Chapter 14
Current Study Designs, Methods, and Future Directions of Genetic Association Mapping

Jami Jackson
North Carolina State University, USA

Alison Motsinger-Reif
North Carolina State University, USA

ABSTRACT

Rapid progress in genotyping technologies, including the scaling up of assay technologies to genome-wide levels and next generation sequencing, has motivated a burst in methods development and application to detect genotype-phenotype associations in a wide array of diseases and other phenotypes. In this chapter, the authors review the study design and genotyping options that are used in association mapping, along with the appropriate methods to perform mapping within these study designs. The authors discuss both candidate gene and genome-wide studies, focused on DNA level variation. Quality control, genotyping technologies, and single-SNP and multiple-SNP analyses have facilitated the successes in identifying numerous loci influence disease risk. However, variants identified have generally explained only a small fraction of the heritable component of disease risk. The authors discuss emerging trends and future directions in performing analysis for rare variants to detect these variants that predict these traits with more complex etiologies.

INTRODUCTION

The identification and characterization of genetic risk factors that underlie diseases and other traits of clinical importance is a central goal of human genetics. While there are many study designs, technologies, and analytical tools to connect genotypic variation to such phenotype variation, association mapping has emerged as a commonly used suite of approaches to begin dissecting the etiology of complex traits in health related traits. Such association studies are performed to address a number of possible goals, typically related to either better understanding the biological process of a disease or trait or find variants that could serve as biomarkers to improve health care/long

DOI: 10.4018/978-1-4666-6611-5.ch014

term outcomes. In this chapter, we will review key concepts and choices in association studies that measure DNA sequence variations in order to identify genetic loci that influence a particular trait of interest. Most association mapping studies are able to find variants that are of relatively high frequency in the population, though recent studies are starting to interrogate less common, or even rare variants for association.

Unlike linkage studies that are commonly used for Mendelian traits using family based studies, association studies rely on linkage disequilibrium (LD) across the genome to associate phenotypes of interest to genetic loci. LD is a property of genetic variants that describes the degree to which an allele of one variant is inherited (correlated) with an allele of another variant within a population. LD is related to the idea of chromosomal linkage (where two markers on a chromosome are physically joined/inherited together across generations), but is a population level (not individual level) construct. Mutation and recombination with both break apart regions of the chromosome, and this linkage "decay" will build up across generations until eventually all alleles in the population are independent (in linkage equilibrium). There are a number of factors that influence the rate of decay of linkage in human populations, resulting in the fact that different human subpopulations have very different degrees and patterns of LD. An excellent review of LD, including a review of measures to quantify it, can be found in (Devlin & Risch, 1995; Reich et al., 2001).

LD is an important concept underlying association mapping, as it allows for association mapping without actually having to directly assay all variants in the genome (indirect association). Because variants are not statistically (or biologically) independent, a reasonable number of markers can contain a substantial amount of information about the variants across a region of the genome. As genotyping technology advances to allow more variants to be assayed, it cannot be assumed that significantly associated variants are the functional (causative) variant, but it is more likely that they are in LD with the true causal variant(s). This needs to be considered in the study design of an association mapping experiment, with follow up fine resolution mapping experiment and replication.

Traditionally, association mapping relies on the Common Disease Common Variant (CDCV) hypothesis. This hypothesis asserts that common, interacting disease variants are the underlying cause of common complex disease (perhaps with environmental factors as well). This hypothesis emerged from both population genetics theory, as well as some early initial success in associating variants with high allele frequencies to diseases such as Alzheimer's and Type II diabetes (Altshuler et al., 2000; Corder et al., 1993). There are a couple of key ramifications of this hypothesis for association mapping. First, if common variation influences disease, then the overall effect size (penetrance) of each of the common variants must be lower that that of the causal variants of rare disorders. This is due to the direct relationship between the allele frequency and the overall prevalence of disease. Second, if such common risk variants have small effect sizes, but the overall heritability (proportion of variation in the trait that is due to genetic variation) of a trait is high, then it follows that the trait etiology must be due to multiple common, low penetrant variants.

Such association analysis can be done at the level of a single locus test (evaluating a single locus) or, with developments in technology making high throughput (bid data) genotyping accessible, it is becoming common to test for associations at loci across the entire genome. In fact, such genome-wide association studies (GWASs) have successfully identified numerous genetic loci at which common variants influence disease risk. As of January 22, 2014, the National Human Genome Research Institute GWAS Catalog of GWAS publications has 1789 publications and