S5-2

550,000 tag SNP markers for Genome-wide association studies

Richard Shen

Illumina, Inc., San Diego, CA

Many of the common SNPs are known and the results of the International HapMap Project have shown that the information from a majority of these SNPs can be captured by genotyping 250,000-500,000 well-chosen tagSNPs (International HapMap Project, 2005). We have developed two standard tagSNP panels that capture the majority of common variation in CEPH, Han Chinese/Japanese, and Yoruba populations. The two tagSNP panels are each contained on a single BeadChip which uses the Infinium™ whole genotyping genome assay (Steemers, et al. 2006) that allows a convenient, easy-to-use, high-throughput, parallelized genotyping method for whole-genome association studies.

The CEPH, Han Chinese/Japanese, and Yoruba populations have approximately 2 million common SNPs each (minor allele frequency =0.05) identified from the HapMap Phase I+II data. To capture this variation, tagSNPs were chosen by an algorithm utilizing the linkage disequilibrium statistic r^2 (Carlson, et al. 2004). An r^2=0.8 threshold was used for common SNPs in or within 10kb of genes or in evolutionarily conserved regions. For all other regions, an r^2=0.7 threshold was used. This panel captures 90%, 87%, and 57% of the HapMap Phase I+II variation in CEPH, Han Chinese/Japanese, and Yoruba populations using pairwise tests at r^2 =0.8, respectively. Ninety-six percent, 90% and 92% of all SNP loci are polymorphic in the CEPH, Han Chinese/Japanese and Yoruba populations, respectively, with average minor allele frequencies of 0.23, 0.21, and 0.22, respectively, in these populations. The average spacing between common SNP loci (MAF =0.05) is 5.5, 6.5, and 6.2kb in the CEPH, Han Chinese/Japanese and Yoruba populations, respectively. We have also included over 4,000 SNPs from recently reported LOH/copy number (CN) regions of the genome for more comprehensive coverage for LOH/CN applications and have confirmed several hundred of these regions using this panel. In addition, we have also included 180 mitochondrial SNPs and over 7,000 non-synonymous SNPs. This tagSNP panel is a valuable resource for both genome-wide association and CN studies and will help identify genetic variation affecting both human health and disease.