Genome-wide association study (GWAS) Past, Present and Perspective

As we knonw, GWAS is a well-known approach to identify the susceptibility genes or loci for complex disease. Generally speaking, diseases related to genetics include monogenic disease which follows the Mendelian law, polygenic or complex diseases show no clear Mendelian inheritance patterns. How to find disease-causing gene for monogenic disease? We used Genome Wide Linkage Study Based on families and positional cloning. In the past 30 years, Scientists have found more than 2000 disease-causing genes, which provides a solid foundation for disease prevention and cure.

Complex disease also called common disease or polygenic disease, resulting from a number of minor genes and environmental factors, which were characterized by genetic heterogeneity, phenotype complexity and ethnic variability. To explore the pathogenesis of the complex disease, we have made many efforts in terms of the Human Genome Project (HGP), HapMap Project which make it feasible for the so called genome wide association study (GWAS). It has been proved very useful to identify common genetic factors that influence health. Now I would like to talk about the Principle of Association Study, which refers to compare the some gene or allele frequency between cases and controls.

So far, there are two main platforms for genome wide SNP genotyping: Affymetrix and Illumina. Advances in genotyping technology have vastly increased the number of variants that can be typed and decreased the per-sample costs. Actually, GWAS WAVE has already come since 2007; the development in GWAS is at an unprecedented speed in the past 5 year. There are many high quality GWAS papers published in Nature, Science, Nature Genetic and The New England Medicine about complex disease. Up to date, more than 200 complex diseases and traits were studied by GWAS approach and lots of susceptibility genes/loci were identified. GWAS has been proved to be an effective way to find susceptibility gene(s) for complex diseases. According to NIH national human genome research institute website, more than 1187 published papers and over 5954 associated SNPs have been identified up to date.

Our research teams have already performed GWAS in eight complex diseases: Psoriasis, SLE, Vitiligo, Leprosy, Atopic Dermatitis, Schiz, High Myopia and Esophageal Carcinoma. To make full use of the genetic resources, we share the same control samples in our series GWAS study regarding to different complex diseases.

After the first stage of GWAS, what we can do beyond the current SNP based GWAS. Actually, nowadays we can further more mine the GWAS dataset by the following: steps. 1. Deeply analysis of GWAS data, perform a replication study in a large number of samples; 2. Imputation SNPs based on HapMap data; 3. Collaboration and Meta analysis study; 4. Sharing susceptibility genes for different diseases; 5. Fine mapping and re-sequencing of established associations; 6. Pathway analysis based on GWAS data; 7. Interaction and epistasis study (Gene - Gene, Gene - Environment).

As we all know, GWAS mainly based on the hypothesis “common disease, common variants”. Although it has successfully identified a large number of genetic variants associated with complex traits, GWAS still have some limitations: 1 More disease-associated SNPs located at intergenic and intronic region, barely SNPs at functional region (such as exon and 5’UTR) 2 The limitation of chip number of variants, genotyping 1/6~1/5 SNPs in genome 3 Only identified some common diseases-associated variants (MAF>5%), Rare variants (MAF<5%) and structural variants (SV) were omitted. How to explore the susceptibility genes for complex diseases in future?

we can rely on GWAS of SNP and CNV, whole genome sequencing, exome sequencing, whole genome epigenetics, whole genome transcript and whole genome expression and so on.

What should we do in genetic research in the long term? Generally speaking, through deep disease genomics research, plus what we learn from the functional studies and mechanism studies, we could get a chance to find clinical predictor and to eluciate disease pathogenesis. It will be very useful for disease diagnosis, prediction, risk evaluation, prevention, and treatment. The ultimate goals are to develop new drug for treatment and finally make personalized medicine come ture.