DNA Variation and Disease

Anthony J. Brookes

University of Leicester, United Kingdom

Following completion of the human genome sequence, many different genomics investigations have been undertaken. By integrating genome sequences from other vertebrates, and by a range of expression and functional investigations, we are starting to understand which DNA stretches are most important in different scenarios. It is convenient to consider these discoveries under the categorizations 'systems biology' (what all the genes do, alone and together) and 'genome variation' (how gene sequence variations impact functional and phenotypic differences between individuals). These two fields will clearly need to be merged in time, but for now the disciplines are largely separate activities. This presentation shall consider progress and challenges relating to genome variation analysis, in particular concerning the role of sequence polymorphism in disease causation and predisposition.

The last decade has seen a phenomenal increase in our knowledge about, and ability to investigate, single nucleotide polymorphism (SNP) in the human genome. The international HapMap has created a basic whole-genome description of SNP frequencies and haplotype patterns across the globe. This is now being leveraged to conduct unbiased genome scans for disease association in many disorders. There is, however, great uncertainty and debate about how productive this will be, at what stage technologies will enable truly comprehensive genome scanning (whole-genome resequencing), how useful that will be, and how we will handle so much data. However, a decade of focussed genetic association analysis has already been conducted, along with a number of whole-genome scans performed by companies. Therefore, we should be able to make a reasonable prediction of what this field will produce, and this will be discussed.

Very recently, the challenge of studying genome variation from the perspective of SNPs was complicated by the realization that copy-number polymorphism (CNP) is abundant in the human genome. Technologies to study CNP genome-wide down to a high-level of resolution are lacking, but much data is nevertheless being generated using various hybridization and genotyping systems. There are many unanswered questions in these early days of exploration into CNP, but already it seems that large structural variation may have a far bigger impact on disease than does SNP. This includes genome regions of instability (causing 'genomic' disorders) and simple polymorphic differences in how many copies of various gene regions we each carry. An update on this shall be given.