PL-1 *[2000 Oct. 19 (Thur) 09:00~09:50/Hall A]*

Microsatellite Mapping of HLA-associated Disease Genes and Human Genome Diversity Project

Hidetoshi Inoko

Department of Molecular Life Science, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan

Now that human genome sequencing has been almost completed, it is an upcoming important subject to carry out identification of disease genes, especially for common diseases such as schizophrenia, atopy, diabetes, hypertension, cerebral infarction, cardiac failure etc) which have multifactorial genetical basis. This "human genome diversity project" is conducted by mapping and association analyses using polymorphic genetic markers. For this purpose, SNP (single nucleotide polymorphism) markers are now being extensively and world-widely collected, and applied to disease mapping in a lot of laboratories. However, SNP is generally bi-allelic, and so polymorphism is not considered to be extensive enough to localize disease genes on the human genome by genome-wide mapping.

Instead, we propose to use microsatellite which displays a high degree of polymorphism in repeat number of repetitious unit and so is expected to serve as a more useful genetic marker for genome-wide mapping. To test this hypothesis, we have focused on mapping of diseases in the HLA region on chromosome 6p21.3, which is divided into class I, class II and class III regions from centromere to telomere. The human MHC (HLA) encompasses a 3.6 Mb (3,600 kb) segment, which is characterized by high gene density (one gene/18 kb), extensive genetic polymorphism (more than 900 alleles at 12 HLA loci) and the presence of susceptible loci for more than 100 diseases (HLA-associated diseases). In these respects, the HLA region provides an excellent mini-genome model region for "human genome diversity project". This area has currently been determined for the complete genomic sequence by the international consortium groups (1). Among these efforts, our group (Tokai University) determined the nucleotide sequence of the 2.2 Mb continuous HLA region including the 1.8 Mb entire class I region as well as the 0.4 Mb class III region (2), in order to elucidate the detailed gene organization in the HLA class I

region and identify susceptible genes for many HLA class I-associated diseases.

The HLA class I region is 1,796,938 bp long, embracing as many as 127 genes (60 known and 67 new genes) with the gene density of one gene every 14.1 kb, which is comparable to that of the gene-rich class III region. The GC content is fairly uniform throughout the class I region, representing 45.8% on average, which corresponds to the isochore H1.

More than 30 pairs of homologous segments containing class I and class I chain-related (MIC) genes were recognized, especially at the 300 kb telomeric end of the HLA class I region. These segments generally consist of MIC - HCGIX - 3.8-1 - P5 - HCGIV - HLA class I - HCGII as an elementary unit. Base on these facts, we could construct a model which explains how through 8 rounds of successive segmental duplications of the above-mentioned elementary unit, the HLA class I region was shaped. That the currently non-essential HLA-F and MICE genes have acted as progenitors to today's immune-competent HLA-ABC and MICA/B genes provides experimental evidence for evolution by "birth and death".

A total of 758 microsatellite repeats were identified in the HLA class I genomic sequence. These consist of 203 di-, 139 tri-, 273 tetra- and 143 penta-nucleotide repeats, yielding an overall density of one microsatellite per 2.3 kb. Among the 758 microsatellite identified here, 70 have been selected and subjected to polymorphism analysis within a Japanese population. As expected, 38 out of these 70 microsatellites are quite polymorphic with an average of 8.9 alleles and 72$B!s(B heterozygosity. As these polymorphic microsatellites are evenly dispersed throughout the class I region, they should serve as much needed genetic markers in linkage and association analyses, enabling investigators to precisely map class I-associated disease susceptibility loci. In fact, by investigation of genetic polymorphisms in 38 microsatellite repeats, we could successfully reduce the critical regions for Beh$B9F(Bt's disease (associated with B51) and psoriasis vulgaris (associated with Cw6) to 50 kb segments, around the HLA-B gene and telomeric of the HLA-C gene, respectively. From these segments, strong candidate genes for both the diseases have been identified.

Thus, refined microsatellites provide useful and valuable genetic markers for precise disease mapping. In contrast, SNP mapping are not easy even for these HLA class I associated diseases at the Mb level mapping, mainly due to a few alleles (two alleles) and low heterozygosity of SNP markers. These results are validated by different lengths of linkage disequilibrium observed for SNP and

microsatellite. Namely, the length of linkage disequilibrium observed for SNP is less than 5 kb, whereas it is as long as 100 kb for microsatellite. This suggests that one microsatellite per 100 kb is enough for genome-wide mapping, but a huge number of SNPs (about one million) are required for genome-wide mapping. Collectively, the efficient method of genome-wide mapping is to first use microsatellites markers which enable to narrow down the critical region to approximately 100 kb and thereafter to employ SNP markers within thus determined critical region for fine mapping to identify a causative gene. Based on this conclusion, we are now collecting 30,000 polymorphic markers (one microsatellite per 100 kb) throughout the human genome which will be subjected to genome-wide mapping of complex disease.

References

(1) The MHC sequencing consortium : Complete sequence and gene map of a human major histocompatibility complex (MHC). Nature 401 : 921-923, 1999.

(2) Shiina T, Tom G, Oka A, Takishima N, Yamagata T, Kikkawa E, Iwata K, Tomizawa M, Okuaki N, Kuwano Y, Watanabe K, Fukuzumi Y, Itakura S, Sugawara C, Ono A, Yamazaki M, Tashiro H, Ando A, Ikemura T, Soeda E, Kimura M, Bahram S, Inoko H : Molecular dynamics of MHC genesis unraveled by sequencing analysis of the 1,796,938 bp HLA class I region. Proc Natl Acad Sci USA 96 : 13282-13287, 1999.