray. Using this method, Hardenbol *et al.* were able to genotype more than 1,000 SNPs per assay. Although this method offers a high level of multiplexing, it requires many enzymatic reaction steps and a complicated probe design.

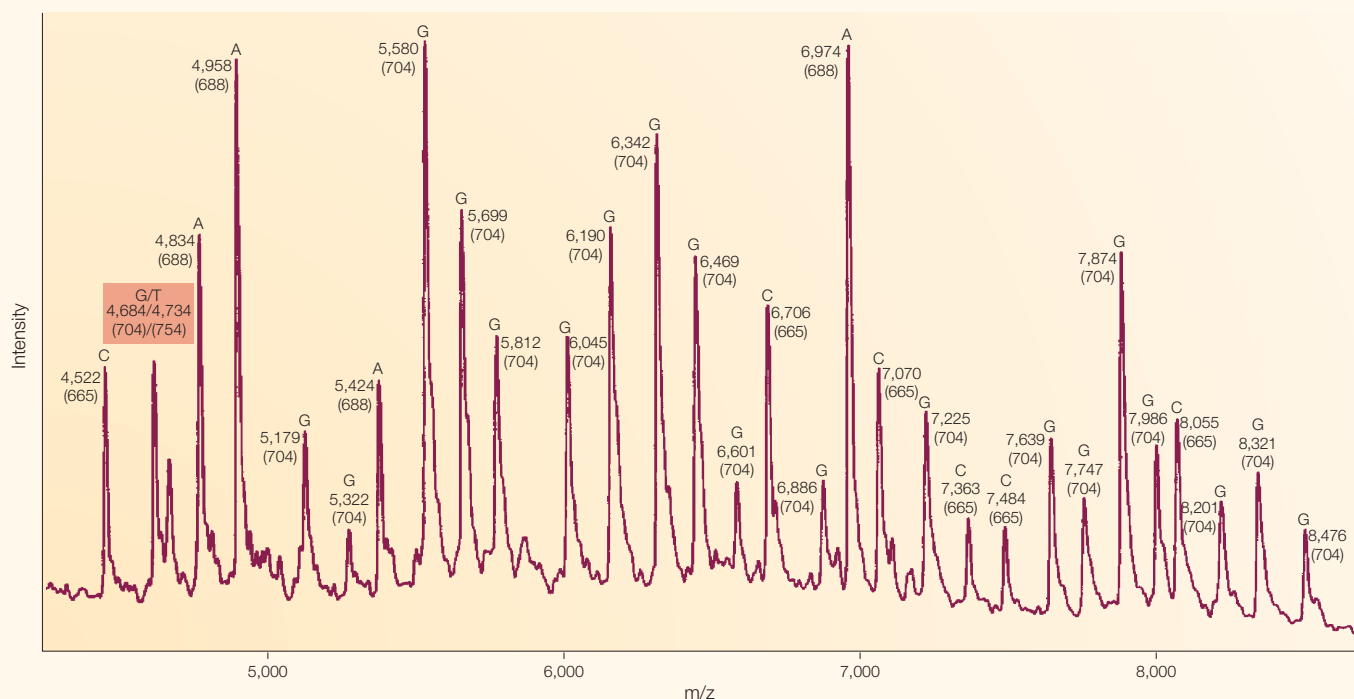


Figure 6 | **Simultaneous detection of nucleotide variations in 30 codons of the p53 gene using SPC–SBE.** Shown here is a mass spectrum from a head and neck tumour, which contains a heterozygous genotype G/T (4,684/4,734 daltons) in codon 157. Each peak represents a different polymorphism, which is labelled with its nucleotide identity and absolute mass value. The values in parentheses, which denote the mass difference between each DNA extension product and its corresponding primer, are used to determine the nucleotide identity. m/z, mass per charge ratio.

The BEAMing approach⁵⁶, named on the basis of its principal components (beads, emulsion, amplification and magnetics), was recently developed for detecting genetic variations. In this method, each DNA template is separately amplified by a large number of primers that are immobilized on a magnetic bead in emulsion. Alleles are distinguished by fluorescent dye-labelled probes and analysed using FLOW CYTOMETRY. Apart from allowing the identification of allelic variations, this approach also offers the ability to quantify these variations. Furthermore, the DNA sample can be recovered from the flow

cytometer for subsequent analysis. However, the many steps of manipulation and difficulties in the accurate determination of allele frequency are important challenges for this technique.

Conclusions

SPC sequencing and SPC–SBE multiplex SNP analysis are two complementary approaches that can be used to determine and characterize the nature of genetic variations. In cases in which the design of the primer immediately upstream of the variable site is not possible owing to nonspecific hybridization, SPC

sequencing can be used to determine the variation by designing a primer several bases away from the variable site. In some cases, the determination of only a single nucleotide using the SPC–SBE method might not suffice to distinguish the variation as a point mutation, insertion or deletion. In such cases, the SPC sequencing could be used to determine the sequence beyond the SBE site, providing unambiguous verification of the nature of the genetic mutation.

It has been reported that MALDI-TOF MS can detect DNA sequencing fragments of up to 100 bp with sufficient resolution if cleavable

Glossary

ANALYTE

A molecule that is of interest for analysis in a particular study.

COMPRESSION

A phenomenon in gel electrophoresis that occurs as a result of the thermodynamic stability of hairpin formations in DNA sequences that terminate in a string of G and C nucleotides.

FLOW CYTOMETRY

The analysis of single cells or subcellular particles by the detection of their light absorption, scattering and/or fluorescence properties as they pass through a laser beam in a directed fluid stream.

INVASIVE CLEAVAGE

The excision of redundant portions of DNA by DNA repair enzymes such as 5' to 3' exonuclease or 5' nuclease. These redundancies are caused when two oligonucleotides, which are hybridized to the same DNA template, overlap along some of their terminal bases. The cleavage of the overlapped portion causes the formation of a nick at the position of redundancy that can later be repaired by ligation.

LASER-INDUCED FLUORESCENCE DETECTION

The measurement of emitted fluorescence signals from molecules that are excited by laser radiation.

PHARMACOGENOMICS

The study of the influence of genetic differences on the variability in the response of individuals to drugs.

PHLEBOTOMY

The removal of blood from a vein for diagnostic therapeutic purposes.

SANGER DNA SEQUENCING

A sequencing method that involves the enzymatic synthesis of DNA chains of different length using dideoxynucleotides, the separation of the DNA fragments by size and the identification of the fragments to reveal the sequences.

TAG SNPS

A small subset of SNPs that is needed to uniquely identify a complete haplotype.

primers are used¹¹. The mass difference between each adjacent DNA sequencing fragment is ~300 daltons. So, by designing each primer to have a mass difference of ~100 daltons, the SPC–SBE method can offer, in principle, the simultaneous analysis of up to 300 SNPs in a single spot of an MS sample plate. Each MS sample plate with 384 spots can identify more than 100,000 SNPs—roughly the number of TAG SNPs that are required to identify all of the haplotypes in the human genome. This level of multiplexing should be achievable by tuning the primer masses with nucleotide analogues that contain stable chemical groups. A master database of primers and the resulting masses of all four possible extension products could be constructed for all SNPs of interest. The experimental data from the MALDI-TOF MS spectra could then be compared with this database to precisely and automatically identify a library of SNPs. The SPC–SBE high-throughput SNP detection approach might therefore be applicable to screening a repertoire of genotypes in candidate genes as potential disease markers.

Sobin Kim, Hameer D. Ruparel and Jingyue Ju are at the Columbia Genome Center and the Department of Chemical Engineering, Columbia University College of Physicians and Surgeons, New York 10032, USA.

T. Conrad Gilliam is at the Columbia Genome Center and the Department of Genetics and Development, Columbia University College of Physicians and Surgeons, New York 10032, USA.

Correspondence to T.C.G. or J.J. at Room 405A, Russ Berrie Medical Science Pavilion, Columbia University College of Physicians and Surgeons, 1150 Saint Nicholas Avenue, New York 10032, USA.
e-mails: tcg1@columbia.edu; dj222@columbia.edu

doi:10.1038/nrg1230

- Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
- Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
- Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N. & Mathies, R. A. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl Acad. Sci. USA* **92**, 4347–4351 (1995).
- Ju, J., Glazer, A. N. & Mathies, R. A. Cassette labeling for facile construction of energy transfer fluorescent primers. *Nucleic Acids Res.* **24**, 1144–1148 (1996).
- Khetarpal, I. *et al.* DNA sequencing using a four-color confocal fluorescence capillary array scanner. *Electrophoresis* **17**, 1852–1859 (1996).
- Bowling, J. M., Bruner, K. L., Cmarik, J. L. & Tibbetts, C. Neighboring nucleotide interactions during DNA sequencing gel electrophoresis. *Nucleic Acids Res.* **19**, 3089–3097 (1991).
- Yamakawa, H. & Ohara, O. A DNA cycle sequencing reaction that minimizes compressions on automated fluorescent sequencers. *Nucleic Acids Res.* **25**, 1311–1312 (1997).
- Fitzgerald, M. C., Zhu, L. & Smith, L. M. The analysis of mock DNA sequencing reactions using matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **7**, 895–897 (1993).
- Roskey, M. T. *et al.* DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Proc. Natl Acad. Sci. USA* **93**, 4724–4729 (1996).
- Kirpekar, F. *et al.* DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acid Res.* **26**, 2554–2559 (1998).
- Monforte, J. & Becker, C. High-throughput DNA analysis by time-of-flight mass spectrometry. *Nature Med.* **3**, 360–362 (1997).
- Fu, D. J. *et al.* Sequencing exons 5 to 8 of the *p53* gene by MALDI-TOF mass spectrometry. *Nature Biotechnol.* **16**, 381–384 (1998).
- Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
- Langer, P. R., Waldrop, A. A. & Ward, D. C. Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc. Natl Acad. Sci. USA* **78**, 6633–6637 (1981).
- Hawkins, T. L., O'Connor-Morin, T., Roy, A. & Santillan, C. DNA purification and isolation using a solid-phase. *Nucleic Acids Res.* **22**, 4543–4544 (1994).
- Uhlen, M. Magnetic separation of DNA. *Nature* **340**, 733–734 (1989).
- Ju, J. DNA sequencing with solid phase capturable dideoxynucleotides and energy transfer primers. *Anal. Biochem.* **309**, 35–39 (2002).
- Edwards, J. R., Itagaki, Y. & Ju, J. DNA sequencing using biotinylated dideoxynucleotides and mass spectrometry. *Nucleic Acids Res.* **29**, 1–5 (2001).
- Prober, J. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336–341 (1987).
- Gebeyehu, G., Rao, P., SooChan, P., Simms, D. A. & Klevan, L. Novel biotinylated nucleotide analogs for labeling and calorimetric detection of DNA. *Nucleic Acid Res.* **15**, 4513–4534 (1987).
- Herman, T., Lefever, E. & Shimkus, M. Affinity chromatography of DNA labeled with chemically cleavable biotinylated nucleotide analogs. *Anal. Biochem.* **156**, 48–55 (1986).
- Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
- Fakhrai-Rad, H., Pourmand, M. & Ronaghi, M. Pyrosequencing™: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* **19**, 479–485 (2002).
- Ruparel, H., Ulz, M. E., Kim, S. & Ju, J. Digital detection of genetic mutations using SPC–sequencing. *Genome Res.* (in the press).
- Friedman, L. S. *et al.* Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genet.* **8**, 399–404 (1994).
- Steele, R. J. C., Thompson, A. M., Hall, P. A. & Lane, D. P. The *p53* tumour suppressor gene. *Br. J. Surgery* **85**, 1460–1467 (1998).
- Kwok, P.-Y. High-throughput genotyping assay approaches. *Pharmacogenomics* **1**, 95–100 (2000).
- Roses, A. Pharmacogenetics and the practice of medicine. *Nature* **405**, 857–865 (2000).
- The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
- Beavis, R. C. & Chait, B. T. Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. *Rapid Commun. Mass Spectrom.* **3**, 436–439 (1989).
- Stoerker, J., Mayo, J. D., Tetzlaff, C. N., Sarracino, D. A., Schwöpe, I. & Richert, C. Rapid genotyping by MALDI-monitored nuclease selection from probe libraries. *Nature Biotechnol.* **18**, 1213–1216 (2000).
- Ross, P. L., Lee, K. & Belgrader, P. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal. Chem.* **69**, 4197–4202 (1997).
- Griffin, T. J., Hall, J. G., Prudent, J. R. & Smith, L. M. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proc. Natl Acad. Sci. USA* **96**, 6301–6306 (1999).
- Lyamichev, V. *et al.* Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnol.* **17**, 292–296 (1999).
- Haff, L. A. & Smirnov, I. P. Multiplex genotyping of PCR products with mass tag-labeled primers. *Nucleic Acids Res.* **25**, 3749–3750 (1997).
- Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl Acad. Sci. USA* **96**, 10016–10020 (1999).
- Ross, P., Hall, L., Smirnov, I. P. & Haff, L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol.* **16**, 1347–1351 (1998).
- Fei, Z., Ono, T. & Smith, L. M. MALDI-TOF mass spectrometry typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res.* **26**, 2827–2828 (1998).
- Griffin, T. J. & Smith, L. M. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* **18**, 77–84 (2000).
- Ding, C. & Cantor, C. R. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption/ionization time-of-flight MS. *Proc. Natl Acad. Sci. USA* **100**, 3059–3064 (2003).
- Ding, C. & Cantor, C. R. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. *Proc. Natl Acad. Sci. USA* **100**, 7449–7453 (2003).
- Kim, S., Edwards, J. R., Deng, L., Chung, W. & Ju, J. Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Res.* **30**, 1–6 (2002).
- Kim, S. *et al.* Multiplex genotyping of the human β_2 -adrenergic receptor gene using solid phase capturable dideoxynucleotides and mass spectrometry. *Anal. Biochem.* **316**, 251–258 (2003).
- Hanson, E. H., Imperatore, G. & Burke, W. HFE gene and hereditary hemochromatosis: A HUGe review. *Am. J. Epidemiol.* **154**, 193–206 (2001).
- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *p53* mutations in human cancers. *Science* **253**, 49–53 (1991).
- Bardelli, A. *et al.* Mutational analysis of the tyrosine kinase in colorectal cancers. *Science* **300**, 949 (2003).
- Kim, S. *et al.* Thirty fold multiplex genotyping of the *p53* gene using solid phase capturable dideoxynucleotides and mass spectrometry. *Genomics* (in the press).
- Malkin, D., Sexsmith, E., Yeger, H., Williams, B. R. G. & Coppes, M. J. Mutations of the *p53* tumor suppressor gene occur infrequently in Wilms' tumor. *Cancer Res.* **54**, 2077–2079 (1994).
- Lahoti, C., Thorne, P., Malkin, D. & Yeger, H. Immunohistochemical detection of *p53* in Wilms' tumors correlates with unfavorable outcome. *Am. J. Pathol.* **148**, 1577–1589 (1996).
- Min, B. M. *et al.* Inactivation of the *p53* gene by either mutation or HPV infection is extremely frequent in human oral squamous cell carcinoma cell lines. *Oral Oncol. Eur. J. Cancer Part B* **30**, 338–345 (1994).
- Popanda, O. *et al.* Mutation analysis of replicative genes encoding the large subunits of DNA polymerase- α and replication factors A and C in human sporadic colorectal cancers. *Int. J. Cancer* **86**, 318–324 (2000).
- Lin-Lee, Y. C., Tatebe, S., Savaraj, N., Ishikawa, T. & Kuo, M. T. Differential sensitivities of the MRP gene family and γ -glutamylcysteine synthetase to prooxidants in human colorectal carcinoma cell lines with different *p53* status. *Biochem. Pharmacol.* **61**, 555–563 (2001).
- Rodi, C. P., Darnhofer-Patel, B., Stanssens, P., Zabeau, M. and van den Boom, D. A strategy for the rapid discovery of disease markers using the massARRAY™ system. *Biotechniques* **32** (Suppl.), 62–69 (2002).
- Chen, X., Levine, L. & Kwok, P.-Y. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* **9**, 492–498 (1999).
- Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnol.* **21**, 673–678 (2003).
- Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variation. *Proc. Natl Acad. Sci. USA* **100**, 8817–8822 (2003).

Acknowledgements

This work was supported by a Packard Fellowship for Science and Engineering (J.J.) and the Columbia University Genomics Initiative.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

LocusLink: <http://www.ncbi.nlm.nih.gov/LocusLink>
BRCA1 | TP53

OMIM: <http://www.ncbi.nlm.nih.gov/omim>

Human hereditary haemochromatosis | squamous carcinoma | Wilms tumour

Access to this interactive links box is free online.

Eörs Szathmáry is a biologist with an interest in how higher and higher levels of evolutionary units emerged from those at lower levels. He has been strongly influenced by John Maynard Smith with whom he published a scholarly book (*The Major Transitions in Evolution*, Freeman, 1995) and a popular book (*The Origins of Life*, 1999) on this subject, which were translated into a dozen languages. One of his stronger research interests concerns theoretical approaches to the origins of life. He is now professor at Eötvös University Budapest and permanent fellow at the Collegium Budapest (Institute for Advanced Study).

URLS

Eörs Szathmáry's web page

<http://www.colbud.hu/fellows/szathmary.shtml>

Scripps Research Institute

<http://www.scripps.edu>

Steven Benner's web page

<http://www.chem.ufl.edu/benner.html>

ONLINE

Sobin Kim, a graduate of the Seoul National University in South Korea, received her Ph.D. in chemical engineering from Columbia University, New York, in 2001. She then joined the Laboratory of DNA Sequencing and Chemical Biology in the Columbia Genome Center as a postdoctoral research scientist, where she co-invented the multiplex genotyping approach using molecular affinity and mass spectrometry. Her research interest is in developing new genomic technologies using engineering principles.

Hameer D. Ruparel obtained his B.S. in chemical engineering in 2000, from the University of Mumbai, India, and his M.S. in chemical engineering in 2002, from Columbia University, New York. He is now a Ph.D. student in the Department of Chemical Engineering at Columbia University under the guidance of Jingyue Ju. His research work has been involved with the development of new mutation-detection approaches using solid-phase-capturable nucleotide analogues, photochemistry and MALDI TOF mass spectrometry.

T. Conrad Gilliam is professor of genetics and development at Columbia University and director of the Columbia Genome Center. His research is focused on the development of new genomic and bioinformatic approaches to study common heritable disorders in humans, and on the use of behavioural models in mice to inform the study of human neuropsychiatric disorders. He was responsible for determining the locations of the genetic defects that lead to spinal muscular atrophy and an X-chromosome-linked form of dystonia, and for identifying genes that mutate to cause many heritable human disorders including Wilson disease, retinitis pigmentosa and two forms of epilepsy.

Jingyue Ju received his Ph.D. in bio-organic chemistry from the University of Southern California in 1993. He was a United States Department of Energy Human Genome Distinguished Postdoctoral Fellow at the University of California at Berkeley from 1993 to 1995, during which time he co-invented the fluorescent energy-transfer labelling technology for DNA sequencing and analysis. From 1995 to 1999, he was a senior scientist and director of chemistry and assay development at Incyte Genomics, California. In 1999, he joined Columbia University as an associate professor of chemical engineering and head of DNA sequencing and chemical biology at the Columbia Genome Center. He is a recipient of a Packard Fellowship in science and engineering. His research interest is in the design and synthesis of new molecular tags for biological labelling and imaging, and developing new technologies for genomic research.

URLS

LocusLink

BRCA1

<http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=672>

TP53

<http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=7157>

OMIM

Human hereditary haemochromatosis

<http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispim?235200>

squamous carcinoma

<http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispim?601400>

Wilms tumour

<http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispim?194070>