Journal of Human Genetics (2010) 55, 726–730

www.nature.com/jhg

ORIGINAL ARTICLE

Genome-wide association analysis of copy number variations in subarachnoid aneurysmal hemorrhage

Joon Seol Bae[1], Hyun Sub Cheong[2], Byung Lae Park[2], Lyoung Hyo Kim[2], Tae Joon Park[1], Jason Yongha Kim[1], Charisse Flerida A Pasaje[1], Jin Sol Lee[1], Tailin Cui[3], Ituro Inoue[3] and Hyoung Doo Shin[1,2]

Subarachnoid aneurysmal hemorrhage (SAH) due to cerebral aneurysm rupture is a very serious disease resulting in high mortality rate. It has been known that genetic factors are involved in the risk of SAH. A recent breakthrough in genomic variation called copy number variation (CNV) has been revealed to be involved in risks of human diseases. In this study, we hypothesized that CNVs can predict the risk of SAH. We used the Illumina HumanHap300 BeadChip (317 503 markers) to genotype 497 individuals in a Japanese population. Furthermore, individual CNVs were identified using signal and allelic intensities. The genetic effect of CNV on the risk of SAH was evaluated using multivariate logistic regression controlling for age and gender in 187 common CNV regions (frequency >1%). From a total of 4574 individual CNVs identified in this study (9.7 CNVs per individual), we were able to discover 1644 unique CNV regions containing 1232 genes. The identified variations were validated using visual examination of the genoplot image, overlapping analysis with the Database of Genomic Variants (73.2%), CNVpartition (72.4%) and quantitative PCR. Interestingly, two CNV regions, chr4:153210505–153212191 (deletion, 4q31.3, P=0.0005, P^{corr} (corrected P-value)=0.04) and chr10:6265006–6267388 (duplication, 10p15.1, P=0.0006, P^{corr}=0.05), were significantly associated with the risk of SAH after multiple testing corrections. Our results suggest that the newly identified CNV regions may contribute to SAH disease susceptibility.
Journal of Human Genetics (2010) **55,** 726–730; doi:10.1038/jhg.2010.97; published online 12 August 2010

Keywords: copy number variation; genetic epidemiology; genome-wide CNV association study; genomics; subarachnoid aneurysmal hemorrhage

INTRODUCTION

Subarachnoid aneurysmal hemorrhage (SAH) due to cerebral aneurysm rupture is a very serious disease that affects 10–15 out 100 000 people. This condition accounts for 6–8% of the total stroke cases and the mortality rate of affected individuals has been reported to be around 40–50%.[1,2] It is known that both environmental and genetic factors are related to SAH, but the exact mechanism that accounts for the development of the disease is still unknown.[3,4]

Copy number variation (CNV) is another form of structural variation. It is defined as a large DNA fragment that ranges from 1 kb to several megabases and it contains deletions and duplications.[5,6] CNVs may contribute to disease susceptibility by influencing the gene expression level.[7,8] Recent studies have revealed significant associations between CNVs and susceptibility or resistance to human diseases, including autism, inflammatory autoimmune disorders, lung cancer and schizophrenia.[9–12] Although it has already been acknowledged that CNVs contribute to various complex human diseases, there were no previous attempts to study the relationship between SAH and CNV.

In our previous study, which was focused on the identification of a multiallelic single-nucleotide polymorphism (SNP) marker, we found 597 SNP markers showing a multiallelic CNV genotype, commonly known as the common deletion polymorphism, within the CNV region, and performed a further case–control association analysis.[13] In this study, however, we used consecutive markers to analyze the identified CNV region covering. To find individual CNVs, we used the pennCNV algorithm, a program that incorporates multiple factors, including signal intensity (log R ratio (LRR)) and allelic intensity (B-allele frequency), to investigate on more robust risk-associated CNVs. This is the first genome-wide association study to investigate the relationship between common CNV and SAH.

MATERIALS AND METHODS

Subjects and whole-genome SNP genotyping

All individuals included in this study were of Japanese ethnic origin. The subjects who were recruited from Tokai University donated blood for genetic tests after signing the informed consent that was approved by the institutional review board. The control subjects were screened for not harboring intracranial

[1]Laboratory of Genomic Diversity, Department of Life Science, Sogang University, Shinsu-dong, Mapo-gu, Seoul, Republic of Korea; [2]Department of Genetic Epidemiology, SNP Genetics Inc., WooLim Lion's Valley, 371-28, Gasan-Dong, Geumcheon-Gu, Seoul, Republic of Korea and [3]Division of Molecular Life Science, School of Medicine, Tokai University, Shimokasuya 143, Isehara, Kanagawa, Japan
Correspondence: Professor HD Shin, Laboratory of Genomic Diversity, Department of Life Science, Sogang University, Shinsu-dong, Mapo-gu, Seoul 121-742, Republic of Korea. E-mail: hdshin@sogang.ac.kr or Professor I Inoue, Division of Molecular Life Science, School of Medicine, Tokai University, Shimokasuya 143, Isehara, Kanagawa 259-1193, Japan. E-mail: ituro@is.icc.u-tokai.ac.jp
Received 7 June 2010; revised 13 July 2010; accepted 14 July 2010; published online 12 August 2010

aneurysm using neuroradiological imaging such as digital subtraction angiography,[14] three-dimensional computed tomography angiography or magnetic resonance angiography. There were no recruitment restrictions with regard to gender. The presence of SAH was confirmed by conventional angiography, three-dimensional computed tomography,[15] magnetic resonance angiography or surgical finding, whichever is applicable. The study group consisted of 191 patients with SAH and 282 controls without SAH. The control group consisted of individuals who did not have a smoking history and had negative findings on brain computed tomography or magnetic resonance imaging. The age of those in the patient group and control group ranged from 26 to 84 years (mean age=54.8 years, s.d.=11.0) and from 42 to 92 years (mean age=64.4 years, s.d.=10.4), respectively. Genome-wide SNP genotyping was performed using the Illumina HumanHap300 BeadChip containing 317 503 markers (Illumina, San Diego, CA, USA). Approximately 750 ng of genomic DNA extracted from the blood of all individuals was used to genotype each sample. The assay procedure has been described in our previous study.[13] The overall SNP genotyping call rate in this study was 99.80%, an indication of high-quality data.

Identification and mapping of individual CNVs

The signal intensity (LRR) and allelic intensity (B-allele frequency) ratios of all samples were exported from the Illumina BeadStudio software. Samples that did not satisfy the following criteria were excluded from the study: (i) call rate $>99.0\%$;[16] (ii) number of identified CNVs <100; (iii) s.d.[14] of LRR <0.24, as samples with LRR s.d. >0.24 are of low quality, resulting in false-positive CNVs. To identify individual CNVs, we incorporated multiple factors including LRR, B-allele frequency, marker distance and population frequency of the B allele using pennCNV.[17,18] For verification purposes of the identified CNVs, the CNVpartition program with default criteria (Illumina) was initially used to identify CNVs, after which, results were compared with those obtained using pennCNV. We used karyoview of Ensemble to draw the maps of the identified genomic variations (http://apr2006.archive.ensembl.org/Homo_sapiens/karyoview). The input data of each variation were prepared following this format: chromosome, start position, end position and status. The input data of all variants were fixed to 'paste file content' menu before the map was drawn. In the case of the X chromosome, automatic clustering using Illumina BeadStudio software was performed in males and females. To generate a gender file, an automatic estimation of gender was performed using Illumina BeadStudio 3.2 software. Thereafter, the '-chrX' argument for identifying individual CNVs in the X chromosome was used.

Validation of CNV by visual examination using genoplot image

Illumina BeadStudio software provides visual genoplot images representing signal intensity (*y* axis) and allelic intensity (*x* axis) simultaneously per marker. To validate the existence of identified CNV regions, we visually inspected the consecutive changes in signal intensity and allelic intensity at each genoplot image of the marker.

Validation of CNV region by real-time quantitative PCR

We designed a specific amplification primer set (forward primer sequence: 5′-GTGTCTGAGTGTGGGAACCA-3′; reverse primer sequence: 5′-TCCTGAACT TTGTCCCATCTGAGA-3′) and TaqMan-specific probe (5′-CCAGGCCCATGA TATTCCAC-3′) to validate the existence of the CNV region within chr4:153210505–153212191, the region commonly associated with the risk of SAH. Copy number determination analysis was performed using the ABI Prism 7900 sequence detection system. The *RNaseP* gene was coamplified with the marker, which was then used as an internal standard. Amplification reactions (10 μl) were carried out using 10 ng of template DNA, 1× TaqMan Universal Master Mix buffer (Applied Biosystems, Foster City, CA, USA), 900 nM of each primer and 250 nM of each fluorogenic probe. Thermal cycling was initiated with 2 min incubation at 50 °C, followed by a first denaturation step of 10 min at 95 °C, and then 40 cycles of 15 s at 95 °C and for 1 min at 60 °C. Three replicate reactions were performed for similar primer pairs and each copy number of individuals was calculated by Copy Caller v1.0 (Applied Biosystems) using the comparative C_t method.

Statistical analysis and gene ontology analysis

We generated input data of both loss (homozygous deletion=0X; hemizygous deletion=1X; normal copy number=2X) and gain (normal copy number=2X; duplication=3X; amplification=$>$4X). In the case of loss, CNV status 0X, 1X and 2X were coded with 'A_A', 'A_B' and 'B_B,' respectively. In the case of gain, CNV status $>$4X, 3X and 2X were coded with 'A_A,' 'A_B' and 'B_B,' respectively. Logistic regression analyses controlling for age (continuous value) and sex (male=0 or female=1) as covariate were used to calculate the *P*-values for case–control analysis. Golden HelixTree software (Golden Helix, Bozeman, MT, USA; http://www.goldenhelix.com) was used for the association analyses and χ^2-tests were used to determine whether the individual variants were in Hardy–Weinberg equilibrium.

To correct for multiple testing, the effective number of independent markers representing common CNVs (frequency $>1\%$) was calculated using the software SNPSpD (http://genepi.qimr.edu.au/general/daleN/SNPSpD/), which is based on the spectral decomposition of matrices of pairwise linear disequilibrium values between markers.[19] The number of independent marker loci was calculated as 89.3639, and this was applied to correct for multiple testing (*P*-value $\times$89.3639). To provide insights into the functional enrichment of CNVs, we performed gene ontology (GO) analysis using GOstat (http://gostat.wehi.edu.au) provided by Tim Beißbarth.[20,21]

RESULTS

In this study, we identified 4574 individual CNVs and 1644 CNV regions in a Japanese population sample (n=473) using the Illumina HumanHap300 BeadChip. The average number of CNVs per sample was 9.7, with an average length and median size of 63.5 and 31.9 kb, respectively (Supplementary Table 1). Furthermore, results revealed that 1644 CNV regions identified in this study contained 1232 genes. Among the common CNV regions, 187 had a frequency of $>1\%$, 66 CNV regions had $>2.5\%$ and 22 CNV regions had $>5\%$. Moreover, the 22 common CNV regions that contained 84 genes along with the rest of the regions were found to overlap with previously reported CNVs in the Database of Genomic Variants (DGVs) (Table 1). A karyotype map that shows the identified CNV regions according to their frequency is shown in Supplementary Figure 1. Common CNV regions with a frequency of over 1% were distributed evenly on the chromosomes, including X chromosomes. Furthermore, we were able to observe that 73.2% of the CNVs identified in this study matched with those found in DGV. All of them overlapped with previously reported CNVs in DGV. Supplementary Figure 2 shows the size distribution of individual CNVs identified in this study. Most of the identified CNVs were distributed within 1–50 kb in range.

To determine the association between the identified CNVs and the risk of SAH, we performed logistic regression analyses controlling for age and sex as covariates in 187 common CNV regions (frequency $>1\%$). Results showed that two CNV regions (chr4:153210505–153212191, P=0.0005, and chr10:6265006–6267388, P=0.0006) were significantly associated with the risk of SAH (Table 2). After performing Hardy–Weinberg equilibrium test on the two CNV regions, we found no significant differences between the distribution of the observed genotypes and the expected distributions ($P>0.05$). In the case of chr4:153210505–153212191, the frequency of deletion (or loss) in the case group was higher than that in the control group. This fact suggests that the deletion allele may be a risk factor for SAH. In the case of chr10:6265006–6267388, the frequency of duplication (or gain) in the case group was higher than that in the control group. This fact indicates that the increase in copy number in the region may influence the onset of SAH. These two regions contained one gene and four nearby genes within 500 kb. Moreover, we observed that these two regions overlapped with the previous identified CNVs on DGV (Supplementary Figure 3).

Table 1 Summary of common CNV regions (frequency >5%)

CNV region	*Length (bp)*	*No. of CNVs*	*Frequency*	*No. of genes*	*Gene*	*Overlapped with DGV*[a]
chr18:64877649–64909977	32328	149	0.314	0		Yes
chr19:20331564–20779390	447826	95	0.200	3	*ZNF626,ZNF737,ZNF826*	
chr4:161993023–162154792	161769	85	0.179	0		Yes
chr3:65166887–65187636	20749	59	0.124	0		Yes
chr11:49483244–51297186	1813942	52	0.109	6	*LOC440040,LOC441601,LOC646813, OR4A5,OR4C12,OR4C13*	Yes
chr16:965984–1484303	518319	40	0.084	20	*BAIAP3,C16orf38,C16orf42,C16orf91,C1QTNF8, CACNA1H,CCDC154,CLCN7,GNPTG,LOC100130430, LOC146336,SOX8,SSTR5,TELO2,TPSAB1,TPSB2, TPSD1,TPSG1,UBE2I,UNKL*	Yes
chr14:43524894–43584372	59478	38	0.080	0		Yes
chr19:642025–1523369	881344	37	0.078	42	*ABCA7,ADAMTSL5,APC2,ARID3A,ATP5D,AZU1,C19orf21, C19orf22,C19orf23,C19orf24,C19orf25,C19orf26,C19orf6, CFD,CIRBP,CNN2,DAZAP1,EFNA2,ELA2,GAMT,GPX4,GRIN3B, HMHA1,KISS1R,MED16,MEX3D,MIDN,MUM1,NDUFS7,PALM, PCSK4,PLK5P,POLR2E,PRG2,PRSSL1,PRTN3,PTBP1,REEP6, RPS15,SBNO2,STK11,WDR18*	Yes
chr12:130374756–130398112	23356	36	0.076	0		Yes
chr5:41608479–41632026	23547	35	0.074	0		Yes
chr7:153118877–153284978	166101	34	0.072	1	*DPP6*	Yes
chr15:45994758–46034246	39488	32	0.067	0		Yes
chr21:16683912–16708090	24178	31	0.065	1	*C21orf34*	Yes
chr8:5316404–5947152	630748	30	0.063	0		Yes
chr6:29188328–29432578	244250	29	0.061	5	*LOC651503,OR14J1,OR2J2,OR2J3,OR5V1*	Yes
chr7:3411837–3431422	19585	29	0.061	1	*SDK1*	Yes
chr21:26116740–26163866	47126	26	0.055	0		Yes
chr8:145050595–145299088	248493	26	0.055		*C8orf30A,CYC1,EXOSC4,GPAA1,GRINA,HEATR7A, KIAA1875,MAF1,OPLAH,PARP10,PLEC1,SHARPIN,SPATC1*	Yes
chr15:28723577–28853522	129945	25	0.053	0		Yes
chr11:5858528–5891679	33151	25	0.053	1	*OR52E4*	Yes
chr6:31458683–31559455	100772	24	0.051	3	*HCG26,HCP5,MICA*	Yes
chr18:3559620–3561217	1597	24	0.051	1	*DLGAP1*	Yes

Abbreviations: CNV, copy number variation; DGV, Database of Genomic Variant.
[a]Database of Genomic Variants (http://projects.tcag.ca/variation).

Table 2 Logistic regression analysis of identified CNV with risk of subarachnoid aneurysmal hemorrhage while controlling age and gender as covariates among case and normal subjects

					P-*value (HWE)*			*Case*			*Control*						
CNV type	*CNV region*[a]	*Cytoband*	*Length (kb)*	*MAF*	*All*	*Cases* (n=*191*)	*Controls* (n=*282*)	*L/L (0X)*	*L/N (1X)*	*N/N (2X)*	*L/L (0X)*	*L/N (1X)*	*N/N (2X)*	P-*value*	P^{corr} [b]	*Gene*	*Nearby genes (kb)*
Loss	chr4: 153210505–153212191	4q31.3	1.7	0.019	0.673	0.572	0.928	0 *G/G (>4X)*	15 *G/N (3X)*	176 *N/N (2X)*	0 *G/G (>4X)*	3 *G/N (3X)*	279 *N/N (2X)*	**0.0005**	**0.04**		PET112 L (308.9), FBXW7 (249.7)
Gain	chr10: 6265006–6267388	10p15.1	2.4	0.01	0.835	0.768	0.976	0	8	183	0	1	281	**0.0006**	**0.05**	PFKFB3[c]	RBM17 (65.6), PFKFB3 (17.5)[d]

Abbreviations: CNV, copy number variation; G/G, gain/gain; G/N, gain/normal; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; L/L, loss/loss; L/N, loss/normal; MAF, minor allele frequency; N/N, normal/normal; N/N, normal/normal; P^{corr}, corrected *P*-value.
Bold values indicate the case of $P<0.05$.
[a]Version: Human Mar. 2006 (NCBI36/hg18) Assembly.
[b]To achieve the optimal correction for multiple testing of markers representing common CNVs (frequency >1%) in LD with each other, the effective number of independent marker loci (89.3639) was calculated using the software SNPSpD (http://genepi.qimr.edu.au/general/daleN/SNPSpD/), on the basis of spectral decomposition (SpD) of matrices of pairwise LD between the markers.
[c]*Homo sapiens* 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3), transcript variant 2, NCBI Reference Sequence: NM_001145443.
[d]*Homo sapiens* 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3), transcript variant 1, NCBI Reference Sequence: NM_004566.

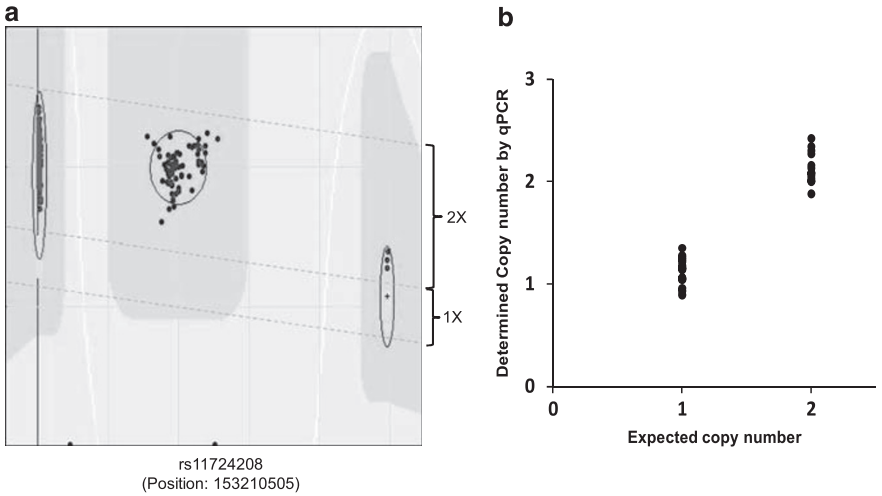

Figure 1 Copy number variation validation by quantitative PCR (qPCR) around rs11724208 within chr4:153210505–153212191. (**a**) Genoplot image of identified deletions (marker name: rs11724208). One genoplot image represents allelic intensity (*x* axis) and signal intensity (*y* axis) of all samples. Two types of copy number (2X and 1X) are depicted. Individuals having hemizygous deletions (copy number: 1X) clustered into two distinct groups (color: yellow). Samples having null copy number are displayed with a black dot at the bottom. (**b**) Validation by qPCR around the rs11724208 within chr4:153210505–153212191. The value of *x* axis (expected copy number) was estimated using the Illumina Genoplot image analysis. The *y* axis indicates the determined copy number by qPCR. The copy number value estimated through visual examination matched with the quantitative measurement value using qPCR. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Figure 1 shows a visualization of identified CNV (CNV type: deletion) within chr4:153210505–153212191 in UCSC Genome Browser and Illumina genoplot images. This genoplot image, which represents signal and allelic intensity of samples per marker, shows that the hemizygous deletions in the first marker (rs11724208; position: 153210505) were also observed in the third marker (rs10029800; position: 153212191).

DISCUSSION

SAH due to rupture of cerebral aneurysm accounts for 6–8% of total stroke cases. As 2–5% of the world's total population has cerebral aneurysm, it is necessary to broaden the study on factors that cause SAH. Early precaution and treatment are very important, as 40% of patients with SAH may die during the onset of the disease.

In this study, we hypothesized that CNV is significantly associated with the risk of SAH. To perform association analysis, it is important to accurately identify individual CNVs. For this purpose, we used pennCNV, a program that is optimized for SNP genotyping BeadChip using multiple factors, namely, signal intensity, allelic intensity and population frequency of B allele.[17,22] This program has been widely used by researchers and the results generated from the program are already publicized in papers.[22–27] Two other ways of verifying the accuracy of the identified CNVs involve comparing the results with the DGV databases in which information on previously reported CNVs is stored, or using another CNV identification program to compare the results with.

To identify reliable CNVs, we used the Illumina HumanHap300 BeadChip, which contains 317 503 SNP/CNV markers. We also compared the results of the individual CNVs that were identified in this study with that of the DGV database and with another CNV identification program called CNVpartition. Our findings revealed that the CNVs identified in this study were in concordance with the DGV database at 73.2% and with the CNVpartition at 72.4% (data not shown), an indication of high-quality CNVs. After comparing the individual CNVs and segmental duplications, we found that many CNVs and common CNV regions (>5%) were located in the segmental duplication region (33.9 and 45.5%, respectively). This observation is similar to the results of other studies[6,28,29] wherein CNVs discovered were closely related with segmental duplication.[14] Further, results from GO analysis showed that cell differentiation displayed distributions similar to those of previous studies. However, sensory perception of smell and neurophysiological processes did not produce a highly enriched region in the GO category (Supplementary Table 2).

To identify disease-susceptible regions for the risk of SAH, we performed a genome-wide CNV association analysis using a logistic regression model. In this study, we found two new CNV regions (chr4:153210505–153212191 and chr10:6265006–6267388) that were significantly associated with risk of SAH. Interestingly, we also observed that *PET112L* (PET112, *Saccharomyces cerevisia*, homolog-like), *FBXW7* (F-box and WD40 domain protein 7), *PFKFB3* (transcript variant 1 of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3) and *RBM17* (RNA binding motif protein 17) genes were located within 500 kb of those CNV regions. PET112L (OMIM no.: 603645) is predominantly expressed in muscle and heart tissues in which oxidative phosphorylation occurs at a high rate. Although this protein has a role in accumulation of cytochrome oxidase subunit II and in the stability of mitochondrial genome in *S. cerevisia*,[30] its function in the human body is still unclear. FBXW7 (OMIM no.: 606278), on the other hand, targets the ligase activity to cyclin E, a protein previously implicated in the regulation of neuronal apoptosis. Reciprocal relation with loss of FBXW7 and deletion or mutation of PTEN has been reported to be associated with human breast cancer cell lines and primary tumors.[31] Furthermore, both *RBM17* and *PFKFB3* genes have been considered to be associated with diabetes, obesity and endocrine disease;[32] however, further studies are required to clarify how those genes may affect an individual's susceptibility to SAH. From our previous study, we found that one SNP marker (rs1242541) within a CNV region was significantly associated with the risk of SAH (P=0.0006),[13] but in this study, no significant association was detected in this region. These discrepancies in the results may be due to the fact that our previous study only analyzed single multiallelic

CNV markers on CNV regions. In other words, the present study, therefore, is different from the previous study, as analyses of CNVs were conducted at a larger scale.

In conclusion, we have described the first genome-wide CNV association analysis on CNV and the risk of SAH. We found significant associations between several CNV regions and the risk of SAH after identifying 4574 individual CNVs from 473 samples using the Illumina HumanHap300 BeadChip. Although future studies should be conducted to determine those susceptible regions that are functional and elaborate on the underlying mechanism of their function, our findings suggest that the newly identified disease susceptible CNV regions in this study may influence the risk of SAH in a Japanese population.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

CNV, copy number variation; DGV, Database of Genomic Variant; SAH, subarachnoid aneurysmal hemorrhage.

ACKNOWLEDGEMENTS

We thank Dr Kai Wang (Department of Genetics, University of Pennsylvania, and Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA) for providing useful scripts for the pennCNV. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (no. 2009-0080157). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (no. 2010-0011206). This work was also supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (no. 2009-0093822).

1 Hop, J. W., Rinkel, G. J., Algra, A. & van Gijn, J. Case-fatality rates and functional outcome after subarachnoid hemorrhage: a systematic review. *Stroke* **28,** 660–664 (1997).
2 Yong-Zhong, G. & van Alphen, H. A. Pathogenesis and histopathology of saccular aneurysms: review of the literature. *Neurol. Res.* **12,** 249–255 (1990).
3 Bromberg, J. E., Rinkel, G. J., Algra, A., Greebe, P., van Duyn, C. M., Hasan, D. *et al.* Subarachnoid haemorrhage in first and second degree relatives of patients with subarachnoid haemorrhage. *Br Med J* **311,** 288–289 (1995).
4 Ruigrok, Y. M., Buskens, E. & Rinkel, G. J. Attributable risk of common and rare determinants of subarachnoid hemorrhage. *Stroke* **32,** 1173–1175 (2001).
5 Cook, E. H. Jr. & Scherer, S. W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455,** 919–923 (2008).
6 Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
7 Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M. *et al.* Copy number variation: new insights in genome diversity. *Genome Res.* **16,** 949–961 (2006).
8 Henrichsen, C. N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nat. Genet.* **41,** 424–429 (2009).
9 Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316,** 445–449 (2007).
10 Cappuzzo, F., Hirsch, F. R., Rossi, E., Bartolini, S., Ceresoli, G. L., Bemis, L. *et al.* Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J. Natl Cancer Inst.* **97,** 643–655 (2005).
11 St Clair, D. Copy number variation and schizophrenia. *Schizophr. Bull.* **35,** 9–12 (2009).
12 Knight, S. J., Regan, R., Nicod, A., Horsley, S. W., Kearney, L., Homfray, T. *et al.* Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* **354,** 1676–1681 (1999).
13 Bae, J. S., Cheong, H. S., Kim, J. O., Lee, S. O., Kim, E. M., Lee, H. W. *et al.* Identification of SNP markers for common CNV regions and association analysis of risk of subarachnoid aneurysmal hemorrhage in Japanese population. *Biochem. Biophys. Res. Commun.* **373,** 593–596 (2008).
14 Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41,** 334–341 (2009).
15 Need, A. C., Attix, D. K., McEvoy, J. M., Cirulli, E. T., Linney, K. L., Hunt, P. *et al.* A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum. Mol. Genet.* **18,** 4650–4661 (2009).
16 Makishima, T., Nakashima, T., Nagata-Kuno, K., Fukushima, K., Iida, H., Sakaguchi, M. *et al.* The highly conserved DAD1 protein involved in apoptosis is required for N-linked glycosylation. *Genes Cells* **2,** 129–141 (1997).
17 Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
18 Wang, K., Chen, Z., Tadesse, M. G., Glessner, J., Grant, S. F., Hakonarson, H. *et al.* Modeling genetic inheritance of copy number variations. *Nucleic. Acids Res.* **36,** e138 (2008).
19 Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74,** 765–769 (2004).
20 Beissbarth, T. Interpreting experimental results using gene ontologies. *Methods Enzymol.* **411,** 340–352 (2006).
21 Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20,** 1464–1465 (2004).
22 Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459,** 528–533 (2009).
23 Cusco, I., Medrano, A., Gener, B., Vilardell, M., Gallastegui, F., Villa, O. *et al.* Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum. Mol. Genet.* **18,** 1795–1804 (2009).
24 Daly, A. K., Donaldson, P. T., Bhatnagar, P., Shen, Y., Pe'er, I., Floratos, A. *et al.* HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* **41,** 816–819 (2009).
25 Diskin, S. J., Hou, C., Glessner, J. T., Attiyeh, E. F., Laudenslager, M., Bosse, K. *et al.* Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459,** 987–991 (2009).
26 Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451,** 998–1003 (2008).
27 Vrijenhoek, T., Buizer-Voskamp, J. E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A. *et al.* Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.* **83,** 504–510 (2008).
28 Cooper, G. M., Nickerson, D. A. & Eichler, E. E. Mutational and selective effects on copy-number variants in the human genome. *Nat. Genet.* **39,** S22–S29 (2007).
29 Perry, G. H., Tchinda, J., McGrath, S. D., Zhang, J., Picker, S. R., Caceres, A. M. *et al.* Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl Acad. Sci. USA* **103,** 8006–8011 (2006).
30 Petruzzella, V., Tiranti, V., Fernandez, P., Ianna, P., Carrozzo, R. & Zeviani, M. Identification and characterization of human cDNAs specific to BCS1, PET112, SCO1, COX15, and COX11, five genes involved in the formation and function of the mitochondrial respiratory chain. *Genomics* **54,** 494–504 (1998).
31 Mao, J. H., Kim, I. J., Wu, D., Climent, J., Kang, H. C., DelRosario, R. *et al.* FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression. *Science* **321,** 1499–1502 (2008).
32 Lowe, C. E., Cooper, J. D., Brusko, T., Walker, N. M., Smyth, D. J., Bailey, R. *et al.* Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat. Genet.* **39,** 1074–1082 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)