Genomics & Informatics Vol. 8(3) 103-107, September 2010

KARE Genomewide Association Study of Blood Pressure Using Imputed SNPs

Kyung-Won Hong[1¶], Ji-Eun Lim[1¶], Young-Jin Kim[2], Nam H. Cho[3], Chol Shin[4] and Bermseok Oh[1*]

[1]Department of Biomedical Engineering, School of Medicine, Kyung Hee University, Seoul 130-701, Korea, [2]Center for Genome Science, National Institute of Health, Seoul, Korea, [3]Department of Preventive Medicine, Ajou University School of Medicine, Suwon 443-749, Korea, [4]Department of Internal Medicine, Korea University Ansan Hospital, Ansan 425-707, Korea

Abstract

The imputation of untyped SNPs enables researchers to validate association findings across SNP arrays and also enables them to test a large number of SNPs to reveal the fine structure of the association peak, facilitating interpretation of the results and the location of causal polymorphisms. In this study, we applied the imputation method to a genomewide association study and recapitulated the previously associated gene loci of blood pressure traits in Korean cohorts. A total of 1,827,004 SNPs were imputed by the IMPUTE program, and we conducted a genomewide association study for systolic and diastolic blood pressure. While no SNPs passed the Bonferroni correction p-value ($p=2.74\times10^{-8}$ for 1,827,004 SNPs), 12 novel loci for systolic blood pressure and 16 novel loci for diastolic blood pressure were detected by imputed SNPs, with $10^{-5}<$ p-value $<10^{-4}$. Moreover, 7 regions (ATP2B1, 10p15.1, ARHGEF12, ALX4, LIPC, 7q31.1, and TCF7L2) out of 14 genetic loci that were previously reported revealed that the imputed SNPs had lower p-values than those of genotyped SNPs. Moreover, a nonsynonymous SNP in the CSMD1 gene, one of the 14 genes, was found to be associated with systolic blood pressure ($p<0.05$). These results suggest that the imputation method can facilitate the discovery of novel SNPs as well as enhance the fine structure of the association peak in the loci.

Keywords: blood pressure, genomewide association study, Korea Association Resource, imputation, single nucleotide polymorphism

Introduction

The human genome contains ~10 million single nucleotide polymorphisms (SNPs), but only a small fraction of them has been assayed using current high-density microarrays, such as Illumina and Affymetrix platforms. To enhance an original association study, the genotypes of untyped SNPs can be imputed based on nearby markers and can be tested for association with phenotypes of interest. This strategy enables researchers to easily replicate and compare previous findings across array types and enables them to test a large number of SNPs to reveal the fine structure of the association peak, facilitating interpretation of the results and the location of causal polymorphisms (Hao *et al.*, 2009). By using the imputed SNPs, Tobin and colleagues identified a significant association of the WNT1 gene with diastolic blood pressure, and Knouff and colleagues identified a significant association between the CETP gene and diastolic blood pressure (Knouff *et al.*, 2008; Tobin *et al.*, 2008).

The first genomewide association study for large cohorts in Korea was conducted by the Korea National Institute of Health (KNIH) and reported associations with 8 quantitative traits, including blood pressure (Cho *et al.*, 2009). The Korean GWAS suggested a total of 14 loci for blood pressure, 2 (ATP2B1 and 10 p15.1) of which were significant association loci (p-value $<10^{-5}$); the remaining 12 (ARHGEF12, CSMD1, ARSG, SGSM1, GJA8, ALX4, 2q31.1, 15q22.1, LIPC, CSK, 7q21.13, 7q31.1, 10q25.2) were suggestive loci ($10^{-5}<$ p-value $<10^{-4}$). To enhance the original association study, a total of 1,827,004 imputed SNPs were reanalyzed for their association with blood pressure in this study.

Methods

Subjects and their genotypes were reported in a previous genomewide association study (Cho *et al.*, 2009). Briefly, subjects came from 2 community-based cohorts, the rural community Ansung and the urban community Ansan, in KyungGi-Do province, near Seoul, Korea. Most DNA samples were isolated from the peripheral blood of participants and genotyped using the Affymetrix Genomewide Human SNP array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA). After the quality control steps, we finally used

¶Hong, K.W. and Lim, J.E. contributed equally to this work.
*Corresponding author: E-mail ohbs@khu.ac.kr
Tel +82-2-961-0617, Fax +82-2-961-5515
Accepted 24 August 2010

7751 individuals for blood pressure traits. The 351,677 genotyped SNPs had a missing gene call rate below 0.1, a minor allele frequency (MAF) higher than 0.01, and no deviation from Hardy-Weinberg Equilibrium (HWE) ($p > 1 \times 10^{-6}$) (WTCCC, 2007).

Table 1. Basic characteristics of study subjects

Variables	Count/ mean±standard deviation
Number of individuals	7,751
Gender [men (%)/women (%)]	3,747 (50)/3,804 (50)
Age	51.44±8.79
BMI	24.4±3.2
SBP	115.65±17.25
DBP	74.21±11.27

The basic characteristics and blood pressures of the subjects are listed in Table 1. For this study, blood pressure measurements were taken 3 times in the supine position. Before the first measurement, participants rested for 5 minutes, and 3 measurements were made at least 30 seconds apart. The average of the 3 measurements was used for this study.

The KARE dataset, comprising 351,677 SNPs for 7751 individuals, was merged with that of International HapMap Phase II JPT (Japanese)+HCB (Chinese) panel 2. The genotypes of the KARE individuals were imputed using IMPUTE (Howie *et al*., 2009).

Most statistical analyses were performed using PLINK, version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) (Purcell *et al*., 2007) and SAS (version 9.1; SAS Institute Inc., Cary, NC, USA). Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were tested for their as-

Table 2. Novel SNP loci identified by the imputed SNPs

Locus	Nearby genes	No. of SNPs[a]	p-value's range	
			Min.	Max.
Systolic blood pressure				
1q25.1	CACYBP/MRPS14/TNN/KIAA0040/TNR	1	9.67×10^{-5}	
TNR		1	9.90×10^{-5}	
ALK		1	5.29×10^{-5}	
2q14.2	TMEM185B/RALB/INHBB/LOC84931	1	8.35×10^{-5}	
3p24.3	**ZNF385D**	**1**	$\mathbf{2.06 \times 10^{-5}}$	
3q25.1	HPS3/CP/TM4SF18/TM4SF1/TM4SF4/WWTR1	1	2.83×10^{-5}	
SLIT2		2	8.30×10^{-5}	9.40×10^{-5}
4q25	PAPSS1	1	6.78×10^{-5}	
4q32.3	NPY1R/NPY5R/TKTL2/C4orf43/MARCH1	1	7.55×10^{-5}	
EBF1		2	7.69×10^{-5}	9.90×10^{-5}
SEMA3E		1	9.76×10^{-5}	
8p12	NRG1	1	7.55×10^{-5}	
Diastolic blood pressure				
14q23.2	HIF1A/SNAPC1/SYT16/FLJ43390	1	8.84×10^{-5}	
17q25.3	SEPT9	2	2.36×10^{-5}	2.48×10^{-5}
18q21.33	CDH20/RNF152	1	7.27×10^{-5}	
20q13.13	SULF2	1	9.32×10^{-5}	
2q33.1	SPATS2L/KCTD18	7	3.35×10^{-5}	8.76×10^{-5}
2q37.1	UGT1A1/3/4/5/6/7/8/9/10/DNAJB3/HJURP/MSL3L2/TRPM8/SPP2	6	6.07×10^{-5}	9.70×10^{-5}
3p12.3	ROBO1	1	8.07×10^{-5}	
3p24.3	**ZNF385D**	**1**	$\mathbf{8.26 \times 10^{-5}}$	
5q12.1	PDE4D/PART1/DEPDC1B/ELOVL7/ERCC8/NDUFAF2	1	3.70×10^{-5}	
8p22	SGCZ/TUSC3	1	1.24×10^{-5}	
8p23.1	TDH/C8orf12/FAM167A/BLK/GATA4/NEIL2/FDFT1/CTSB	1	7.19×10^{-5}	
BCR		2	3.62×10^{-5}	4.58×10^{-5}
ELOVL7		1	1.91×10^{-5}	
LOC100188947		1	5.29×10^{-5}	
PTPRT		1	2.48×10^{-5}	
TRPM8		2	1.74×10^{-5}	2.27×10^{-5}

Boldface highlightsthe locus associated with both systolic and diastolic blood pressure, [a]Number of significant imputed SNPs in the locus.

Table 3. Comparison of the association results of genotyped and imputed SNPs in the reported gene regions

Candidate region			Number of SNPs			Top SNPs				LD	
Cho et al., Table 1	Locus	Franking	Total	Genotyped	Imputed	Reported	p-value	Imputed	p-value	r^2	D'
Systolic blood pressure											
ATP2B1	12q21.33	±20 kb	29	14	15	rs17249754	3.4×10^{-7}	rs12579302	2.4×10^{-7}	1.0	1.0
rs715987	10p15.1	±500 kb	967	284	683	rs715987	1.1×10^{-5}	rs10795186	9.3×10^{-6}	1.0	1.0
Diastolic blood pressure											
ATP2B1	12q21.33	±20 kb	29	14	15	rs17249754	3.8×10^{-7}	rs12579302	4.7×10^{-7}	1.0	1.0
Cho *et al.*, Supplementary Table 6											
Systolic blood pressure											
ARHGEF12	11q23.3	±20 kb	81	33	48	rs10790381	5.4×10^{-6}	rs17123861	4.9×10^{-6}	1.0	1.0
CSMD1	8p23.2	±20 kb	3,335	989	2,346	rs995322	1.3×10^{-5}	rs6558796	2.5×10^{-5}	1.0	1.0
ARSG	17q24.2	±20 kb	81	41	40	rs12945290	3.8×10^{-5}	rs12451531	4.2×10^{-5}	0.9	1.0
SGSM1	22q11.23	±20 kb	174	31	143	rs7287595	4.4×10^{-5}	rs5760712	3.3×10^{-2}	0.0	0.1
GJA8	1q21.1	±20 kb	12	5	7	rs7544630	3.1×10^{-5}	rs12028407	4.0×10^{-2}	0.0	0.3
ALX4	11p11.2	±20 kb	60	23	37	rs879238	1.7×10^{-4}	rs7481493	1.0×10^{-5}	0.8	1.0
Diastolic blood pressure											
rs1006815	2q31.1	±500 kb	652	208	444	rs1006815	2.0×10^{-4}	rs1371441	2.4×10^{-4}	1.0	1.0
LIPC	15q22.1	±20 kb	145	73	72	rs11631342	3.4×10^{-5}	rs11636642	2.8×10^{-5}	1.0	1.0
CSK	15q24.1	±20 kb	39	8	31	rs1378942	4.6×10^{-5}	rs4886410	7.2×10^{-5}	1.0	1.0
rs2564438	7q21.13	±500 kb	618	156	462	rs2564438	7.1×10^{-5}	rs4577890	1.3×10^{-4}	0.5	1.0
rs2940371	7q31.1	±500 kb	729	169	560	rs2940371	2.1×10^{-4}	rs7799583	1.1×10^{-4}	0.9	1.0
TCF7L2	10q25.2	±20 kb	76	38	38	rs10787472	3.6×10^{-5}	rs10885405	9.6×10^{-6}	0.4	1.0

Top SNPs: SNPs that have the lowest p-value in the region, as evaluated by linear regression analysis, controlling for area, age, sex, and BMI LD: linkage disequilibrium between reported and imputed Top SNPs.

sociation by linear regression analysis with an additive model (1-d.f.) after adjustments for area, age, sex, and BMI.

Results and Discussion

Using a filtering scheme of INFO ≥0.5 and posterior probability ≥0.9, a total of 1,827,004 SNPs were imputed and analyzed for the genomewide association study of blood pressure traits. While no SNP passed the Bonferroni correction p-value ($p=2.74\times10^{-8}$ for 1,827,004 SNPs), 12 novel loci for systolic blood pressure (SBP) and 16 loci for diastolic blood pressure (DBP) were detected by the imputed SNPs with 10^{-5} < p-value < 10^{-4} (Table 2). Most of the loci had lower than 2 SNPs of significance near the region, except 20q13.13 and 2q33.1, which had 7 and 6 SNPs of significance, respectively. Notably, the SNP (rs1027345) near the ZNF385D gene in the 3p24.3 region was significant for both SBP and DBP ($p=2.06\times10^{-5}$ and 8.26×10^{-5}, respectively).

In order to compare the signal p-values of the imputed and typed SNPs, SNPs that were located in the 14 previously reported genetic loci in the KARE study were investigated (Table 3). A total of 7483 SNPs were extracted from the 1,827,004 imputed SNPs and analyzed for blood pressure. The most significant SNPs and their gene names are described in Table 3. For the extraction, the candidate regions were defined as±20 kb for the gene region and±500 kb for the intergenic region. Among the 14 loci, 7 loci (ATP2B1, 10p15.1, ARHGEF12, ALX4, LIPC, 7q31.1, and TCF7L2) showed enhanced association with blood pressure by the imputed SNPs.

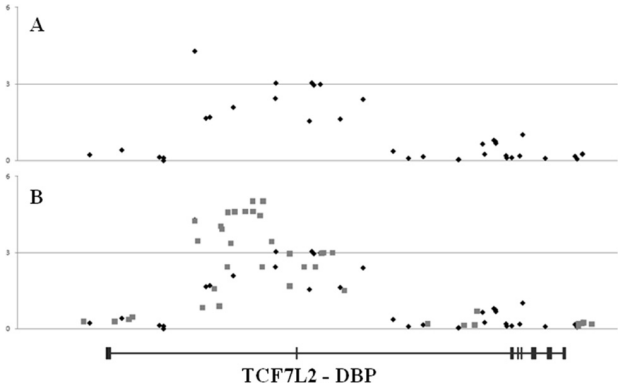

Fig. 1. Signal (-logP) plots of the association between SNPs of the TCF7L2 region and diastolic blood pressure (DBP). A. genotyped SNPs. B. Both genotyped and imputed SNPs.

Imputation methods cover the candidate gene region

Table 4. Nonsynonymous SNPs among the imputed SNPs and the association results

Gene	SNP	Protein Change	Systolic blood pressure		Diastolic blood pressure	
			$\beta \pm$s.e.	p	$\beta \pm$s.e.	p
CSMD1	*rs11984691*	*S3207N*	*2.687±1.105*	*0.015*	*0.847±0.749*	*0.259*
ALX4	rs3824915	R35T	0.007±0.281	0.980	0.147±0.192	0.442
CSK	rs3803568	R105Q	0.506±0.331	0.126	0.319±0.224	0.154
LIPC	rs3829462	F356L	0.104±0.458	0.821	-0.125±0.310	0.687
ARSG	rs3213690	N30D	0.201±0.309	0.515	0.279±0.209	0.184
SGSM1	rs2073201	R873K	-0.172±0.308	0.576	-0.326±0.208	0.118

densely, which enables one to test a large number of SNPs to reveal the fine structure of the association peak. The association results of the TCF7L2 gene could fall under this case (Fig. 1). While only one SNP with a p-value lower than 10^{-4} was present in the peak of the association by the genotyped SNPs, the imputed SNPs showed multiple SNPs (p-value $<10^{-4}$) that were associated around the peak, indicating that the imputation methods enhance the fine structure of the GWAS results. Moreover, 6 nonsynonymous SNPs in the 14 genes were identified from the imputed SNPs, and the rs11984691 of CSMD1 revealed an association signal with systolic pressure (p-value=0.015) (Table 4). *CSMD1* (CUB and Sushi multiple domains 1) is located on chromosome 8p23.2 and is known as a putative suppressor of squamous cell carcinomas of the head and neck (Sun *et al.*, 2001). The gene encodes a 389-KDa transmembrane protein, named for its repeat CUB and sushi domains. Both CUB and sushi domains are found in many other proteins and are thought to be sites of either protein-protein or protein-ligand interactions (Kristiansen *et al.*, 1999; Lau and Scholnick, 2003). The association between the CSMD1 gene and blood pressure was replicated in our previous report (Hong *et al.*, 2010), and the nonsynonymous rs11984691 might be a causative variation of the CSMD1 gene for the regulation of blood pressure, which waits for its functional analysis in the future.

In conclusion, a total of 1,827,004 SNPs were imputed based on International HapMap Asian data, and a genomewide association study of KARE for systolic and diastolic blood pressure was performed. Even though no SNPs were found to be statistically significant over the Bonferroni corrected p-value, 12 novel loci for systolic blood pressure and 16 novel loci for diastolic blood pressure were identified in the range of $10^{-5}<$p-value$<10^{-4}$. Additionally, the imputed SNPs revealed better p-values in 7 out of 14 genetic loci that were previously reported by the genotyped SNPs. These results suggest that the imputation method can facilitate the discovery of associated SNPs as well as enhance the fine structure of the association in the loci.

Acknowledgments

The Consortium for Large Scale Genome Wide Association Study was supported by genotyping data (Genome Wide association analysis of community based cohort study; 2007) from the Korean Genome Analysis Project (4845-301) and the Korea National Institute of Health (Korea Center for Disease Control, Ministry for Health, Welfare and Family Affairs), Republic of Korea.

References

Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M., Cha, S.H., Kim, J.W., Han, B.G., Min, H., Ahn, Y., Park, M.S., Han, H.R., Jang, H.Y., Cho, E.Y., Lee, J.E., Cho, N.H., Shin, C., Park, T., Park, J.W., Lee, J.K., Cardon, L., Clarke, G., McCarthy, M.I., Lee, J.Y., Oh, B., and Kim, H.L. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* 41, 527-534.

Hao, K., Chudin, E., McElwee, J., and Schadt, E.E. (2009). Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. *BMC Genet.* 10, 27.

Hong, K.W., Go, M.J., Jin, H.S., Lim, J.E., Lee, J.Y., Han, B.G., Hwang, S.Y., Lee, S.H., Park, H.K., Cho, Y.S., and Oh, B. (2010). Genetic variations in ATP2B1, CSK, ARSG and CSMD1 loci are related to blood pressure and/or hypertension in two Korean cohorts. *J. Hum. Hypertens.* 24, 367-372.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5, e1000529.

Knouff, C.W., Lim, N., Song, K., Yuan, X., Walker, M.C., Townsend, R., Waeber, G., Matthews, P.M., Vollenweider, P., Waterworth, D.M., and Mooser, V. (2008). Pharmacological effects of lipid-lowering drugs recapitulate with a larger amplitude the phenotypic effects of common variants within their target genes. *Pharmacogenet Genomics* 18, 1051-1057.

Kristiansen, M., Kozyraki, R., Jacobsen, C., Nexo, E., Verroust, P.J., and Moestrup, S.K. (1999). Molecular dissection of the intrinsic factor-vitamin B12 receptor, cubi-

lin, discloses regions important for membrane association and ligand binding. *J. Biol. Chem.* 274, 20540-20544.

Lau, W.L., and Scholnick, S.B. (2003). Identification of two new members of the CSMD gene family small star, filled. *Genomics* 82, 412-415.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575.

Sun, P.C., Uppaluri, R., Schmidt, A.P., Pashia, M.E., Quant, E.C., Sunwoo, J.B., Gollin, S.M., and Scholnick, S.B. (2001). Transcript map of the 8p23 putative tumor suppressor region. *Genomics* 75, 17-25.

Tobin, M.D., Timpson, N.J., Wain, L.V., Ring, S., Jones, L.R., Emmett, P.M., Palmer, T.M., Ness, A.R., Samani, N.J., Smith, G.D., and Burton, P.R. (2008). Common variation in the WNK1 gene and blood pressure in childhood: the Avon Longitudinal Study of Parents and Children. *Hypertension* 52, 974-979.

WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-78.