Chapter 5
Single Nucleotide Polymorphism and its Application in Mapping Loci Involved in Developing Human Diseases and Traits

Rui-Ru Ji
Bristol-Myers Squibb, USA

ABSTRACT

Common diseases or traits in humans are often influenced by complex interactions among multiple genes as well as environmental and lifestyle factors rather than being attributable to a genetic variation within a single gene. Identification of genes that confer disease susceptibility can be facilitated by studying DNA markers such as single nucleotide polymorphism (SNP) associated with a disease trait. Genome-wide association approaches offers a systematic analysis of the association of hundreds of thousands of SNPs with a quantitative complex trait. This method has been successfully applied to a wide variety of common human diseases and traits, and has generated valuable findings that have improved the understanding of the genetic basis of many complex traits. This chapter outlines the general mapping process and methods, highlights the success stories, and describes some limitations and challenges that lie ahead.

INTRODUCTION

SNP, or single nucleotide polymorphism, is a genetic variation in a person's DNA sequence that occurs when a single nucleotide is replaced by one of the other three nucleotides. SNPs are very common in the human population, occurring in the genome more than one percent of the time (http://www.ncbi.nlm.nih.gov/About/primer/snps.html). Since only three to five percent of the human genome encodes protein sequences, the majority of SNPs are outside of the so called

DOI: 10.4018/978-1-60960-491-2.ch005

coding regions. SNPs within a coding region are of particular interest to researchers because they are more likely to alter the biological function of a protein. Because of their high prevalence in the human genome, SNPs can be used as genetic markers to pinpoint the susceptibility loci for a disease in an association study.

Botstein and colleagues (Botstein et al., 1980) first proposed the genome-wide association approach. They suggested that the naturally occurring DNA sequence polymorphisms could be used as genetic markers to examine systematically the transmission of phenotypes in families. When a polymorphism shows significant linkage to a disease, additional markers can be genotyped in the region, termed fine mapping, to identify the responsible gene or variant.

In Mendelian traits and diseases, the underlying gene can usually be mapped unambiguously and precisely to a small chromosomal interval because of the strong correlation between genotype and phenotype. Subsequently, discovery of coding sequence variants in one of a small number of candidate genes in affected individuals usually provides sufficient evidence to establish the identity of the causal gene (Botstein & Risch, 2003). The same certainties do not apply to complex disease and traits, which are polygenic in nature. It is much more difficult to identify genes that contribute to complex traits because of low penetrance, epistasis, locus heterogeneity, and variable expressivity. The candidate gene approach was shown to be woefully inadequate as many disease genes were completely unsuspected based on prior knowledge of biological pathways.

A possible path forward was suggested by advances in population genetics and genomics: instead of mapping disease genes by tracing phenotype in families, one could identify them by association studies by comparing the frequencies of genetic variants among case and control individuals (Altshuler et al., 2008).

The completion of the human genome sequence (International Human Genome Sequencing Consortium, 2004) and the provision of an initial catalog of human genetic variations and a haplotype map (known as the International Human HapMap Project; International HapMap Consortium, 2005) have made it possible to perform genome-wide association study (GWAS) utilizing SNPs as genetic markers. The rapid technology development in high throughput genotyping platforms and data analysis approaches have now permitted GWAS to be undertaken in a large number of samples (McCarthy et al., 2008).

The underlying rationale for GWAS is the so called "common disease, common variant" hypothesis, which predicts that common variants (classically defined as having a minor allele frequency larger than 1%) in the human populations manifest a common disease (Risch & Merikangas, 1996; Collins et al., 1997). In the past few years we have witnessed the success of GWA studies in identifying hundreds of common genetic variants associated with common diseases and traits (Goldstein, 2009; Hirschhorn, 2009; Kraft & Hunter, 2009). An updated list of published GWA studies can be found at the National Human Genome Research Institute's catalog of published genome-wide association studies (http://www.genome.gov/GWAStudies/).

The findings from GWAS have provided valuable novel insights into the complex allelic architecture of common diseases and traits. However, for most conditions studied to date, the implicated genetic variants only explain a fraction of the familial aggregation, limiting the early application potential for predicting individual risk. Much work remains to obtain a complete catalog of susceptibility loci and to elucidate the molecular mechanisms through which these variants operate. As such, it remains a distant objective to translate these findings into clinical practice.

The purpose of this chapter is to review the current status of the application of SNPs as genetic markers in mapping quantitative complex traits in humans. We will outline the general mapping process and the analytical methods, highlight the