G&I Genomics & Informatics

pISSN 1598-866X eISSN 2234-0742
Genomics Inform 2012;10(3):194-199
http://dx.doi.org/10.5808/GI.2012.10.3.194

ORIGINAL ARTICLE

Effect of Combining Multiple CNV Defining Algorithms on the Reliability of CNV Calls from SNP Genotyping Data

Soon-Young Kim[1,2], Ji-Hong Kim[1,2], Yeun-Jun Chung[1,2]*

[1]Integrated Research Center for Genome Polymorphism, The Catholic University of Korea School of Medicine, Seoul 137-701, Korea, [2]Department of Microbiology, The Catholic University of Korea School of Medicine, Seoul 137-701, Korea

In addition to single-nucleotide polymorphisms (SNP), copy number variation (CNV) is a major component of human genetic diversity. Among many whole-genome analysis platforms, SNP arrays have been commonly used for genomewide CNV discovery. Recently, a number of CNV defining algorithms from SNP genotyping data have been developed; however, due to the fundamental limitation of SNP genotyping data for the measurement of signal intensity, there are still concerns regarding the possibility of false discovery or low sensitivity for detecting CNVs. In this study, we aimed to verify the effect of combining multiple CNV calling algorithms and set up the most reliable pipeline for CNV calling with Affymetrix Genomewide SNP 5.0 data. For this purpose, we selected the 3 most commonly used algorithms for CNV segmentation from SNP genotyping data, PennCNV, QuantiSNP; and BirdSuite. After defining the CNV loci using the 3 different algorithms, we assessed how many of them overlapped with each other, and we also validated the CNVs by genomic quantitative PCR. Through this analysis, we proposed that for reliable CNV-based genomewide association study using SNP array data, CNV calls must be performed with at least 3 different algorithms and that the CNVs consistently called from more than 2 algorithms must be used for association analysis, because they are more reliable than the CNVs called from a single algorithm. Our result will be helpful to set up the CNV analysis protocols for Affymetrix Genomewide SNP 5.0 genotyping data.

Keywords: CNV defining algorithm, DNA copy number variations, SNP array

Introduction

Human genome variation has facilitated the understanding of inter-individual phenotypic differences [1, 2]. In addition to single-nucleotide polymorphisms (SNPs), it is widely accepted that large-scale DNA structural variation, termed copy number variation (CNV), is a major component of human genetic diversity [2, 3]. Genomewide SNP genotyping data can be used for CNV calling [4]; therefore, if we can get reliable CNV calls from the SNP genotyping data, a CNV-based genomewide association study (GWAS) can be realized. Among many whole-genome CNV analysis platforms, SNP arrays have been suggested as a resource for CNV discovery due to their ubiquitous genome coverage and relatively advantageous resolution. However, despite the importance of CNV-disease association analysis, CNV calling from SNP genotyping data has not been well established. Affymetrix Genomewide SNP 5.0 is one of the commonly used SNP array platforms for SNP-GWAS as well as CNV analysis [5]. We previously validated the accuracy and reproducibility of CNVs called from Affymetrix SNP array 5.0 data by comparing the CNV calls from 3 different array platforms using NEXUS software: Affymetrix SNP array 5.0, Agilent 2X244K CNV array, and NimbleGen 2.1M CNV array [6].

Recently, a number of CNV defining algorithms have been developed, which have facilitated the CNV-based GWAS [7-14]. However, due to the fundamental limitation of SNP genotyping data for the measurement of signal intensity, there are still concerns regarding the possibility of false discovery or low sensitivity for detecting CNVs [15, 16]. Indeed, CNV calling is dependent on the types of array

Received August 10, 2012; Revised August 20, 2012; Accepted August 23, 2012

***Corresponding author:** Tel: +82-2-2258-7343, Fax: +82-2-537-0572, E-mail: yejun@catholic.ac.kr

platforms and analytic tools. Each platform and calling algorithm has its own advantages and disadvantages; so, one single algorithm or array platform is not always best for determination of CNVs [17-19]. Recently Pinto *et al*. [20] showed that different analytic tools applied to the same raw data typically yielded CNV calls with <50% concordance and, using multiple algorithms, minimize the number of false discoveries. To remedy the potential limitations of SNP array for CNV detection, more than one way of CNV calling by using several different segmentation algorithms are performed, and overlapped calls are used for GWAS analysis [21-24].

In this study, we tried to verify the effect of adopting multiple CNV calling algorithms and set up the most reliable pipeline for CNV calling with Affymetrix Genomewide SNP 5.0 data. We selected the 3 most commonly used algorithms for CNV segmentation from SNP genotyping data, PennCNV, QuantiSNP, and BirdSuite. After defining the CNV loci using the 3 different algorithms, we assessed how many of them overlapped with each other, and we also validated the CNVs by genomic quantitative PCR (qPCR). Finally we concluded that CNVs that were consistently called from more than 2 different calling algorithms are more reliable than the CNVs called from a single algorithm. Our result will be helpful to set up the CNV analysis protocols for Affymetrix Genomewide SNP 5.0 genotyping data.

Methods

Study materials

We used Affymetrix Genomewide SNP 5.0 genotyping data provided by the Korea Association Resource (KARE) consortium, Korean Genome Epidemiology Study (KoGES). As a control, we purchased a HapMap cell line, GM10851, from Coriell Institute for Medical Research (Camden, NJ, USA) and extracted genomic DNA using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany).

Pre-processing SNP array data

Before CNV calling procedures, all required pre-processing procedures, including allele correction, summarization, and background correction, were performed as described previously by using the software provided by Affymetrix [5]. In brief, background-corrected data were normalized using quantile normalization and summarized by median polish.

CNV calling

For this study, Affymetrix 5.0 SNP array (Affymetrix, Santa Clara, CA, USA) data of 10 subjects were randomly selected from the KARE dataset for CNV calling. Affymetrix 5.0 data of NA10851 was used as a control. For defining CNVs, we choose 3 different segmentation algorithms, PennCNV [25], QuantiSNP [26] and BirdSuite [27]. PennCNV implements a hidden Markov model (HMM) and considers SNP allelic ratio distribution in addition to signal intensity. We only used CNVs from PennCNV with samples that had a standard deviation of log R ratio (LRR) smaller than 0.2, drift values of b-allele frequency (BAF_drift) smaller than 0.01, and waviness factor between −0.05 and 0.05. For QuantiSNP, we did not apply any criteria for sample filtering, since QuantiSNP uses an Objective Bayes Hidden-Markov Model (OB-HMM) for calibrating the model for fixing the false positive error rate and maximum marginal likelihood to set other hyperparameters. It was originally developed for detecting CNVs from BeadArray SNP genotyping data of Illumina, but Affymetrix data also could be processed as long as they have probe signal intensities and B-allele frequencies. Birdsuite uses exclusive copy number analysis routine (Canary) for CNP locus and HMM for finding rare CNVs. It generates a logarithm of the odds ratio (LOD) value for each CNV that describes the likelihood of a CNV relative to no CNV over a given interval, including flanking sequences. Of the BirdSuite calls, only the CNVs that have an LOD value smaller than 2 were used for further analysis. All 3 methods were used with default values, and no other option was added or changed.

Validation

To estimate the false discovery rate of our CNV-calling algorithm, CNVs were validated by genomic real-time qPCR. For this purpose, we randomly selected 25 CNV loci and performed genomic qPCR using DNAs from study subjects who showed corresponding CNVs on that loci. The PCR primers used in this study were designed using the PrimerQuest program (http://www.idtdna.com/Scitools/Applications/Primerquest). To verify the specificity of the PCR reactions under the unified denaturation temperature (60°C), we performed PCR and agarose gel electrophoresis for each primer set. We also screened the University of California Santa Cruz (UCSC) database (http://genome.ucsc.edu/) to confirm the unique sequence without any repeat sequences in the primers. Sequence information of the primers for genomic qPCR validation is listed in Table 1.

Ten microliters of the reaction mixture contained 10 ng of genomic DNA, SYBR Premix Ex Taq TM II (TaKaRaBio, Shiga, Japan), 1× ROX (Toyobo, Osaka, Japan), and 10 pmol of primers. Thermal cycling conditions consisted of 1 cycle of 10 s at 95°C, followed by 40 cycles of 5 s at 95°C, 10 s at 61°C, and 20 s at 72°C. All PCR experiments were repeated twice, and amplification efficiencies for both target and reference genes were evaluated using a standard curve over 1 : 5 serial

Table 1. Sequence information of the primers for genomic qPCR validation

Chr	Start	End	Length (bp)	Forward	Reverse
1	108,538,343	108,538,461	119	AGGAGGTTGCACCATGGTTAGTCA	GGCCACAGCACATCTTGTGAAACA
1	150,842,026	150,842,185	160	AACCATGGACTTCGTGGGTAGTCA	CATGCTCCATGCATTGTGGTGGAA
3	259,837	259,952	116	CAAATGGAACAGCAGGGTCAGCAA	TGCTGTGTCCAGCATCCTATGTGT
4	69,073,047	69,073,213	167	TTGTTGGAGGAACAAAGCCCAACC	TGGCTGGTGTCTGTTCTGATTGGT
5	92,610,728	92,610,891	164	AGGTTGATGAGCCACACAGGGTAT	TGCTCCTGAATTCCTCAGCTTCCA
5	178,045,652	178,045,850	199	AGGCAAGAGGTAGCCCACCTTAAT	AAAGCAGGAGCTGAGAGGCAGAAA
7	141,704,582	141,704,737	156	AAACAGACAGGCACTGGTCCATCT	ATGGCATAACCTCCATCCCACTCA
7	141,711,867	141,712,039	173	TGAGACTGTGGATCTTTGGCCACT	TAATTCCACATGTCCAGGCCCACT
7	154,025,113	154,025,228	116	TGCAATGGCACGATCTTGTCTCAC	AGGCATGATGGTGGGTGCCTATAA
11	55,162,545	55,162,631	87	TCTATCACGTGCACCCAGCTCATT	TGTGGATGTGTAGCAAAGGTCGGA
11	55,207,777	55,207,961	185	TGCACTACACCATCATCACGACCA	ATCAATGCGAGCCAACTTCAGCAG
14	81,569,884	81,569,974	91	TGCATGTTAGGAGGCTGTGGATCA	TGAGGCAGAACAATGTGGCCTCTA
14	81,569,895	81,569,974	80	AGGCTGTGGATCAATACGGGTTCA	TGAGGCAGAACAATGTGGCCTCTA
16	54,363,707	54,363,878	172	AGCCGCATCTGTAGTCCTGAAAGT	GTTCCCTCCAAAGCTGGCAATGTT
5	814,195	814,318	124	ACCTCGGCCGGATTCTGGATTAAA	ACCTTCATGGCAGGTGAGAAGACA
9	44,685,893	44,685,973	81	ATGACAGACAGGACCCAACACCAT	TCAACAATAGGGCAGAGGAAGCCA
15	19,145,532	19,145,684	153	GGCGCAGTGGTTCATGCTTGTAAT	TGCGTACCACCATGCCTAGCTAAA
15	19,833,989	19,834,167	179	TGCCTAAGCTGTGTTACTCTGCCA	CGCAAAGGGTTACAGATGGCAACA
20	1,531,337	1,531,532	196	ATCACCCAATTGCGGACTCCTCTT	ACAGACTCTACGGCGTTGGCTTTA
22	22,702,643	22,702,799	157	GCCTGACTTCGAAATGGTGGCAAA	TGGTTGCCTGGTTTCTAGCCCTAT
2	87,557,596	87,557,722	127	TGAGAGGCAGGTGGATTTGGATGT	TGGAAGACACACAGCGAACCTCTT
4	70,207,254	70,207,437	184	ACCTCAAATTTCAGTGCCGAAGGC	TCTTCTGTGCTGGCTGTGGATTCT
15	22,237,790	22,237,886	97	AGCTCAGGAGATGAAAGGGCACAT	TCTGCCTGAAGCAAGTGTACCTGT
15	19,843,004	19,843,164	161	AGACTTGCCTTCTTGTCAGCCTCA	AGACAGGGCAGGAAGAACTTTCCA
19	48,112,193	48,112,355	163	TCCTGCATCCTCCTGTGTGACATT	TGGCTACATCTGGTACAAAGGGCA

qPCR, quantitative PCR; Chr, chromosome.

Table 2. General characteristics of the copy number variation (CNV) calls from 3 algorithms

		No filtration			>7 consecutive proves		
		PennCNV	QuantiSNP	BirdSuite	PennCNV	QuantiSNP	BirdSuite
Total no. of calls		814	1,087	1,362	295	311	385
Size	Average	44,672.6	39,911	15,466.5	104,264	103,782	34,463.9
	Median	12,981	7,599	7,964	50,799	50,597	23,528
Small CNV	<1 kbp	49 (6.0%)	272 (25.0%)	193 (14.2%)	0	0	0

dilutions. The copy number of each target was defined as $2^{-\Delta\Delta CT}$, where ΔCt is the difference in threshold cycles for the sample in question normalized against the reference gene (RNaseP) and expressed relative to the value obtained by calibrator DNA (NA10851 and Promega DNA), as described elsewhere [28].

Results and Discussion

The general characteristics of the CNV calls from the 3 algorithms are shown in Table 2. The total number of CNV calls from the 3 algorithms was similar-range 814 to 1,362. The median size of the CNVs called by PennCNV (12.9 kb) were bigger than the other 2 algorithms (7.5 kb and 7.9 kb). The size of the CNVs consistently detected by 2 or more algorithms was generally bigger than that of the CNVs detected by a single method. For example, the average sizes of the CNVs that were detected from all 3 algorithms, 2 algorithms, and 1 method were 107 kb, 73.2 kb, and 15.9 kb, respectively (Table 3). Fig. 1 shows the examples of different sizes of CNVs according to the LRR distribution of gain- and loss-type CNV regions, depending on how many algorithms detected the CNV in that region. These data suggest that larger-sized CNVs are generally more prominent; so, they can be relatively easily detected by any segmentation method, regardless of the algorithm. But, detection of the smaller-sized CNVs can be affected more easily by the characteristics of the CNV defining algorithm; so, they can

Table 3. Number of copy number variations (CNVs) consistently defined from 3 different algorithms

	No filtration					>7 consecutive proves				
	Size	CNV	PennCNV	QuantiSNP	BirdSuite	Size	CNV	PennCNV	QuantiSNP	BirdSuite
All overlap	107.2	40	40	42	40	208	14	14	15	15
2 overlap	73.2	425	356	424	91	130	231	212	223	14
Unique	15.9	2,270	418	621	1,231	40	498	69	73	356

Size: average size (kb).

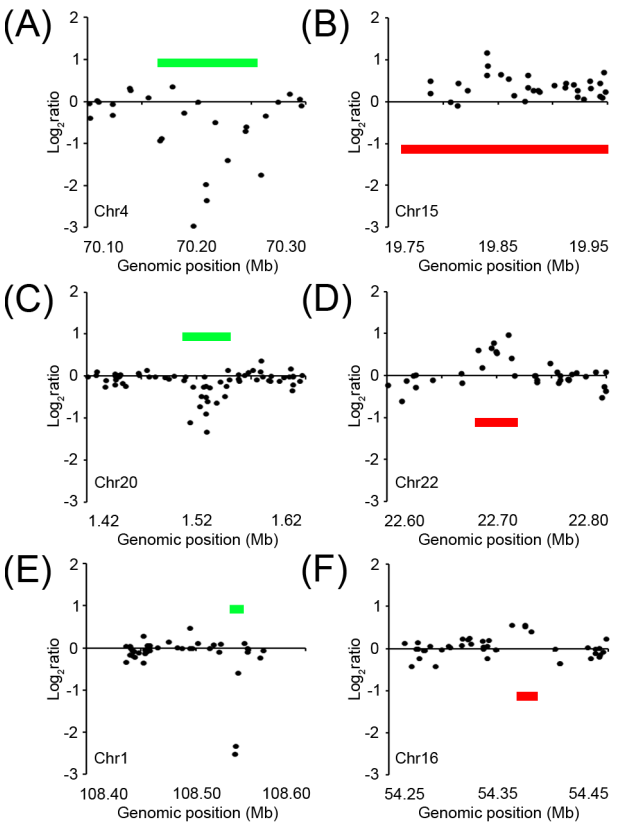

Fig. 1. Signal intensity (Log R ratio) plots of the copy number variation (CNV) regions. (A, B) CNVs detected from all 3 algorithms. (C, D) CNVs detected from 2 of the 3 algorithms. (E, F) CNVs uniquely detected by a single algorithm. Green bars, copy number-loss CNV regions; red bars, copy number-gained CNV regions.

be detected by 1 particular algorithm.

A recent report examining the impact of inaccuracy of CNV detection from SNP genotyping data revealed that CNVs, defined as the copy number changes of >7 consecutive probes, were fairly reliable in case of deletion type CNVs [29]. In our previous study exploring the CNV profiles of Koreans using an Affymetrix array, we suggested the filtering condition of CNV call as >6 consecutive probes to be reliable [5]. Therefore, to make more reliable conditions in this study, we filtered the CNV calls as >7 consecutive probes. Under this criterion, the number of CNV calls from the 3 algorithms was reduced down from 295 to 385, while the median size became bigger-50 kb by PennCNV and QuantiSNP and 23.5 kb by BirdSuite (Table 2). In this condition, same as above, the size of the CNVs consistently detected by 2 or more algorithms was generally bigger than that of the CNVs detected by a single method (Table 3).

In terms of the number of CNVs consistently identified by different algorithms, of the CNV calls using the PennCNV algorithm, 48.6% (396/814 CNVs) was detected by 2 or more algorithms, and only 4.9% (40/814 CNVs) was detected by all 3 algorithms. Similarly, of the CNV calls using QuantiSNP, 42.9% (466/1,087 CNVs) was detected by 2 or more algorithms and only 3.9% (42/1,087 CNVs) was detected by all 3 algorithms. However, in the case of the BirdSuite algorithm, although the number of CNV calls was the biggest, only 9.6% (131/1,362 CNVs) was detected by 2 or more algorithms and 2.9% (40/1,362 CNVs) was detected by all 3 algorithms. Taken together, only 17% of the total CNV calls (465/2,735 CNVs) were defined by more than 2 algorithms, and only 1.5% (40/2,735 CNVs) was defined by all 3 algorithms (Table 3). In the filtering condition of >7 consecutive probes, the number of CNVs consistently identified by different algorithms was improved. For example, of the CNV calls using the PennCNV algorithm, 76.6% (226/295 CNVs) was detected by 2 or more algorithms and only 23.4% (69/295 CNVs) was detected uniquely by PennCNV. QuantiSNP calls showed a similar trend. However, even under this condition, 92.5% (356/385 CNVs) of BirdSuite calls were unique (Table 3). All the CNVs consistently detected by 2 algorithms were from PennCNV and QuantiSNP, while no CNVs that were detected by the BirdSuite algorithm were consistently detected by 2 algorithms, with only 1 exception. These data suggest that CNV calls from SNP genotyping depend substantially on the characteristics of calling algorithms, and only part of the CNV calls from each algorithm seems to be reliable. Therefore, CNVs identified from SNP genotyping data must be validated as completely as possible, and especially in the case of using a single calling algorithm, the CNVs must be validated more carefully. Filtering the CNV

Table 4. Genomic qPCR validation of the CNV calls from SNP genotyping data

Subject no.	Unique		2 overlaps		3 overlaps	
	CNV call	qPCR	CNV call	qPCR	CNV call	qPCR
No filtration						
1	9	3	2	1	0	0
2	0	0	3	2	1	1
3	4	1	4	4	0	0
4	7	3	2	2	3	2
5	4	1	4	3	0	0
6	3	3	5	4	0	0
7	4	2	3	2	1	1
8	6	3	5	1	1	1
9	4	1	5	0	0	0
10	6	1	0	0	1	0
	47	18 (38.3%)	33	19 (57.6%)	7	5 (71.4%)
>7 consecutive proves						
1	6	3	1	1	0	0
2	0	0	3	2	1	1
3	1	1	4	4	0	0
4	3	2	1	1	3	2
5	2	0	3	3	0	0
6	0	0	5	4	0	0
7	1	1	3	2	1	1
8	2	2	5	1	1	1
9	3	1	4	0	0	0
10	5	1	0	0	0	0
	23	11 (47.8%)	29	18 (62.1%)	6	5 (83.3%)

qPCR, number of consistently detected CNVs by quantitative PCR; CNV call, number of copy number variation calls; SNP, single-nucleotide polymorphism.

calls by the number of consecutive probes can improve the reliability of CNV calls.

To validate the CNVs identified from 3 different algorithms in this study, we randomly selected 25 CNV regions across the whole chromosomes. The genomic qPCR validation results are listed in Table 4. In case of the CNVs identified by all 3 algorithms, 71.4% (5/7) of the consistency was observed between the CNV call and genomic qPCR result. Of the CNVs identified by more than 2 algorithms, 57.6% (19/33) was consistent. Of the unique CNV calls defined by just a single algorithm, only 38.3% (18/47) was consistent, but the other 61.7% (28/46) was not consistent. In the case of the CNVs defined as >7 consecutive probes, the consistency was generally improved-47.8% (unique CNVs), 62.2% (CNVs identified by more than 2 algorithms), and 83.3% (CNVs identified by all 3 algorithms). These results indicate that CNV calls from 2 or more algorithms are more reliable than those from single algorithms. CNV calls from all 3 algorithms are, of course, the most reliable, but this is too stringent; so, the number of CNVs is not applicable for GWAS analysis.

In this study, we aimed to verify the effect of adopting multiple CNV calling algorithms and set up the most reliable pipeline for CNV calling with Affymetrix Genomewide SNP 5.0 data. We found that CNVs defined by a single CNV calling algorithm may not reliable enough for further GWAS study, regardless of the types of algorithms. Based on our findings, we propose that for reliable CNV-based GWAS using SNP array data, CNV calls must be performed with at least 3 different algorithms, and the CNVs consistently called from more than 2 methods must be used for association analysis.

Acknowledgments

This study was supported by a grant from the Korea Healthcare Technology R&D Project (A092258) and the Korea Health 21 R&D Project (A040002), Ministry of Health and Welfare, Republic of Korea.

References

1. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;39(7 Suppl):S37-S42.
2. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006;7:85-97.
3. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 2007;3:1787-1799.
4. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, *et al*. Global variation in copy number in the human genome. *Nature* 2006;444:444-454.
5. Yim SH, Kim TM, Hu HJ, Kim JH, Kim BJ, Lee JY, *et al*. Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum Mol Genet* 2010; 19:1001-1008.
6. Kim JH, Jung SH, Hu HJ, Yim SH, Chung YJ. Comparison of the Affymetrix SNP Array 5.0 and oligoarray platforms for defining CNV. *Genomics Inform* 2010;8:138-141.
7. Lai WR, Johnson MD, Kucherlapati R, Park PJ. Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* 2005;21:3763-3770.
8. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res* 2010;38:e105.
9. Barnes C, Plagnol V, Fitzgerald T, Redon R, Marchini J, Clayton D, *et al*. A robust statistical method for case-control association testing with copy number variation. *Nat Genet* 2008;40: 1245-1252.

10. Forer L, Schönherr S, Weissensteiner H, Haider F, Kluckner T, Gieger C, *et al*. CONAN: copy number variation analysis software for genome-wide association studies. *BMC Bioinformatics* 2010;11:318.
11. Pique-Regi R, Cáceres A, González JR. R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics* 2010;11:380.
12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.
13. Subirana I, Diaz-Uriarte R, Lucas G, Gonzalez JR. CNVassoc: association analysis of CNV data using R. *BMC Med Genomics* 2011;4:47.
14. Kim JH, Hu HJ, Yim SH, Bae JS, Kim SY, Chung YJ. CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics* 2012;28:1790-1792.
15. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic* 2009;8: 353-366.
16. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, *et al*. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;453:56-64.
17. Baumbusch LO, Aarøe J, Johansen FE, Hicks J, Sun H, Bruhn L, *et al*. Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* 2008;9:379.
18. Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, *et al*. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* 2009;10:588.
19. Hester SD, Reid L, Nowak N, Jones WD, Parker JS, Knudtson K, *et al*. Comparison of comparative genomic hybridization technologies across microarray platforms. *J Biomol Tech* 2009; 20:135-151.
20. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, *et al*. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 2011;29:512-520.
21. Ramayo-Caldas Y, Castelló A, Pena RN, Alves E, Mercadé A, Souza CA, *et al*. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics* 2010; 11:593.
22. Degenhardt F, Priebe L, Herms S, Mattheisen M, Mühleisen TW, Meier S, *et al*. Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *Am J Med Genet B Neuropsychiatr Genet* 2012;159B:263-273.
23. Marenne G, Rodríguez-Santiago B, Closas MG, Pérez-Jurado L, Rothman N, Rico D, *et al*. Assessment of copy number variation using the Illumina Infinium 1M SNP-array: a comparison of methodological approaches in the Spanish Bladder Cancer/EPICURO study. *Hum Mutat* 2011;32:240-248.
24. Kawamura Y, Otowa T, Koike A, Sugaya N, Yoshida E, Yasuda S, *et al*. A genome-wide CNV association study on panic disorder in a Japanese population. *J Hum Genet* 2011;56: 852-856.
25. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al*. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665-1674.
26. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, *et al*. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007;35:2013-2025.
27. Korn JM, Kuruvilla FG, McCarroll SA, 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* 2008;40:1253-1260.
28. Yim SH, Chung YJ, Jin EH, Shim SC, Kim JY, Kim YS, *et al*. The potential role of *VPREB1* gene copy number variation in susceptibility to rheumatoid arthritis. *Mol Immunol* 2011;48: 1338-1343.
29. Wineinger NE, Tiwari HK. The impact of errors in copy number variation detection algorithms on association results. *PLoS One* 2012;7:e32396.