**Single-Nucleotide Polymorphism (SNP) Analysis**

R M Twyman, University of York, York, UK

© 2009 Elsevier Ltd. All rights reserved.

**Introduction**

Single-nucleotide polymorphisms (SNPs) are individual nucleotide positions in the genome that vary at polymorphic levels in human populations. Most SNPs are diallelic (i.e., the SNP site may be occupied by either of two different bases), and the less common base must have a frequency of at least 1%, else it is classed as a rare mutation rather than a polymorphism. SNPs represent the most abundant form of genetic variation in humans, accounting for more than 90% of all differences between unrelated people. SNPs influence many central nervous system phenotypes, including susceptibility to neurological disorders. For example, an important genetic determinant in Alzheimer's disease is the APOE gene, which encodes apolipoprotein E (ApoE). The presence of two SNPs in this gene yields three common variants – APOE ε2, ε3, and ε4, which produce the proteins ApoE2, E3, and E4. Each of these variants is associated with a different level of disease susceptibility, APOE ε4 conferring the greatest risk and APOE ε2 the lowest. APOE ε4 homozygotes are the highest-risk group for Alzheimer's disease.

As well as affecting disease susceptibility, SNPs also affect our responses to drugs and other xenobiotics. Extending the example presented above, SNPs in the APOE gene contribute not only to the risk of Alzheimer's disease but also to drug responses, with the high-risk APOE ε4 homozygotes showing the poorest response to tacrine, an acetylcholinesterase inhibitor. SNPs therefore play a major role in all stages of the drug development process, from target identification through to clinical trials. The analysis of SNPs may help scientists tailor drugs and drug regimes to particular genotypes, which is the underlying principle of pharmacogenomics. In this context, it has been necessary to develop techniques for the rapid, high-throughput genotyping of SNPs in large samples.

**Principles of SNP Genotyping**

All SNP genotyping methods combine two distinct steps – one to discriminate between alternative alleles at the SNP locus and one to report the nature of the allele by generating a signal that can be detected, in an automated fashion if possible.

There are three general allele discrimination methods: hybridization/annealing (sometimes with a downstream enzymatic discrimination step), primer extension, and enzyme cleavage. In each case, the discrimination assay may take place in solution (so-called homogeneous assays) or in a solid phase immersed or coated in liquid, such as a microarray (so-called heterogeneous assays). In many of these assays, it is necessary to amplify the genomic DNA prior to detection of the allele, and amplification may indeed be integrated into the detection step. In other cases, the detection assay is sensitive enough to work directly on nonamplified genomic DNA or complementary DNA.

While allele discrimination methods come in three broad varieties, there are many different ways to generate and detect the signal, some based on fluorescence, others based on enzymatic reactions, and a few based on direct quantification. Most follow the fate of a label, either in real time or at the assay end point. Some methods based on mass spectrometry (MS) measure the masses of allele-specific oligonucleotides, so a label is not required.

**Allele Discrimination**

**Allele-Specific Hybridization**

The simplest method for discriminating between alleles at a SNP locus is hybridization using allele-specific oligonucleotide (ASO) probes. Two probes are required, one specific for each allele, and stringency conditions are employed such that a single-base mismatch is sufficient to prevent hybridization of the nonmatching probe (Figure 1). The two probes can be combined in labeled/unlabeled pairs, in which case two separate assays are required for each SNP (one in which probe 1 is labeled and probe 2 is unlabeled, and one in which these roles are reversed). Alternatively, and preferably in the case of high-throughput assays, distinct labels can be used for each probe (e.g., different fluorophores or mass tags) so that each SNP can be genotyped in a single reaction.

Allele-specific polymerase chain reaction (PCR) is an extension of the allele-specific hybridization principle, in which discrimination is achieved by allele-specific primer annealing followed by PCR amplification rather than the direct detection of a hybridized probe (Figure 2). Three primers are required, two to match alternative alleles at the SNP site and one to anneal to the opposite strand. Two reactions are set up, each containing one of the SNP primers together with the reverse primer. Although the variable base position can be sited somewhere in the middle of the SNP primers, in a manner directly analogous to the ASO strategy, this method tends to allow the extension of a mismatched primer–template because the
stringency of annealing cannot be optimized (annealing has to favor both primers in the reaction and is therefore difficult to optimize for the allele-specific primer alone). The preferred approach is to place the 3’ end of one of the primers at the variable base position since extension is dependent on perfect complementarity at the 3’ end of the primer.

A further extension of the allele-specific hybridization approach is allele-specific ligation, in which two oligonucleotides are annealed at adjacent sites and DNA ligase is used to join them together. One of the oligonucleotides has two allele-specific variants, and the discriminatory base is at the most 3’ position. If there is perfect complementarity between the allele-specific oligonucleotide and the template, it will be successfully ligated to the common oligonucleotide. If there is a mismatch, then ligation will fail, and the two oligonucleotides will remain separate. The assay benefits from the inclusion of additional steps such as the ligase chain reaction (in which the ligation reaction is reiterated with further oligonucleotides complementary to the first ligation product) or rolling circle amplification technology, in which two allele-specific oligonucleotides about 80 nt in length are designed to form a circle on the template, generating a closed loop (or padlock probe) when ligated. Primers annealing to this circle can be extended with a strand-displacing DNA polymerase so that when the nascent strand completes the circle and encounters itself, it is continually displaced, generating a long concatemer that is easy to detect using fluorescence methods.

**Minisequencing**

Minisequencing is also known as allele-specific primer extension. Like allele-specific PCR, it involves the extension of a primer annealing to the template DNA, but the primer in minisequencing is nondiscriminatory. The primer anneals one base upstream of the polymorphic site, and discrimination is achieved by detecting which base is incorporated in the first polymerization step (Figure 3). Allele-specific primer extension methods are more adaptable than hybridization/annealing assays because a much greater diversity of labeling strategies can be used. The free nucleotides in solution can be labeled with four different fluorescent tags, mass tags, or haptens, allowing the same mix to be used in the detection of many SNPs in a highly multiplexed fashion, such as on a microarray. Primer extension methods for genotyping have been developed commercially, for example as Genetic Bit Analysis technology by Orchid Biosciences, and they form the basis of many popular commercial genotyping systems, such as SNaPshot (Applied Biosystems).

**Allele-Specific Enzymatic Cleavage**

The use of enzymes to distinguish between alternative alleles is the basis of one of the earliest genotyping
methods – the analysis of restriction fragment length polymorphisms (RFLPs). An RFLP is generated when a single nucleotide polymorphism occurs at a restriction endonuclease recognition sequence, and one allele preserves the sequence while the other destroys it. In a sequence of genomic DNA with three adjacent restriction sites, the middle one containing a SNP, then digestion of amplified genomic DNA with the appropriate restriction endonuclease will produce either a single large fragment (if the central restriction site is absent) or two smaller fragments (if the central restriction site is present and cleavage occurs).

Traditional RFLP analysis cannot be applied on a large scale because it relies on gel electrophoresis and Southern blot hybridization, which are cumbersome to use in high-throughput experiments. Modern takes on RFLP analysis which combine restriction enzymes and PCR amplification can accelerate this process and facilitate moderate multiplexing (e.g., amplified fragment length polymorphisms and cleaved amplified polymorphic sites). However, these have been used more for linkage analysis than for genotyping. One enzymatic assay developed for genotyping is the Invader assay, which employs a unique method of allelic discrimination involving overlapping probes and an enzyme that specifically recognizes the resulting ‘flap’ (Figure 4). An advantage of the Invader assay is that it requires no PCR amplification. Two signal probes are used, one recognizing each allele, plus an invader probe. The signal and invader probes hybridize in tandem, and the signal probe overlaps the invader probe, generating a flap that is recognized by an enzyme called cleavase, a modified form of the thermostable FEN-1 (for flap endonuclease) enzyme. A flap with the appropriate structure is generated only if the signal probe completely matches the template and there is a one-base invasion. The cleaved flap can then be used in a second-round invader assay to generate a fluorescent signal, or the cleaved flap can be detected by MS.

### Allele Detection

#### Fluorescence Methods

**Direct detection of incorporated fluorescent labels** The most straightforward way to detect an allele-specific product is to incorporate one or more nucleotides conjugated to a fluorescent dye. Direct fluorescence detection is generally used with heterogeneous assay formats (microarrays, bead arrays) and is particularly suitable for minisequencing methods since single-base extension can be carried out with four nucleotides bearing different fluorophores, which allows the immediate detection of alleles on the basis of the emission wavelength of the extension product. Direct fluorescence is also often used when gel or capillary electrophoresis is used to separate allele-specific products.

**Direct detection of intercalated fluorescent labels** Intercalating fluorescent dyes such as ethidium bromide fluoresce much more brightly in the presence of double-stranded DNA than single-stranded DNA. This is the basis of a unique detection method known as dynamic allele-specific hybridization, which involves a single (unlabeled) allele-specific oligonucleotide probe that is added to a PCR product containing an SNP site and allowed to anneal when ethidium bromide is present. The mixture is then heated, and fluorescence is measured in real time as the temperature rises. The fluorescence decreases as denaturation (melting) occurs, and the thermal melting profile for each allele will be different because only one is perfectly complementary to the probe and the other contains a mismatch.

**Fluorescence resonance energy transfer** Fluorescence resonance energy transfer (FRET) occurs when two fluorophores are in proximity and one of them (the donor) has an emission spectrum that overlaps the excitation spectrum of the other (the acceptor). When a lone donor fluorophore is excited, light is produced with a characteristic emission spectrum. However, when the donor fluorophore is excited in proximity to the acceptor, energy is transferred to the acceptor fluorophore with the result that the intensity of emission from the donor is reduced (quenched) while that of the acceptor is enhanced.

Although FRET enhancement can be used for SNP genotyping, two of the most popular homogeneous assays in current use employ FRET quenching. The TaqMan assay exploits the intrinsic 5’ nuclease activity of Taq DNA polymerase to generate a fluorescent
signal from a short allele-specific oligonucleotide probe. Two probes are required, one specific for each allele. Each probe contains a unique donor fluorophore and a common acceptor fluorophore and is short enough for the donor to be quenched when the probe is intact (either hybridized or in solution). When the probe hybridizes to the PCR template, however, the 5′ exonuclease activity of Taq DNA polymerase digests it, thus releasing the two fluorophores into solution and eliminating the quenching effect (Figure 5). The molecular beacon assay involves the use of longer probes that have self-complementary ends labeled with the donor and acceptor fluorophores. Nonhybridized probes will self-anneal, bringing the donor and acceptor fluorophores together, resulting in quenching. Where hybridization occurs, the donor and acceptor fluorophores are separated such that the quenching effect is eliminated and a fluorescent signal is produced. As with the TaqMan assay, two probes labeled with different donor fluorophores are required, one specific for each allele (Figure 6).

Fluorescence polarization When a fluorophore is excited by plane polarized light, the fluorescence emitted by the dye is also polarized, a phenomenon termed fluorescence polarization (FP). Complete FP occurs only when a dye molecule is stationary. The degree of FP thus depends on the movement of molecules in solution, which in turn depends on various factors, including the molecular mass. SNP genotyping assays can change the molecular mass of a target oligonucleotide, for example by primer extension, probe hydrolysis, or invasive cleavage, and this is the basis of commercial genotyping systems such as SNaPshot. Such changes can be recorded as changes in FP as long as the temperature and viscosity of the solution remain the same.

Pyrosequencing

Pyrosequencing is a novel method for sequencing short stretches of DNA on the basis of the detection of pyrophosphate, a normal by-product of DNA synthesis. Although similar in principle to primer extension allele discrimination methods, pyrosequencing is suitable not only for typing SNPs but also for scoring entire haplotypes (groups of linked SNPs). Pyrosequencing works on the basis that pyrophosphate can be used to generate adenosine triphosphate (ATP), which then stimulates luciferase activity, causing the emission of a chemiluminescent signal. To achieve this, the primer extension reaction must include APS and the enzyme ATP sulfurylase, which converts adenosine 5′ phosphosulfate into ATP in the presence of pyrophosphate. Also present are luciferin (the substrate of luciferase) and the enzyme apyrase, which continuously degrades unincorporated deoxyribonucleotide triphosphates (dNTPs) and excess ATP. The dNTPs are added to the reaction one by one. Only if the incoming dNTP is complementary to the template will it extend the primer and release pyrophosphate, resulting in the production of an equimolar amount of ATP. Visible light is generated in proportion to the amount of ATP and is detected as a peak on a pyrogram. When the degradation of the present dNTP is completed by apyrase, the next is added. A thioderivative of dATP must be used to avoid constant stimulation of luciferase activity.
Mass Spectrometry

Genotyping assays involving MS are distinct from those discussed above in that signal detection is direct, based on differing molecular weights of small DNA fragments rather than the behavior of a label. The analysis of DNA by MS requires soft ionization (i.e., without fragmentation) and is usually achieved by matrix-assisted laser desorption/ionization (MALDI) followed by time of flight (TOF) analysis. The MALDI procedure involves mixing the allele-specific products of the discrimination assay with a matrix compound on a metal plate. The mixture is then heated with a short laser pulse, causing it to expand into the gas phase, where ionization is achieved by application of a strong potential difference. Ions are accelerated toward the detector, and the TOF (the time taken to reach the detector) is measured, allowing the mass–charge ratio to be calculated.

High-resolution mass spectrophotometers can accurately discriminate between alternative alleles in DNA fragments of three to 20 nt in length. The smallest difference in mass between alternative allelic products is about 9 Da, representing an adenosine–thymidine polymorphism. Advantages of MS-based methods include accuracy, as long as the sample is very pure, and the fact that each genotyping reaction takes a fraction of a second, which allows thousands of reactions to be carried out serially in a single day.

The majority of current MS genotyping platforms involve the use of DNA chips because the chip can be used directly as a MALDI plate. Thus far, the technique has been used to determine the masses of allele-specific hybridization probes, allele-specific primer extension products, and the products of invasileavage reactions. Commercial systems utilizing MS include MassArray (Sequenom) and PinPoint (PerSeptive Biosystems).

Another promising MS genotyping technology involves the use of cleavable mass tags (chemical moieties of known molecular mass) as labels to replace fluorophores. The advantage of mass tags is that, because of the accuracy of MS-based mass determination, the range of labels that can be used is virtually limitless, so ultrahigh-throughput genotyping is possible. The MassCode system (Qiagen Genomics) uses this technology.

Summary

High-throughput SNP genotyping has helped facilitate association studies aiming to find patterns among the 0.1% of variation in the human genome (mostly represented by SNPs) corresponding to differences in complex neurological phenotypes such as susceptibility to neuropathophysiological disorders, psychiatric disorders, and responses to drugs. Many of the technologies described above facilitate highly multiplexed analysis, although MS stands out as a platform in which increased throughput is realized by ultrarapid serial processing.

The sensitivity of most genotyping techniques reflects some form of amplification step either before, during, or after the actual allele discrimination assay. In the majority of cases, this amplification step relies on the PRC, but novel formats such as the invader assay and rolling circle amplification technology are also becoming popular. While homogeneous assay formats are versatile, array-based technology platforms have the advantage of higher throughput. Recent developments include the use of electrostringent hybridization arrays, in which electric currents are used to increase the speed and sensitivity of hybridization, and the detection of allele-specific products on the basis of their electrochemical properties when hybridized to oligonucleotide probes attached to minute electrodes. Another emerging technology, bead arrays, combines the flexibility of the homogeneous assay format with the convenience and throughput of chip-based detection systems. Bead array systems that employ fluorescence-based detection and sorting are commercially available from Lynx Therapeutics and Illumina.

See also: Alzheimer’s Disease: Molecular Genetics; Brain Glucose Metabolism: Age, Alzheimer’s Disease and ApoE Allele Effects; Fluorescence Microscopy in the Neurosciences; Genomics of Brain Aging: Apolipoprotein E; Single Cell Genomic DNA Analysis.

Further Reading