Genomics & Informatics Vol. 6(1) 18-28, March 2008

Chromosome 22 LD Map Comparison between Korean and Other Populations

The Korean HapMap Consortium, Jong-Eun Lee[1]*, Hye-Yoon Jang[1], Sook Kim[1], Yeon-Kyeong Yoo[1], Jung-Joo Hwang[2]*, Hyojung Jun[2], Kyusang Lee[2], Okkyung Son[2], Jun-Mo Yang[3]*, Kwang-Sung Ahn[3], Eugene Kim[3], Hye-Won Lee[3], Kyuyoung Song[4]*, Hie-Lim Kim[4], Seong-Gene Lee[4], Yongsook Yoon[4], Kuchan Kimm[5], Bok-Ghee Han[5], Bermseok Oh[5], Chang-Bae Kim[6]*†, Hoon Jin[6], Kyoung O. Choi[6], Hyojin Kang[6] and Young J. Kim[6]*

The Korean HapMap Consortium (Participants are in alphabetical order by institution, principal investigator, and then last name)
[1]Genotyping Centers: DNA Link, Inc., [2]Samsung Advanced Institute of Technology, [3]Sungkyunkwan University School of Medicine, [4]University of Ulsan College of Medicine, [5]Community Engagement/Public Consultation and Sample Collection Group: the Genomic Research Center, Korean National Institute of Health, [6]Analysis Group: Medical Genomics Research Center

Abstract

Single nucleotide polymorphisms (SNPs) are the most abundant forms of human genetic variations and resources for mapping complex genetic traits and disease association studies. We have constructed a linkage disequilibrium (LD) map of chromosome 22 in Korean samples and compared it with those of other populations, including Yorubans in Ibadan, Nigeria (YRI), Centre d'Etude du Polymorphisme Humain (CEPH) reference families (CEU), Japanese in Tokyo (JPT) and Han Chinese in Beijing (CHB) in the HapMap database. We genotyped 4681 of 111,448 publicly available SNPs in 90 unrelated Koreans. Among genotyped SNPs, 4167 were polymorphic. Three hundred and five LD blocks were constructed to make up 18.6% (6.4 of 34.5 Mb) of chromosome 22 with 757 tagSNPs and 815 haplotypes (frequency $\geq$ 5.0%). Of 3430 common SNPs genotyped in all five populations, 514 were monomorphic in Koreans. The CHB + JPT samples have more than a 72% overlap with the monomorphic SNPs in Koreans, while the CEU + YRI samples have less than a 38% overlap. The patterns of hot spots and LD blocks were dispersed throughout chromosome 22, with some common blocks among populations, highly concordant between the three Asian samples. Analysis of the distribution of chimpanzee-derived allele frequency (DAF), a measure of genetic differentiation, Fst levels, and allele frequency difference (AFD) among Koreans and the HapMap samples showed a strong correlation between the Asians, while the CEU and YRI samples showed a very weak correlation with Korean samples. Relative distance as a quantitative measurement based upon DAF, Fst, and AFD indicated that all three Asian samples are very proximate, while CEU and YRI are significantly remote from the Asian samples. Comparative genome-wide LD studies provide useful information on the association studies of complex diseases.

Keywords: haplotype, HapMap, Korean, LD, populations, SNP

*Corresponding author: E-mail jonglee@dnalink.com, jungjoo.hwang@samsung.com, jmyang@smc.samsung.co.kr, kysong@amc.seoul.kr, changbae@kribb.re.kr, yjkim8@kribb.re.kr
Tel +82-42-879-8118, Fax +82-42-879-8119
†Present address: Chang-Bae Kim, Major in Life Science, College of Natural Sciences, Sangmyung University, Seoul 110-743, Korea
Accepted 11 March 2008

Introduction

Vast amounts of information on single nucleotide polymorphisms (SNPs) and progress in high-throughput genotyping technology have generated a great deal of interest in establishing genome-wide linkage disequilibrium (LD) maps for genetic studies of complex traits (Chakravarti 2001; The International HapMap Consortium 2003; Myers and Bottolo 2005). LD is known to occur in a block-like structure across the genome, with conserved haplotype blocks of tens to hundreds of kilobases punctuated by "hot spots" of recombination (Daly *et al.* 2001). Since the concept of whole genome association studies using SNPs was introduced (Risch and Merikangas 1996), an optimal number of SNPs required for association studies has been center of extensive debate (Kruglyak 1999). Initial studies have focused on average LD levels and the variability in processes that generate LD (Cardon and Abecasis 2003). Although a single chromosome could carry many haplotypes in LD blocks, recent studies suggest that haplotypic variation may be much lower than previously imagined (Jeffreys *et al.* 2001; Patil *et al.* 2001; Gabriel *et al.* 2002). Patil's group identified haplotype blocks on chromosome 21

for which over 80% of chromosomes were represented by a few common haplotypes (Patil *et al*. 2001). In the analysis of human chromosome 22 with a marker density of one SNP per 15 kb, Dawson's group reported a highly variable pattern of LD along the chromosome, in which extensive regions of complete LD of up to 804 kb in length were interspersed with regions of no detectable LD (Dawson *et al*. 2002). Although differences of LD patterns between populations have been reported (Abecasis *et al*. 2002; Reich *et al*. 2001, Zavattari *et al*. 2002), little information is available on the haplotype structure in different populations other than the recent study by S.B. Gabriel, *et al*. (Gabriel *et al*. 2002). On the other hand, haplotype analysis has been widely employed in linkage studies for narrowing down the location of disease susceptibility genes (Zhang *et al*. 2004; Park 2007).

The International HapMap Project was launched to develop a haplotype map of the human genome, the HapMap, which will describe the common patterns of human DNA sequence variation among four population samples: 30 trios from Yoruba in Ibadan, Nigeria (YRI), 45 unrelated Japanese in Tokyo, Japan (JPT), 45 unrelated Han Chinese in Beijing, China (CHB), and 30 trios in a Utah, US population with Northern and Western European ancestry (CEU) from the CEPH collection (The International HapMap Consortium 2003; 2004; 2007). As the International HapMap Project releases a validated SNP map of 1 marker per kb for the HapMap samples, the general applicability of the HapMap data needs to be confirmed in samples from related populations. Recent comparative studies of LD patterns have shown a high degree of concordance among various populations (Gabriel *et al*. 2002; Shifman *et al*. 2003; Stenzel *et al*. 2004; Mueller *et al*. 2005). As the HapMap samples include Japanese and Chinese, it was our interest to test whether significant differences in LD exist between Koreans and the two other Asian samples.

In this paper, we measured the LD pattern along chromosome 22 in Korean samples and compared the Korean data with those of the four HapMap samples. We were interested in exploring how the HapMap data could be used to estimate the genomic structure of Koreans. We expect that this study will contribute to the development of proper strategies for association studies of common complex diseases in Koreans using the HapMap data.

Methods

SNP Selection

A total of 111,448 reference SNPs from chromosome 22 in the dbSNP (http://www.ncbi.nlm.nih.gov/SNP, build 116) were collected. To maximize cost effectiveness of genotyping, SNPs were selected based on the following criteria: 1) markers with even spacing, 2) verified SNPs, 3) coding SNPs. The SNPs were scored for the selection of the study using the following strategies. First, it was most important in mapping chromosomal LD blocks to have relatively equal spaces between SNP markers. Second, verified SNP markers (validation status was scored as 0 to 4 in the dbSNP) that had higher scores were chosen to prevent or reduce genotyping failure. Also, repeated sequence regions were excluded by repeat masking with Primer3 software (Rozen and Skaletsky 2000). Third, to be useful for a further study, protein coding SNPs had higher scores. A total of 12,674 genotyping experiments were conducted by four Genotyping Centers, and a final set of 4681 markers passed the stringent quality control procedure (The International HapMap Consortium 2003).

DNA Samples

Genomic DNA from 90 unrelated Korean individuals without family histories of major diseases was obtained from the Genomic Research Center in the Korean National Institute of Health (KNIH). The KNIH samples were collected as part of an epidemiological project and represent urban and rural regions in the south of Seoul. The sex ratio was 0.5 and the mean age was 50. Informed consent from all participating subjects was obtained through KNIH, and research approval came from the relevant ethical committees. DNA was isolated from peripheral blood leukocytes according to standard procedures with proteinase K-RNase digestion, followed by phenol-chloroform extraction.

Genotyping

For each SNP, we chose a set of three primers: two PCR primers to amplify a product of 100-200 bps under standard conditions and an optimized extension primer to be complementary to the sequence immediately to a SNP site.

For genotyping, we employed three platforms-6063 SNP genotypings were done using the Orchid Bioscience SNP-IT^{TM} assay (Princeton, NJ), 984 SNP genotypings using the PerkinElmer Life Sciences FP-TDI assay (Boston, MA), and 5627 SNP genotypings using the Sequenom MassARRAY^{TM} (San Diego, CA).

Statistical Analysis

A genotype frequency for each SNP was checked for

consistency between the observed values and those expected from the Hardy-Weinberg equilibrium test in each assay. Haploview version 3.2 (Barrett *et al.* 2004), based on the expectation-maximization (EM) method (Excoffier and Slatkin 1995), was used to infer haplotype phase and population frequency and to estimate the Lewontin's coefficients D' (Lewontin 1998), LOD, and correlation coefficient r (Hill and Robertson 1968). PHASE v2.1 was used to estimate the recombination parameters (Li and Stephens 2003; Crawford *et al.* 2004) and assess the statistical significance of haplotype profile differences and individual haplotype fre-

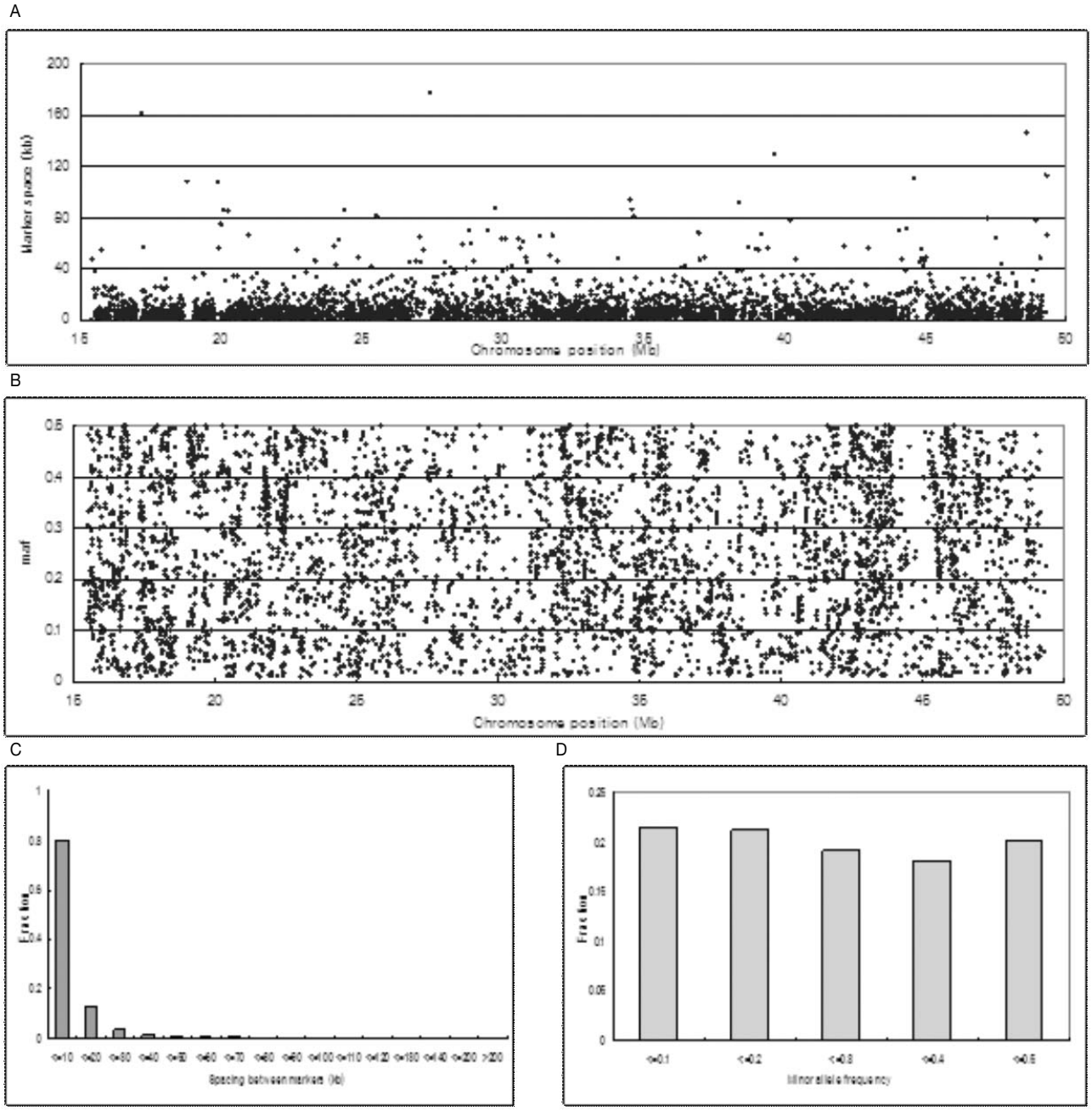

Fig. 1. Marker space and minor allele frequency distribution on chromosome 22 (average marker space, 7.3 kb; median marker space 3.8 kb). (A) The scatter plot of spacing between markers for the 4681 Korean markers in this study. (B) The scatter plot of minor allele frequency. (C) The distribution of spacing between 4681 markers in Koreans. (D) The distribution of minor allele frequency.

quency differences between the two samples (Stephens and Donnelly 2003). Hot spots were evaluated to have a recombination rate greater than or equal to 10 (Crawford *et al*. 2004). Derived allele frequency (DAF) was estimated by comparing major human alleles with the Chimpanzee reference allele downloaded from the UCSC database and by considering human minor allele frequency as DAF, when both alleles are the same. DAF correlation between populations was calculated using the Pearson product-moment correlation implemented in R language. Fst was calculated according to Wright's F-statistics (Wright 1951). Korean variation patterns of SNPs were compared with those of the HapMap samples including CEU, CHB, JPT, and YRI in the HapMap database. Neighbor-joining trees of average Fst and correlation of DAF and AFD among populations were constructed by PhyloDraw (Choi *et al*. 2000).

Quality Control

Each year, one quality assessment exercise was carried out to evaluate each platform and laboratory protocol. Quality control followed the International HapMap Project protocol (The International HapMap Consortium 2003): 1) pass rate of 75% or more per-plate genotype (more than 68 samples out of 90); 2) no more than 1 reproducibility error per plate (consistency across 5 duplicate genotypes ((1 discrepancy); 3) Hardy-Weinberg goodness-of-fit test ($p > 0.001$).

Results

Genomic variation analysis

For Build 116 of the dbSNP, 111,448 SNPs in human chromosome 22 were available. In order to select an evenly spaced subset of SNPs that maximized the power of calculating LD maps, we screened the SNP set to remove ambiguous positions (2642), non bi-allelic SNPs (5904), SNPs located in repeat region (10,067), and also SNPs with extreme GC contents. We developed bioinformatic parametric optimization and heuristic methods for selecting SNPs from these SNPs (Kim *et al*. 2003). We placed a higher score on those SNPs from the Japanese SNP database, SNPs with frequency information, SNPs with validation scores, and SNPs with functionally important positions such as non-synonymous SNPs, in decreasing order. By using the in silico selection protocols, we selected approximately 20,000 SNPs to have an approximately 1.5 kb spacing. Four Genotyping Centers selected further approximately 12,000 SNPs of these SNPs to conduct a series with their own platform genotyping experiments. Approximately 50% of genotyping reactions failed because of either PCR or extension reaction failure, and approximately 400 SNPs were discarded because their physical position had been changed since Build 116. Of 12,000 genotyped SNPs, 4681 SNPs met the series of stringent quality control procedures (The International HapMap Consortium 2003) and were used for the construction of LD maps. The complete list of 4681 SNPs, their reference sequence numbers (rs number), positional information, minor allele frequencies in 90 individuals, and other information are freely available at http://www.khapmap.org.

The 4681 SNP markers were evenly spaced along chromosome 22 (average space, 7.3 kb; median space, 3.8 kb) except for several long gaps (Fig. 1). As shown in Fig. 1B, the experimented SNPs had a randomly distributed minor allele frequency. The distribution of minor allele frequencies of 4681 SNPs was relatively even, as shown in Fig. 1D and were similar across five population samples (data not shown). Of 4681 SNPs, 514 SNPs were monomorphic in Korean samples. The CHB + JPT had more than 72% overlap with the monomorphic SNPs in Koreans, while CEU and YRI had less than 38% overlap (Table 1). Of those SNPs that were polymorphic in the HapMap samples but monomorphic in Koreans, more than 33% of polymorphic markers showed a minor allele frequency less than 0.1: CEU (45.1%), JPT (48.7%), CHB (59.3%), and YRI (33.7%).

Comparison of LD patterns and recombination rate

An overall LD (Linkage Disequilibrium) pattern covering the entire length of chromosome 22 (Recombinant rate, D' and r^2, LD block, Derived allele, Fst, and Allele frequency difference) is shown in Fig. 2. Linkage disequilibrium is the non-random association of alleles at two or more loci, not necessarily on the same chromo-

Table 1. Distribution of 514 Korean monomorphic SNPs in European, Japanese, Chinese and African populations

MAF	CEU	JPT	CHB	YRI
Monomorphic SNPs	195 (37.9%)	401 (78.0%)	374 (72.8%)	164 (31.9%)
Polymorphic SNPs	319 (62.1%)	113 (22.0%)	140 (27.2%)	350 (68.1%)
Sum	514	514	514	514
Number of SNPs with maf < 0.1 out of Polymorphic SNPs	144 (45.1%)	55 (48.7%)	83 (59.3%)	118 (33.7%)

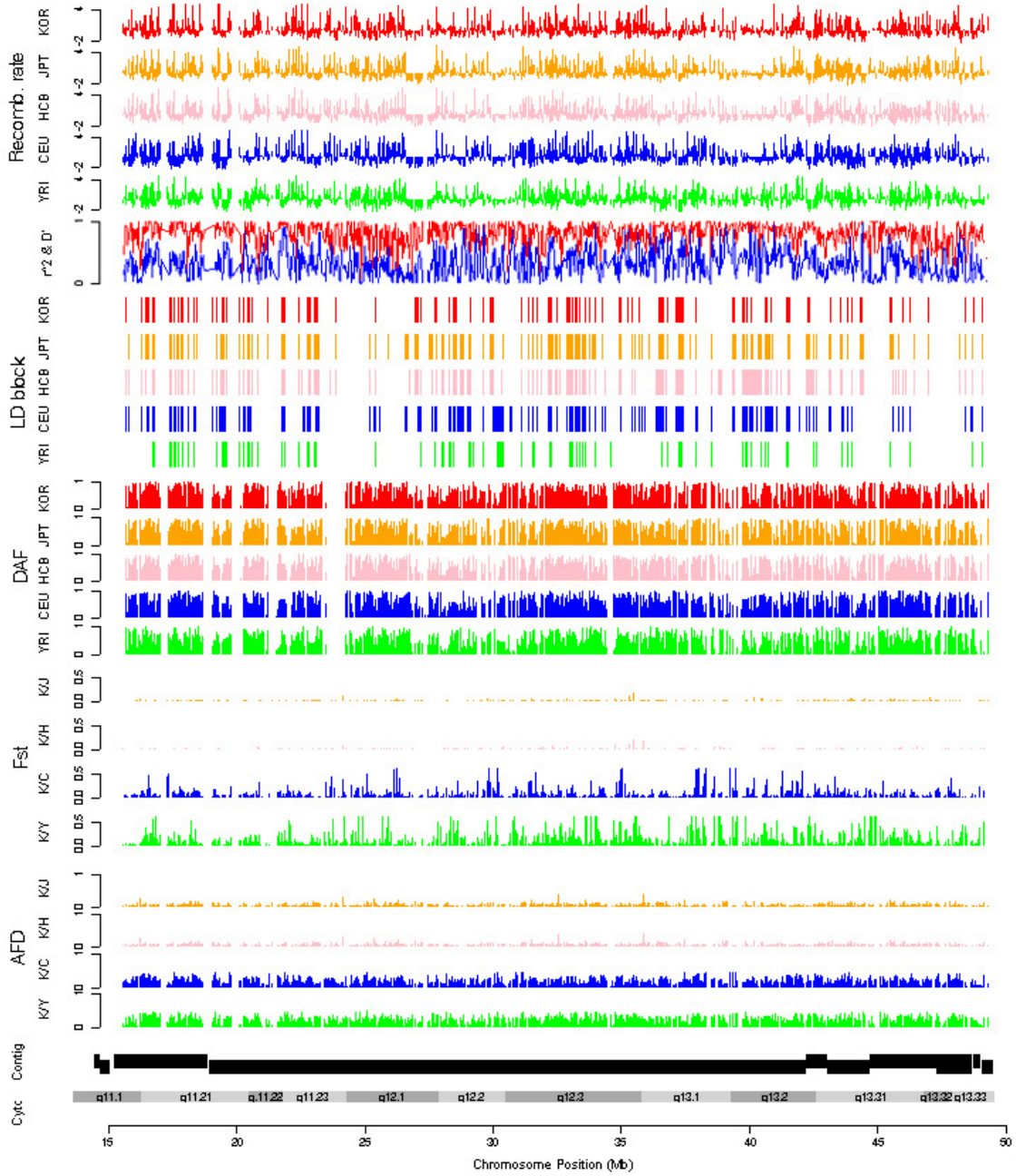

Fig. 2. Patterns of LD on chromosome 22 (15 Mb∼50 Mb). In each panel, the top five sections present recombination rate (taken logarithm of base 2) for Korean (KOR), Japanese (JPT), Chinese (CHB), European (CEU), and African (YRI) populations. Beneath the graph, sliding window plots of r^2 (blue) and |D'| (red) coefficients for common allele markers (overlap of ±5 kb, markers with minor allele frequency ≥0.02) are shown. The next five sections show LD blocks of KOR, JPT, CHB, CEU, and YRI, which were constructed by the Haploview 3.2 program (Gabriel *et al.* 2002), Beneath the bar graph, derived allele frequency (DAF) from chimpanzee is shown for KOR, JPT, CHB, CEU, and YRI. The next four sections present Fst for KOR vs. JPT (K/J), KOR vs. CHB (K/H), KOR vs. CEU (K/C), and KOR vs. YRI (K/Y). Allele frequency differences (AFD) for K/J, K/H, K/C, and K/Y are provided, wherein the CEU minor allele is the reference for the analysis. The chromosome 22 sequence contigs for NCBI release build 35 are shown (successively offset to show breakpoints), followed by the chromosome 22 cytogenetic bands and the chromosomal position. KOR is designated by red, JPT by orange, CHB by pink, CEU by blue, and YRI by green.

some. Of the various ways to define haplotype blocks, we utilized a stringent block definition by Gabriel *et al*. (2002).

An estimated number of hot spots (recombination rate >10), the number of LD blocks, the total LD length, the number of tagSNPs, and the number of haplotypes in five population samples are summarized in Table 2. When the two measures of LD (D' and r^2) were investigated between all pair-wise combinations of markers, the patterns of LD across five population samples were similarly complex (Fig. 2, only the Korean data shown). The patterns of hot spots and LD blocks were dispersed throughout chromosome 22, with some common blocks among population samples, especially highly concordant between Koreans, Japanese, and Chinese. The largest block size in the Korean sample was 273 kb, with an average block size of 20.3 kb and a median block size of 9.7 kb. The position of hot spots was relatively well conserved across all 5 population samples, considering the LD block patterns that were estimated by the Haploview program (Table 2). Total length of LD blocks, number of tagSNPs, and haplotypes were quite similar between Koreans and other populations, except that the YRI sample that showed much lower values. It seemed that African LD blocks were so scattered that they built fewer blocks, and thereby had less total LD block lengths than other populations. The presence of relatively well conserved hot spots across all five population samples indicates that recombination events may not be random along chromosome 22.

Table 2. Comparison of LD patterns and recombination rates in Korean, European, Chinese, Japanese and African populations

Populations*	KOR	CEU	CHB	JPT	YRI
Number of hot spots	161	237	161	165	158
Number of LD blocks	305	314	293	297	238
Total LD length (kb)	6,391.4	8,099.8	8,139.6	8,121.1	4,020.5
Number of tagSNPs	757	749	779	773	577
Number of haplotypes	815	881	772	809	695

*2,528 Overlapped SNPs in five populations were used.

Analyses of human population evolution

To investigate the patterns of population allele evolution from the chimpanzee reference sequence, Korean genotype data were compared to calculate the derived allele frequency (DAF). Of major Korean alleles, 61.4% were identical to chimpanzee reference alleles, while 37.5% were flipped into minor alleles, and 1.2% did not match the chimpanzee reference; therefore, the allele with the lower MAF showed higher similarity in allele frequency (Table 3).

The results of Korean, CEU, and YRI DAF on chromosome 22 are plotted in Fig. 2, and the correlation of derived alleles between Koreans and the four populations are depicted in Fig. 3. The correlation of derived alleles between other populations is presented in Fig. 4. A strong correlation between the Asian samples was clearly seen (Pearson's correlation coefficient >0.80), while the weakest correlation (Pearson's correlation coefficient <0.5) was between Koreans and YRI, as shown in Table 4.

Variation among populations

In Fig. 2, the measurement of Fst level (Weir 1996) provides information on the difference between Koreans against four populations by each different comparison. The results indicate that the comparisons between Asian samples are more similar to each other than between those of non-Asian groups. The average Fst levels between Asians were less than 0.01, whereas the average Fst values between YRI and other population samples were always greater than 0.1 (Table 4).

Table 3. Comparison of Korean SNP allele with the chimpanzee reference sequence

Korean minor allele frequency	H_major = Chimp ref.*	Percent	H_minor = Chimp ref.†	Percent	H_allele != Chimp ref.‡	Percent
maf ≤ 0.1	644/4,012	16.1%	221/4,012	5.5%	11/4,012	0.3%
0.1 < maf ≤ 0.2	533/4,012	13.3%	303/4,012	7.6%	7/4,012	0.2%
0.2 < maf ≤ 0.3	486/4,012	12.1%	266/4,012	6.6%	9/4,012	0.2%
0.3 < maf ≤ 0.4	418/4,012	10.4%	303/4,012	7.6%	11/4,012	0.3%
0.4 < maf ≤ 0.5	381/4,012	9.5%	411/4,012	10.2%	8/4,012	0.2%
Total	2,462/4,012	61.4%	1,504/4,012	37.5%	46/4,012	1.2%

*Human major allele is the same as the chimpanzee reference allele, †Human minor allele is the same as the chimpanzee reference allele, ‡Human allele is not the same as the chimpanzee reference allele.

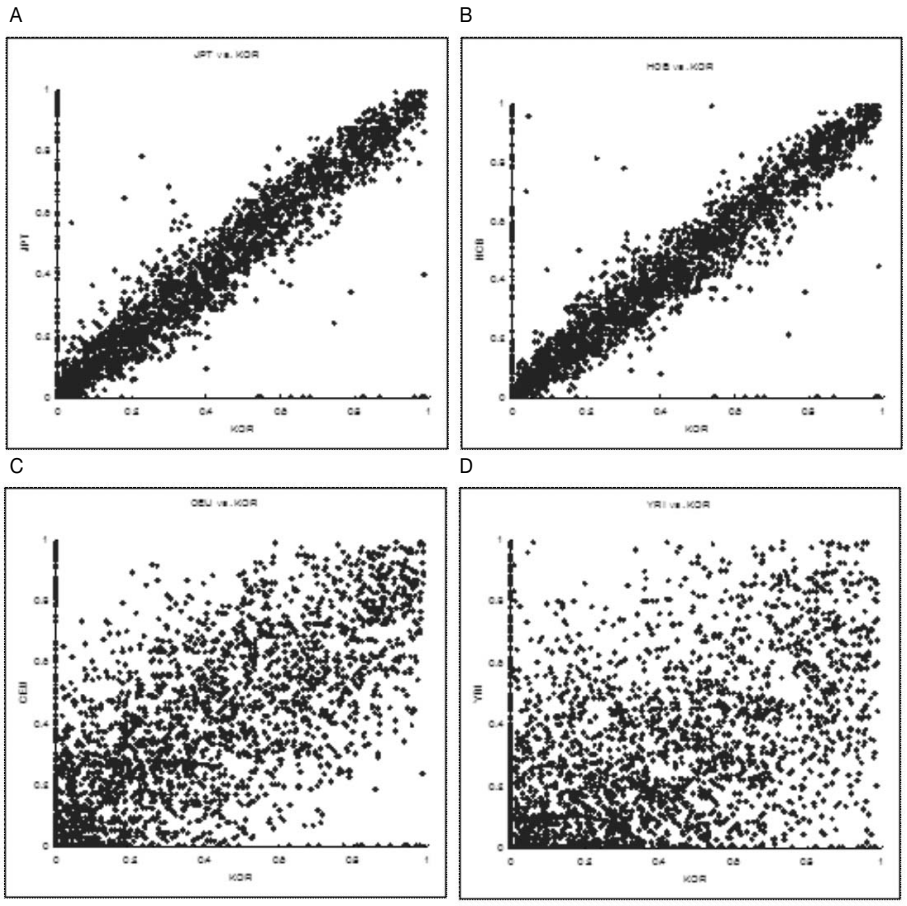

Fig. 3. Correlation of Koreans (KOR) and other populations for the derived allele frequencies (DAFs) from the chimpanzee allele as the reference. (A) The DAF distribution of Japanese (JPT) and KOR. (B) The DAF distribution of Chinese (CHB) and KOR. (C) The DAF distribution of Europeans (CEU) and KOR. (D) The DAF distribution of Africans (YRI) and KOR. Pearson's product-moment correlation is shown in Table 4.

The allele frequency difference (AFD) among the five population samples was analyzed by taking the absolute difference of allele frequency at a locus using an European minor allele as the reference (Fig. 2). The comparison of Koreans to other Asian samples indicates shorter peaks than non-Asian populations, by AFD measurement.

These AFDs were well correlated with Pearson's product-moment correlation (Table 4). Again, the correlation between Asians was greater than 0.93, whereas the correlation between Africans and other populations was always less than 0.35.

All three parameters-- Pearson's product-moment correlation of DAF, the AFD correlation with the European minor allele as the reference, and the average Fst between all four populations-- generated the same qualitative neighbor-joining tree as shown in Fig. 5A (only the tree of DAF is shown). The smaller Fst values and the higher the correlation of DAF and AFD mean that populations are more closely connected. The distance-mediated neighbor-joining tree suggests that Asians make a very tight cluster in which the distance between Koreans and Japanese is the shortest (Fig. 5B). The distance between Asian and African clusters is longer than the distance between European and Asian clusters. Relative distance as a quantitative measure, considering Europeans as 1, indicated that all three Asian populations were very proximate, while Europeans and Africans were significantly remote from Asian populations (Table 5).

If one population has an opposite minor allele compared with the European minor allele at a locus, the allele is designated as flipped. Table 6 summarizes the number of SNPs that were flipped on chromosome 22. Considering CEU as the reference, JPT had the least SNPs flipped (23.3%), while YRI had the most (31.1%).

Discussion

Recent progress in the human genome linkage disequilibrium (LD) map will facilitate identification and characterization of genetic variants that contribute to human phenotypes in general and human diseases, in particular (Dawson *et al*. 2002). Genome-wide maps of common haplotype blocks enable us to maximize information content and to minimize the number of SNPs

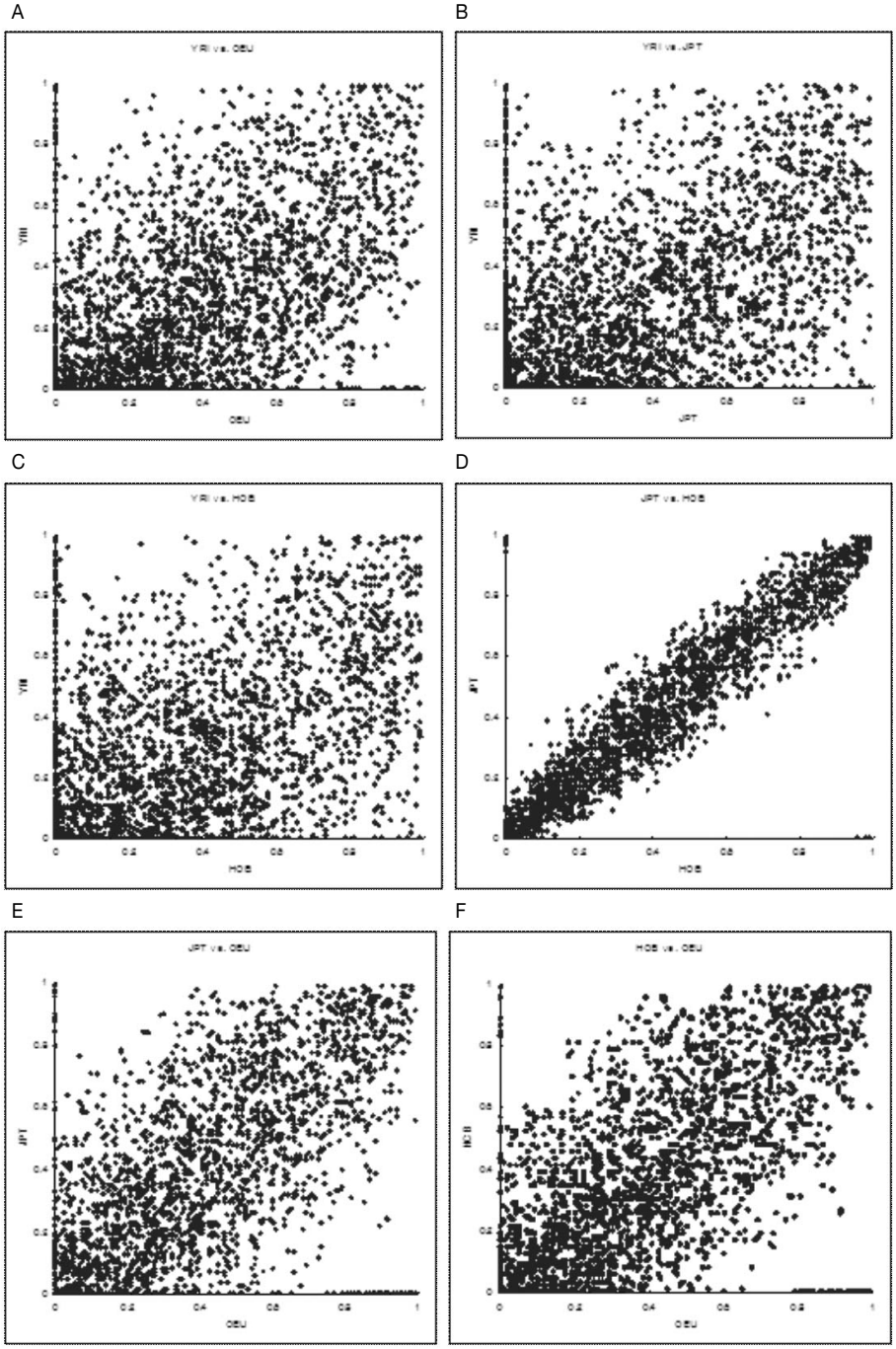

Fig. 4. Correlation of four populations for the derived allele frequencies (DAFs) from chimpanzee allele as the reference. (A) The DAF distribution of Africans (YRI) and Europeans (CEU). (B) The DAF distribution of YRI and Japanese (JPT). (C) The DAF distribution of YRI and Chinese (CHB). (D) The DAF distribution of JPT and CHB. (E) The DAF distribution of JPT and CEU. (F) The DAF distribution of CHB and CEU. Pearson's product-moment correlation is shown in Table 4.

that are required for whole genome-wide screening. However, it is not clear whether its applicability will extend beyond the populations chosen and be limited to the few samples studied (Lai *et al*. 2002).

We have studied the LD and haplotype architecture of chromosome 22 using 90 unrelated healthy Koreans as a part of the Korean HapMap Project. Chromosome 22 was chosen due to its relatively small size. When this work was carried out in 2003, the number of validated SNPs in dbSNP was much less compared with the following years, which is one of the reasons that we failed in PCR and extension reactions of the SNP genotyping experiments.

Comparing 3430 common SNPs among five populations, 514 SNPs were monomorphic in Koreans. Of those, more than 72% was found to be shared in Asian populations, while Europeans and Africans had less than a 38% overlap, reflecting strong genetic affinities among

Table 4. Pearson's product-moment correlation of derived allele frequency (DAF), allele frequency difference (AFD) correlation as the European minor allele as the reference, and average Fst between European, Chinese, Japanese, Korean and African populations

(AFD) Average Fst	DAF				
	CEU	CHB	JPT	KOR	YRI
CEU	0	0.780 (0.416)	0.762 (0.406)	0.616 (0.399)	0.689 (0.332)
CHB	0.063	0	0.947 (0.935)	0.806 (0.950)	0.622 (0.294)
JPT	0.065	0.009	0	0.806 (0.948)	0.603 (0.271)
KOR	0.067	0.006	0.007	0	0.486 (0.293)
YRI	0.103	0.115	0.117	0.115	0

Table 5. Relative distance of European, Chinese, Japanese, Korean and African populations from Europeans and their standard deviation (Europeans are referenced as 1)

Populations	Relative distance from European	Std. dev
CEU	1.000	0.026
CHB	0.757	0.019
JPT	0.786	0.024
KOR	0.784	0.028
YRI	1.428	0.051

Table 6. Number of SNPs flipped in European, Chinese, Japanese, Korean and African populations when referenced to European minor allele in chromosome 22

Populations	SNPs flipped	Total number of SNPs	Percent
CEU	0	2,528	Reference
CHB	618	2,528	24.4%
JPT	590	2,528	23.3%
KOR	629	2,528	24.9%
YRI	787	2,528	31.1%

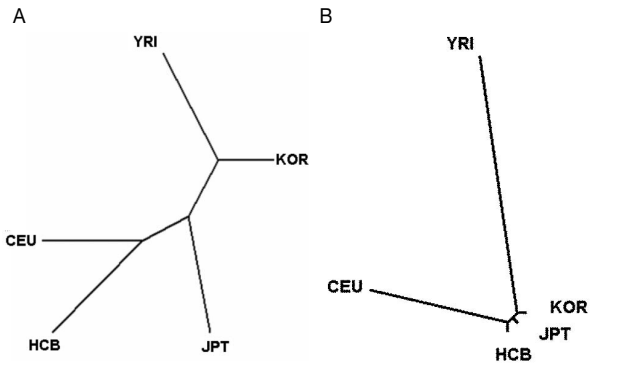

Fig. 5. Network tree diagram of correlations based on three parameters, Pearson's product-moment correlation of derived allele difference (DAF); allele frequency difference (AFD) correlation as the European minor allele as the reference; and average Fst between European (CEU), Japanese (JPT), Chinese (CHB), Korean (KOR), and African (YRI) populations. (A) A qualitative neighbor-joining tree of DAF. (B) Distance-mediated neighbor-joining tree.

the three Asian populations. Many comparative studies of LD patterns at various levels showed a high degree of concordance among various populations (Gabriel *et al.* 2002; Stenzel *et al.* 2004; De La Vega *et al.* 2005). The patterns of hot spots and LD blocks in this study also showed dispersed common blocks throughout chromosome 22, especially highly overlapped among Asian populations. The numbers of hot spots seem more conservative than the number of LD blocks (Table 2). Because recombination hotspots are ubiquitous features of the human genome and create clear block structures of different populations, they delimit strong cold spots of LD blocks (McVean *et al.* 2004).

One approach to understand what makes humans unique as a species is to perform structural and functional comparisons between the genomes of humans and our closest evolutionary relatives, the great apes (Thomson *et al.* 2000). Recently, since the draft sequence of the chimpanzee genome has been available, derived allele frequency (DAF) can be determined by comparing human alleles with chimpanzee, wherein DAF reflects the evolutionary direction of an allele toward human. A strong correlation between Asians was clearly seen, while Europeans and Africans showed very weak correlation (Fig. 3).

Furthermore, analysis of the distribution of Fst levels and allele frequency difference (AFD) between Koreans and the four other populations confirmed a strong correlation between Asians, while Europeans and Africans showed very weak correlation with Koreans. The distance-mediated neighbor-joining tree that considered these three factors showed that Asians and Africans are more remotely connected and that Europeans are closer to the Asian populations (Fig. 5B). Among Asian clusters, Koreans and Japanese are the closest. These results reflect the "Out of Africa" theory, which claims that our roots are in Africa, and most of the oldest non-African lineages are Asian (Thomson *et al.* 2000). The short relational distance of Asians is partly responsible for geographical adjacency, historical factors, and migration of people.

In summary, the Korean LD map of chromosome 22 was constructed and compared with other populations. Our observation from allelic characteristics of Koreans, including genetic variations, indicated that all three Asian populations are very proximate, while Europeans and Africans are significantly remote from Asian populations. The fact that the Asian cluster is tightly related is implied from the genome-wide allelic comparison. However, because chromosome region-specific differences in LD patterns could have occurred, association studies in a given population will require further genotyping experiments and analyses across the whole human genome.

Acknowledgements

The authors thank Aravinda Chakravarti, Peter Chen, and Carl Kashuk at McKusick-Nathans Institute of Genetic Medicine in Johns Hopkins University for valuable advice and comments on the manuscript. The authors gratefully acknowledge the partial financial support of the Korean Ministry of Science and Technology, National Research and Development Program, and the Korean Haplotype Information Development Program.

References

Abecasis, G.R., Noguchi, E., Heinzmann, A., Traherne, J.A., Bhattacharyya, S., *et al.* (2001). Extent and distribution of linkage disequilibrium in three genome regions, *Am. J. Hum. Genet.* 68, 191-197.

Bansal, A., Van den Boom, D., Kammerer, S., Honisch, C., Adam, G., *et al.* (2005). Association testing by DNA pooling: an effective initial screen, *Proc. Natl. Acad. Sci. USA* 99, 16871-16874.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2004). HaploView: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Cardon, L.R., and Abecasis, G.R. (2003). Using haplotype blocks to map human complex trait loci. *Trends in Genetics* 19, 135-140.

Chakravarti, A. (2001). To a future genetic medicine. *Nature* 409, 822-823.

Choi, J.H., Jung, H.Y., Kim, H.S., and Cho, H.G. (2000). PhyloDraw: a phylogenetic tree drawing system. *Bioinformatics* 16, 1056-1058.

Crawford, D.C., Bhangale, T., Hellenthal, N. Li, G., Rieder, D., *et al.* (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* 36, 700-706.

Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J., and Lander, E.S. (2001). High resolution haplotype structure in the human genome. *Nat. Genet.* 29, 229-232.

Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., *et al.* (2002). A first generation linkage disequilibrium map of human chromosome 22. *Nature* 418, 544-548.

De La Vega, F.M., Isaac, H., Collins, A., Scafe, C.R., Halldorsson, B.V., *et al.* (2005). The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res.* 15, 454-462.

Excoffier, L. and Slatkin, M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12, 921- 927.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., *et al.* (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225-2229.

Genomme, G.A., and Van Oene, M. (2005). High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* 45, 660-666.

Hill, W.G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226-231.

Jeffreys, A.J., Kauppi, L., and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* 29, 217-222.

Kim, Y.J., Choi, K.O., Kang, H.J., Kim, C.B., Yu, U.S., *et al.* (2003). Parametric optimization in the selection of SNP loci for hapMap project. *The 2nd Annual Conference of the Korean Society for Bioinformatics*, Daejeon, Korea.

Kruglyak, L. (1999). Prospect for whole genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22, 151-157.

Lai, E., Bowman, C., Bansal, A., Hughes, A., Mosteller, M., and Roses, A.D. (2002). Medical applications of haplotype- based SNP maps: learning to walk before we run. *Nat. Genet.* 32, 353-354.

Lewontin, R.C. (1998). On measures of gametic disequilibrium. *Genetics,* 120, 849-852.

Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using SNP data. *Genet.* 165, 2213-2233.

McVean, G.A.T., Myers, S., Hunt, S., Deloukas, P., Bentley, D., *et al.* (2004). The fine-scale structure of recombination rate variation in the human genome. *Science* 304, 581-584.

Mueller, J.C., Lohmussaar, E., Magi, R., Remm, M., Bettecken, T., *et al.* (2005). Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am. J. Hum. Genet.* 76, 387-398.

Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310, 321-324.

Park, L.Y. (2007). Controlling Linkage Disequilibrium in Association Tests: Revisiting APOE Association in Alzheimer's Disease. *Genomics & Informatics* 5(2), 61-67.

Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., *et al.* (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294, 1719-1723.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., *et*

al. (2001). Linkage disequilibrium in the human genome. *Nature* 411, 199-204.
Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516-1517.
Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365-386.
Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., and Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum. Mo. Genet.* 12, 771-776.
Stenzel, A., Lu, T., Koch, W.A., Hampe, J., Guenther, S.M., *et al*. (2004). Patterns of linkage disequilibrium in the human MHC region on human chromosome 6p. *Hum. Genet.* 14, 377-385.
Stephens, M., and Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* 73, 1162-1169.
The International HapMap Consortium. (2003). The International HapMap Project. *Nature* 426, 789-796.
The International HapMap Consortium. (2004). Integrating ethics and science in the International HapMap Project. Nature Reviews: *Genetics* 5, 467-475.
The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-861.
Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., and Feldman, M.W. (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* 97, 7360-7365.
Weir, M. (1996). Genetic data analysis II. *Sinauer*, Sunderland.
Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.* 15, 323-354.
Zavattari, P., Deidda, E., Whalen, M., Lampis, R., Mulargia, A., *et al*. (2000). Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.* 9, 2947-2957.
Zhang, W., Collins, A., and Morton, N.E. (2004). Does haplotype diversity predict power for association mapping of disease susceptibility?. *Hum. Genet.* 115, 157-164.

Website References

The UCSC Genome Bioinformatics Site. *http://genome.ucsc.edu/*
The SNP Consortium (TSC) Ltd. *http://snp.cshl.org/*
A database of Japanese Single Nucleotide Polymorphisms. *http://snp.ims.u-tokyo.ac.jp/*
Korean HapMap Project homepage. *http://www.khapmap.org/*
International HapMap Project homepage. *http://www.hapmap.org/*
Korean National Institute of Health (KNIH). *http://www.mohw.go.kr/index.jsp/*
Single Nucleotide Polymorphism database. *http://www.ncbi.nlm.nih.gov/SNP/*