

SNP identification, linkage disequilibrium, and haplotype analysis for a 200-kb genomic region in a Korean population

Kwang Joong Kim^{a,b}, Hye-Ja Lee^a, Mi-Hyun Park^a, Seung-Hun Cha^a, Kyung-Seon Kim^a,
Hung-Tae Kim^a, Kuchan Kimm^a, Bermseok Oh^a, Jong-Young Lee^{a,*}

^a Center for Genome Sciences, Korean National Institute of Health, 5 Nokbun-Dong, Eunpyung-Gu, Seoul 122-701, Korea

^b Department of Molecular Bioscience, Kangwon National University, Chuncheon, Korea

Received 14 July 2005; accepted 4 March 2006

Available online 17 August 2006

Abstract

Understanding patterns of linkage disequilibrium (LD) across genomes may facilitate association mapping studies to localize genetic variants influencing complex diseases, a recognition that led to the International Haplotype Mapping Project (HapMap). Divergent patterns of haplotype frequency and LD across global populations require that the HapMap database be supplemented with haplotype and LD data from additional populations. We conducted a pilot study of the LD and haplotype structure of a genomic region in a Korean population. A total of 165 SNPs were identified in a 200-kb region of 22q13.2 by direct sequencing. Unphased genotype data were generated for 76 SNPs in 90 unrelated Korean individuals. LD, haplotype diversity, and recombination rates were assessed in this region and compared with the HapMap database. The pattern of LD and haplotype frequencies of Korean samples showed a high degree of similarity with Japanese data. There was a strong correlation between high LD and low recombination frequency in this region. We found considerable similarities in local LD patterns between three Asian populations (Han Chinese, Japanese, and Korean) and the CEPH population. Haplotype frequencies were, however, significantly different between them. Our results should further the understanding of distinctive Korean genomic features and assist in designing appropriate association studies.

© 2006 Elsevier Inc. All rights reserved.

Keywords: SNP identification; linkage disequilibrium; haplotype; Korean population

In combination with the completion of the human genome project, recent advances in identifying and genotyping genetic variants are providing a more detailed understanding of global patterns of genetic variation. The primary motivation for the construction of any linkage disequilibrium (LD) map in the human genome is to facilitate identification and characterization of genetic variants related to common complex diseases. Previous studies of LD in humans have shown a high degree of variability, indicating that LD may extend from a few to several hundred kilobases [1,2]. For practical applications, however, the local patterns of haplotype conservation are of primary interest. Patterns of genetic diversity and LD are shaped by many factors, including mutation, recombination, selection, population demography, and genetic drift [3]. Studies of LD between single-nucleotide polymorphisms (SNPs) and micro-

satellites show greater LD in Eurasians than in Africans [4,5]. Mutations that are involved in disease may also differ in frequency across ethnic groups because of historical selection. Many common diseases such as hypertension, diabetes, and obesity, which differ in prevalence across ethnic groups, may also be influenced by genes that have been under natural selection. The divergent pattern of haplotype frequencies and LD across global populations, as well as high levels of substructure in regions such as Africa, indicates the importance of characterizing haplotypes and LD across ethnically diverse populations [6–8].

The International Haplotype Mapping Project (HapMap) aims to characterize the distribution and extent of LD across the entire human genome [9]. However, there are some critical unknowns in association mapping with the HapMap database. First, it is unclear whether one or a few haplotype maps can provide useful information for populations not represented in the HapMap populations. Therefore, the HapMap project, which

* Corresponding author. Fax: +82 2 354 1063.

E-mail addresses: leejy63@nih.go.kr, leejy63@hotmail.com (J.-Y. Lee).

Table 1
Summary of SNP identification for a 200-kb region of 22q13.2 in a Korean sample population

Gene	Promoter ^a	5'UTR ^b	Syn	Nonsyn	3'UTR	Intron	Total
<i>TNFRSF13C</i>	–	–	–	–	–	2	2
<i>C22orf18</i>	1	1	3	–	1	4	10
<i>SEPT3</i>	1	1	–	1	3	15	21
<i>MGC26816</i>	1	–	–	1	–	25	27
<i>NAGA</i>	3	1	–	–	1	8	13
<i>LOC150368</i>	1	–	3	1	2	4	11
<i>LOC91689</i>	–	–	–	–	–	3	3
<i>NDUFA6</i>	–	–	–	–	1	2	3
<i>LOC246785</i>	3	–	–	–	–	–	3
Intergenic region	–	–	–	–	–	–	72
Total	10	3	6	3	8	63	165

^a Promoter, SNPs within 2 kb of the 5' end of a gene feature, but not in the transcript for the gene.

^b UTR, untranslated region; Syn, synonymous cSNP; Nonsyn, nonsynonymous cSNP.

proposes to characterize LD in a small subset of ethnic groups, will need to be expanded with information about other populations. Second, undetected genomic structure of a population in case–control association studies can yield false positive association [10,11]. Thus, knowledge of ethnicity and statistical tests of substructure are important for the proper design of case–control association studies and for identifying disease-predisposing alleles that may differ across ethnic groups.

Here we present a comparison study of the LD and haplotype structure for a genomic region in chromosome 22 that is designed to evaluate whether haplotype maps from the HapMap project are likely to be useful for the Korean population. To find the best map for such an association study, we investigated the genomic structure of a 200-kb region in an unrelated Korean sample population as a pilot study. In July 2004, there were no available genotype data in the HapMap database for this 200-kb region of Asian samples. To exclude SNPs that are representative of monomorphic or rare alleles in the Korean population from genotyping, we selected 115 SNPs (validated in the discovery stage with >0.05 minor allele frequency (MAF)) for the subsequent genotyping stage rather than using the HapMap or dbSNP database. For SNP discovery, the 200-kb region was

screened by direct sequencing. Unphased genotype data were generated from 90 unrelated Korean individuals by pyrosequencing. LD, haplotype diversity, and recombination rates in this region were assessed and compared with the HapMap database.

Results

SNP discovery and genotyping

To find useful markers for analyzing the genomic structure of Korean population, a 200-kb genomic region encompassing nine genes was screened by direct sequencing of DNA from 24 unrelated Korean individuals. A total of 165 SNPs, including 10 in promoter regions, 11 in untranslated regions, 9 in coding regions, 63 in introns, and 72 in intergenic regions, were identified (Table 1). Of these 165 SNPs, 47 were determined to be novel by comparing our data with the dbSNP database (build 124; <http://www.ncbi.nlm.nih.gov/projects/SNP/>). Of 165 SNPs, 115 were selected for genotyping from 90 unrelated Korean individuals using the pyrosequencing method. Of 115 SNPs, 89 were successfully genotyped; assay failures of the remaining SNPs were due to PCR failure (9 SNPs) and genotyping failure (17 SNPs). Of 89 SNPs, 13 (including 8 SNPs violating Hardy–Weinberg equilibrium and 5 SNPs showing <0.05 MAF) were excluded from the subsequent analysis. As a result, a total of 76 SNPs were used in analyzing the pattern of LD and haplotype from the Korean study population. The average spacing of 76 SNPs was 2.6 kb, but they were not evenly spaced (Fig. 1). All data for the SNPs discovered and genotyped from these Korean samples have been deposited with the KSNP database (<http://www.ngri.re.kr/SNP/>).

Linkage disequilibrium

To investigate patterns of LD in the described 200-kb region of chromosome 22, two measures of LD (D' and r^2) were estimated between all pair-wise combinations of markers, using Haploview version 3.2 (<http://www.broad.mit.edu/mpg/haploview/>). The distributions and the list of SNPs analyzed in the study are shown in Fig. 1 and in Supplementary Table 1, respectively. Fig. 2a shows LD patterns of the 200-kb region in

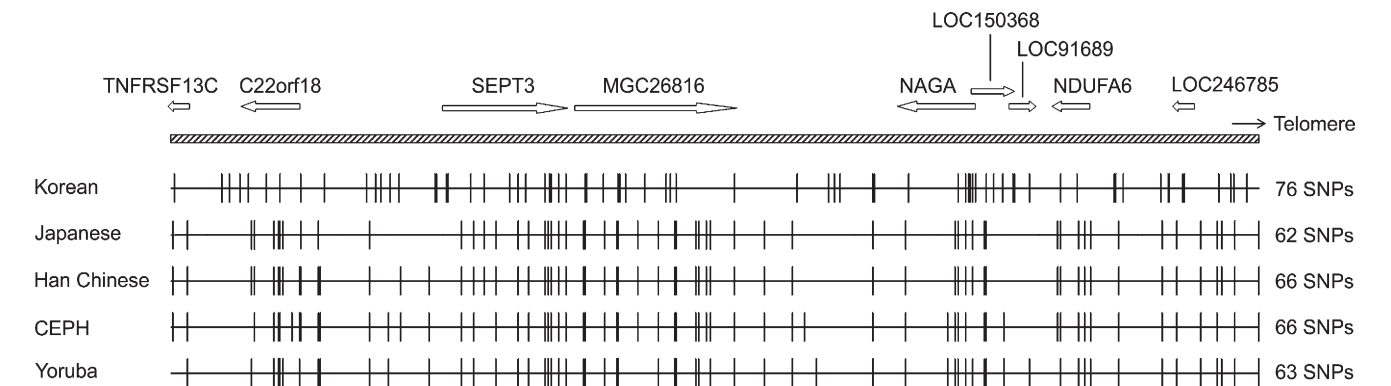


Fig. 1. Genomic organization of a 200-kb region of chromosome 22q13.2 and distribution of SNP markers in five populations. CEPH, a group of Utah residents with ancestry from northern and western Europe collected by the Centre d'Etude du Polymorphisme Humain.

Download English Version:

<https://daneshyari.com/en/article/5907981>

Download Persian Version:

<https://daneshyari.com/article/5907981>

[Daneshyari.com](https://daneshyari.com)