Research Article

Copy-Number Variations Observed in a Japanese Population by BAC Array CGH: Summary of Relatively Rare CNVs

Yasunari Satoh,[1] Keiko Sasaki,[1] Yuko Shimoichi,[1] Keiko Sugita,[2] Hiroaki Katayama,[2] and Norio Takahashi[1]

[1] *Department of Genetics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan*
[2] *Department of Information Technology, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan*

Correspondence should be addressed to Norio Takahashi, takahash@rerf.or.jp

Copy-number variations (CNVs) may contribute to genetic variation in humans. Reports regarding existence and characteristics of CNVs in a large apparently healthy Japanese cohort are quite limited. We report the data from a screening of 213 unrelated Japanese individuals using comparative genomic hybridization based on a bacterial artificial chromosome microarray (BAC aCGH). In a previous paper, we summarized the data by focusing on highly polymorphic CNVs (in ≥5.0 % of the individuals). However, rare variations have recently received attention from scientists who espouse a hypothesis called "common disease and rare variants." Here, we report CNVs identified in fewer than 10 individuals in our study population. We found a total of 126 CNVs at 52 different BAC regions in the genome. The CNVs observed at 27 of the 52 BAC-regions were found in only one unrelated individual. The majority of CNVs found in this study were not identified in the Japanese who were examined in the other studies. Family studies were conducted, and the results demonstrated that the CNVs were inherited from one parent in the families.

1. Introduction

Genomes vary from one another in multifarious ways, and the totality of this genetic variation underpins the heritability of human traits. Various recent studies show the landscape of genetic variation and allow estimation of the relative contributions of sequence (base substitutions) and structural variation (indels, insertions, or deletions; copy-number variations [CNVs], and inversions). Among DNA sequence variations in the human genome, CNVs directly contribute to gene expression changes through gene dosage effects; the distribution of CNVs is now considered to be wider than previously thought [1–6]. Genome resequencing studies have shown that CNVs, involving the gain or loss of several hundreds of bases to several hundred kilobases (kb) of the genome, can be an important source of genetic variation among human populations of different ethnic groups as well as among individuals. Following the development of methodologies and the introduction of new research platforms [7–12], information regarding the nature and pattern of CNVs

from representative populations has been accumulated. The functional impact of CNVs has been demonstrated across the full range of biology [13], from cellular phenotypes, such as gene expression [14], to all classes of human disease with an underlying genetic basis: sporadic, Mendelian, complex, and infectious (reviewed in [15]).

Clinical geneticists need to discriminate pathogenic from benign CNVs in their patients and have made extensive use of data from CNV surveys of apparently healthy individuals [16]. The mere presence or absence of a variant in such control data sets is only partially informative, as determination of the pathogenicity of inherited CNVs is at present limited by a lack of information on CNV frequency and combinations in apparently healthy individuals. Based on these considerations, examinations of a relatively large number of individuals from various specific ethnic groups have recently been conducted using different array platforms, such as BAC arrays [6, 17, 18], oligo arrays [19–21], and others [22]. The results are not always consistent, and it is likely that different human populations bear different inherited CNVs. The

numbers of Japanese individuals examined to date are smaller than those in studies of other ethnicities [4]. Polymorphic CNVs have received considerable attention since they might play an important role in the etiology of common diseases. Therefore, more data regarding CNVs should be accumulated from Japanese populations. We summarized the available data by focusing on highly polymorphic CNVs (in ≥5.0% of the individuals) in a previous paper [23]. However, rare variations have recently received attention from scientists who espouse a hypothesis called "common disease and rare variants," in contrast to the "common disease-common variants" hypothesis that underpins most genome-wide association studies. In this paper, we focus on CNVs observed at relatively low frequency (<5.0% of the individuals) in a population residing in Hiroshima and Nagasaki, Japan, using CGH with BAC clones as targets.

2. Materials

The study was conducted in two stages: Stage (1) 80 unrelated Japanese individuals were examined by BAC-aCGH with an array having 2,241 BAC clones; Stage (2) 133 unrelated Japanese individuals were examined by BAC-aCGH containing 2,622 BAC clones.

2.1. Target BAC Clones. The majority of the clones used in Stage 1 of this study were selected from the set of cytogenetically mapped P1 artificial chromosome (PAC) clones and bacterial artificial chromosome (BAC) clones reported by the BAC Resource Consortium [26] and obtained from either the Children's Hospital Oakland Research Institute (Oakland, CA, USA) or from Invitrogen Inc., Co. (Carlsbad, CA, USA). In Stage 2, together with the BAC clones in Stage 1, additional 381 BAC clones were used, a majority of which were collaboratively obtained from Dr. Matsumoto of Yokohama City University. A total of 2,241 clones of chromosomal fragments from chromosome 1 to chromosome 22 were used in Stage 1 and 2,622 clones in Stage 2. The clones were distributed every 1.2 Mb across all human autosomes in Stage 1 and every 1.1 Mb in Stage 2. In addition to autosomal clones, four kinds of X-chromosomal clones were used as internal references. With respect to the examinations in Stage 1, three sets of arrays were constructed as described in the previous paper [27]. For Stage 2, all of the clones were printed onto one glass slide.

2.2. Genomic DNA Samples. The genomic DNA samples used in this study were reported in a previous study [23, 27]. In brief, the DNA used for the population studies and the family studies was extracted from lymphoblastoid cell lines obtained from the offspring of atomic-bomb survivors and their family members. Lymphoblastoid cell lines were derived from a cryopreserved archive of families consisting of father, mother, and offspring from Hiroshima and Nagasaki for whom permanent cell lines have been established by Epstein-Barr (EB) virus transformation. The genomic DNA was isolated using conventional methods as described in detail elsewhere [28]. A total of 305 of the offsprings were screened initially. Since the offspring in some cases included

siblings, we selected one representative offspring to create a group of "unrelated individuals," thereby avoiding any duplication of CNVs from families containing two or more siblings, in accordance with the rules described in a previous report [23]. In this study, 80 individuals (40 from each city) were included in examinations in Stage 1 and 133 (84 from Hiroshima, 49 from Nagasaki) in Stage 2. The individuals provided their informed consent prior to the study. The Institutional Review Board of our foundation approved this study.

3. Methods

3.1. Array Preparation. The arrays were prepared as described in the previous paper [23]. In brief, with respect to Stage 1, cloned DNA was digested by *Not*I. On the other hand, in Stage 2, cloned DNA was digested with *Mse*I, and the fragmented DNAs were amplified by ligation-mediated PCR carried out as described by Snijders et al. [29]. In both stages, the target DNAs ($0.5\,\mu g/\mu L$) were dissolved in 50% dimethylsuloxide and printed in triplicate onto the glass slides (Matsunami Glass Co. Ltd.) using the Affymetrix 417 Arrayer (Affymetrix).

3.2. Screening by the Array CGH Method. The screenings of both stages were conducted following the procedures described previously [23]. In brief, for labeling DNA, test and reference genomic DNA ($1.25\,\mu g$ each) was cut by *Bam*HI and labeled by a random-priming method with Cyanine-5- and Cyanine-3-labeled dUTP (Cy5- and Cy3-dUTP; PerkinElmer Life Sciences, Wellesley, MA, USA). Human *Cot*I DNA was employed to block repetitive sequences of the labeled probes.

Prehybridization was conducted to block repetitive sequence binding of target DNA on the arrays and to prevent nonspecific binding of probe DNA to the targets. Following the initial incubation, the prehybridization solution was removed, and hybridization solution with Cy-labeled DNA (prepared as described above) was added. The prehybridization and the hybridization were conducted with continuous mixing overnight at 37°C. After hybridization, the arrays were washed by the procedures reported previously [27]. All of the procedures were conducted using the GeneTAC Hybridization Station (Genomic Solutions Inc., Ann Arbor, MI, USA).

After washing procedures, fluorescent images of the hybridized arrays were obtained using a ScanArray 5000 confocal laser scanner (PerkinElmer Life Sciences). ArraySuite (Scanalytics Inc., Fairfax, VA, USA) in Stage 1 and Gene Pix (Axon Instruments, Sunnyvale, CA) in Stage 2 were used to quantify the fluorescence of each spot on the array images. We then processed the data using software specifically developed to identify "CNVs" [27].

3.3. Procedures of Population Studies. The population studies were conducted in accordance with accepted procedures as described in the previous paper [27]. In brief, hybridization was performed as follows: mixtures of (a) Cy5-labeled reference DNA and Cy3-test DNA, and (b) inverse labeling with
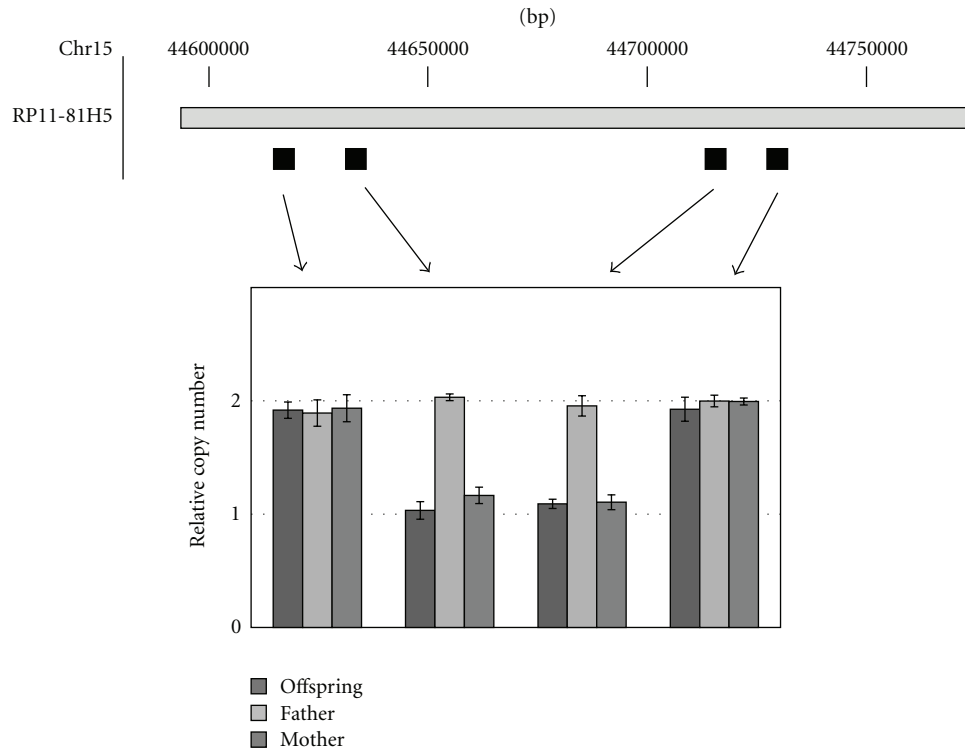
FIGURE 1: Characterization of a loss-type CNV detected on a BAC-clone RP11-81H5. A genome map of the RP11-81H5-containing region was obtained from the UCSC Genome Browser website. The distance (bp) from the p-terminal of chromosome 15 is shown in the uppermost part. The positions of PCR primers examined are shown by solid boxes. The experimental conditions are described more precisely in the text. The relative copy number of each site was normalized using the quantity of the amplified segment of a portion of a different chromosome (Chromosome 16) as a standard (relative copy number = 2) and was shown as mean $\pm$ SD ($n = 3$).

Cy3-reference DNA and Cy5-test DNA were applied to the slides. The two complementary hybridizations (i.e., (a) and (b)) were conducted for each individual. The variations in a given sample from a single individual that were identified consistently in both complementary hybridizations ((a) and (b)) were regarded as putative CNVs. Since only DNA segments detected as putative CNVs in two individuals or more were regarded as true CNVs, further analyses were not conducted. On the contrary, when CNVs were detected in only one individual, all were confirmed with repeated examinations using arrays.

3.4. The qPCR Procedure for CNVs Identified in Offspring. The qPCR procedure was performed using SYBR premix EX Taq (Takara-Bio, Ohtsu, Japan) and the Light Cycler System (Roche Diagnostics Japan, Tokyo, Japan), according to the procedures described in the previous paper [27]. The results were analyzed with Light Cycler Data Analysis software using a second derivative maximum model. Relative DNA content of variants for every amplimer was calculated using a "normal" individual without CNVs as controls. The relative copy number of each site was normalized using quantity of the amplified segment of a portion of a different chromosome (chromosome 16:78888541–78888690; this position empirically showed the smallest deviation among many samples in our experiments) as a standard (relative copy number = 2) and was indicated as mean $\pm$ SD ($n = 3$).

3.5. Procedure of Family Studies for Confirmation of Inheritance of CNVs Identified Once in This Study. For the CNVs identified in only one offspring, further examinations were carried out for confirmation of inheritance. The family studies using qPCR for the CNVs were conducted using DNAs from the mother, the father and sibling of offspring, if available. The procedures of qPCR were the same as above.

4. Results

4.1. Population Studies. We examined 213 unrelated individuals. The main purpose of this paper is to report the data accumulated about CNVs found in <5.0% of the individuals. (The number of CNVs was less than 10 in each BAC region.) The results obtained from the population studies are shown in Tables 1 and 2. In Table 1, the CNVs identified in two individuals or more are summarized. On the other hand, in Table 2, the CNVs identified in only one individual through this study are summarized. A total of 126 CNVs were identified at 52 different BAC regions in the genome. The CNVs observed at 27 of 52 BAC regions were found in only one unrelated individual, and all of them were confirmed by repetitive analysis using the same array system as in the population studies. The family studies by qPCR were conducted for 23 of 27 CNVs, and the results revealed that those 23 CNVs were inherited from one of the parents. The inheritance of CNVs from the remaining two offsprings was not
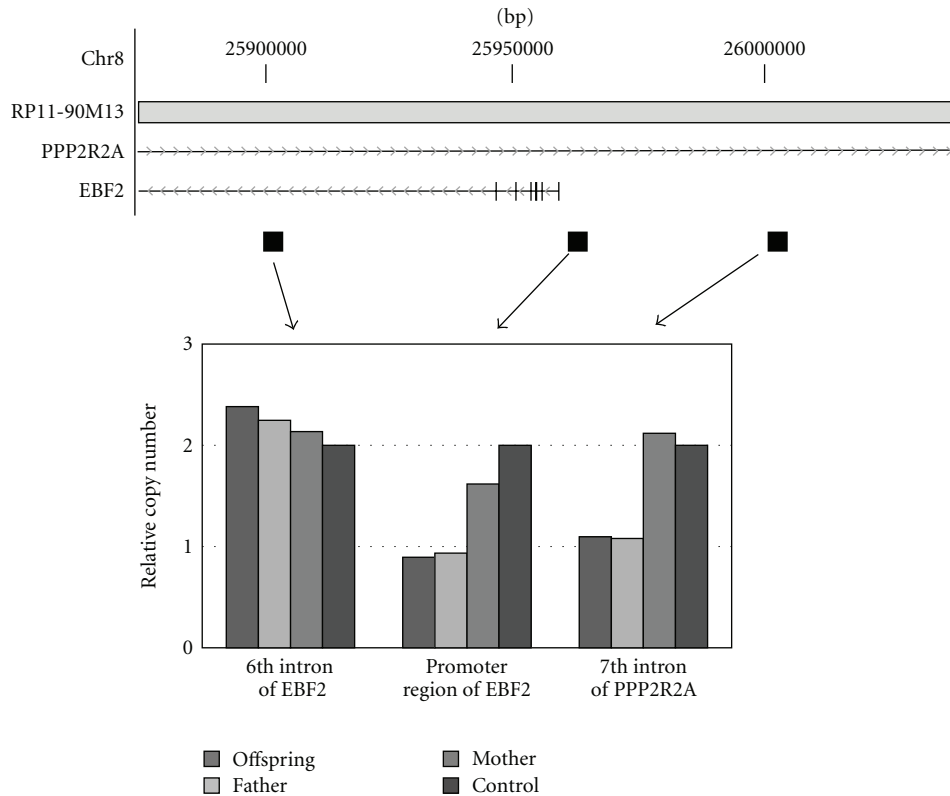
FIGURE 2: Characterization of a loss-type CNV detected on a BAC-clone RP11-90M13. A genome map of the RP11-90M13-containing region was obtained from the UCSC Genome Browser website. The distance (bp) from the p-terminal of chromosome 8 is shown in the uppermost part. That region contains two genes, protein phosphatase 2, regulatory subunit B, alpha (*PPP2R2A*) and early B-cell factor 2 (*EBF2*). The positions of PCR primers examined are shown by solid boxes. The individual has variant fragments with a deletion that contains a part of the 7th intron of the *PPP2R2A* gene and the promoter region of the *EBF2* gene.

confirmed, since DNA from the parents was not obtained. On the other hand, the CNVs on two BAC clones (RP11-88B18, RP11-86B6) were identified in only one unrelated individual. In fact, however, they were observed in individuals who were not selected as "unrelated" individuals. The numbers of individuals who expressed the CNVs are described in the parentheses in Table 2. For that reason, we did not conduct family studies for those CNVs.

4.2. The Typical Patterns of Three Family Studies of CNVs on BAC Clones by qPCR Analysis. As typical examples of family studies, the results of qPCR carried out for the CNVs observed in three BAC-clone regions are described in Figure 1 (RP11-81H5), Figure 2 (RP11-90M13), and Figure 3 (RP11-90O18 and RP11-90C3). For the first case, a single 180,264 bps deletion was observed in an offspring. The family study demonstrated that the CNV was inherited from the mother. There is no gene in this BAC clone region. For the second case, the BAC-clone (RP11-90M13) contains two genes, protein phosphatase 2, regulatory subunit B, alpha (*PPP2R2A*) and early B-cell factor 2 (*EBF2*). The deletion-type CNV was confirmed by qPCR, and the CNV was found to be inherited from the father. The deletion covered part of intron 7 of *PPP2R2A* and the promoter region of *EBF2*. For the last case, as shown in Figure 3, the CNV was covered by

two BAC clones (RP11-90O18 and RP11-90C3). This result was revealed by qPCR, and the CNV was found to be inherited from the father. The CNV contained the series of zinc finger protein families (*ZNF 107, 138, 273, 117, 92*).

4.3. Comparing Our Data with Published Data. Many segmental duplications have already been summarized in public databases, such as the Database of Genomic Variants (DGV Database; http://projects.tcag.ca/variation/), UCSC Human Genome Browser (UCSC Database. http://genome.ucsc.edu/index.html), and NCBI Map Viewer (NCBI Database; http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi). The known CNVs have also been reported in the same databases. However, when the CNVs reported in the databases or papers were too small to be detected by our BAC-aCGH methods, we assumed that they were different from our CNVs. The CNV data obtained in our studies, including an indication of the presence or absence of CNVs already reported in other databases, are summarized in Tables 1 and 2.

4.4. Comparing Our Data with the Data of Japanese Published by Other Groups. Compared to Caucasian populations, very little information about CNVs in Japanese populations has been systematically screened. We compared our data with the data of Japanese populations from the HapMap project as
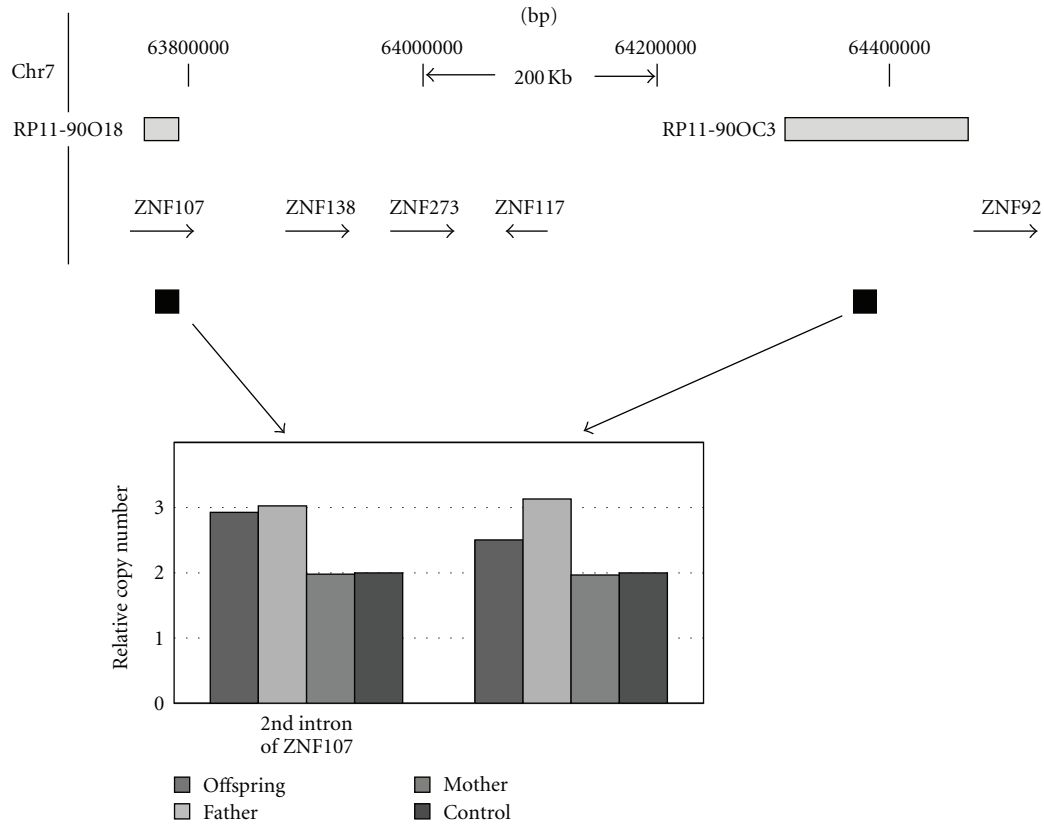
Figure 3: Characterization of a gain-type CNV detected on a BAC-clone RP11-90O18 and RP11-90C3. A genome map of both RP11-90 O18- and RP11-90C3-containing regions was obtained from the UCSC Genome Browser website. The distance (bp) from the p-terminal of chromosome 7 is shown in the uppermost part. That region contains zinc finger print protein gene families, that is, *ZNF 107, 138, 273, 117, and 92*. The positions of PCR primers examined are shown by the solid boxes.

reported by Redon et al. [4], in which CNVs were examined by tiling BAC array. In addition to the data by BAC array, three groups, including Redon et al. [4], reported the data of Japanese populations based on examinations conducted using oligo-array methods, although each group used a different array platform [24, 25]. These results are also summarized in Tables 1 and 2. In the case where a multiple number of CNVs were identified in our study (Table 1), the CNVs in 16 of 25 BAC regions were reported in the other reports. Moreover, regarding one individual, the CNVs identified in nine of 27 BAC regions were reported in the data reported by the others, although we should emphasize again that the CNVs identified in our BAC region are not exactly the same as those reported by the others. The majority of CNVs identified in our study tend not to be identified in the other reports. Thus, they can be considered novel CNVs in the general Japanese population.

4.5. Genes and Disease-Related Genes. We summarized some of the genes and disease-related genes that overlap with the BAC-clone region for our CNVs (Table 3). In addition, mRNAs also have been reported in the databases, but they are too numerous to describe here. All BAC-clone regions contained at least one mRNA, although the functions of a majority of the mRNAs are not yet known (data not shown).

5. Discussion

We examined 213 unrelated Japanese individuals using BAC-aCGH and found a total of 126 CNVs on 52 different BAC regions in the genomes. The CNVs observed on 27 of the 52 BAC-regions were found in only one unrelated individual. The family studies were conducted for 23 CNVs, and the results demonstrated that these were inherited from one of the parents. A fraction of the regions involved in the CNVs observed in our study (i.e., 34 of 126 or about 27%; Tables 1 and 2) were not reported previously in other studies listed in the database. Thus, they are considered novel CNVs. In contrast, the data of highly polymorphic CNVs [23] demonstrated that only 8% (55 of 680) had not been reported. This result showed that a significantly large number of novel CNVs have been identified among relatively less frequent CNVs ($\chi^2 = 36.64$, $P = 5.10 \times 10^{-10}$). This finding suggests that the structural rearrangements for creating less frequent CNVs are evolutionarily recent. We thus paid attention to differences in the number of CNVs identified in Hiroshima compared to Nagasaki. With respect to the CNVs

TABLE 1: Summary of the CNVs detected in plural unrelated individuals.

Clone number	Chrom. number	Position (Mb)	Variants Amp[1]	Delete	Known variant	Segmental duplication	Hiroshima[2]	Nagasaki[3]	Redon[4] [4]	Redon[5] [4]	Park et al. [24]	Ikeda et al. [25]
RP11-81B15	1	150.7	9	0	Pre[6]	Abs[8]	4	5	Abs[10]	Abs	Pre	Abs
RP11-91G12	1	194.9	2	0	Pre	Pre[9]	2	0	Abs	Abs	Pre	Pre
RP11-118M12	2	241.2	3	0	Pre	Pre	0	3	Pre[11]	Pre	Abs	Abs
RP11-91F17	3	167.5	2	0	Pre	Abs	2	0	Abs	Abs	Abs	Abs
RP11-101J14	4	8.1	2	0	Pre	Abs	0	2	Abs	Abs	Abs	Abs
RP11-80P12	4	132.3	2	0	Pre	Abs	1	1	Abs	Abs	Abs	Abs
RP11-23D23	7	0.84	4	0	Pre	Abs	3	1	Abs	Abs	Abs	Abs
RP11-90C3	7	64.3	1	1	Pre	Abs	1	1	Pre	Abs	Pre	Pre
RP11-35P20	7	71.6	2	0	Pre	Pre	0	2	Pre	Pre	Pre	Abs
RP11-45M12	8	3.9	0	2	Pre	Abs	2	0	Pre	Pre	Pre	Abs
RP11-113B7	8	4.2	0	4	Pre	Abs	4	0	Pre	Pre	Abs	Abs
RP11-89I12	8	4.4	0	4	Pre	Abs	4	0	Pre	Pre	Pre	Abs
RP11-79I19	8	3.7	0	4	Pre	Abs	4	0	Pre	Pre	Abs	Abs
RP11-310C13	9	113.5	6	1	Abs[7]	Abs	5	2	Abs	Abs	Abs	Abs
RP11-79M9	10	77.7	6	0	Abs	Abs	4	2	Abs	Abs	Abs	Abs
RP11-326B4	13	66.8	8	0	Abs	Abs	5	3	Abs	Abs	Abs	Abs
RP11-89A14	13	82.4	5	0	Pre	Abs	4	1	Abs	Abs	Abs	Abs
RP11-531P20	13	84.4	2	0	Pre	Abs	2	0	Abs	Pre	Abs	Abs
RP11-143O10	13	88.5	4	0	Pre	Abs	2	2	Abs	Abs	Pre	Abs
RP11-91C17	14	43.2	5	0	Abs	Abs	3	2	Abs	Abs	Abs	Abs
RP11-194H7	15	32.5	2	0	Pre	Pre	2	0	Pre	Abs	Abs	Abs
RP11-152F13	15	80.9	0	9	Pre	Pre	5	4	Pre	Abs	Abs	Abs
RP11-88B2	18	61.8	0	2	Pre	Abs	0	2	Pre	Abs	Pre	Abs
RP11-13J15	21	24.1	0	4	Pre	Abs	2	2	Abs	Abs	Abs	Pre
CTA-526G4	22	20.6	1	2	Pre	Abs	2	1	Pre	Abs	Abs	Pre
Total of CNVs			66	33			63	36				

[1] Amplification. Number of CNVs observed in Hiroshima individual [2] and in Nagasaki individual [3]. [4] Radon's BAC data. [5] Radon's oligo-array data. CNVs present [6] or absent [7] in previous papers. Segmental duplication absent [8] or present [9] in BAC-clones on which CNVs were observed. CNVs absent [10] or present [11] in Japanese.

observed multiple times, 12 of 25 were identified in one city. These CNVs appear to be so-called "private polymorphisms" occurring more recently than the others. On the other hand, quite a small number (21 of 126; about 17%) of the CNVs in this report were found on the BAC clones that overlapped with segmental duplication (Tables 1 and 2). A majority (about 63%) of highly polymorphic CNVs, however, were observed on the BAC-clone overlapping with segmental duplication (see Table 1 in [23]). This result suggests that segmental duplication might play a significant role in the creation of highly polymorphic CNVs. These observations are supported by previous data from Sharp et al. [17], who reported the sharing of CNVs among several populations, meaning that the specific genomic imbalances either predated the migration of modern humans from Africa or arose independently in different populations.

On the other hand, the data representing Japanese populations are based on reports from three groups. We compared our CNV data to that of these other reports [4, 24, 25]. The CNVs found in 27 of 52 BAC regions (about 52%) were not identified by the other groups. Speculation as to the reason for this includes the possibility that since the CNVs were private polymorphisms as mentioned above, they were not identified in the Japanese populations used in the other works.

In humans, it is estimated that around 23% of currently known CNVs reside in known or putative genes. Many CNVs that lie outside genic regions of the genome may still have a significant influence on gene expression by affecting gene regulatory elements [30]. Another important area of exploration is linkage disequilibrium (LD), that is, nonrandom association between alleles at different loci, between CNVs and other genomic variants (especially SNPs), which may provide important insight into the genomic evolution (e.g., the recombination patterns) of CNVs [31, 32]. However, studies of LD between CNVs and SNPs are hindered by the fact that CNVs generally reside in genomic regions associated with segmental duplications and/or repeat-rich regions, which are difficult to sequence and use for detection of SNPs.

TABLE 2: Summary of the CNVs detected in single unrelated individuals.

Clone number	Chrom. number	Position (Mb)	Delete or Amp[1]	Known variant	Segmetal duplication	Known inheritance[2]	Redon [4]	Redon[3] [4]	Park et al. [24]	Ikeda et al. [25]
RP5-895L2	1	231.9	Deletion	Abs[4]	Abs[6]	MO	Abs[8]	Abs	Abs	Abs
RP11-204C23	3	3	Delete	Pre[5]	Abs	FA	Abs	Abs	Abs	Abs
RP11-80D22	4	43.7	Amp	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-22O8	4	141.1	Deletion	Abs	Abs	ND	Abs	Abs	Abs	Abs
RP11-104N13	6	165	Deletion	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-90O18	7	63.8	Amp	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-79I19	8	4	Delete	Pre	Abs	MO	Abs	Pre	Abs	Abs
RP11-90J21	8	8.8	Amp	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-90M13	8	25.9	Deletion	Abs	Abs	FA	Abs	Abs	Abs	Abs
RP11-91G3	8	91.1	Deletion	Pre	Abs	ND	Pre[9]	Pre	Abs	Abs
RP11-91H19	10	58.2	Deletion	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-176H12	10	65.7	Amp	Pre	Abs	MO	Abs	Abs	Abs	Abs
RP11-18B9	11	40.3	Delete	Pre	Abs	MO	Abs	Abs	Abs	Abs
RP11-75H24	11	58.6	Amp	Pre	Pre[7]	MO	Pre	Abs	Pre	Abs
RP11-90N7	12	11.4	Amp	Pre	Pre	MO	Abs	Abs	Pre	Abs
RP11-86C3	13	89	Deletion	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-80H14	15	20.5	Amp	Pre	Abs	FA	Abs	Abs	Abs	Pre
RP11-81H5	15	44.7	Delete	Abs	Abs	MO	Abs	Abs	Pre	Abs
RP11-90E5	15	98.4	Amp	Pre	Abs	FA	Abs	Abs	Abs	Abs
RP11-79I15	16	16.6	Deletion	Pre	Pre	MO	Pre	Pre	Abs	Abs
RP11-89L6	16	75.7	Amp	Abs	Abs	MO	Abs	Abs	Abs	Abs
CTB-15E16	16	79.0	Amp	Pre	Abs	MO	Abs	Abs	Abs	Abs
RP11-265I6	18	36.1	Delete	Abs	Abs	FA	Abs	Abs	Abs	Abs
RP11-89N1	18	76.0	Amp	Pre	Abs	MO	Abs	Abs	Abs	Pre
RP5-908M14	20	60.4	Amp	Pre	Abs	MO	Pre	Abs	Pre	Abs
Remarks		Each CNV as below was detected in three individuals in different families, but two of them were detected in individuals who were not selected as an unrelated individual. Both CNVs were not identified in three other studies of Japanese								
RP11-88B18	10	56.6	Deletion	Pre	Abs	ND	1 (3)			
RP11-86B6	13	104	Amp	Abs	Abs	ND	1 (3)			

[1] Amplification. [2] CNVs were inherited from the mother (MO) or the father (FA). [3] Data from Affymetrix oligo array. CNVs absent [4] or present [5] in previous papers. Segmental duplication absent [6] or present [7] in BAC clones on which CNVs were observed. CNVs absent [8] or present [9] in Japanese data from three types of oligo arrays.

Recent association studies have attempted to link CNVs with expression profiles, diseases, and known phenotypic differences. On the other hand, distribution of CNVs differs depending on genetic background, and population-specific CNVs might account for the divergence of some physiological traits and disease prevalence among populations [31]. Like SNPs, CNVs also show variability among populations; therefore, comprehensive CNV information is needed when applying CNVs as genetic markers in disease or trait studies.

For the reasons cited above, we made plans to accumulate Japanese data in which we provide information about the relation between Japanese CNV data and phenotypic differences correlated with reported diseases. In our previous paper [23], we summarized such data by focusing on highly polymorphic CNVs found in ≥5.0% of the individuals.

More recent studies based on advanced molecular technologies, such as genome-wide association studies [33, 34] and next generation sequencing [35, 36], reported that many genes appear to play important roles in the etiology of common diseases. Depending on newly developed technologies, rare variations have recently received attention from scientists who espouse a hypothesis called "common disease and rare variants." In this paper, we summarized CNVs that were identified in fewer than 10 individuals in our previous population study. We focused on genes reported in the DGV, UCSC, and NCBI databases (Table 3), although many mRNAs were also listed in them. Since these genes are good candidate markers for enabling us to examine the etiology of common diseases and phenotypical heterogeneities among individuals, our CNVs have the potential to become useful markers in future studies.

6. Conclusion

In this study, 34 CNVs were new, indicating that CNV coverage of the human genome is still incomplete and that there is

TABLE 3: Genes in BAC clones with CNVs.

BAC-clone number	Chrom. number	Position (Mb)	Genes
RP11-81B15	1	150.7	Late cornified envelope 5A (LCE5A), 3A(LCE3A), 3B(LCE3B), 3C(LCE3C), 3D(LCE3D), 3E(LCE3E), cysteine-rich C-terminal 1 (CRCT1)
RP11-91G12	1	194.9	Complement factor H (CFH). Complement factor H-related 3 (CFHR3)
RP11-118M12	2	241.2	Aquaporin 12B (AQP12B), aquaporin 12A (AQP12A), kinesin family member 1A (KIF1A)
RP11-101J14	4	8.1	Actin binding LIM protein family, member 2 (ABLIM2)
RP11-23D23	7	0.84	ArfGAP with dual PH domains 1 (ADAP1)
RP11-90C3	7	64.3	Integrator complex subunit 4-like 1 (INTS4L1)
RP11-35P20	7	71.6	tRNA-yW synthesizing protein 1 homolog B (S.cerevisiae) (TYW1B)
RP11-45M12	8	3.9	CUB and Sushi multiple domains 1 (CSMD1)
RP11-113B7	8	4.2	CUB and Sushi multiple domains 1 (CSMD1)
RP11-89I12	8	4.4	CUB and Sushi multiple domains 1 (CSMD1)
RP11-79I19	8	3.7	CUB and Sushi multiple domains 1 (CSMD1)
RP11-310C13	9	113.5	Chromosome 9 open reading frame 84 (C9orf84)
RP11-79M9	10	77.7	Chromosome 10 open reading frame 11 (C10orf11)
RP11-326B4	13	66.8	Protocadherin 9 (PCDH9)
RP11-91C17	14	43.2	Rho GTPase activating protein 5 (ARHGAP5)
RP11-194H7	15	32.5	Golgin A8 family, member A (GOLGA8A)
RP11-152F13	15	80.9	Ubiquitin-conjugating enzyme E2Q family member 2 pseudogene 2 (UBE2Q2P2), golgin A6 family-like 9 (GOLGA6L9), cytoplasmic polyadenylation element binding protein 1, golgin A2 pseudogene (LOC727849), cytoplasmic polyadenylation element binding protein 1(CPEB1)
CTA-526G4	22	20.6	Topoisomerase (DNA) III beta (TOP3B)
RP11-90O18	7	63.8	Zinc finger protein 107 (ZNF107),zinc finger protein 138 (ZNF138)
RP11-79I19	8	4	CUB and Sushi multiple domains 1 (CSMD1)
RP11-90J21	8	8.8	Exoribonuclease 1 (ERI1)
RP11-90M13	8	25.9	Early B-cell factor 2 (EBF2)
RP11-91G3	8	91.1	Calbindin 1, 28 kDa (CALB1)
RP11-90N7	12	11.4	Proline-rich protein BstNI subfamily 4 (PRB4), proline-rich protein BstNI subfamily 1 (PRB1), proline-rich protein BstNI subfamily 2 (PRB2)
RP11-80H14	15	20.5	Tubulin, gamma complex-associated protein 5 (TUBGCP5), cytoplasmic FMR1 interacting protein 1 (CYFIP1)
RP11-81H5	15	44.7	MicroRNA 548a-3 (MIR548A3)
RP11-90E5	15	98.4	ADAM metallopeptidase with thrombospondin type 1 motif, 17 (ADAMTS17)
RP11-89N1	18	76.0	Ribosome binding factor A (putative) (RBFA), ADNP homeobox 2 (ADNP2), hypothetical LOC100130522 (LOC100130522), par-6 partitioning defective 6 homolog gamma (*C. elegans*) (PARD6G)
RP5-908M14	20	60.4	Ribosomal protein S21 (RPS21), Cdk5 and Abl enzyme substrate 2 (CABLES2), chromosome 20 open reading frame 151 (C20orf151)

diversity between the Japanese and other ethnic populations. Moreover, the CNVs found in about two-thirds of the BAC regions examined in our study were not identified in the studies in which three groups examine Japanese populations using different array platforms. The newly identified CNVs extend the coverage of CNVs in the human genome (also the general Japanese population). Some CNVs contained genes that might be related to phenotypic heterogeneity among individuals. Moreover, it is expected that the CNVs could now be taken into consideration when genetic studies, for example, CNV association studies, are conducted.

Authors' Contribution

Y. Satoh and N. Takahashi equally contributed to this work as first authors.

Acknowledgments

References

[1] J. L. Freeman, G. H. Perry, L. Feuk et al., "Copy number variation: new insights in genome diversity," *Genome Research*, vol. 16, no. 8, pp. 949–961, 2006.

[2] A. J. Iafrate, L. Feuk, M. N. Rivera et al., "Detection of large-scale variation in the human genome," *Nature Genetics*, vol. 36, no. 9, pp. 949–951, 2004.

[3] M. Jakobsson, S. W. Scholz, P. Scheet et al., "Genotype, haplotype and copy-number variation in worldwide human populations," *Nature*, vol. 451, no. 7181, pp. 998–1003, 2008.

[4] R. Redon, S. Ishikawa, K. R. Fitch et al., "Global variation in copy number in the human genome," *Nature*, vol. 444, no. 7118, pp. 444–454, 2006.

[5] J. Sebat, B. Lakshmi, J. Troge et al., "Large-scale copy number polymorphism in the human genome," *Science*, vol. 305, no. 5683, pp. 525–528, 2004.

[6] K. K. Wong, R. J. deLeeuw, N. S. Dosanjh et al., "A comprehensive analysis of common copy-number variations in the human genome," *American Journal of Human Genetics*, vol. 80, no. 1, pp. 91–104, 2007.

[7] D. Pinkel, R. Segraves, D. Sudar et al., "High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays," *Nature Genetics*, vol. 20, no. 2, pp. 207–211, 1998.

[8] A. S. Ishkanian, C. A. Malloff, S. K. Watson et al., "A tiling resolution DNA microarray with complete coverage of the human genome," *Nature Genetics*, vol. 36, no. 3, pp. 299–303, 2004.

[9] J. R. Pollack, C. M. Perou, A. A. Alizadeh et al., "Genome-wide analysis of DNA copy-number changes using cDNA microarrays," *Nature Genetics*, vol. 23, no. 1, pp. 41–46, 1999.

[10] E. Tuzun, A. J. Sharp, J. A. Bailey et al., "Fine-scale structural variation of the human genome," *Nature Genetics*, vol. 37, no. 7, pp. 727–732, 2005.

[11] S. A. McCarroll, T. N. Hadnott, G. H. Perry et al., "Common deletion polymorphisms in the human genome," *Nature Genetics*, vol. 38, no. 1, pp. 86–92, 2006.

[12] D. F. Conrad, T. D. Andrews, N. P. Carter, M. E. Hurles, and J. K. Pritchard, "A high-resolution survey of deletion polymorphism in the human genome," *Nature Genetics*, vol. 38, no. 1, pp. 75–81, 2006.

[13] M. E. Hurles, E. T. Dermitzakis, and C. Tyler-Smith, "The functional impact of structural variation in humans," *Trends in Genetics*, vol. 24, no. 5, pp. 238–245, 2008.

[14] B. E. Stranger, M. S. Forrest, M. Dunning et al., "Relative impact of nucleotide and copy number variation on gene phenotypes," *Science*, vol. 315, no. 5813, pp. 848–853, 2007.

[15] J. A. Buchanan and S. W. Scherer, "Contemplating effects of genomic structural variation," *Genetics in Medicine*, vol. 10, no. 9, pp. 639–647, 2008.

[16] C. Lee, A. Iafrate, and A. Brothman, "Copy number variations and clinical cytogenetic diagnosis of constitutional disorders," *Nature Genetics*, vol. 39, supplement 7, pp. S48–S54, 2007.

[17] A. Sharp, D. Locke, S. McGrath et al., "Segmental duplications and copy-number variation in the human genome," *American Journal of Human Genetics*, vol. 77, no. 1, pp. 78–88, 2005.

[18] T. D. de Ståhl, J. Sandgren, A. Piotrowski et al., "Profiling of copy number variations (CNVs) in healthy individuals from three ethnic groups using a human genome 32 K BAC-clone-based array," *Human Mutation*, vol. 29, no. 3, pp. 398–408, 2008.

[19] A. J. Iafrate, L. Feuk, M. N. Rivera et al., "Detection of large-scale variation in the human genome," *Nature Genetics*, vol. 36, no. 9, pp. 949–951, 2004.

[20] D. Pinto, C. Marshall, L. Feuk, and S. W. Scherer, "Copy-number variation in control population cohorts," *Human Molecular Genetics*, vol. 16, no. 2, pp. R168–R173, 2007.

[21] G. Zogopoulos, K. Ha, F. Naqib et al., "Germ-line DNA copy number variation frequencies in a large North American population," *Human Genetics*, vol. 122, no. 3-4, pp. 345–353, 2007.

[22] S. J. White, L. Vissers, A. Geurts et al., "Variation of CNV distribution in five different ethnic populations," *Cytogenetic and Genome Research*, vol. 118, no. 1, pp. 19–30, 2007.

[23] N. Takahashi, Y. Satoh, K. Sasaki, Y. Shimoichi, K. Sugita, and H. Katayama, "Characteristics of highly polymorphic segmental copy-number variations observed in Japanese by BAC-array-CGH," *Journal of Biomedicine and Biotechnology*, vol. 2011, Article ID 820472, 11 pages, 2011.

[24] H. Park, J. Kim, Y. Ju et al., "Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing," *Nature Genetics*, vol. 42, no. 5, pp. 4000–4005, 2010.

[25] M. Ikeda, B. Aleksic, Y. Kinoshita et al., "Genome-wide association study of schizophrenia in a Japanese population," *Biological Psychiatry*, vol. 169, no. 5, pp. 472–478, 2010.

[26] The BAC Resource Consortium, "Integration of cytogenetic landmarks into the draft sequence of the human genome," *Nature*, vol. 409, no. 6822, pp. 953–958, 2001.

[27] N. Takahashi, N. Tsuyama, K. Sasaki et al., "Segmental copy-number variation observed in Japanese by array-CGH," *Annals of Human Genetics*, vol. 72, no. 2, pp. 193–204, 2008.

[28] T. Maniatis, E. Fritsch, and J. Sambrook, "Isolation of high-molecular-weight, eukaryotic DNA from cells grown in tissue culture," in *Molecular Cloning. A Laboratory Manual*, pp. 280–281, Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 1982.

[29] A. Snijders, N. Nowak, R. Segraves et al., "Assembly of microarrays for genome-wide measurement of DNA copy number," *Nature Genetics*, vol. 29, no. 3, pp. 263–264, 2001.

[30] P. Cahan, Y. Li, M. Izumi, and T. A. Graubert, "The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells," *Nature Genetics*, vol. 41, no. 4, pp. 430–437, 2009.

[31] G. H. Perry, N. J. Dominy, K. G. Claw et al., "Diet and the evolution of human amylase gene copy number variation," *Nature Genetics*, vol. 39, no. 10, pp. 1256–1260, 2007.

[32] D. Conrad, D. Pinto, R. Redon et al., "Origins and functional impact of copy number variation in the human genome," *Nature*, vol. 464, no. 7289, pp. 704–712, 2010.

[33] Y. Tanaka, N. Nishida, M. Sugiyama et al., "Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C," *Nature Genetics*, vol. 41, no. 10, pp. 1105–1109, 2009.

[34] S. C. Greenway, A. C. Pereira, J. C. Lin et al., "De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot," *Nature Genetics*, vol. 41, no. 8, pp. 931–935, 2009.

[35] H. M. Wood, O. Belvedere, C. Conway et al., "Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens," *Nucleic Acids Research*, vol. 38, no. 14, Article ID gkq510, pp. e151–e151, 2010.

[36] E. T. Cirulli and D. B. Goldstein, "Uncovering the roles of rare variants in common disease through whole-genome sequencing," *Nature Reviews Genetics*, vol. 11, no. 6, pp. 415–425, 2010.