ARTICLE

Genomics & Informatics
Vol. 9(4) 152-160, December 2011

http://dx.doi.org/10.5808/GI.2011.9.4.152
pISSN 1598-866X eISSN 2234-0742

Genome-wide Survey of Copy Number Variants Associated with Blood Pressure and Body Mass Index in a Korean Population

Sanghoon Moon, Young Jin Kim, Yun Kyoung Kim, Dong-Joon Kim, Ji Young Lee, Min Jin Go, Young-Ah Shin, Chang Bum Hong and Bong-Jo Kim*

Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, Osong 363-951, Korea

Abstract

Hypertension is the major factor of most death and high blood pressure (BP) can lead to stroke, myocardial infarction and cardiac failure. Moreover, hypertension is strongly correlated with body mass index (BMI). Although the exact causes of hypertension are still unclear, some of genetic loci were discovered from genome-wide association study (GWAS). Therefore, it is essential to study genetic variation for finding more genetic factor affecting hypertension. The purpose of our study is to conduct a CNV association study for hypertension-related traits, BP and BMI, in Korean individuals. We identified 2,206 CNV regions from 3,274 community-based Korean participants using the Affymetrix Genome-Wide Human SNP Array 6.0 platform and performed a logistic regression analysis of CNVs with two hypertension-related traits, BP and BMI. Moreover, the 4,692 participants in an independent cohort were selected for respective replication analyses. GWAS of CNV identified two loci encompassing previously known hypertension-related genes: *LPA* (lipoprotein) on 6q26, and *JAK2* (Janus kinase 2) on 9p24, with suggestive p-values (0.0334 for *LPA* and 0.0305 for *JAK2*). These two positive findings, however, were not evaluated in the replication stage. Our result confirmed the conclusion of CNV study from the WTCCC suggesting weak association with common diseases. This is the first study of CNV association study with BP and BMI in Korean population and it provides a state of CNV association study with common human diseases using SNP array.

*Corresponding author: E-mail kbj6181@cdc.go.kr
Tel +82-43-719-8881, Fax +82-43-719-8908
Received 11 November 2011, Revised 28 November 2011, Accepted 9 December 2011

Keywords: copy number variation, hypertension-related traits, BMI, blood pressure, genome-wide association study

Introduction

Copy number variation (CNV) is defined as a loss or gain of long segments of DNA, and it is relatively common in the human genome (Freeman *et al.*, 2006; McCarroll *et al.*, 2008; Redon *et al.*, 2006). Because some chromosomal duplications and deletions are known to be associated with disease susceptibility and gene dosage, CNVs are considered major sources of genomic variation along with single nucleotide polymorphisms (SNPs) (Freeman *et al.*, 2006; McCarroll *et al.*, 2008). A CNV study has reported population-dependent CNVs that inconsistently occur between ethnic groups (Redon *et al.*, 2006).

To date, a number of genomic variants associating SNP loci with susceptibility to common complex human diseases have been identified through genome-wide association studies (GWASs) (The WTCCC, 2007; Estivill and Armengol, 2007; Hong *et al.*, 2009; Cho *et al.*, 2009). Although some CNV studies have also reported associations with several complex disorders, the corresponding CNV data are scarce compared with SNP data. Moreover, previous GWASs reported associations of SNPs with anthropometric traits including blood pressure (BP) and body mass index (BMI) in the Korean population (Hong *et al.*, 2009; Cho *et al.*, 2009), but no CNV association studies with these traits have been reported.

The aims of our current study were as follows: 1) discover Korean CNV regions (CNVRs) in large-scale cohort samples; 2) examine additional loci that may be associated with hypertension-related traits; 3) compare association results of CNVs with those of previously known risk loci. Toward these goals, we performed a three-stage analysis. In the first stage, we conducted a genome-wide survey of CNVs in 3,274 healthy Korean participants as part of the Korean Genome Epidemiology Study (KoGES) using the Affymetrix Genome-Wide Human SNP Array 6.0 platform. Using stringent quality control, we ascertained 2,206 CNVRs. In the second stage, we carried out a logistic regression analysis of CNVs with two hypertension-related traits, BP and BMI.

Of the 3,274 participants in this study, 1,725 (602 cases, 1,123 controls) were selected for association analysis with BP; for BMI, 3,212 participants (1,112 cases, 2,100 controls) were similarly selected. In the third stage, of the 4,692 participants genotyped with the Affymetrix Genome-Wide Human SNP array 5.0 platform in an independent cohort, we selected 3,507 (1,561 cases, 1,946 controls) and 4,592 (2,095 cases, 2,497 controls) participants for inclusion in the respective replication analysis datasets for these two CNVRs.

In conclusion, our result is consistent with that of the Wellcome Trust Case Control Consortium (WTCCC) CNV study (2010). Thus, our result confirmed the conclusion of WTCCC CNV study suggesting weak association with complex traits. Although we also found no Asian-specific CNVs associated with hypertension-related traits, our study is the first to examine the contribution of CNVs to common complex diseases in a Korean population and will provide new insight into the genetics associated with complex traits having underlying CNVs in the Korean population.

Methods

Study participants

Study participants were enrolled from a population-based cohort called the Health Examinee (HEXA) cohort, which is a part of KoGES. DNA was extracted from peripheral blood with informed consent. Through each step of SNP quality control procedures (sample call rate $\geq$ 95%, gender inconsistency check, exclusion of cancer patients, population stratification check, etc.), the 3,703 healthy individuals (aged between 40 and 70 years) were genotyped on the Affymetrix Human Genome-Wide SNP array 6.0. This dataset was composed of 44.6% male and 55.4% female samples. Files with raw signal intensity (CEL files) from the 3,703 samples were converted to normalized log R intensity ratios.

CNV detection stage

Pooled reference

For the Affymetrix Genome-Wide Human SNP array 6.0, pre-processing procedures (background subtraction, normalization, summarization of probe set) were performed with the apt-probeset-summarize application (Affymetrix Power Tools; http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx). After pre-processing procedures, the signal intensity ratio between the test and the reference sample of each probe was transformed to $\log_2$ scale with the chromosomal coordinate of the probes (UCSC version hg18/ NCBI Build 36). The technical artifact, spatial autocorrelation or 'wave', which occurs in a large dataset used to determine the location of CNV across the genome, were removed by WaveNorm software (Marioni *et al*., 2007).

DNA from a single individual such as NA10851 or NA15510 was used for CNV discovery in some reported CNV analyses (Redon *et al*., 2006; Komura *et al*., 2006; Olsen and Venkatraman, 2004). However, the probability that the reference CNV region was homozygously deleted may lead to detection of false-positive CNVs (Marioni *et al*., 2007; Komura *et al*., 2006). Therefore, we assumed that the reference effect of a homozygously deleted CNV region could be minimized by using the pooled reference signal intensity. Finally, we used the average intensity of 100 samples that were randomly selected from among 3,703 cohort participants rather than a single individual's intensity (Fig. 1).

CNV detection

To increase the robustness of CNV detection, two different algorithms, DNAcopy (Olsen and Venkatraman, 2004) and the Genome Alteration Detection Analysis (GADA) (Pique-Regi *et al*., 2008), were used. To define the threshold, different threshold Ts were tested from 3 to 4.5. Finally, we ran DNAcopy by default and the GADA R-package with T=4.5, α=0.2, and MinSegLen=6 on the 3,603 individuals (excluding 100 samples for pooled reference). In addition, the average $\log_2$ ratio for defining CNV segments was set to $\pm$0.3. Eventually, we chose 3,274 samples with both a median of the absolute value of all pairwise differences (MAPD) of $<$0.3 and the number of CNV segments within the outlier detection limits (Q1$-$1.5$\times$IQR$\sim$Q3+1.5$\times$IQR) (see supplementary methods: participant pruning; supplementary Fig. S2). After subsequent stringent quality-control processes, 2,206 CNVRs were used for further studies (see supplementary methods: CNV segment selection and CNVR; supplementary Fig. S1; supplementary Fig. S3).

Disease association stage

The basic characteristics of the participants are listed in Table 1. In this study, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were taken from the averages of three recumbent measurements. The 602 cases were defined as having SBP$\geq$140 mmHg or DBP $\geq$90 mmHg. The 1,123 controls were defined as having SBP$<$120 mmHg and DBP$<$80 mmHg. Average SBP and DBP of study participants were 119.15$\pm$18.25 and 75.85$\pm$12.70 mmHg, respectively (Table 1). In addition, we defined BMI as the individual' s weight in kilograms divided by the height in meters squared, according to the World Health Organization BMI thresholds. The

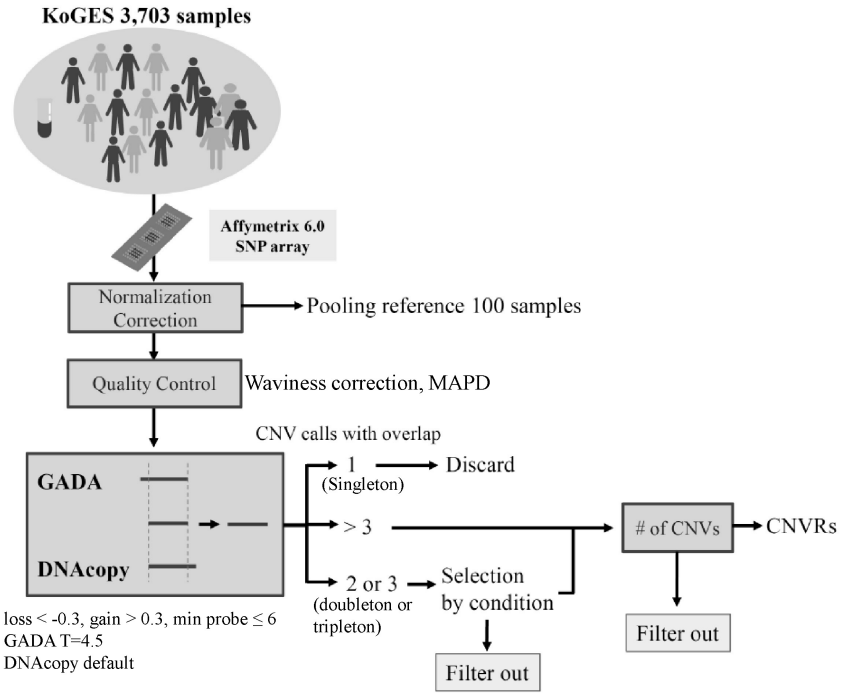

Fig. 1. Overall scheme for CNV detection. The average intensity of 100 samples was used for the reference signal intensity rather than that of the single individual. To exclude erroneous CNV signals, we applied a stringent rule from singletons to tripletons. All detected singletons were discarded. For doubletons and tripletons, overlapping regions were considered a CNVR only if >90% of the CNV segment of each individual reciprocally overlapped with each other (see supplementary Fig. S3). For CNV segments occurring in more than four individuals, CNV segments with any sized overlapping region were considered a CNVR.

Table 1. Basic characteristics of study participants

Variable	HEXA cohort			Ansung-Ansan cohort	
	Total	BMI[a]	BP	BMI	BP
N	3,274	3,212	1,725	4,592	3,507
Control/Case	-	2,100/1,112	1,123/602	2,497/2,095	1,946/1,561
Age (years)	53.23±8.37	53.24±8.35	52.54±8.26	53.95±9.01	53.55±9.04
Gender (male/female)	1,470/1,804	1,444/1,768	708/1,017	2,145/2,447	1,664/1,843
BMI (kg/m^2)	23.95±2.88	24.08±2.76	23.77±2.90	24.85±3.06	24.72±3.20
SBP (mmHg)	121.76±14.25	121.80±14.27	119.15±18.25	119.07±18.63	116.66±20.76
DBP (mmHg)	77.18±9.85	77.20±9.86	75.85±12.70	76.30±11.83	74.84±13.07

[a]The body mass index (BMI) is the weight in kilograms divided by the square of the height in meters.
HEXA, Health Examinee (HEXA) cohort; BP, blood pressure; SBP, systolic blood pressure; DBP, diastolic blood pressure.

1,112 cases were defined as having a BMI≥25 kg/m^2, and the 2,100 controls were defined having 18.5≤BMI <25 kg/m^2. The average BMI of all study participants was 24.08±2.76 (Table 1). For the regression analysis, copy number deletion, normal and duplication were encoded to −1, 0 and +1, respectively. Logistic regression analyses were conducted by SAS software, and all models were adjusted for gender and age.

Replication stage

Two positive CNV loci were evaluated in an independent cohort. For replication analysis, we used 4,692 participants from the Ansung-Ansan cohort, which is also a part of KoGES. These participants were genotyped with the Affymetrix Genome-Wide Human SNP Array 5.0 and passed each step of the SNP quality control procedures (sample call rate ≥96%, heterozygosity ≥30%, gender inconsistency check, exclusion of patients with tumors,

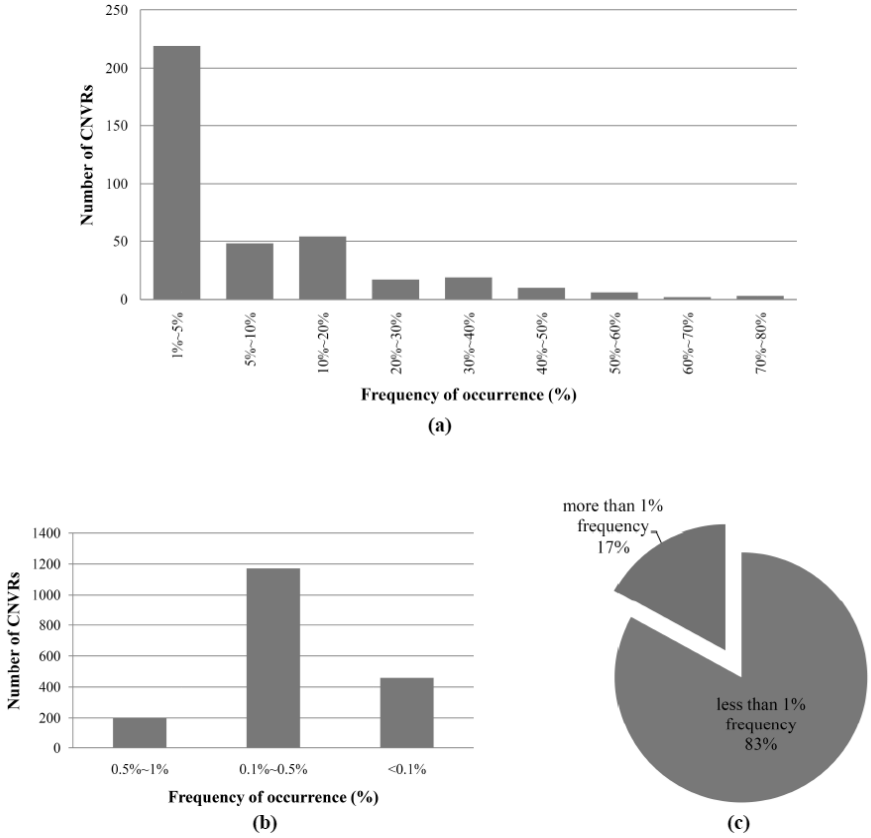

Fig. 2. CNV frequencies for 2,206 detected CNVRs in 3,274 Korean individuals. We surveyed the occurrence of each region above and below a frequency rate of 1%. (a) Among the frequencies >1%, the majority of CNVRs had frequencies of 1～5% (uncommon variant). (b) Of those with a frequency of <1%, variants with a frequency of 0.1～0.5% were the most abundant owing to our stringent quality control rule. (c) Approximately four times as many CNVRs with a frequency of <1% were seen compared to CNVRs with a frequency of >1%.

and population stratification check). Using the same definition as for the association stage, we selected 3,507 (1,561 cases, 1,946 controls) and 4,592 (2,095 cases, 2,497 controls) individuals for BP and BMI, respectively. The DNAcopy was used for CNV detection. Logistic regression analyses were also conducted.

Results

Characteristics of CNVs and CNVRs

As mentioned above, 145,689 of 163,829 CNV segments were selected for further study. These segments consisted of 53,817 gains and 91,872 losses. Finally, 2,206 CNVRs were defined from among these CNV segments. Of these CNVRs, 69.2% (1,527 of 2,206) overlapped by ≥1 bp with 14,411 CNVRs in the Database of Genomic Variants (DGVs; http://projects.tcag.ca/variation/). Conversely, 10.7% (1,538 of 15,411) of CNVRs in DGV matched with our results (see supplementary Fig. S4). The average and median length of all CNVRs was 72.1 kb and 23.0 kb, respectively (see supplementary Table 1). Moreover, an average of 44.5 CNV segments was found in each individual (see supplementary Fig. S5).

Frequency analysis

We examined the distribution of the 2,206 CNVRs by frequency of occurrence. Fig. 2A and Fig. 2B show the distribution of CNVRs with a frequency of more than 1% and less than 1%, respectively. Among the CNVRs with a frequency of >1%, most of the detected CNVRs had a frequency of 1～5% (Fig. 2A). On the other hand, because we excluded all singleton CNV segments as well as some of the doubletons and tripletons according to the stringent rule (see supplementary methods: CNV region), most CNVRs with a frequency of <1% were included in a frequency range of 0.1～0.5% (Fig. 2B). The

Table 2. Logistic regression results of CNVRs associated with hypertension-related traits (p-value < 0.05, adjusted for age and gender)

	Chromosome	Start	End	Status	p-value	OR[a]	CI[b] (95%)		Overlapping Gene (s)
BP	chr22	16272371	16292154	GAIN	0.0009	5.31	1.98	14.25	*CECR2*
	chr14	105960424	106019613	GAIN	0.0066	2.04	1.22	3.40	
	chr9	66786380	67848642	GAIN/LOSS	0.0078	2.14	1.22	3.75	*ANKRD20A1, ANKRD20A3, AQP7P1*
	chr2	46542552	46559652	LOSS	0.0203	3.36	1.21	9.33	
	chr4	18697503	18736911	LOSS	0.0203	0.44	0.22	0.88	
	chr9	4513668	4527834	LOSS	0.0203	0.60	0.39	0.93	*SLC1A1*
	chr3	36228796	36312767	LOSS	0.0239	3.62	1.19	11.07	
	chr9	36431	182129	GAIN/LOSS	0.0251	3.92	1.19	12.98	*CBWD1, CBWD5, FOXD4*
	chr5	32130968	32208250	GAIN	0.0324	1.78	1.05	3.01	*GOLPH3, PDZD2*
	chr6	160951451	160988957	GAIN/LOSS	0.0334	0.50	0.27	0.95	*LPA*
	chr8	92178937	92255245	LOSS	0.0431	3.09	1.04	9.23	*LRRC69*
	chr4	63349523	63377599	GAIN/LOSS	0.0432	0.81	0.66	0.99	
	chr15	31687811	31699225	LOSS	0.0436	0.53	0.28	0.98	*RYR3*
	chr17	36359072	36420182	GAIN	0.0445	0.21	0.05	0.96	*KRT39, KRT-40, KRTAP3-1, KRTAP3-2, KRTAP3-3*
BMI	chr2	24448832	24473356	LOSS	0.0074	0.61	0.43	0.88	
	chr2	49386546	49634617	LOSS	0.0082	0.63	0.45	0.89	
	chr5	5457366	5505883	GAIN	0.0110	0.53	0.32	0.86	*KIAA0947*
	chr14	71871968	71884317	LOSS	0.0137	0.65	0.46	0.92	*RGS6*
	chr2	52605073	52646886	GAIN/LOSS	0.0145	1.16	1.03	1.30	
	chr16	16197996	16930836	GAIN/LOSS	0.0208	0.64	0.43	0.93	*ABCC6, LOC339047, NOMO3*
	chr17	38510611	38512906	GAIN/LOSS	0.0212	2.07	1.11	3.86	*BRCA1*
	chr1	232762674	232779235	LOSS	0.0219	2.11	1.11	4.01	
	chr14	25786353	25884384	LOSS	0.0269	0.50	0.27	0.92	
	chr9	5055372	5074708	GAIN/LOSS	0.0305	1.90	1.06	3.40	*JAK2*
	chr11	67225909	67526847	LOSS	0.0307	1.60	1.04	2.46	*UNC93B1*
	chr7	19136421	19147444	LOSS	0.0375	1.89	1.04	3.44	
	chr2	129342617	129369760	GAIN/LOSS	0.0409	0.82	0.68	0.99	
	chr15	41632795	41801547	GAIN/LOSS	0.0459	0.58	0.34	0.99	*CATSPER2, CKMT1A, CKMT1B, HISPPD2A, STRC*

[a]OR, Odds ratio.
[b]CI, confidence interval.

number of CNVRs with a frequency of <1% was approximately four times as many as the number with a frequency of >1% (Fig. 2C).

Gene ontology analysis

We surveyed RefSeq genes partially or entirely harboring 2,206 CNVRs and found 1,556 genes. The PANTHER classification system (Thomas *et al.*, 2003) was used to assess the functional implications of these CNVRs. PANTHER results showed genes involved in metabolic processes, cellular processes and communication, immune system processes, etc (see supplementary Table 2).

Disease association and replication study

We tested the association between CNVs and two kinds of applicable anthropometry traits, BP and BMI. In all, 14 CNVRs that overlapped by ≥1 bp with 18 genes were significantly associated with BP and hypertension with p-value < 0.05. Regarding an association between BMI and CNV, 14 CNVRs that overlapped by ≥1 bp with 13 genes showed association, with p-value < 0.05 (Table 2). To compare all 31 genes with previously reported genetic associations, we used Genopedia in HuGE Navigator (Yu *et al.*, 2008). Two CNVRs encompassing known hypertension-related genes, *LPA* and *JAK2*, were observed with suggestive p-values (0.0334 for *LPA* and 0.0305 for *JAK2*). CNV loci overlapping *LPA* were registered in DGV. These two CNVRs were selected for further replication studies.

For the replication stage, CNV loci overlapping *LPA* and *JAK2* were observed with non-significant p-values of 0.8687 (odds ratio: 1.041, 95% confidence interval: 0.644~1.684) and 0.2782 (odds ratio: 0.775, 95% confidence interval: 0.489~1.229), respectively (Fig. 3).

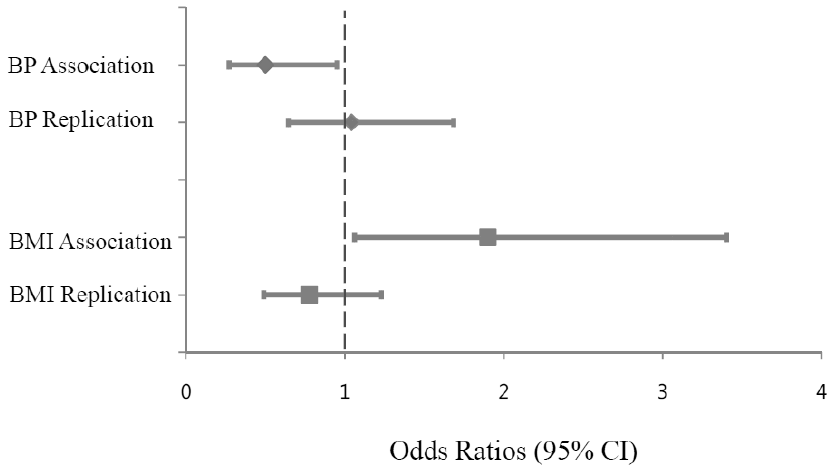

Fig. 3. Odds ratios of two positive findings (*LPA* and *JAK2*) in the association stage with 95% confidence intervals (CI).

Discussion

Hypertension refers to a clinically significant increase in BP, and it is an important risk factor for cardiovascular disease; further, excess body weight is a major risk factor for hypertension (The WTCCC, 2007). Development of hypertension, however, hypertension has a multifactorial etiology, including both genetic and environmental factors. Therefore, the distinctive causes and factors of hypertension are still not fully understood.

In this study, we found that two CNVRs overlapped with previously known risk loci, namely *LPA* on 6q26 and *JAK2* on 9p24. Some association studies have been reported regarding these loci. For example, common variants in *LPA* are associated with myocardial infarction (Deo *et al*., 2011). In particular, higher copy numbers of the kringle (IV) domain in this protein are associated with lower lipoprotein levels (van der Hoek *et al*., 1993). Moreover, Wong *et al*. (2007) conducted a CNV analysis using a whole-genome array comparative genomic hybridization (aCGH). Among the detected CNVs, they found CNV loci overlapping *LPA*. Association of BP with *JAK2* has also been reported. Guilluy *et al*. (2010) showed that inhibition of *JAK2* prevents angiotensin II-mediated hypertension in wild-type mice.

In contrast with the results of the association stage, the two positive findings were not confirmed in the replication stage. CNV loci overlapping *LPA* and *JAK2* were observed with nonsignificant p-values. In the association stage, other well-established risk loci such as *ATP2B1, FTO, MC4R, PPYR1, ST7L-CAPZA1, FIGN-GRB14, ENPEP, NPR3*, and *TBX3* that were observed in previous GWAS or CNV studies (Sha *et al*., 2009; Levy *et al*., 2009; Newton-Cheh *et al*., 2009; Kato *et al*., 2011; Frayling *et al*., 2007; Loos *et al*., 2008; Glessner *et al*., 2010; Dorajoo *et al*., 2011) were not observed in our current study. There are some potential reasons for why these loci were not significantly observed in our study. First, risk loci affecting hypertension may differ when studying CNVs and SNPs. Alternatively, rare CNV variants may affect the etiology of hypertension, or hypertension may have few common risk alleles that each has a relatively large effect. Finally, some of the controls may have been misclassified as cases owing to limitations of current CNV genotyping technology.

Similar results supporting our study have been reported. The WTCCC performed an association study of common CNVs with eight common diseases including hypertension using the original WTCCC study samples (WTCCC, 2010) and identified three CNV loci that were associated with common diseases. They concluded, however, that none of the three loci was believed to be a functional variant and that common CNVs may not play a major role in the genetic basis of common diseases (WTCCC, 2010). Similar to the conclusion of WTCCC, few association studies of complex diseases with CNVs as newly discovered genetic markers have been reported. For example, Charchar *et al*. (2010) conducted a genome-wide survey of CNVs in the spontaneously hypertensive rat; they identified and validated 16 CNVs associated with the spontaneously hypertensive phenotype. In contrast with the association in the spontaneously hypertensive rat, Johnson *et al*. (2009) found no association of CNVs in *BMPR2* with pulmonary arterial hypertension in human. Our data are consistent with those of the WTCCC CNV study, which suggests that functional variants and common CNVs may not play a major role in the genetic basis of common diseases (WTCCC, 2010).

Moreover, the absence of a positive result in the rep-

lication stage of our study may be due to the frequency of occurrence. That is, because recurrence rates of uncommon (or less common) CNV loci with frequencies of 1～5% may vary between populations, our two positive findings (2.9% for *LPA* and 1.9% for *JAK2*) may not be replicated in the association stage. Alternatively, similar to the association stage, the limitation of platform resolution may lead to replication failure. Consequently, to resolve obstacles underlying a CNV association study such as the complexity of the genetic architecture for hypertension and the limitation of the current CNV genotyping techniques, a more comprehensive study with a higher-resolution platform will be necessary to identify genetic variants that affect hypertension. There are some limitations associated with our approach. As we used the Affymetrix 5.0 for the replication stage, we could not address whether two signals were not replicated or were false positive signals. Therefore, validation experiment such as quantitative PCR (qPCR) is needed for the accurate analysis.

In conclusion, our result is consistent with that of the WTCCC CNV study. Thus, our study confirmed the conclusion of WTCCC CNV study suggesting no association with complex traits. In addition, this is the first CNV association to examine the contribution of CNVs to hypertension in a Korean population. Therefore, our study will also provide new insight into the genetics associated with complex traits having CNVs in the Korean population.

Supplementary materials

Supplementary data including a figure and two tables can be found with this article online at http://www.genominfo.org/html/UploadFile/article2_201112_SP.pdf.

Acknowledgements

This work was supported by intramural grant from the Korea National Institute of Health (2008-N73002-00) and grants from the Korea Centers for Disease Control and Prevention (4845-301, 4851-302, 4851-307).

References

Charchar, F.J., Kaiser, M., Bingham, A.J., Fotinatos, N., Ahmady, F., Tomaszewski, M., and Samani, N.J. (2010), Whole-genome survey of copy number variation in the Spontaneously Hypertensive Rat: Relationship to quantitative trait loci, gene expression, and blood pressure. *Hypertension* 55, 1231-1238.

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., Lee, J.K., 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.

Deo, R.C., Wilson, J.G., Xing, C., Lawson, K., Kao, W.H., Reich, D., Tandon, A., Akylbekova, E., Patterson, N., Mosley, T.H., Jr., Boerwinkle, E., and Taylor, H.A., Jr. (2011). Single-nucleotide polymorphisms in LPA explain most of the ancestry-specific variation in Lp(a) levels in African Americans. *PLoS One* 6, e14581.

Dorajoo, R., Blakemore, A.I., Sim, X., Ong, R.T., Ng, D.P., Seielstad, M., Wong, T.Y., Saw, S.M., Froguel, P., Liu, J., and Tai, E.S. (2011). Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int. J. Obes.* [Epub ahead of print].

Estivill, X. and Armengol, L. (2007). Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 3, 1787-1799.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., Shields, B., Harries, L.W., Barrett, J.C., Ellard, S., Groves, C.J., Knight, B., Patch, A.M., Ness, A.R., Ebrahim, S., Lawlor, D.A., Ring, S.M., Ben-Shlomo, Y., Jarvelin, M.R., Sovio, U., Bennett, A.J., Melzer, D., Ferrucci, L., Loos, R.J., Barroso, I., Wareham, N.J., Karpe, F., Owen, K.R., Cardon, L.R., Walker, M., Hitman, G.A., Palmer, C.N., Doney, A.S., Morris, A.D., Smith, G.D., Hattersley, A.T., and McCarthy, M.I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316, 889-894.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., Carter, N.P., Scherer, S.W., and Lee, C. (2006). Copy number variation: new insights in genome diversity. *Genome Res.* 16, 949-961.

Glessner, J.T., Bradfield, J.P., Wang, K., Takahashi, N., Zhang, H., Sleiman, P.M., Mentch, F.D., Kim, C.E., Hou, C., Thomas, K.A., Garris, M.L., Deliard, S., Frackelton, E.C., Otieno, F.G., Zhao, J., Chiavacci, R.M., Li, M., Buxbaum, J.D., Berkowitz, R.I., Hakonarson, H., and Grant, S.F. (2010). A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *Am. J. Hum. Genet.* 87, 661-666.

Guilluy, C., Bregeon, J., Toumaniantz, G., Rolli-Derkinderen, M., Retailleau, K., Loufrani, L., Henrion, D., Scalbert, E., Bril, A., Torres, R.M., Offermanns, S., Pacaud, P., and Loirand, G. (2010). The Rho exchange factor Arhgef1 mediates the effects of angiotensin II on vascular tone and blood pressure. *Nat. Med.* 16, 183-190.

Hong, K.W., Jin, H.S., Cho, Y.S., Lee, J.Y., Lee, J.E., Cho, N.H., Shin, C., Lee, S.H., Park, H.K., and Oh, B. (2009). Replication of the Wellcome Trust genome-wide association study on essential hypertension in a Korean population. *Hypertens. Res.* 32, 570-574.

Johnson, J.A., Vnencak-Jones, C.L., Cogan, J.D., Loyd, J.E., and West, J. (2009). Copy-number variation in BMPR2 is not associated with the pathogenesis of pul-

monary arterial hypertension. *BMC Med. Genet.* 10, 58.
Kato, N., Takeuchi, F., Tabara, Y., Kelly, T.N., Go, M.J., Sim, X., Tay, W.T., Chen, C.H., Zhang, Y., Yamamoto, K., Katsuya, T., Yokota, M., Kim, Y.J., Ong, R.T., Nabika, T., Gu, D., Chang, L.C., Kokubo, Y., Huang, W., Ohnaka, K., Yamori, Y., Nakashima, E., Jaquish, C.E., Lee, J.Y., Seielstad, M., Isono, M., Hixson, J.E., Chen, Y.T., Miki, T., Zhou, X., Sugiyama, T., Jeon, J.P., Liu, J.J., Takayanagi, R., Kim, S.S., Aung, T., Sung, Y.J., Zhang, X., Wong, T.Y., Han, B.G., Kobayashi, S., Ogihara, T., Zhu, D., Iwai, N., Wu, J.Y., Teo, Y.Y., Tai, E.S., Cho, Y.S., and He, J. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* 43, 531-538.
Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E., Lee, C., Scherer, S.W., Jones, K.W., Shapero, M.H., Huang, J., and Aburatani, H. (2006). Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res.* 16, 1575-1584.
Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., Aulchenko, Y., Lumley, T., Kottgen, A., Vasan, R.S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D.E., Mitchell, G.F., Mattace-Raso, F.U., Smith, A.V., Taylor, K., Scharpf, R.B., Hwang, S.J., Sijbrands, E.J., Bis, J., Harris, T.B., Ganesh, S.K., O'Donnell, C.J., Hofman, A., Rotter, J.I., Coresh, J., Benjamin, E.J., Uitterlinden, A.G., Heiss, G., Fox, C.S., Witteman, J.C., Boerwinkle, E., Wang, T.J., Gudnason, V., Larson, M.G., Chakravarti, A., Psaty, B.M., and van Duijn, C.M. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 41, 677-687.
Loos, R.J., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., Inouye, M., Freathy, R.M., Attwood, A.P., Beckmann, J.S., Berndt, S.I., Prostate, L.C., Ovarian Cancer Screening, T., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Bergmann, S., Bennett, A.J., Bingham, S.A., Bochud, M., Brown, M., Cauchi, S., Connell, J.M., Cooper, C., Smith, G.D., Day, I., Dina, C., De, S., Dermitzakis, E.T., Doney, A.S., Elliott, K.S., Elliott, P., Evans, D.M., Sadaf Farooqi, I., Froguel, P., Ghori, J., Groves, C.J., Gwilliam, R., Hadley, D., Hall, A.S., Hattersley, A.T., Hebebrand, J., Heid, I.M., KORA, Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Herrera, B., Hinney, A., Hunt, S.E., Jarvelin, M.R., Johnson, T., Jolley, J.D., Karpe, F., Keniry, A., Khaw, K.T., Luben, R.N., Mangino, M., Marchini, J., McArdle, W.L., McGinnis, R., Meyre, D., Munroe, P.B., Morris, A.D., Ness, A.R., Neville, M.J., Nica, A.C., Ong, K.K., O'Rahilly, S., Owen, K.R., Palmer, C.N., Papadakis, K., Potter, S., Pouta, A., Qi, L., Nurses' Health, S., Randall, J.C., Rayner, N.W., Ring, S.M., Sandhu, M.S., Scherag, A., Sims, M.A., Song, K., Soranzo, N., Speliotes, E.K., Diabetes Genetics, I., Syddall, H.E., Teichmann, S.A., Timpson, N.J., Tobias, J.H., Uda, M., Sardi, N.I.A.S., Vogel, C.I., Wallace, C., Waterworth, D.M., Weedon, M.N., Wellcome Trust Case Control, C., Willer, C.J., Fusion, Wraight, Yuan, X., Zeggini, E., Hirschhorn, J.N., Strachan, D.P., Ouwehand, W.H., Caulfield, M.J., Samani, N.J., Frayling, T.M., Vollenweider, P., Waeber, G., Mooser, V., Deloukas, P., McCarthy, M.I., Wareham, N.J., Barroso, I., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Kraft, P., Hankinson, S.E., Hunter, D.J., Hu, F.B., Lyon, H.N., Voight, B.F., Ridderstrale, M., Groop, L., Scheet, P., Sanna, S., Abecasis, G.R., Albai, G., Nagaraja, R., Schlessinger, D., Jackson, A.U., Tuomilehto, J., Collins, F.S., Boehnke, M., and Mohlke, K.L. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* 40, 768-775.
Marioni, J.C., Thorne, N.P., Valsesia, A., Fitzgerald, T., Redon, R., Fiegler, H., Andrews, T.D., Stranger, B.E., Lynch, A.G., Dermitzakis, E.T., Carter, N.P., Tavare, S., and Hurles, M.E. (2007). Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* 8, R228.
McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., and Altshuler, D. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40, 1166-1174.
Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., Papadakis, K., Voight, B.F., Scott, L.J., Zhang, F., Farrall, M., Tanaka, T., Wallace, C., Chambers, J.C., Khaw, K.T., Nilsson, P., van der Harst, P., Polidoro, S., Grobbee, D.E., Onland-Moret, N.C., Bots, M.L., Wain, L.V., Elliott, K.S., Teumer, A., Luan, J., Lucas, G., Kuusisto, J., Burton, P.R., Hadley, D., McArdle, W.L., Wellcome Trust Case Control, C., Brown, M., Dominiczak, A., Newhouse, S.J., Samani, N.J., Webster, J., Zeggini, E., Beckmann, J.S., Bergmann, S., Lim, N., Song, K., Vollenweider, P., Waeber, G., Waterworth, D.M., Yuan, X., Groop, L., Orho-Melander, M., Allione, A., Di Gregorio, A., Guarrera, S., Panico, S., Ricceri, F., Romanazzi, V., Sacerdote, C., Vineis, P., Barroso, I., Sandhu, M.S., Luben, R.N., Crawford, G.J., Jousilahti, P., Perola, M., Boehnke, M., Bonnycastle, L.L., Collins, F.S., Jackson, A.U., Mohlke, K.L., Stringham, H.M., Valle, T.T., Willer, C.J., Bergman, R.N., Morken, M.A., Doring, A., Gieger, C., Illig, T., Meitinger, T., Org, E., Pfeufer, A., Wichmann, H.E., Kathiresan, S., Marrugat, J., O'Donnell, C.J., Schwartz, S.M., Siscovick, D.S., Subirana, I., Freimer, N.B., Hartikainen, A.L., McCarthy, M.I., O'Reilly, P.F., Peltonen, L., Pouta, A., de Jong, P.E., Snieder, H., van Gilst, W.H., Clarke, R., Goel, A., Hamsten, A., Peden, J.F., Seedorf, U., Syvanen, A.C., Tognoni, G., Lakatta, E.G., Sanna, S., Scheet, P., Schlessinger, D., Scuteri, A., Dorr, M., Ernst, F., Felix, S.B., Homuth, G., Lorbeer, R., Reffelmann, T., Rettig, R., Volker, U., Galan, P., Gut, I.G., Hercberg, S., Lathrop, G.M., Zelenika, D., Deloukas, P., Soranzo, N., Williams, F.M., Zhai, G., Salomaa, V., Laakso, M., Elosua, R., Forouhi, N.G., Volzke, H., Uiterwaal, C.S., van der Schouw, Y.T., Numans, M.E., Matullo, G., Navis, G.,

Berglund, G., Bingham, S.A., Kooner, J.S., Connell, J.M., Bandinelli, S., Ferrucci, L., Watkins, H., Spector, T.D., Tuomilehto, J., Altshuler, D., Strachan, D.P., Laan, M., Meneton, P., Wareham, N.J., Uda, M., Jarvelin, M.R., Mooser, V., Melander, O., Loos, R.J., Elliott, P., Abecasis, G.R., Caulfield, M., and Munroe, P.B. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* 41, 666-676.

Olsen, A.B., and Venkatraman, E.S. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557-572.

Pique-Regi, R., Monso-Varona, J., Ortega, A., Seeger, R.C., Triche, T.J., and Asgharzadeh, S. (2008). Sparse representation and Bayesian detection of genome copy number alterations from microarray data. *Bioinformatics* 24, 309-318.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature* 444, 444-454.

Sha, B.Y., Yang, T.L., Zhao, L.J., Chen, X.D., Guo, Y., Chen, Y., Pan, F., Zhang, Z.X., Dong, S.S., Xu, X.H., and Deng, H.W. (2009). Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J. Hum. Genet.* 54, 199-202.

The Welcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129-2141.

van der Hoek, Y.Y., Wittekoek, M.E., Beisiegel, U., Kastelein, J.J., and Koschinsky, M.L. (1993). The apolipoprotein(a) kringle IV repeats which differ from the major repeat kringle are present in variably-sized isoforms. *Hum. Mol. Genet.* 2, 361-366.

Wellcome Trust Case Control Consortium, Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., Holmes, C., Marchini, J.L., Stirrups, K., Tobin, M.D., Wain, L.V., Yau, C., Aerts, J., Ahmad, T., Andrews, T.D., Arbury, H., Attwood, A., Auton, A., Ball, S.G., Balmforth, A.J., Barrett, J.C., Barroso, I., Barton, A., Bennett, A.J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O.J., Braund, P.S., Bredin, F., Breen, G., Brown, M.J., Bruce, I.N., Bull, J., Burren, O.S., Burton, J., Byrnes, J., Caesar, S., Clee, C.M., Coffey, A.J., Connell, J.M., Cooper, J.D., Dominiczak, A.F., Downes, K., Drummond, H.E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D.M., Evans, G., Eyre, S., Farmer, A., Ferrier, I.N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J.A., Freathy, R.M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C.J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G.A., Hocking, L., Howard, E., Howard, P., Howson, J.M., Hughes, D., Hunt, S., Isaacs, J.D., Jain, M., Jewell, D.P., Johnson, T., Jolley, J.D., Jones, I.R., Jones, L.A., Kirov, G., Langford, C.F., Lango-Allen, H., Lathrop, G.M., Lee, J., Lee, K.L., Lees, C., Lewis, K., Lindgren, C.M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D.C., McArdle, W.L., McGuffin, P., McLay, K.E., Mentzer, A., Mimmack, M.L., Morgan, A.E., Morris, A.P., Mowat, C., Myers, S., Newman, W., Nimmo, E.R., O'Donovan, M.C., Onipinla, A., Onyiah, I., Ovington, N.R., Owen, M.J., Palin, K., Parnell, K., Pernet, D., Perry, J.R., Phillips, A., Pinto, D., Prescott, N.J., Prokopenko, I., Quail, M.A., Rafelt, S., Rayner, N.W., Redon, R., Reid, D.M., Renwick, Ring, S.M., Robertson, N., Russell, E., St Clair, D., Sambrook, J.G., Sanderson, J.D., Schuilenburg, H., Scott, C.E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B.M., Simmonds, M.J., Smyth, D.J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H.E., Stone, M.A., Su, Z., Symmons, D.P., Thompson, J.R., Thomson, W., Travers, M.E., Turnbull, C., Valsesia, A., Walker, M., Walker, N.M., Wallace, C., Warren-Perry, M., Watkins, N.A., Webster, J., Weedon, M.N., Wilson, A.G., Woodburn, M., Wordsworth, B.P., Young, A.H., Zeggini, E., Carter, N.P., Frayling, T.M., Lee, C., McVean, G., Munroe, P.B., Palotie, A., Sawcer, S.J., Scherer, S.W., Strachan, D.P., Tyler-Smith, C., Brown, M.A., Burton, P.R., Caulfield, M.J., Compston, A., Farrall, M., Gough, S.C., Hall, A.S., Hattersley, A.T., Hill, A.V., Mathew, C.G., Pembrey, M., Satsangi, J., Stratton, M.R., Worthington, J., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W., Parkes, M., Rahman, N., Todd, J.A., Samani, N.J., and Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464, 713-720.

Wong, K.K., deLeeuw, R.J., Dosanjh, N.S., Kimm, L.R., Cheng, Z., Horsman, D.E., MacAulay, C., Ng, R.T., Brown, C.J., Eichler, E.E., and Lam, W.L. (2007). A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* 80, 91-104.

Yu, W., Gwinn, M., Clyne, M., Yesupriya, A., and Khoury, M.J. (2008). A navigator for human genome epidemiology. *Nat. Genet.* 40, 124-125.