Special Article

Genetic Polymorphisms and Implications for Human Diseases

Orapan Sripichai PhD*,
Suthat Fucharoen MD*

** Thalassemia Research Center, Institute of Science and Technology for Research and Development, Mahidol University, Phutthamonthon, Nakhonpathom*

After the sequencing of the human genome is done, enormous genomic information and high-throughput profiling technologies are used. Increased attention has been paid to applying this knowledge to get better understanding of inherited diseases and complex disorders. Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide in the genome sequence is altered. SNPs are an important tool for the study of the human genome. Application of SNPs analysis to human disease permits exploration of the influence of genetic polymorphisms on disease susceptibility, drug sensitivity/resistance, and ultimately health care. Databases of SNPs provide a powerful resource for association studies that try to establish a relationship between a phenotype and regions of the genome.

Genomic approaches have garnered so much attention and investment because they offer the potential to provide better understanding of genetic factors in human health and disease, as well as more-precise definitions of the non-genetic factors involved.

Keywords: *SNPs, Single nucleotide polymorphism, Association studies*

J Med Assoc Thai 2007; 90 (2): 394-8
Full text. e-Journal: *http://www.medassocthai.org/journal*

Genetic and environment factors are the two keys that make human phenotype variations[1]. When the genomic DNA sequences on equivalent chromosome of any two individuals are compared, there is substantial variation in the sequence at many points throughout the genome. There are many forms of these genetic variations. The simplest type results from a single base mutation substitutes one nucleotide for another, and is known as single nucleotide polymorphism (SNP) (Box 1). Many other variations may result from the insertion or deletion of a section of DNA. The most common insertion/deletion events occur in repetitive sequences elements, where the repeated nucleotide patterns or variable number tandem repeat polymorphisms (VNTRs) expands or contacts as a result of insertion or deletion[2,3]. These DNA sequences variations are sometimes described as *mutations* and sometimes as polymorphisms. A mutation is defined as any change in a DNA sequence away from normal. This implies there is a normal allele that is prevalent in the population and that the mutation changes it to a rare and abnormal variant. In contrast, a polymorphism is a DNA sequence variation that is common in the population. In this case, no single allele is regarded as the standard sequence. Instead, there are two or more equally acceptable alternatives. The arbitrary cut-off point between a mutation and a polymorphism is 1 percent. That is, to be considered as a polymorphism, the variation must have a frequency of 1 percent or greater in a given population. If an allele occurs at a frequency lower than 1 percent, the allele is regarded as a mutation[4].

Although more than 99 percent of human DNA sequences are the same across the population, variations in DNA sequences may have a major impact on how human beings respond to disease, bacteria, viruses, toxins, chemicals, drugs, and other therapies[5].

Correspondence to : Fucharoen S, Institute of Science and Technology for Research and Development, Mahidol University, Phutthamonthon, Nakhonpathom 73170, Thailand. Phone: 0-2889-2558, Fax: 0-2889-2559

Many clinical phenotypes observed in diseases seem to have considerable genetic components. The presence of a specific genetic variation allele may be implicated as a causative factor in human genetic disorders. Therefore, screening for such allele in an individual might enable the detection of a genetic predisposition to disease. In the mean time, some polymorphisms sequences variants may merely serve as a modifying risk for some phenotype. Many polymorphisms may be found within genes and may influence characteristics such as height and hair color rather than medical importance while some does contribute to disease susceptibility and can influence drug responses. However, many polymorphisms are found outside the genes and are completely neutral in effect[3].

Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are genetic variations that occur when a single nucleotide: adenine (A), thymine (T), cytosine (C) or guanine (G) in the genome sequences is altered (Box 2). SNPs are mostly biallelic polymorphisms, that is, the nucleotide identity at these polymorphic positions is generally constrained to one of two possibilities in human, rather than the three or four nucleotide possibilities that could occur. SNPs are the most common form of genetic variation, accounting for 90 percent of all human genetic variations. Their density is estimated to occur approximately every 500-1,000 bases in the entire human DNA sequences, leading to a total of several millions SNPs in the overall human population (Box 3)[6].

SNPs may occur in both coding and non coding regions of the genome. Because only about 3 to 5 percent of a human DNA sequences code for the production of proteins, changes in non-coding sequences is more common than coding sequences changes. Synonymous changes in coding sequence are more common than non-synonymous changes. SNPs found within a coding sequence are of particular interest to scientists because they are more likely to alter the biological function of a protein. Occasionally, a SNP may actually cause a disease, and can be used to search for and isolate the disease-causing gene.

Many SNPs have no effect on cell function, but it is believed that they could predispose people to disease or influence their response to drug. However, SNPs are not absolute indicators of disease development. A good example is the genes associated with the late onset Alzheimer's, apolipoprotein E or ApoE[7]. This gene contains two SNPs that result in three possible alleles for this gene: ε2, ε3, and ε4. An individual who inherits at least one ε4 allele will have a greater chance of getting Alzheimer's. In addition, those who have inherited two ε4 alleles may never develop Alzheimer's,

Single Nucleotide Polymorphisms (SNPs)

SNP Facts:

- Human share about 99.9% sequence identity. The other 0.1% are mostly SNPs.
- SNPs occur about every 1000 bases. There are hot spots of SNPs.
- Most SNPs have only 2 alleles (biallelic SNP).
- Most SNPs not in coding regions (99% not in genes).
- SNPs can cause silent, harmless, harmful, latent changes.

Box 1.

Genetic Alphabetic Order 4 Letters
A, T, C, G

.....ATCGGGTACCGTACAACCTTAGCTAGGCTAAGCCC.....
.....ATCGGGTACTGTACAGCCTTAGCTAGGCAAAGCCC.....
↑ ↑ ↑
SNP SNP SNP

Box 2.

Total Number of SNPs in PHASE II HapMap Project

• Chromosome 1	452,581	• Chromosome 13	229,548
• Chromosome 2	529,264	• Chromosome 14	173,602
• Chromosome 3	388,015	• Chromosome 15	150,653
• Chromosome 4	364,437	• Chromosome 16	170,126
• Chromosome 5	347,177	• Chromosome 17	142,081
• Chromosome 6	386,131	• Chromosome 18	164,710
• Chromosome 7	302,572	• Chromosome 19	102,133
• Chromosome 8	301,965	• Chromosome 20	181,411
• Chromosome 9	257,037	• Chromosome 21	86,218
• Chromosome 10	307,570	• Chromosome 22	97,026
• Chromosome 11	297,481	• Chromosome X	183,229
• Chromosome 12	279,403	• Chromosome Y	104

Total 5,894,474

Box 3.

while another who has inherited two ε2 alleles may. Most SNPs are not responsible for a disease state. Instead, they serve as biological markers for pinpointing a disease on the human genome map, because they are usually located near a gene found to be associated with certain disease[3,5].

The importance of SNPs in genetic studies comes from at least three different categories. First, SNPs can be used to reconstruct the history of genome. This is due to their abundance and most are inherited from one generation to the next, evolutionary stable, making them easier to follow in population studies. Studying the frequency and distribution of SNPs can lead to information on the evolution of the species. Second, SNPs can be directly responsible for genetic diseases since they may alter the genetic sequence of gene or of a regulatory region. Finally, SNPs may be utilized as markers to build the high-density genetic maps need to perform association studies, which locate and identify genes of functional importance. It has been proposed that a set of 3,000 biallelic SNP markers would be sufficient for whole-genome-mapping studies in humans; a map of 100,000 or more SNPs has been proposed as an ultimate goal to enable effective genetic-mapping studies in large populations[8].

As previously described, SNPs may also be associated with the responses to therapeutic agent. The best examples are genetic polymorphisms of drug-metabolizing enzymes, which affect about 30% of all drugs. An effective treatment proven in one patient may become ineffective in the others. Therefore, the most appropriate drug for an individual (*personalized medicine*) could be determined in advance before treatment by analyzing a patient's SNP profile[9].

SNPs Based Association Study

A trial to establish a relationship between a phenotype (usually a disease) and one or more regions of the genome is known as association studies[3]. This is possible by measuring the frequencies of a number of SNPs in two populations that differ for the presence of the phenotype, and by detecting SNPs that show a significant difference in frequency. If the factor contributes to an increased risk for one phenotype occurrence, then that factor should be found at a higher frequency in individuals with that phenotype compared to non-phenotypic controls. This type of association is often called linkage disequilibrium (LD). The polymorphism being tested may have direct functional consequences bearing on the phenotype state. Alternatively, knowing the genomic position of those SNPs may lead to identifying gene involved in the phenotype under consideration since the groups of closely linked allele at nearby SNP sites tend to be inherited together as haplotypes[8].

The genome-wide association study approach is difficult because researchers have to contend with recombination events that have accumulated over, perhaps, tens of thousands of years. In a sense, recombination has had time to shuffle the DNA into smaller blocks, and that requires more markers. Therefore, the success of an association study is a direct

function of the number, density, and quality of the SNPs used. To cover the entire genome, anywhere from 100,000 to 1,000,000 SNPs markers might be needed for an association study[6,8]. One way to address the cost, time, and labor that are involved in large-scale genotyping is to carry out analyses not on individual DNA samples, but on pools made up of DNA from many individuals[10].

The other approach to perform association studies more effectively is the candidate-gene approach to limit the number of SNPs tested for pathogenic effect by careful pre-selection[3,5]. The basis of the selection might focus on biologically defined candidate genes, genes suggested by differential display experiments, or positional candidates from prior linkage investigations. Employing SNPs, which are more likely to have functional consequences, such as nonsynonymous coding SNPs and promoter variants, is obviously sensible.

A number of association studies, focused on candidate genes, regions of linkage to a disease. More large-scale studies have already led to the discovery of genetic risk factors for common diseases. Examples include Alzheimer's disease (APOE)[7], type 1 diabetes (human leukocyte antigen (HLA)[11], type 2 diabetes (PPARG)[12], deep vein thrombosis (factor V)[13], myocardial infarction (LTA)[14], stroke (PDE4D)[15], and asthma (ADAM33)[16]. However, in some studies, the future approach, including the multi-cohort collaboration, will be needed to clarify the preliminary associations and identify other potential candidate genes.

SNPs Based Association Study in Thalassemia

Beta-thalassemia/HbE disease is one of the most common thalassemias in Southeast Asia. The clinical expression of this disease is remarkably variable, ranging from nearly asymptomatic to severe, transfusion-dependent disease. A number of additional genetic modifier genes may account for this variability. To identify genetic modifiers influencing severity among 1060 β^o-thalassemia/HbE patients, the authors used a MassARRAY spectrometry to conduct a genome-wide association study involving approximately 110,000 gene-based SNPs. This assay panel corresponds to SNPs with a median spacing of 10.4 kilobases, in approximately 99% of all known and predicted human genes. DNA from approximately 200 regionally matched patients representing the extremes of mild and severe cases β^o-thalassemia/HbE were included in each DNA pool. To determine precise allele and genotype frequencies, more than 600 SNPs that showed reproducible allelic differences at $p < 0.05$ after two pools DNA analyses were selected for individual genotyping. A number of SNPs showed evidence for association with disease severity, including several in reported quantitative trait loci (QTLs) associated with fetal hemoglobin (HbF) levels. However, the most strongly associated SNPs were within a region on chromosome 11 distinct from the β-globin gene cluster, within which most analysis to date has focused. Further study is needed to identify disease modifier genes in thalassemia.

References

1. Weatherall DJ. Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases. BMJ 2000; 321: 1117-20.
2. Bentley DR. The Human Genome Project - an overview. Med Res Rev 2000; 20: 189-96.
3. Barnes MR, Gray IC. Bioinformatics for geneticists. United Kingdom: Wiley; 2003.
4. Brookes AJ. The essence of SNPs. Gene 1999; 234: 177-86.
5. Chakravarti A. To a future of genetic medicine. Nature 2001; 409: 822-3.
6. Riva A, Kohane IS. SNPper: retrieval and analysis of human SNPs. Bioinformatics 2002; 18: 1681-5.
7. Strittmatter WJ, Roses AD. Apolipoprotein E and Alzheimer's disease. Annu Rev Neurosci 1996; 19: 53-77.
8. The International HapMap Project. Nature 2003; 426: 789-96.
9. Eichelbaum M, Ingelman-Sundberg M, Evans WE. Pharmacogenomics and individualized drug therapy. Annu Rev Med 2006; 57: 119-37.
10. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. Nat Rev Genet 2002; 3: 862-71.
11. Dorman JS, LaPorte RE, Stone RA, Trucco M. Worldwide differences in the incidence of type I diabetes are associated with amino acid variation at position 57 of the HLA-DQ beta chain. Proc Natl Acad Sci U S A 1990; 87: 7370-4.
12. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 2000; 26: 76-80.
13. Dahlback B. Resistance to activated protein C caused by the factor VR506Q mutation is a common risk factor for venous thrombosis. Thromb Haemost 1997; 78: 483-8.

14. Ozaki K, Tanaka T. Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. Cell Mol Life Sci 2005; 62: 1804-13.
15. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 2003; 35: 131-8.
16. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002; 418: 426-30.

การศึกษาความสัมพันธ์ระหว่างความหลากหลายทางพันธุกรรม (genetic polymorphisms) กับโรคต่าง ๆ

อรพรรณ ศรีพิชัย, สุทัศน์ ฟู่เจริญ

สนิป (SNPs = Single nucleotide polymorphisms) คือความหลากหลายทางพันธุกรรมของดีเอ็นเอระหว่างมนุษย์ในแต่ละคนที่เกิดจากความแตกต่างของนิวคลีโอไทด์เพียงตำแหน่งเดียว แต่อาจจะมีความสัมพันธ์กับลักษณะทางกายภาพที่แสดงออกมาของสิ่งมีชีวิตต่าง ๆ

จากความรู้เรื่องจีโนมมนุษย์ พบว่าเมื่อนำลำดับเบสของมนุษย์แต่ละคนมาเปรียบเทียบกันจะมีความเหมือนกันประมาณ 99.9 เปอร์เซ็นต์ ส่วนที่แตกต่างกันออกไปเพียง 0.1 เปอร์เซ็นต์ เป็นบริเวณสำคัญที่ทำให้แต่ละคนมีความแตกต่างกัน เช่น สีตา สีผิว สีผม ความสูง ความแข็งแรง การเป็นโรค การตอบสนองต่อยา เป็นต้น จะเห็นได้ว่าความแตกต่างทางพันธุกรรมของคนแต่ละคน หรือแต่ละกลุ่มประชากรในประเทศหนึ่ง ๆ เกิดจากผลของ "สนิป" นั้นเอง ด้วยเหตุนี้สนิปจึงสามารถใช้เป็นเครื่องหมายทางชีวภาพ (biomarker) สำหรับสร้างแผนที่พันธุกรรม ซึ่งช่วยในการค้นหาจีนที่ทำให้เกิดโรคต่าง ๆ และยังใช้ในการทำนายความเสี่ยงต่อการเกิดโรคในแต่ละคนได้ นอกจากนี้ในอนาคตแพทย์อาจนำข้อมูลสนิปมาใช้ในการตัดสินใจให้การรักษาที่ตรงจุด หรือเลือกใช้ยาได้อย่างมีประสิทธิภาพ