Journal of Human Genetics (2011) 56, 852–856

www.nature.com/jhg

ORIGINAL ARTICLE

A genome-wide CNV association study on panic disorder in a Japanese population

Yoshiya Kawamura[1,2], Takeshi Otowa[3], Asako Koike[4], Nagisa Sugaya[5], Eiji Yoshida[6], Shin Yasuda[6], Ken Inoue[7], Kunio Takei[8], Yoshiaki Konishi[9], Hisashi Tanii[9], Takafumi Shimada[3], Mamoru Tochigi[3], Chihiro Kakiuchi[3], Tadashi Umekage[10], Xiaoxi Liu[11], Nao Nishida[11], Katsushi Tokunaga[11], Ryozo Kuwano[12], Yuji Okazaki[13], Hisanobu Kaiya[1,2,6] and Tsukasa Sasaki[8,14]

Family and twin studies have indicated that genetic factors have an important role in panic disorder (PD), whereas its pathogenesis has remained elusive. We conducted a genome-wide copy number variation (CNV) association study to elucidate the involvement of structural variants in the etiology of PD. The participants were 2055 genetically unrelated Japanese people (535 PD cases and 1520 controls). CNVs were detected using Genome-Wide Human SNP array 6.0, determined by Birdsuite and confirmed by PennCNV. They were classified as rare CNVs (found in $<1\%$ of the total sample) or common CNVs (found in $\geqslant 5\%$). PLINK was used to perform global burden analysis for rare CNVs and association analysis for common CNVs. The sample yielded 2039 rare CNVs and 79 common CNVs. Significant increases in the rare CNV burden in PD cases were not found. Common duplications in 16p11.2 showed Bonferroni-corrected *P*-values <0.05. Individuals with PD did not exhibit an increased genome-wide rare CNV burden. Common duplications were associated with PD and found in the pericentromeric region of 16p11.2, which had been reported to be rich in low copy repeats and to harbor developmental disorders, neuropsychiatric disorders and dysmorphic features.
Journal of Human Genetics (2011) **56,** 852–856; doi:10.1038/jhg.2011.117; published online 20 October 2011

Keywords: 16p11.2; copy number variation; deletion; duplication; genome-wide association study; global burden analysis; Japanese; panic disorder

INTRODUCTION

Panic disorder (PD) is characterized by recurrent and unexpected panic attacks, subsequent anticipatory anxiety and phobic avoidance, with the frequent development of agoraphobia.[1] The lifetime prevalence of PD is 4.7% with female preponderance.[2] PD frequently takes a chronic course, with many remissions and relapses.[3] Genetic epidemiological researches including family and twin studies have shown that genetic as well as environmental factors have an important role in the pathogenesis of PD. First-degree relatives of proband with PD have a six-fold increased risk of developing the condition.[4] Twin studies show that about 40% of liability toward PD relates to heritable factors.[5] Although several linkage studies and many association studies on more than 350 candidate genes for PD have been concluded, the results were inconsistent, negative or not clearly replicated.[6]

Copy number variations (CNVs) are quantitative structural variants; they are deletions and duplications of DNA segments ranging from a kilobase to several megabases. CNVs cover 12% of the human genome,[7] which is much more than single-nucleotide polymorphisms (SNPs). As recent studies have found evidence for the potential involvement of CNVs in neuropsychiatric conditions,[8–10] we explored genome-wide CNV association with PD in a Japanese population.

MATERIALS AND METHODS

The following methods are shown using a simplified block diagram in Figure 1.

Sample recruitment

Participants were 2212 unrelated, ethnically Japanese sample of 595 PD cases (189 males and 406 females; age=38.9 ± 10.9 years, mean ± s.d.) and 1617 controls (706 males and 911 females; age=39.0 ± 11.0 years, mean ± s.d.) recruited in the vicinity of Tokyo, Nagoya or Niigata, Japan. These areas are known to belong to the main cluster of the Japanese population structure.[11] The diagnosis was confirmed according to the Diagnostic and Statistical Manual, 4th edition criteria,[1] using the Mini-International Neuropsychiatric

[1]Outpatient Clinic for Anxiety Disorders, Yokohama Clinic, Yokohama, Japan; [2]Research Center for Panic Disorder, Nagoya Mental Clinic, Nagoya, Japan; [3]Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [4]Central Research Laboratory, Hitachi Ltd, Tokyo, Japan; [5]Research Project for Addictive Substances, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; [6]Outpatient Clinic for Anxiety Disorders, Akasaka Mental Clinic, Tokyo, Japan; [7]Department of Public Health, Fujita Health University School of Medicine, Toyoake, Japan; [8]Department of Psychiatry, Office for Mental Health Support, University of Tokyo, Tokyo, Japan; [9]Department of Psychiatry, Graduate School of Medicine, Mie University, Tsu, Japan; [10]Division for Environment, Health and Safety, University of Tokyo, Tokyo, Japan; [11]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [12]Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; [13]Department of Neurology, Tokyo Metropolitan Matsuzawa Hospital, Tokyo, Japan and [14]Department of Developmental Sciences, Graduate School of Education, University of Tokyo, Tokyo, Japan
Correspondence: Dr Y Kawamura, Outpatient Clinic for Anxiety Disorders, Yokohama Clinic, 1-2-10 Kitasaiwai, Nishi, Yokohama, Kanagawa 220-0004, Japan.
E-mail: yoshiya-tky@umin.ac.jp
Received 29 July 2011; revised 9 September 2011; accepted 21 September 2011; published online 20 October 2011

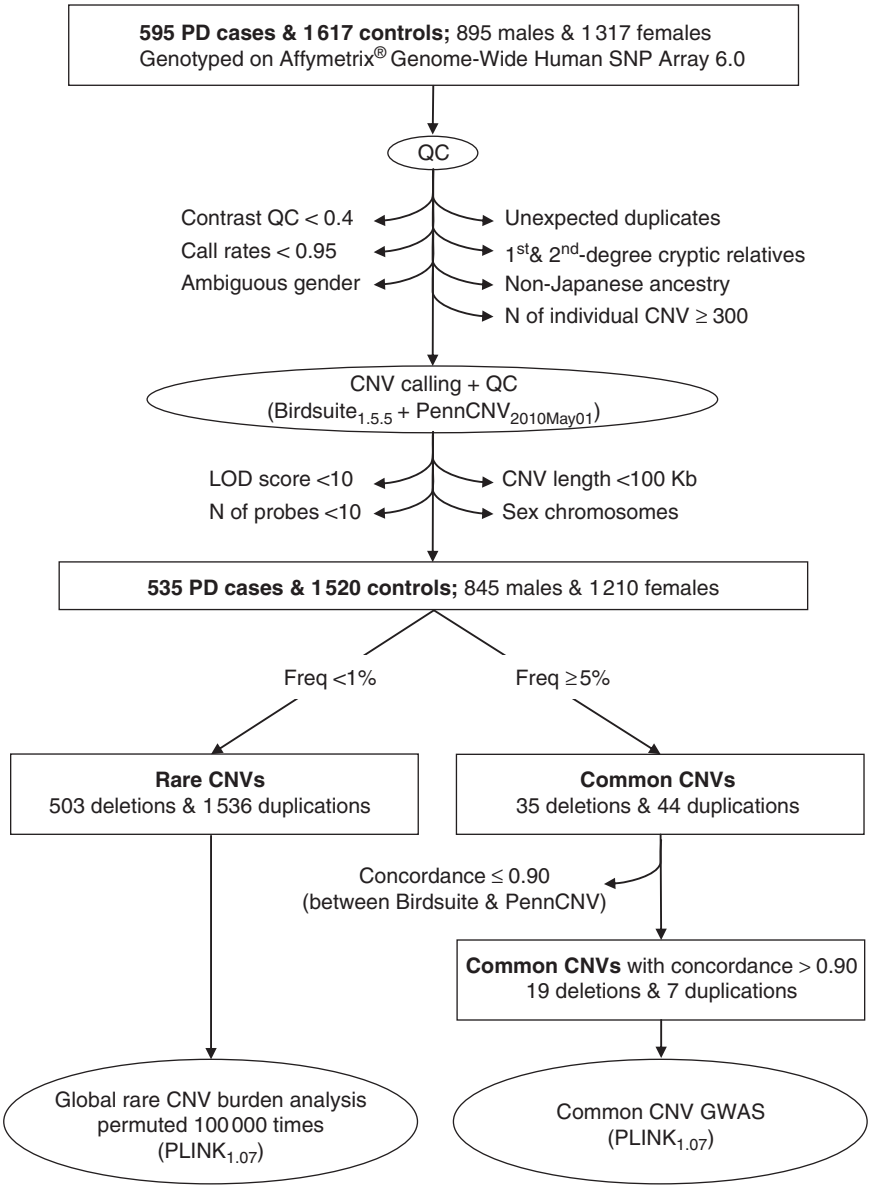

Figure 1 Block diagram of materials and methods used for genome-wide rare and common CNV association analysis of panic disorder.

Interview[12] and clinical records. Details of the recruitment of cases[13] and controls[13–16] are described elsewhere.

The aims of the present study were clearly explained to all subjects. They gave written, informed consent. The study was approved by Ethical Committees of the Faculties of Medicine at the University of Tokyo, Mie University and Niigata University, and conformed to the provisions of the Declaration of Helsinki.

CNV detection and determination

Genomic DNA was isolated from leukocytes in whole blood using the Wizard genomic DNA purification kit (Promega Corporation, Madison, WI, USA). SNPs and CNVs were detected using Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA, USA). The results of the SNP genome-wide association study are described elsewhere.[17] CNVs were determined with Birdsuite software[18] version 1.5.5 over the thresholds of logarithm of odds $\geqslant 10$,[18,19] the number of probes in CNV $\geqslant 10$[18,20,21] and individual CNVs <300. A total of 906 600 SNP probes and 946 000 CNV non-polymorphic probes on the chip were genotyped at the University of Tokyo, Graduate School of Medicine, according to the manufacturer's protocol. PennCNV software[22] version 1 May 2010 was also used as an alternative algorithm to Birdsuite, to establish stringent CNV data; common CNVs were evaluated if they showed concordance rates of >0.90 between the algorithms.[18,23]

Quality control

A total of 157 subjects were excluded from further analysis, because they did not pass the quality control criteria for contrast quality control $\geqslant 0.4$,[24] sample-wise call rate $\geqslant 0.95$,[25] ambiguous gender, unexpected duplicates, cryptic relatives or population stratifications. Ambiguous gender was assessed by sex check function in PLINK software[26] version 1.07. Unexpected duplicates and first- or second-degree cryptic relatives were assessed using the identity-by-state/identity-by-descent estimation function of PLINK to estimate the cryptic relatedness of each pair of subjects. To detect population outliers, the subjects were assessed through comparison with 45 JPT, 45 CHB, 45 CEU and 45 YRI HapMap[27] subjects without relationships using the EIGENSTRAT software.[28] Supplementary Figure 1 indicates the principal component analysis plots before and after filtering for quality control.

Table 1 Subjects after quality controls

	Cases			*Controls*		
	Males and females	*Males*	*Females*	*Males and females*	*Males*	*Females*
Number (%)						
Total	535 (100.0)	174 (100.0)	361 (100.0)	1520	671	849
With comorbidity						
Agoraphobia	311 (62.6)	90 (56.3)	221 (65.6)	NA	NA	NA
Major depressive disorder	212 (44.8)	57 (38.5)	155 (47.7)	NA	NA	NA
Bipolar disorders	124 (23.9)	51 (30.4)	73 (20.9)	NA	NA	NA
Mean age (s.d.)						
Age at recruitment	38.2 (11.7)	38.3 (10.4)	38.1 (12.2)	39.8 (10.8)	41.6 (10.2)	38.6 (11.0)
Age at onset	25.8 (13.2)	26.9 (13.9)	25.3 (12.9)	NA	NA	NA

Abbreviation: NA, not applicable.

Filtering for quality control identified 2055 samples: 535 cases aged 38.2 ± 11.7 years (174 males and 361 females) and 1520 controls aged 39.8 ± 10.8 years (671 males and 849 females; Table 1).

Statistical analysis

CNVs with segment sizes of <100 Kb were excluded by PLINK, because the platforms had limited power to detect smaller CNVs.[19,23,29] CNVs were classified as rare (found in <1% of the total sample) or common (found in ⩾5%). They were considered to be colocalized if they overlapped by at least 50% of their length, as implemented in PLINK. Common CNVs were also excluded when the lengths of the consensus regions shared by all overlapping CNV segments were equivalent to one base pair, which meant these regions included only one probe.

PLINK was used to perform global burden analysis for rare CNVs on the basis of 100 000 permutations and association analysis for common CNVs between cases and controls at each consensus region. Sex chromosomes were removed from the analysis, as they are incorrectly segmented by Birdsuite[19] and PennCNV.[21] Bonferroni-corrected *P*-values <0.5 were considered significant.

RESULTS

The sample yielded 2039 rare CNVs (503 deletions and 1536 duplications) and 79 common CNVs (35 deletions and 44 duplications; Figure 1).

As for rare CNVs, neither the number of segments per person (RATE) nor the proportion of samples with one or more segments (PROP) showed significant differences between PD cases and controls (Table 2). The same analysis was repeated for CNVs showing values of <5% in the sample, and resulted in 550 deletions and 1646 duplications. The trends between cases and controls were virtually identical to those in Table 2 (data not presented).

Among common CNVs, 19 deletions and 7 duplications had concordance rates of >0.90 between Birdsuite and PennCNV. Five common duplications on the locus of 31 901 431–33 149 454 bp in chromosome 16 (16p11.2) showed corrected *P*-values of <0.05; the statistical power to detect these CNVs was estimated to be more than 0.8 for α=0.001 when the frequency of CNVs was 0.21, the penetrance of CNVs was 0.01 and the CNV relative risk was 1.4 for 535 cases and 1520 controls.[30–32] Other common CNVs with concordance rates of >0.90 did not show *P*-values of <0.05 for either deletions or duplications (Supplementary Table 1).

Table 2 Global rare CNV burden analysis

	Deletions			*Duplications*			*Deletions and duplications*		
Tests	*Cases*	*Controls*	P-*values*[a]	*Cases*	*Controls*	P-*values*[a]	*Cases*	*Controls*	P-*values*[a]
RATE[b]	0.022	0.029	0.82	0.015	0.012	0.37	0.036	0.029	0.27
PROP[c]	0.022	0.028	0.81	0.015	0.011	0.27	0.036	0.028	0.24

Abbreviation: CNV, copy number variation.
Rare CNVs exhibited no significant differences in global burden analysis. The values of RATE and PROP were lower than those of the previous study.[19] CNVs were considered to be colocalized and identical if they overlapped by at least 50% of their length. The number of rare CNVs was consequently estimated to be comparatively low in the present study.
[a]*P*-values were based on the difference between cases and controls, one-tailed and permuted 100 000 times.
[b]Number of segments per person.
[c]Proportion of samples with one or more segments.

DISCUSSION

The present study explored genome-wide rare and common CNV associations between 535 PD cases and 1520 controls in a Japanese population. Rare CNVs exhibited no significant differences in global burden analysis. Individuals with PD did not have an increased genome-wide rare CNV burden. The number of rare CNVs was estimated to be smaller than that of the previous study[19] on the basis of global burden analysis, because CNVs were considered to be colocalized and identical in the present study if they overlapped by at least 50% of their length.

The common CNVs found in this study had all been previously reported. Common duplications in 16p11.2 showed Bonferroni-corrected significance values of $P<0.0019$ (=0.05/(19 deletions and 7 duplications)) in association analysis. These included several genes of *IGH*, *HERC2P4*, *TP53TG3*, *SLC6A8* and *SLC6A10P* in their union region for the total distance of 31 901 431–33 149 454 bp spanned by all duplication segments. No genes in their consensus regions were shared by all overlapping duplication segments (Table 3 and Supplementary Figure 2).

Pseudogenes and non-coding RNAs (including microRNAs and small nucleolar RNAs (snoRNAs)) were sought in the union region using existing databases (Pseudogene.org,[33] miRBase,[34] snoRNABase[35] and NONCODE[36]), as Salmena *et al.*[37] recently emphasized their important roles in pathological conditions. Although no miRNAs or snoRNAs were found, 50 pseudogenes (listed in Supplementary Table 2) and 241 PIWI-interacting RNAs (piRNAs; shown in Supplementary Figure 2) were detected in the union region. No pseudogenes and three piRNAs (n138582, n119652 and n127652) were found in the consensus regions (Table 3). piRNAs are particularly abundant in the

Table 3 Results of genome-wide association study for common CNVs in 16p11.2

Chr	*Regions (bp)*[a]	*Types*	*Analysed regions (bp)*[b]	*OR*	P-*values*	*Genes included*
16	31 901 431–33 149 454	Duplication				*IGH*, *HERC2P4*, *TP53TG3*, *SLC6A8*, *SLC6A10P*
			32 288 452–32 293 616	2.02	5.0×10^{-4}	None
			32 472 013–32 480 977	2.33	4.1×10^{-6}	None
			32 522 376–32 522 419	2.35	3.5×10^{-6}	None
			32 716 092–32 723 322	2.49	4.7×10^{-4}	None
			32 747 828–32 747 898	2.37	8.4×10^{-4}	None

Abbreviations: Chr, chromosome; CNV, copy number variation; OR, odds ratio.
[a]Union regions spanned by all CNV segments.
[b]Consensus regions shared by all overlapping CNV segments.

Table 4 Subjects harboring significant common duplications in 16p11.2 (16: 31 901 431–33 149 454 bp)

	Cases					*Controls*
Consensus regions (bp)[a]	*Freq_case*	*Mean age at onset (s.d.)*	*Freq_AGO*	*Freq_MDD*	*Freq_BD*	*Freq_control*
16: 32 288 452–32 293 616	17.2	18.8 (15.8)	57.8	46.7	15.3	9.3
	[17.2, 17.2]	[18.4 (16.1), 19.0 (15.8)]	[53.8, 59.6]	[34.8, 51.9]	[19.2, 13.6]	[10.7, 8.2]
16: 32 472 013–32 480 977	22.2	19.5 (15.9)	58.3	45.5	17.1	10.9
	[22.4, 22.2]	[18.9 (16.3), 19.8 (15.7)]	[52.9, 60.8]	[29.0, 52.9]	[22.9, 14.5]	[12.5, 9.7]
16: 32 522 376–32 522 419	22.2	19.5 (15.9)	58.3	45.5	17.1	10.9
	[22.4, 22.2]	[18.9 (16.3), 19.8 (15.7)]	[52.9, 60.8]	[29.0, 52.9]	[22.9, 14.5]	[12.5, 9.5]
16: 32 716 092–32 723 322	10.3	18.8 (16.2)	62.0	37.2	10.0*	4.4
	[9.8, 10.5]	[21.5 (15.2), 17.6 (16.6)]	[53.3, 65.7]	[21.4, 44.8]	[6.7*, 11.4]	[4.5, 4.4]
16: 32 747 828–32 747 898	10.3	18.5 (16.2)	61.2	36.6	10.2*	4.6
	[10.3, 10.2]	[21.2 (14.8), 17.2 (16.9)]	[56.2, 63.6]	[21.4, 44.4]	[12.5, 9.1]	[4.5, 4.7]

Abbrebiations: AGO, agoraphobia; BD, bipolar disorders; Freq, frequency; MDD, major depressive disorder.
Frequencies mean the percentage ratio of subjects with CNVs in the subgroups. Figures in square brackets are those in males and females, respectively. Frequencies of comorbid bipolar disorders are significantly lower in all and male cases with some duplications than those without them.
[a]Regions shared by all overlapping CNV segments.
*$P<0.05$.

male germline and adult testes,[38] and are involved in developmental regulation and genomic imprinting through DNA methylation.[39,40] PD development may be influenced by common duplications in 16p11.2 through piRNA regulation.

Table 4 shows information on subjects with significant common duplications in 16p11.2. No significant differences were observed in the mean age at the onset of PD, the frequency of cases, controls, comorbid agoraphobia or comorbid major depressive disorder among subjects with and without duplications, except for lower frequencies of comorbid bipolar disorders in all and male cases showing duplication with the consensus regions of 32 716 092–32 723 322 bp (χ^2=5.91, df=1, P=0.02 in all; χ^2=4.37, df=1, P=0.04 in males) and in all cases showing duplication with the consensus regions of 32 747 828–32 747 898 bp (χ^2=5.61, df=1, P=0.02).

The pericentromeric region of 16p11.2 was reported to be rich in low copy repeats,[41] causing non-allelic homologous recombination and the harboring of novel genomic disorders.[42] It has been suggested that rare CNVs in the same region of 16p11.2 (but 2 Mb upstream of the region discussed here), are associated with developmental and neuropsychiatric disorders or dysmorphic manifestations such as mental retardation, autism spectrum disorders, attention-deficit/hyperactivity disorder, schizophrenia, abnormal head size or syringomyelia.[20,43–45] This is the first study to demonstrate significant associations of CNVs in 16p11.2 with PD.

Recently, the Wellcome Trust Case–Control Consortium identified biological artifacts of common CNVs typed on existing platforms that lead to false-positive associations.[21] It might be appropriate to suggest that care should be taken when determining common CNVs using currently available SNP-based microarray and algorithm software, as differences in performance between algorithms have been reported.[46,47] In our study, we took a rigorous and conservative analytical approach to minimize false-positive findings. We set strict thresholds for logarithm of odds ($\geqslant$10), the number of probes in CNV ($\geqslant$10) and CNV segment sizes ($\geqslant$100 Kb). In addition, we used two different algorithms to identify the CNV difference between cases and controls.

There are some limitations to our study. First, the case sample size (595 PD patients) might not be adequate in comparison with the control sample size (1617 subjects). A larger case sample is required to detect CNVs with lower *P*-values. Second, CNVs in sex chromosomes were not analyzed, because Birdsuite and PennCNV cannot determine them sufficiently. An improved algorithm is required for the determination of whole-genome CNVs, including those in sex chromosomes. Finally, our results may include false negatives due to the strict analytical parameters used. In particular, we focused on relatively large CNVs ($>$100 Kb) in this study. Given that the median size of CNVs in humans has been estimated as up to be 20 Kb,[22,48] a significant proportion of CNVs might not have been analyzed.

In spite of these limitations, this is the first study to our knowledge that explores genome-wide CNV association with PD.

In conclusion, (i) individuals with PD did not exhibit an increased genome-wide rare CNV burden and (ii) common duplications in 16p11.2 were associated with PD in a Japanese population.

ACKNOWLEDGEMENTS

We thank Assoc Prof K Iwamoto of the Department of Molecular Psychiatry in the Graduate School of Medicine at the University of Tokyo for his suggestions. Some of the data used for this research are available on the DB Development for Medical Application from Disease Analysis website (https://gwas.lifescience-cedb.jp/index.html), which is part of the Integrated Database Project funded by Japan's Ministry of Education, Culture, Sports, Science and Technology. This study was supported by Grants-in-Aid for Scientific Research (number 17019029; number 21300242) from Japan's Ministry of Education, Culture, Sports, Science and Technology (to T Sasaki).

1 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn, (American Psychiatric Press: Washington, DC, 1994).
2 Kessler, R., Stang, P., Wittchen, H., Ustun, B., Roy-Byrne, P. & Walters, E. Lifetime panic-depression comorbidity in the National Comorbidity Survey. *Arch. Gen. Psychiatry* **55,** 801–808 (1998).
3 Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G. *et al.* The cross-national epidemiology of panic disorder. *Arch. Gen. Psychiatry* **54,** 305–309 (1997).
4 Goldstein, R. B., Wickramaratne, P. J., Horwath, E. & Weissman, M. M. Familial aggregation and phenomenology of 'early'-onset (at or before age 20 years) panic disorder. *Arch. Gen. Psychiatry* **54,** 271–278 (1997).
5 Hettema, J. M., Neale, M. C. & Kendler, K. S. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am. J. Psychiatry* **158,** 1568–1578 (2001).
6 Maron, E., Hettema, J. M. & Shlik, J. Advances in molecular genetics of panic disorder. *Mol. Psychiatry* **15,** 681–701 (2010).
7 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).
8 Cook, E. H. Jr & Scherer, S. W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455,** 919–923 (2008).
9 Lesch, K. P., Selch, S., Renner, T. J., Jacob, C., Nguyen, T. T., Hahn, T. *et al.* Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. *Mol. Psychiatry* **16,** 491–503 (2011).
10 Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43,** 838–846 (2011).
11 Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83,** 1–12 (2008).
12 Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E. *et al.* The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59** (Suppl 20), 22–33 Quiz, 34–57 (1998).
13 Otowa, T., Sugaya, N., Inoue, K., Shimada, T., Kawamura, Y., Tochigi, M. *et al.* Replication of a genome-wide association study of panic disorder in a Japanese population. *J. Hum. Genet.* **55,** 91–96 (2010).
14 Koike, A., Nishida, N., Inoue, I., Tsuji, S. & Tokunaga, K. Genome-wide association database developed in the Japanese Integrated Database Project. *J. Hum. Genet.* **54,** 543–546 (2009).
15 Kawamura, Y., Liu, X., Akiyama, T., Shimada, T., Otowa, T., Sakai, Y. *et al.* Association between oxytocin receptor gene OXTR polymorphisms and affective temperaments as measured by TEMPS. *J. Affect. Disord.* **127,** 31–37 (2010).
16 Adachi, S., Tajima, A., Quan, J., Haino, K., Yoshihara, K., Masuzaki, H. *et al.* Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. *J. Hum. Genet.* **55,** 816–821 (2010).
17 Otowa, T., Kawamura, Y., Nishida, N., Sugaya, N., Yoshida, E., Inoue, K. *et al.* Meta-analysis of genome-wide association studies followed by a replication study implicates novel loci for panic disorder. *Mol. Psychiatry* ref type: unpublished work submitted for publication.
18 Korn, J. M., Kuruvilla, F. G., McCarroll, S. A., Wysoker, A., Nemesh, J., Cawley, S. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40,** 1253–1260 (2008).
19 Zhang, D., Cheng, L., Qian, Y., Alliey-Rodriguez, N., Kelsoe, J. R., Greenwood, T. *et al.* Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol. Psychiatry* **14,** 376–380 (2009).
20 Weiss, L. A., Shen, Y., Korn, J. M., Arking, D. E., Miller, D. T., Fossdal, R. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
21 Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., Robson, S. *et al.* Genome-wide association study of CNVs in 16000 cases of eight common diseases and 3000 shared controls. *Nature* **464,** 713–720 (2010).
22 Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. *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).
23 McCarroll, S. A. Extending genome-wide association studies to copy-number variation. *Hum. Mol. Genet.* **17,** R135–R142 (2008).
24 Affymetrix. Quality control assessment in genotyping console, revision 1. *Affymetrix White Paper* 1–12 (2008).
25 Miyagawa, T., Nishida, N., Ohashi, J., Kimura, R., Fujimoto, A., Kawashima, M. *et al.* Appropriate data cleaning methods for genome-wide association study. *J. Hum. Genet.* **53,** 886–893 (2008).
26 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
27 International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
28 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
29 Bassett, A. S., Scherer, S. W. & Brzustowicz, L. M. Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *Am. J. Psychiatry* **167,** 899–914 (2010).
30 Ohashi, J., Yamamoto, S., Tsuchiya, N., Hatta, Y., Komata, T., Matsushita, M. *et al.* Comparison of statistical power between 2 * 2 allele frequency and allele positivity tables in case-control studies of complex disease genes. *Ann. Hum. Genet.* **65,** 197–206 (2001).
31 Kim, W., Gordon, D., Sebat, J., Ye, K. Q. & Finch, S. J. Computing power and sample size for case-control association studies with copy number polymorphism: application of mixture-based likelihood ratio test. *PLoS One* **3,** e3475 (2008).
32 Ohashi, J. A practical case-control association test for detecting a susceptibility allele at a copy number variation locus. *J. Hum. Genet.* **54,** 169–173 (2009).
33 Zhang, Z., Harrison, P. M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13,** 2541–2558 (2003).
34 Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39,** D152–D157 (2011).
35 Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34,** D158–D162 (2006).
36 He, S., Liu, C., Skogerbo, G., Zhao, H., Wang, J., Liu, T. *et al.* NONCODE v2.0: decoding the non-coding. *Nucleic Acids Res.* **36,** D170–D172 (2008).
37 Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146,** 353–358 (2011).
38 Yan, Z., Hu, H. Y., Jiang, X., Maierhofer, V., He, L., Hu, Y. *et al.* Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* **39,** 6596–6607 (2011).
39 Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–747 (2007).
40 Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S. *et al.* Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. *Science* **332,** 848–852 (2011).
41 Inoue, K. & Lupski, J. R. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3,** 199–242 (2002).
42 Sharp, A. J., Hansen, S., Selzer, R. R., Cheng, Z., Regan, R., Hurst, J. A. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.* **38,** 1038–1042 (2006).
43 Bijlsma, E. K., Gijsbers, A. C. J., Schuurs-Hoeijmakers, J. H. M., van Haeringen, A., Fransen van de Putte, D. E., Anderlid, B. M. *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. *Eur. J. Med. Genet.* **52,** 77–87 (2009).
44 McCarthy, S., Makarov, V., Kirov, G., Addington, A., McClellan, J., Yoon, S. *et al.* Microduplications of 16p11.2 are Associated with Schizophrenia. *Nat. Genet.* **41,** 1223–1227 (2009).
45 Schaaf, C. P., Goin-Kochel, R. P., Nowell, K. P., Hunter, J. V., Aleck, K. A., Cox, S. *et al.* Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: three patients with syringomyelia. *Eur. J. Hum. Genet.* **19,** 152–156 (2011).
46 Dellinger, A. E., Saw, S. M., Goh, L. K., Seielstad, M., Young, T. L. & Li, Y. J. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.* **38,** e105 (2010).
47 Koike, A., Nishida, N., Yamashita, D. & Tokunaga, K. Comparative analysis of copy number variation detection methods and database construction. *BMC Genet.* **12,** 29 (2011).
48 Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464,** 704–712 (2010).

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