SNP and its Applications in Treating Human Diseases

1S. Abdolhamid Angaji, 2K Akashi, 3S. Darvishani, 4M. Sabeti Azad

1,2,3,4Department of Biology, Tarbiat Moallem University, Tehran, Iran.

Abstract: Single nucleotide polymorphism (SNP) is a single base mutation in DNA. SNPs are the simplest form and also the most common sources of genetic polymorphism in the human genome. They have been used in homogeneity testing, plant genetics, pharmacogenetic studies and so on. They have also been used to identify and map complex and common diseases such as high blood pressure, diabetes, and heart disease and etc. Application of SNP in human diseases is studied in this work. There are several methods for SNP genotyping as classified in four categories in this work: Hybridization-based, PCR-based, Restriction site cleavage and Combinations methods. The diseases-associated SNP which are detected by any of these methods is investigated. Finally, the application of some specific methods in detecting SNP for a certain disease was explained. These include the application of SNP array in diabetes disease, the application of Pyrosequencing in Schizophrenia disease, the application of DHPLC in breast cancer disease and the application of melting curve in Parkinson disease.

Key words: Single base mutations, Genotyping, Hybridization.

INTRODUCTION

Various scientific endeavors had already started even before the completion of the first human genome reference sequence to identify unique genetic differences between individuals. 99.9% of one individual DNA sequences will be identical to that of another person. Of the 0.1% difference, over 80% will be single nucleotide polymorphisms (Elmer).

A SNP is a source variance in a genome (Gobbi, 2006) which was discovered by Dr. Steve Ligget. A SNP (single nucleotide polymorphism) is a single base mutation in DNA. SNPs are the most simple form and most common source of genetic polymorphism in the human genome (Gobbi, 2006). SNPs are not uniformly distributed over the entire human genome, neither over all chromosomes and nor within a single chromosome. There are one third as many SNPs within coding regions as non-coding region SNPs (Gobbi, 2006). Current estimates are that SNPs occur as frequently as every 100-300 bases (Elmer). There are variations between human populations, so a SNP allele that is common in one geographical or ethnic group may be much rarer in another (SNP). Some SNPs may be specific for a certain population. In addition, groups of SNPs that are associated with each other (linkage disequilibrium (LD) block) are not the same in different populations (Wang, 2001).

Single nucleotides may be changed (substitution), removed (deletions) or added (insertion) to a polynucleotide sequence SNPs are associated with diversity in the population, individuality, susceptibility to diseases, and individual response to medicine. Recently, it has been suggested that SNPs can be used for homogeneity testing and pharmacogenetic studies and to identify and map complex, common diseases such as high blood pressure, diabetes, and heart disease etc (Shastry, 2002; Chanock, 2003).

Applications of Different Methods in Detecting SNP-Associated Diseases:

Genotyping provides a measurement of the genetic variation between members of a species. SNP are one of the most common types of genetic variation. Because SNPs are evolutionarily conserved, they have been proposed as markers for use in quantitative trait loci (QTL) analysis and in association studies in place of microsatellites. SNPs can also provide a genetic fingerprint for use in identity testing.

The increase in interest in SNPs has been reflected by the furious development of a diverse range of SNP genotyping methods. SNPs present a potentially vast arena for the detection of genetic alterations that seem
to relate to medically important differences in disease susceptibility and drug response (Lohmann, 2000).

In this part some SNP genotyping methods which have been arranged in four groups; Hybridization-based methods, PCR-based methods, Restriction site cleavage methods and Combinations methods and their applications in human diseases are studied.

In the Hybridization based methods (Table 1), specify of sequences are determinant. Several applications have been developed that interrogate SNPs by hybridizing complementary DNA probes to the SNP site through hybridization based methods. The challenge of this approach is reducing cross-hybridization between the allele-specific probes. This challenge is generally overcome by manipulating the hybridization stringency conditions and alleles can be distinguished by hybridizing complementary oligonucleotide sequence to a target sequence (Rapley, 2004). There are several kinds of methods which are classified as Hybridization based methods. In this study, we explained the application of SNP Array, Allele specific oligonucleotide ligation, Allele specific hybridization, Array hybridization and fluorescence detection, Single-feature polymorphism (SFP), Invader probe (Invader PCR) and MutS protein binding assays methods, in human diseases, as subclasses of Hybridization based methods.

In the PCR-based methods (Table 2), the methods operate on the basis of the specific amplification of a target allele by the polymerase chain reaction with extension primers designed such that their 3’ end is placed at the mutation site. When this base is complementary to that of the specific allele, the DNA segment is amplified; when it is not complementary, the polymerase chain reaction cannot proceed (Angaji, 2009). There are several kinds of methods which are classified as PCR-based methods. In this study, we explained the application of Primer extension, Minisequencing, Single base primer extension (SBE) method, method, Pyrosequencing and Allele-specific extension methods, in human diseases, as subclasses of PCR-based methods.

Restriction site cleavage methods (Table 3) are marker techniques that rely on the recognition site of a specific restriction enzyme that can maximally detect all the corresponding restriction sites within a genome. In contrast, every SNP in context with its surrounding genomic sequence is unique. SNPs can mark functionally important allelic differences and SNPs that flag individual alleles of known genes have been used widely as molecular markers (Angaji, 2009). This technique is used when an SNP occurs at a restriction endonuclease recognition sequence, and one allele preserves the sequence while the other destroys it. If we consider any DNA fragment with three adjacent restriction sites, with the middle one containing an 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) (Twyman, 2005). There are several kinds of methods which are classified in the restriction site cleavage methods. In this study, we explained the application of Invasive cleavage, Cleavage fragment length polymorphism (CFLP), Denaturing Gradient Gel Electrophoresis (DGGE), Denaturing high performance liquid chromatography (DHPLC), Single strand conformation polymorphism (SSCP) and Chemical cleavage mismatch methods, in human diseases, as subclasses of Restriction site cleavage methods.

In combinations methods (Table 4), methods have been gathered, centered on a statistical approach (Kirk, 2002; Suha, 2005; Winkler, 2006; Tokuhiro, 2003; Gail, 2008; Hawn, 2005; Jiang, 2004; Toyota, 2000; Pham, 2005; Pont-Kingdon, 2003) or followed through multiple approach and instruments (Gerhardus, 2007) that finally result in detecting single nucleotide polymorphisms in genome.

### Table 1: Hybridization-based methods and their applications in treating human diseases.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Allele specific hybridization</td>
<td>Breast cancer (Gerhardus, 2007).</td>
</tr>
</tbody>
</table>
In statistical approach first, with one detecting method SNPs are identified in genome and then the results are analyzed with statistic analysis. There are several kinds of methods which are classified as combinations methods. In this study, we explained the application of Constructing haplotypes in Linkage-Disequilibrium-block, Case-Control association study, Melting curve analysis and Peptide nucleic acid (PNA) probes, an optically amplifying conjugated polymer (CP), and S1 nuclease enzyme methods, in human diseases, as subclasses of Combinations methods.

Table 2: PCR-based methods and their applications in treating human diseases.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Primer extension</td>
<td>Coronary artery disease and diabetes (Broadbent, 2008).</td>
</tr>
<tr>
<td>1.1. Minisequencing</td>
<td>Imbalanced allelic expression in foetal liver or kidney tissues (Liljedahl, 2004)</td>
</tr>
<tr>
<td></td>
<td>Detection of alternatively spliced transcripts in leukemia cell lines (Milani, 2006)</td>
</tr>
<tr>
<td></td>
<td>Chronic obstructive pulmonary (Hersh, 2006).</td>
</tr>
<tr>
<td>1.3. Allele-specific extension</td>
<td>Breast cancer (Li, 2007).</td>
</tr>
<tr>
<td>2. single base primer extension (SBE method)</td>
<td>Broad range of analytical necessities: single-cell analysis, molecular diagnosis of monogenic diseases, forensic mitochondrial DNA analysis on highly degraded human remains, high-throughput SNP screening for population studies (Podini, 2009).</td>
</tr>
</tbody>
</table>

Table 3: Restriction site cleavage methods their applications in treating human diseases.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Invasive cleavage</td>
<td>10 mutations in the cystic fibrosis transmembrane conductance regulator gene (Chen, 2005).</td>
</tr>
<tr>
<td>2. DGGE</td>
<td>AIDS (Petersen, 2005).</td>
</tr>
</tbody>
</table>

Table 4: Combinations methods their applications in treating human diseases.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Peptide nucleic acid (PNA) probes, an optically amplifying conjugated polymer (CP), and S1 nuclease enzyme</td>
<td>Neurodegenerative disease (Gaylord, 2005).</td>
</tr>
</tbody>
</table>

Case Studies:
Here the applications of four methods in detecting SNP-associated diseases are discussed. From each of Hybridization-based methods, PCR-based methods, Restriction site cleavage methods and Combinations methods groups, one method is chosen and its application in a certain diseases is studied.

An SNP array is a useful tool to study the whole genome. The most important application of SNP array is in determining disease susceptibility and consequently, in pharmacogenomics by measuring the efficacy of drug therapies specifically for the individual.
Framingham Heart Study (FHS) family members were genotyped on the Affymetrix 100K single nucleotide polymorphism array and examined for association with incident diabetes and six diabetes-related quantitative traits (Florez, 2007).

In another study a dense genomewide linkage search of the family was undertaken using a first generation 10K single nucleotide polymorphism chip containing 10,044 markers. A region of homozygosity harboring the neonatal diabetes disease gene on chromosome 10p12.1-p13 was identified. It is likely that chromosome 10p12.1-p13 may harbor maturity-onset diabetes of the young or type 2 diabetes genes (Sellick, 2003).

Type 2 diabetes, a complex disease that is characterized by insulin resistance and impaired β-cell function, represents a serious global public health problem, with more than 100 million people affected worldwide. In another study DNA from 124 type 2 diabetic case subjects and 295 control subjects with normal glucose tolerance were genotyped on the Affymetrix 100K single nucleotide polymorphism array. Recently, four GWAS studies of type 2 diabetes have identified variants at several novel loci, including SLC30A8, IGF2BP2, CKDAL1, CDKN2A/CKDN2B, and HHEX/IDE, that show strong replicated association with type 2 diabetes (Rampersaud, 2007).

Trace amines and their receptors may be implicated in the pathogenesis of psychiatric disorders. Previous studies have reported association of the trace amine associated receptor 6 (TAAR6) genes with susceptibility to schizophrenia. In one study in Korean patients, five single nucleotide polymorphisms (SNPs: rs4305745; rs8192625; rs7452939; rs6903874 and rs6937506) were genotyped in the TAAR6 gene and in the 3′ regulatory region, using pyrosequencing. SNP rs6903874 was significantly associated with schizophrenia (p = 0.012) and bipolar disorder (p = 0.004). A three SNP haplotype consisting of alleles GCT from SNPs rs7452939, rs6903874 and rs6937506, respectively, was significantly over-represented in patients with schizophrenia (p = 0.0003) and bipolar disorder (p = 0.00002). A second three SNP haplotype (GTT) derived from the same SNPs was significantly under-represented in patients with bipolar disorder (p = 0.001).

In another study, they found a highly significant association between schizophrenia and a COMT haplotype (P=9.5×10^{-5}). To test the COMT-schizophrenia association, they examined 12 SNPs in the COMT gene using the quantitative pyrosequencing technology. Of the 12 SNPs studied, 4 were found not to be polymorphic in the population they studied, and an additional SNP failed to undergo PCR amplification. Among the remaining seven SNPs, they found significant allele-frequency differences between patients and control individuals for five SNPs. Using 12 SNPs distributed in the 27 kb of the COMT gene and testing 7 of them with a very large sample of Ashkenazi Jews, they have been able to confirm a complex association between the COMT gene and schizophrenia. Their results also suggest the possibility of more than one functional polymorphism affecting susceptibility to schizophrenia at the COMT locus. This is supported by the finding that SNP rs165599 affects primarily women, whereas rs737865 affects both sexes, although in different ways (Shifman, 2002).

In another study, they genotyped 100 Japanese schizophrenics and 100 controls recruited from the Kyushu area for 11 SNP markers distributed in the SLCA12 region using the direct sequencing and pyrosequencing methods, and examined allele, genotype and haplotype association with schizophrenia. They found significant differences in genotype and allele frequencies of SNP2 between cases and controls (P = 0.013 and 0.008, respectively). After Bonferroni corrections, the two significant differences disappeared. They tested haplotype associations for all possible combinations of SNP pairs. SNP2 showed significant haplotype associations with the disease (P = 9.4 × 10^{-5}, P = 0.0052 with Bonferroni correction, at the lowest) in 8 combinations. They concluded that at least one susceptibility locus for schizophrenia is probably located within or nearby SLCA12 in the Japanese population (Deng, 2004).

In another study, using restriction fragment length polymorphism and pyrosequencing methods, they genotyped two TNFA gene promoter SNPs (-G308A, -G238A) and analyzed the haplotype structure in 24 Canadian families of primarily Celtic origin. Their results demonstrate that after correction for multiple testing based on simulations of 10 000 replicates of unlinked/unassociated data, there is evidence for association (P=0.026) of a specific haplotype (-308A, -238G) with schizophrenia and schizophrenia spectrum disorders with a family-based trimmed haplotype linkage disequilibrium test (Saviouk, 2005).

In another study, they investigated the gene for dystrobrein-binding protein 1 (DTNBP1), or dysbindin, which has been strongly suggested as a positional candidate gene for schizophrenia, in three samples of subjects with schizophrenia and unaffected control subjects of German (418 cases, 285 controls), Polish (294 cases, 113 controls), and Swedish (142 cases, 272 controls) descent. They analyzed five single-nucleotide polymorphisms (P1635, P1325, P1320, P1757, and P1578) and identified significant evidence of association in the Swedish sample but not in those from Germany or Poland.

Their results suggest that genetic variation in the dysbindin gene is particularly involved in the development of schizophrenia in cases with a familial loading of the disease (Bogaert, 2003).

Breast cancer predisposition genes identified to date (e.g., BRCA1 and BRCA2) are responsible for less than 5% of all breast cancer cases. Many studies have shown that the cancer risks associated with individual commonly occurring SNPs are incremental. However, polygenic models suggest that multiple commonly occurring low to modestly penetrant SNPs of cancer related genes might have a greater effect on a disease when considered in combination (Ona, 2006).

Ten studies included report on 12 different scanning methods for the detection of BRCA1 and BRCA2 mutations. Among them, only DHPLC can be considered to be used as a scanning test in a routine diagnostic setting. The DHPLC is a chromatography-based method and is able to distinguish between homoduplcs – in the case of two wild-type alleles- and heteroduplces in the case of a heterozygous sequence alteration (SNP). The advantage of DHPLC is that it enables the rapid, sensitive, and accurate identification of polymorphisms and mutations in an automated fashion (Atencio, 2001). It is used to detect mutations with a high sensitivity in BRCA2 and BRCA1.

In one study 71 breast and breast-ovarian cancer (HBC/HBOC) families along with 95 control individuals from a wide range of ethnicities were analyzed by means of denaturing high-performance liquid chromatography (DHPLC) and direct sequence analysis. In the coding (10 257 bp) and non-coding (2799 bp) sequences of BRCA2, 82 sequence variants were identified (Wagner, 1999).

Indirect evidence suggests that, in some cases, the mitochondrial dysfunction in idiopathic Parkinson disease (PD) may be due to mutations in mitochondrial DNA (mtDNA). A SNP leading to alterations in complex I activity may play a role in the pathogenesis of PD. The activity of complex I is reduced within brains of patients with PD. Related groups of polymorphic sites in mtDNAs are called haplogroups. Genotyping of SNPs by using fluorescent probes offers several advantages. This technology should prove useful for population-based studies on the interaction between genetic factors and environmental exposures and the risk of PD. In one study, an assay system for rapidly genotyping and haplogrouping mitochondrial polymorphisms based on hybridization probe technology and melting curve analysis was developed. The result was that, in comparison to individuals carrying the most common haplo-group, haplogroup H, individuals classified as haplogroup J and K demonstrate less risk of developing PD. This protective effect is strongly associated with the SNP 10398G that defines haplogroups J and K. The absence of 10398G characterizes haplogroup U and may be associated with an increased risk of PD. SNP 9055A may be associated with a protective effect for women (Jiang, 2004).

In the another study To investigate the H1-SNP 13, they designed variant-specific probes and carried out a melting curve analysis on the LightCycler system with the melting temperature for the G allele of 63°C and for A of 54°C. As a result, an association of the H1 haplotype and subhaplotypes in the microtubule-associated protein Tau (MAPT) gene with PD has been reported (Winkler, 2006).

REFERENCES


