Genomics & Informatics Vol. 5(4) 152-160, December 2007

Characterization of Single Nucleotide Polymorphisms in 55 Disease-Associated Genes in a Korean Population

Seung Ku Lee[1], Hyoun Geun Kim[1], Jason J Kang[2], Wonil Oh[3], Bermseok Oh[4] and KyuBum Kwack[1]*

[1]Medical Genomics Lab, CHA Research Institute, Pochon CHA University, Seongnam, Gyeonggi 463-836, Korea, [2]Macrogen, Inc., Seoul 153-781, Korea, [3]Medipost, Co., Seoul 137-874, Korea, [4]Center for Genome Science, NIH, KCDC, Seoul 122-701, Korea

Abstract

Most common diseases are caused by multiple genetic and environmental factors. Among the genetic factors, single nucleotide polymorphisms (SNPs) are common DNA sequence variations in individuals and can serve as important genetic markers. Recently, investigations of gene-based and whole genome-based SNPs have been applied to association studies for marker discovery. However, SNPs are so population-specific that the association needs to be verified. Fifty-five genes and 384 SNPs were selected based on association with disease. Genotypes of 337 SNPs in candidate genes were determined using Illumina Sentrix Array Matrix (SAM) chips by an allele-specific extension method in 364 unrelated Korean individuals. Allelic frequencies of SNPs were compared with those of other populations obtained from the International HapMap database. Minor allele frequencies, linkage disequilibrium blocks, tagSNPs, and haplotypes of functional candidate SNPs in 55 genetic disease-associated genes were provided. Our data may provide useful information for the selection of genetic markers for gene-based genetic disease-association studies of the Korean population.

Keywords: disease-associated gene; single nucleotide polymorphism (SNP); haplotype; linkage disequilibrium; Korean

Introduction

Single nucleotide polymorphisms (SNPs) are the most common type of variant in the human genome and the main source of phenotypic differences between individuals (Sachidanandam *et al.*, 2001). SNPs may influence complex diseases by a variety of mechanisms. By changing the amino acid sequence encoded by DNA, they may reduce, abolish, or increase functions such as DNA binding, protein-protein interactions, enzyme catalytic activity, or receptor-ligand contact (In *et al.*, 1997; Kammerer *et al.*, 2004; Knight *et al.*, 2004). SNPs also can affect aspects of RNA and protein processing, including stability, transcription, translation, and translational modification (Ozaki *et al.*, 2004; Skoog *et al.*, 1999; Wilson *et al.*, 1997).

Genome-wide association studies have shown that variations in human genome sequences in close physical proximity often are strongly correlated (Gabriel *et al.*, 2002). The correlation structure, or linkage disequilibrium (LD), is complex and varies from one region of the genome to another, as well as between different populations (Hinds *et al.*, 2005). The selection of a maximally informative subset of common SNPs for use in association studies is required to assess the causal roles of common DNA variations in complex human traits (Patil *et al.*, 2001). There have been many studies indicating that identification of genetic variants in association studies is valuable in the development of pharmaceutical products and medical diagnostics (Cardon and Abecasis, 2003; Cargill *et al.*, 1999; Knight, 2003; Knight, 2005). The National Center for Biotechnology Information public SNP database (dbSNP) (Sherry *et al.*, 2001), the Genetic Association Database (GAD) (Becker *et al.*, 2004), the SNP Consortium website (Thorisson and Stein, 2003), and the International HapMap database (Thorisson *et al.*, 2005) are very valuable resources for information on genomic polymorphisms. SNP databases and programs are being developed to search for disease-candidate SNPs. Because gene-based SNP research is being pursued vigorously, selection of SNP candidates from SNP databases is thought to be essential in gene-based association studies.

We tested a hypothesis that selection of candidate genes and SNPs using bioinformatics software and servers is an efficient means of functional SNP discovery. Fifty-five genes were selected from a database that is an archive of human genetic association studies of complex diseases and disorders. Most of these genes have significant association with various diseases from other populations, but not in Koreans. SNPs were selected

*Corresponding author: E-mail kbkwack@cha.ac.kr, kbkwack@gmail.com
Tel +82-31-725-8376, Fax +82-31-725-8350
Accepted 10 December 2007

based on their functional and positional importance using databases. Allelic frequency, LD, and haplotype frequency of SNPs were determined using healthy, unrelated Koreans compared with other ethnic groups using the International HapMap database. The results of this study may provide information helpful for the selection of functional candidate SNPs in gene-based disease-association studies.

Materials and Methods

Reagent

Reagents for the Sentrix Array Matrix were obtained from Illumina Inc (San Diego, California, USA). Other reagents were purchased Sigma-Aldrich (St. Louis, MO).

Genomic DNA samples

The 364 unrelated Korean umbilical cord blood cohort samples used in this study were obtained from Medipost, Inc (Seoul, Korea). The samples and informed consents received Institutional Review Board (IRB) approval. Information on all of the samples, including gender, neonatal body weight, maternal age, and disease pedigree were obtained from Medipost, Inc. Genomic DNA was purified using Gentra DNA purification kits (Gentra, Inc. Minneapolis, USA). The Centre d'Etude du Polymorphisme Humain (CEPH) parents and offspring trio genomic DNA samples were obtained from the Illumina Inc. for genotype testing.

Selection of disease-associated genes and SNPs

Fifty-five candidate genes for common diseases were selected from the Genetic Association Database (GAD), which is an archive of human genetic association studies of complex diseases and disorders (Becker *et al.*, 2004). Polymorphisms in the 55 disease-associated genes were significantly associated with human diseases such as Alzheimer disease, asthma, atopy, cancer, diabetes, heart disease, hypertension, and osteoarthritis. Selected genes were described in Bioinformatic Harvester (http://harvester.fzk.de/harvester/) and the Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/ omim/) (Table 1). SNPs of the 55 genes may be good candidates for association studies of complex diseases in Koreans because SNPs of their candidate genes were significantly associated with diseases in other ethnic populations.

We used bioinformatics servers and software to identify SNPs in genes. Selection of SNPs in promoter regions was achieved using Alibaba software that predicted transcription factor binding sites affected by the nucleotide exchange generated by SNPs. SNPs in introns and exons were selected within the <500 bp exon boundary, and nonsynonymous SNPs were predicted in exons. The SNPs selected for genotyping were located in exons (n = 104), introns (n = 139), and untranslated regions (UTRs; n = 141).

Genotyping using Sentrix Array Matrix chips

Genotyping was performed using the Sentrix Array Matrix (SAM) system (Illumina, Inc., San Diego, California, USA) as described by Fan and co-workers (Fan *et al.*, 2003). DNA samples were assayed according to the SAM protocols supplied by Illumina, Inc. The procedure was similar to the one described previously. Briefly, genomic DNA (250 ng) was immobilized on streptavidin-coated magnetic beads. Allele-specific extension was carried out by ramping the temperature from 70ºC to 30ºC over 16 hr. After extension, excess oligonucleotides were washed out with 50 uL of MEL (master mix for extension/ligation) 1 buffer and 50 uL of UB (universal buffer) 1. The master mix for ligation (37 uL) was added to the extension products and incubated for 15 min at 45ºC. The products were washed and resuspended in 35 uL of IP (inoc PCR) 1 buffer and heated at 95ºC for 1 min. The products were incubated on a magnetic plate for 2 min. The Illumina-recommended Titanium Taq DNA Polymerase mixture (30 uL, BD-Clontech) was added to the uracil DNA glycosylase mixture (UDG, Illumina, Inc.) in new plates. The supernatants (30 uL) from the magnetic plates were used for PCR with a reaction profile of 35 s at 95ºC, 35 s at 56ºC, and 2 min at 72ºC, for 34 cycles. Purification of the PCR products was accomplished using 96-well filter plates (Millipore, USA). All samples were transferred to a 384-well microplate. The arrays were hydrated in UB2, 0.1 N NaOH, and UB2 buffer. Hybridizations were conducted at 60ºC for 30 min and at 45ºC for 16 hr. After hybridization, the arrays were washed twice with UB2 and once with IS (Image SAM) 1 buffer at room temperature with mild agitation, and imaged at a resolution of 0.8 um using a BeadArray Reader (Illumina, Inc., San Diego, CA, USA).

Statistical analysis

Data analysis was performed by Illumina's GenCall software (Version 6.1.3.24) for genotyping distribution. MAF, LD, and haplotype frequencies for genes were estimated using HapAnalyzer (Jung *et al.*, 2004) and Haploview software (Barrett *et al.*, 2005). Ethnic differences in allele frequency were calculated by taking the absolute of the subtraction of the minor allele of groups at each SNP site.

Results

SNPs in the Korean population

Analysis of genotype data was performed using Illumina

Table 1. Candidate genes and genotyped SNPs

Gene Symbol	Gene	Flanking 5UTR	5UTR	coding	intron	3UTR	Flanking 3UTR	Total	Broad phenotype
ACE	NM_000789	0	0	5	6	0	0	11	Alzheimer disease, Coronary artery disease, Hypertension, Myocardial infarction
ADRB2	NM_000024	0	0	1	0	0	0	1	Asthma, High total IgE, Hypertension
AGT	NM_000029	6	1	5	0	1	2	15	Cardiovascular, Hypertension, Myocardial infarction
AGTR1	NM_000685	0	0	0	1	0	0	1	Cardiovascular, Coronary artery disease, Hypertension
APC	NM_000038	0	1	1	2	0	0	4	Adenomatous polyposis, Colorectal adenocarcinomas, Late onset of familial adenoma
APOA1	NM_000039	0	0	1	0	0	0	1	Alzheimer Disease
APOB	NM_000384	4	0	18	3	0	0	25	Cardiovascular, Coronary artery disease, Obesity
APOE	NM_000041	0	0	1	1	0	0	2	Alzheimer disease, Cardiovascular disease, Multiple sclerosis, Myocardial infarction,
AR	NM_000044	0	0	4	0	0	0	4	Prostate cancer
BRCA1	NM_007296	0	0	4	0	0	0	4	Breast cancer, Ovarian cancer
CCL25	NM_005624	0	0	1	0	0	0	1	Inflammatory
CCR5	NM_000579	1	0	0	0	0	0	1	Asthma, HCV infection, HIV infection, Rheumatoid Arthritis
CD14	NM_000591	0	0	1	0	0	0	1	Alcoholic liver disease, Atopy, Total IgE
COMT	NM_000754	0	0	0	3	0	0	3	Alcoholism, Bipolar disorder, Obsessive compulsive disorder, Schizophrenia
CPAMD8	NM_015692	0	0	1	0	0	0	1	Innate immunity
CTLA4	NM_005214	2	0	0	1	0	1	4	Asthma, Graves disease, Rheumatoid arthritis, Type 1 diabetes
CYP11B1	NM_001026213	0	0	1	0	0	0	1	Hypertension, Hyperaldosteronism
CYP11B2	NM_000498	0	0	1	0	0	0	1	Hypertension, End-stage renal disease, Cardiovascular
CYP17A1	NM_000102	0	0	1	1	0	0	2	Breast cancer, Endometriosis, Prostate cancer
CYP1A1	NM_000499	0	0	1	0	0	0	1	Breast cancer, Lung cancer, Prostate cancer
CYP1B1	NM_000104	0	0	1	0	0	0	1	Breast cancer, Prostate cancer, Lung cancer
CYP2E1	NM_000773	1	0	0	0	0	0	1	Adult brain tumors, Parkinson disease
DRD2	NM_000795	0	0	1	5	1	0	7	Alcoholism, Schizophrenia, Type 2 diabetes
DRD3	NM_033663	0	0	1	8	0	0	9	Bipolar disorder, Schizophrenia
DRD4	NM_000797	1	0	1	0	1	0	3	Attention Deficit Hyperactivity Disorder , Attention problems, Major psychoses
F2	NM_000506	0	0	0	0	1	0	1	Behcet disease, Bilateral iliac vein thrombosis, Elevated plasma prothrombin level
FGA	NM_000508	0	0	1	0	1	0	2	Post-trauma fibrinogen increase, Renal amyloidosis, Plasma fibrinogen levels
GNB3	NM_002075	0	0	1	1	0	0	2	Coronary artery disease, Hypertension, Seasonal affective disorder
HFE	NM_000410	0	0	2	0	0	0	2	Cardiovascular death, Colon cancer, Hereditary hemochromatosis, Type 2 diabetes
HTR2A	NM_000621	1	0	0	10	1	0	12	Attention deficit hyperactivity disorder, Alzheimer disease, Schizophrenia
IFNG	NM_000619	0	0	0	2	0	0	2	Asthma, Type 1 Diabetes
IL10	NM_000572	4	0	0	8	2	1	15	Asthma, Eosinophilia, Rheumatoid arthritis, Type 1 diabetes
IL1RN	NM_173842	0	3	0	7	0	0	10	Acute graft-versus-host disease, Alcoholism, Rheumatoid arthritis, Ulcerative colitis
IL4	NM_000589	0	0	1	0	0	0	1	Atopic asthma, Increased IgE. Rheumatoid arthritis, Type 1 diabetes
IL4R	NM_000418	1	0	8	6	0	1	16	Asthma, Atopy, Total serum IgE
IL6	NM_000600	0	1	2	2	0	0	5	Ageing, Bone mineral density, Coronary artery disease, Total serum IgE , Diabetes
INS	NM_000207	0	0	1	0	0	0	1	Atherosclerosis, Obesity, Polycystic ovary syndrome, Prostate cancer, Type 1 diabetes
LIPC	NM_000236	0	0	2	0	0	0	2	Carotid artery stenosis, Heritable lipolytic deficiency, Increased concentration of HDL, low hepatic lipase activity

Table 1. Continued

Gene Symbol	Gene	Flanking 5UTR	5UTR	coding	intron	3UTR	Flanking 3UTR	Total	Broad phenotype
LPL	NM_000237	0	0	0	4	0	1	5	Brain Infarction, Cardiovascular, Coronary artery disease, Elevated HDL cholesterol
LTA	NM_000595	0	0	0	0	3	3	6	Myocadial infarction
MPO	NM_000250	0	0	1	1	0	0	2	Coronary artery disease, Lung squamous cell carcinoma, Alzheimer disease
MTHFR	NM_005957	0	0	3	1	3	0	7	Coronary artery disease, Type 2 diabetes, Carotid atherosclerosis
NAT2	NM_000015	0	0	3	0	0	0	3	Breast cancer, Endometriosis, Parkinson`s disease
NOS3	NM_000603	0	0	5	15	0	0	20	Alzheimer disease, Hypertension, Myocardial infarction, Type 2 diabetes
PON1	NM_000446	0	0	3	0	0	0	3	Bone mineral density, Cardiovascular disease, Parkinson disease
PPARG	NM_138711	0	0	0	27	0	0	27	Body weight, Bone mineral density, Obesity, Type 2 diabetes
PSEN1	NM_000021	0	0	0	4	0	0	4	Alzheimer disease
SLC6A4	NM_001045	0	1	2	12	1	1	17	Alcoholism, Anxiety symptoms, Autism, Cardiovascular, Personality traits, Suicide
TGFB1	NM_000660	0	0	0	2	0	0	2	Asthma severity, Atopic dermatitis, Cardiovascular, Periodontitis
TH	NM_199292	0	0	1	0	0	0	1	Hypertension, Mood disorder, Schizophrenia
TNF	NM_000594	4	1	2	5	0	1	13	Alzheimer disease, Asthma, Bone mineral mass, Psoriasis, Type 2 diabetes
TNFRSF11A	NM_003839	0	0	1	0	0	0	1	Type 1 diabetes
TP53	NM_000546	0	0	1	2	0	0	3	Bladder cancer, Colorectal cancer, Gastric cancer, Lung cancer, Smoking
TTR	NM_000371	0	0	1	2	0	0	3	Amyloid cardiomyopathy, Familial amyloid polyneuropathy
VDR	NM_001017535	3	0	1	30	5	2	41	Addison disease, Osteoporosis, Prostate cancer, Type 1 diabetes
WDR79	NM_018081	0	0	2	1	0	0	3	-
	Total	28	8	95	173	20	13	337	

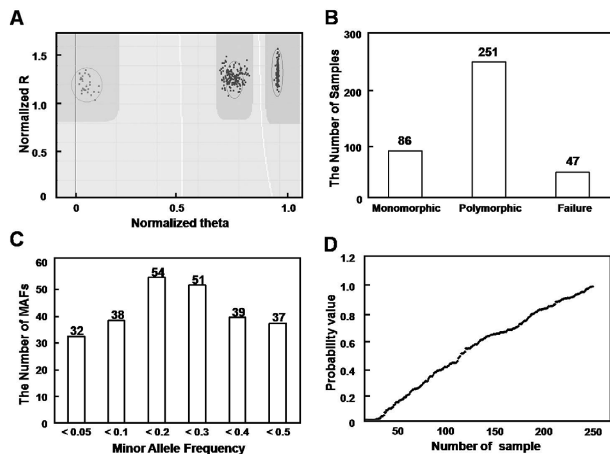

Fig. 1. Summary of SNP genotyping in the Korean population. Distribution of SNPs. Genotyping plots were created by GenCall software (A). Of 337 SNPs, genotype patterns revealed 239 polymorphic SNPs (62.2%), 83 monomorphisms (21.6%), and 62 failed SNPs (16.1%) (B). Distribution of minor allelic frequencies was determined for 364 samples (C). The Hardy-Weinberg equilibrium (HWE) was calculated by the chi-square test for 251 SNPs (D).

GenCall software (Fig. 1A). Cross-contamination, accuracy, and heritability errors were tested using a blank sample and genomic DNA from six families of Centre d'Etude du Polymorphisme Humain (CEPH). Each family was composed of four sets of parents and offspring trios. Successful genotyping was accomplished for 337 SNPs (87.8%) compared with 47 failed SNPs (12.2%) (Fig. 1B). SNP failure occurred because of inexact clustering, revealing inheritability error in the data. The observed concordance in genotyping was 99.6% (122,151 out of 122,668). Two hundred nineteen of the 337 SNPs were polymorphic (65%, MAF >0.05), while the remaining 118 SNPs (35%) were either monomorphic (86 SNP; 25.5%) or of low frequency (32 SNP; 9.5%, MAF < 0.05) (Fig. 1B and C). Successfully genotyped 337 SNPs were located as follows: 36 in promoter regions (5'-UTR flanking and 5'-UTR), 95 in exons (non-synonymous), 173 in introns (exon-intron boundary), and 33 in the 3'-UTR and flanking sequences (Table 1). The chi-square test indicated a deviation from the Hardy-Weinberg equilibrium (HWE) of 21 among 251 polymorphic SNPs (p <0.05) (Fig. 1D). Genotype frequencies and HWE of other genes are shown on our

website (http://cafe.daum.net/Medigenome: Supplementary Table 1).

Differentiation of SNP allelic frequencies in ethnic populations

The minor allele frequencies of the 242 SNPs were compared pairwise with those of other ethnic populations including Japanese (JPT), Chinese (CHB), Caucasian (CEU), and African (YRI), which have been genotyped in the International HapMap Project (Thorisson *et al.*, 2005) (Fig. 2A). MAFs were found to be very similar between KOR, CHB, and JPT, but showed differences between KOR, CEU, and YRI. Differences in allele frequencies were the absolute differences of the minor allele frequency in Koreans from those of the same allele of other ethnic populations (Fig. 2B, C). The differences in allele frequencies varied depending on ethic group (Table 2). Patterns of different SNP frequencies of the ethnic populations are described in Table 3. rs6078 and rs5896 were monomorphic or high-frequency in YRI and CEU, whereas the same sites were polymorphic in KOR at frequencies of 0.324 and 0.398, respectively. In contrast, rs662 was monomorphic in KOR whereas it was polymorphic in YRI (0.217). rs1899951, rs2287499, rs6356, rs4305, and rs1875796 had higher frequencies in YRI than in KOR, whereas rs2972164 had lower frequencies in YRI than in KOR. rs4135329 was only polymorphic in YRI. Frequencies of other SNPs are available on our website (Supplementary Table 2).

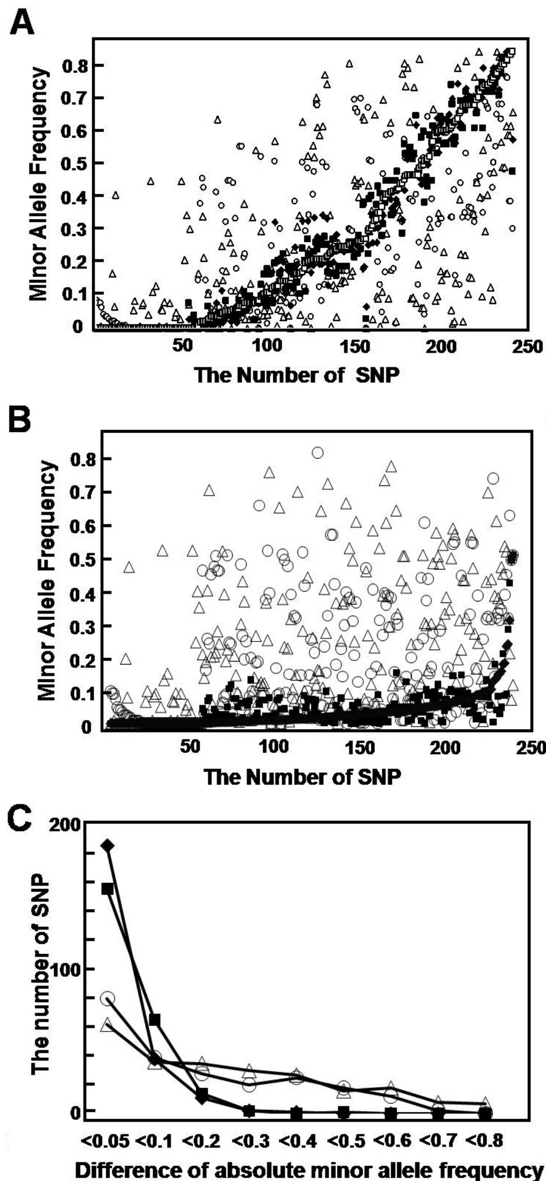

Fig. 2. Comparison of minor allelic frequencies. Minor allele frequencies of 242 SNPs for each population from SNPs database (A, KOR; □, YRI; Δ, CEU; ○, CHB; ■, and JAP; ◆). Absolute differences in allele frequencies of 242 SNPs distributed differences between Koreans and each population (B and C); |KOR-YRI|, Δ; |KOR-CEU|, ○; |KOR-CHB|, ■; and |KOR-JAP|, ◆).

Table 2. Distribution of absolute allele differences in Korean with other ethnic populations

Absolute Frequency	Number (%)			
	CEU-KOR	YRI-KOR	CHB-KOR	JPT-KOR
<0.05	80 (35.7)	62 (25.9)	156 (65.5)	186 (78.2)
<0.1	39 (17.4)	36 (15.1)	65 (27.3)	38 (16.0)
<0.2	28 (12.5)	35 (14.6)	14 (5.9)	11 (4.6)
<0.3	20 (8.9)	30 (12.6)	2 (0.8)	2 (0.8)
<0.4	25 (11.2)	27 (11.3)	0 (0.0)	1 (0.4)
<0.5	18 (8.0)	16 (6.7)	1 (0.4)	0 (0.0)
<0.6	12 (5.4)	18 (7.5)	0 (0.0)	0 (0.0)
<0.7	2 (0.9)	8 (3.3)	0 (0.0)	0 (0.0)
<0.8	0 (0.0)	7 (2.9)	0 (0.0)	0 (0.0)
Total	224	239	238	238

Pairwise linkage disequilibrium mapping and frequencies

Haplotype frequencies for each pair of SNPs and for all SNPs within a gene were estimated within the Korean population using the expectation-maximization (EM) algorithm. To consider patterns of LD, we examined genes in the Korean population containing at least 2 SNPs. Measures of pairwise LD were calculated using the confidence interval method in Haploview software (version 3.2) for all pairs of SNPs in each gene (Gabriel *et al.*, 2002).

Vitamin D (1,25-dihydroxyvitamin D3) receptor (VDR) had 4 haplotype blocks including a first block of 16.6 kb containing SNPs rs11574143, rs11574143, rs3847987, rs739837, rs7975232, rs11574113, rs10875692, rs7305032,

rs11168266, and rs11168267; a second block of 17.6 kb containing SNPs rs11168268, rs2238140, rs2248098, and rs987849; a third block containing SNPs rs1540339, rs11574070, rs2189480, rs3819545, rs3782905, and rs2239186; and the remaining block of 0.2 kb containing SNPs rs11168275 and rs11574050 (Fig. 3A). Three haplotype blocks of peroxisome proliferator–activated receptor gamma (PPARG) were revealed from the LD data for 26 SNPs. PPARG had three haplotype blocks, including a first block of 41 kb containing SNPs rs6809631, rs12636461, rs10510410, rs10510411, rs10510412, rs880663, rs2028760, and rs10510417; a second block of 8 kb containing SNPs rs10510418 and rs4135247; and the remaining block of 43.2 kb containing SNPs rs1373640, rs2028759, rs2972162, rs2938395, rs2959268, rs2938392, rs2959273, rs1875796, and rs1151996 (Fig. 3B). LD blocks and haplotypes of the candidate genes are available on our website (Supplementary Fig. 1).

Discussion

Genetic variations influence the development of cancers and other diseases, and are important for population genetics, genetic disease research, and drug development (Weiss *et al.*, 2001). Because of many genes are involved in the development of diseases and many SNPs are responsible for altered binding sites in promoters or altered splicing regions or altered protein functions, association studies using SNPs are finding increasing numbers of disease target genes (Knight *et al.*, 2004; Ozaki *et al.*, 2004). SNP research is necessary for understanding the role of candidate genes in disease. Recently, information was published on several hundred genes from many different ethnic groups that had statistically significant associations with various diseases (Becker *et al.*, 2004). Results of many studies using the same ethnic groups led

Table 3. Comparison of minor allelic frequencies of Korean SNPs with those of other ethnic groups in the HapMap database

SNP	Minor Allele	YRI	CEU	CHB	JPT	KOR
rs6078	A	0.000	0.000	0.033	0.080	0.324
rs5896	C	1	0.95	0.545	0.466	0.398
rs662	A	0.217	0.642	0.433	0.318	0
rs1824152	A	0.025	0.692	0.433	0.455	0.42
rs1899951	A	0.758	0.075	0.022	0.057	0.043
rs2287499	C	0.917	0.158	0.330	0.239	0.250
rs6356	G	0.933	0.6	0.119	0.314	0.276
rs4305	A	0.925	0.383	0.289	0.466	0.409
rs1875796	C	0.908	0.517	0.231	0.263	0.446
rs2972164	C	0.102	0.509	0.854	0.932	0.896
rs4135329	G	0.534	0.000	0.000	0.000	0.000

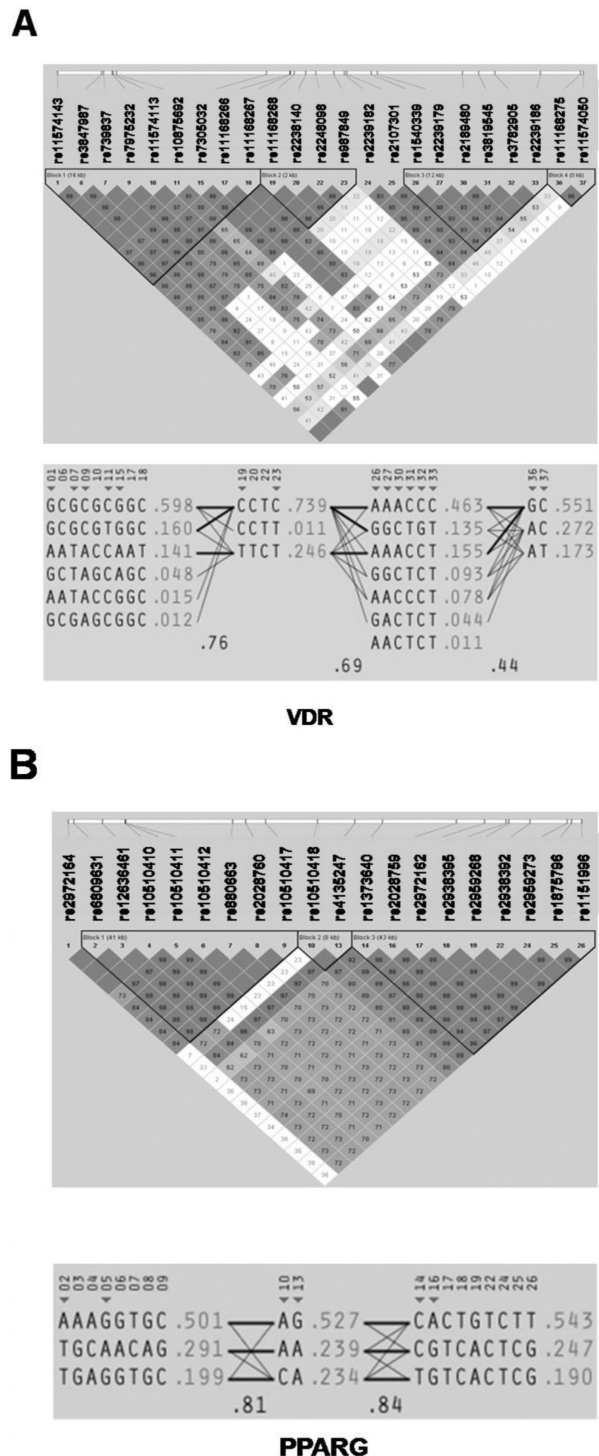

Fig. 3. Pairwise linkage disequilibrium and haplotype frequencies in VDR (A) and PPARG (B) in Korean populations. For common SNPs, samples were removed by p <0.05 and minor allelic frequency <0.1. LD blocks (|D'|) and r^2 of SNPs were described using Haploview software. Regions of high LD (D'=1 and LOD >2) are shown in dark gray.

to controversial conclusions.

We investigated allele frequencies of 332 SNPs in disease-associated genes in Koreans compared with Caucasian, African, Japanese, and Chinese populations. Genes were selected that had a strong association with disease, as determined by statistical significance and repeated publications. The 332 SNP sites in the selected genes were chosen by a similar strategy. Fifty SNP sites that might be involved in transcription factor binding were selected by prediction using Alibaba software (data not shown). Sixty-nine nonsynonymous SNPs were selected on the basis that these SNPs might have effects on protein structure and function. These operationally important SNPs—common functional SNPs—were selected using bioinformatic tools that were available from the internet.

Genotyping was performed on 322 functional SNPs from 55 disease-associated genes from unrelated individuals in the Korean population. Haploview and HapAnalyzer, which were available from the internet, were used for Hardy-Weinberg equilibrium testing and determination of SNP frequencies, LD blocks, haplotypes, and statistical analysis (Fig. 1 and 3; Supplementary Table 1).

The common disease-common variant hypothesis suggests that common diseases are influenced by susceptibility alleles at a few loci that occur with high frequency across ethnically diverse populations (Hirschhorn *et al*., 2002; Tishkoff and Verrelli, 2003). In the Korean population, 158 of the 322 SNPs (49%) had minor allelic frequencies greater than 10% at the 0.05 probability level (Fig. 1). These results indicate that at least 158 SNPs may confirm this hypothesis.

Differences in alleles of several SNPs were found between different ethnic populations (Table 2 and 3, Fig. 2). Also, our results indicate that African and Caucasian populations have similar frequencies in most of the variants, as opposed to the Asian populations (Korean, Japanese, and Chinese). Data from the Korean, Japanese, and Chinese populations revealed similar allelic frequencies among these Asian groups (Fig. 2). Because difference between populations, SNPs should be carefully selected from SNPs database for association studies. In other words, our data indicate that the SNP database is limited in information on MAFs of certain populations. As a consequence, it is necessary to conduct more SNP studies in each population.

Haplotypes are combinations of alleles that differ among different populations (Tishkoff and Verrelli, 2003). LD plays a central role in the current and proposed methods for mapping complex disease genes. Haplotype blocks indicate high LD and low haplotype diversity (Pritchard and Cox, 2002). The discovery of functional variations is important for understanding normal phenotypic variations and susceptibility to genetic disease (Cardon and Abecasis, 2003; Crawford *et al*., 2004; Tanaka *et al*., 2003). Vitamin D is an important factor for bone development and maintenance of bone mass, and it is the principal factor required for control of normal calcium and phosphate homeostasis (Uitterlinden *et al*., 2002). VDR gene polymorphism has been reported to influence calcium intake with regard to bone mineral density (Kiel *et al*., 1997), and growth and parameters of body composition (Lorentzon *et al*., 2000). PPARG2 has been associated with insulin resistance and type 2 diabetes, and inconsistently with obesity and overall adiposity (Ghoussaini *et al*., 2005). We have described the LD block structure, haplotype frequencies, and tagSNPs in these disease-associated genes in samples from unrelated individuals (Fig. 3). These tagSNPs may be used for the design and analysis of genetic association studies, testing variants that are common in the population and that may have a role in complex traits and common diseases.

In summary, we have described SNP frequencies and haplotype frequencies for functional candidate SNPs in 55 candidate genes in unrelated individuals from the Korean population. Our data proved that SNPs are ethnospecific and should be selected very carefully. Our data may help guide selection of SNPs from 55 candiate genes tested in this study for association studies. Additional case-control studies will be necessary to determine the precise effects and function of our selected SNPs. Our results may also contribute information to the fields of medicine, pharmaceutical development, functional genomics, and human evolutionary biology.

Acknowledgments

This work was funded by a grant (0405-BC02-0604-0004) from the Ministry of Health and Welfare, Republic of Korea,

References

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Becker, K.G., Barnes, K.C., Bright, T.J., and Wang, S.A. (2004). The genetic association database. *Nat. Genet.* 36, 431-432.

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

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A.,

Warrington, J., Lipshutz, R., Daley, G.Q., and Lander, E.S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231-238.

Crawford, D.C., Carlson, C.S., Rieder, M.J., Carrington, D.P., Yi, Q., Smith, J.D., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. (2004). Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. *Am. J. Hum. Genet*. 74, 610-622.

Fan, J.B., Oliphant, A., Shen, R., Kermani, B.G., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, S.L., Deloukas, P., Galver, L., Hunt, S., McBride, C., Bibikova, M., Rubano, T., Chen, J., Wickham, E., Doucet, D., Chang, W., Campbell, D., Zhang, B., Kruglyak, S., Bentley, D., Haas, J., Rigault, P., Zhou, L., Stuelpnagel, J., and Chee, M.S. (2003). Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol*. 68, 69-78.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., and Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225-2229.

Ghoussaini, M., Meyre, D., Lobbens, S., Charpentier, G., Clement, K., Charles, M.A., Tauber, M., Weill, J., and Froguel, P. (2005). Implication of the Pro12Ala polymorphism of the PPAR-gamma 2 gene in type 2 diabetes and obesity in the French population. *BMC Med. Genet.* 6, 11.

Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072-1079.

Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet. Med.* 4, 45-61.

In, K.H., Asano, K., Beier, D., Grobholz, J., Finn, P.W., Silverman, E.K., Collins, T., Fischer, A.R., Keith, T.P., Serino, K., Kim, S.W., De Sanctis, G.T., Yandava, C., Pillari, A., Rubin, P., Kemp, J., Israel, E., Busse, W., Ledford, D., Murray, J.J., Segal, A., Tinkleman, D., and Drazen, J.M. (1997). Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription. *J. Clin. Invest.* 99, 1130-1137.

Jung, H.Y., Park, J.S., Park, Y.J., Kim, Y.J., Kimm, K.C., and Koh, I.S. (2004). HapAnalyzer: Minimum Haplotype Analysis System for Association Studies. *Genomics & Informatics* 2, 107-109.

Kammerer, S., Roth, R.B., Reneland, R., Marnellos, G., Hoyal, C.R., Markward, N.J., Ebner, F., Kiechle, M., Schwarz-Boeger, U., Griffiths, L.R., Ulbrich, C., Chrobok, K., Forster, G., Praetorius, G.M., Meyer, P., Rehbock, J., Cantor, C. R., Nelson, M. R., and Braun, A. (2004). Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus. *Cancer Res.* 64, 8906-8910.

Kiel, D.P., Myers, R.H., Cupples, L.A., Kong, X.F., Zhu, X.H., Ordovas, J., Schaefer, E.J., Felson, D.T., Rush, D., Wilson, P.W., Eisman, J.A., and Holick, M.F. (1997). The BsmI vitamin D receptor restriction fragment length polymorphism (bb) influences the effect of calcium intake on bone mineral density. *J. Bone Miner Res.* 12, 1049-57.

Knight, J.C. (2003). Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin. Sci. (Lond)* 104, 493-501.

Knight, J.C. (2005). Regulatory polymorphisms underlying complex disease traits. *J. Mol. Med.* 83, 97-109.

Knight, J.C., Keating, B.J., and Kwiatkowski, D.P. (2004). Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. *Nat. Genet.* 36, 394-399.

Lorentzon, M., Lorentzon, R., and Nordstrom, P. (2000). Vitamin D receptor gene polymorphism is associated with birth height, growth to adolescence, and adult stature in healthy caucasian men: a cross-sectional and longitudinal study. *J. Clin. Endocrinol. Metab.* 85, 1666-1670.

Ozaki, K., Inoue, K., Sato, H., Iida, A., Ohnishi, Y., Sekine, A., Sato, H., Odashiro, K., Nobuyoshi, M., Hori, M., Nakamura, Y., and Tanaka, T. (2004). Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. *Nature* 429, 72-75.

Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., Nguyen, B.T., Norris, M.C., Sheehan, J.B., Shen, N., Stern, D., Stokowski, R.P., Thomas, D.J., Trulson, M.O., Vyas, K.R., Frazer, K.A., Fodor, S.P., and Cox, D.R. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294, 1719-1723.

Pritchard, J.K. and Cox, N.J. (2002). The allelic architecture of human disease genes: common disease-common variant...or not? *Hum. Mol. Genet*. 11, 2417-2423.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D., and

International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928-933.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308-311.

Skoog, T., van't Hooft, F. M., Kallin, B., Jovinge, S., Boquist, S., Nilsson, J., Eriksson, P., and Hamsten, A. (1999). A common functional polymorphism (C→A substitution at position -863) in the promoter region of the tumour necrosis factor-alpha (TNF-alpha) gene associated with reduced circulating levels of TNF-alpha. *Hum. Mol. Genet.* 8, 1443-1449.

Tanaka, T., Ikari, K., Furushima, K., Okada, A., Tanaka, H., Furukawa, K., Yoshida, K., Ikeda, T., Ikegawa, S., Hunt, S.C., Takeda, J., Toh, S., Harata, S., Nakajima, T., and Inoue, I. (2003). Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. *Am. J. Hum. Genet.* 73, 812-822.

Thorisson, G.A., Smith, A.V., Krishnan, L., and Stein, L.D. (2005). The International HapMap Project Web site. *Genome Res.* 15, 1592-1593.

Thorisson, G.A. and Stein, L.D. (2003). The SNP Consortium website: past, present and future. *Nucleic Acids Res.* 31, 124-127.

Tishkoff, S.A., and Verrelli, B.C. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics. Hum. Genet.* 4, 293-340.

Uitterlinden, A.G., Fang, Y., Bergink, A.P., van Meurs, J.B., van Leeuwen, H. P., and Pols, H. A. (2002). The role of vitamin D receptor gene polymorphisms in bone biology. *Mol. Cell. Endocrinol.* 197, 15-21.

Weiss, S.T., Silverman, E.K., and Palmer, L.J. (2001). Case-control association studies in pharmacogenetics. *Pharmacogenomics J.* 1, 157-158.

Wilson, A.G., Symons, J.A., McDowell, T.L., McDevitt, H.O., and Duff, G.W. (1997). Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. *Proc. Natl. Acad. Sci. U S A* 94, 3195-3199.