

Research Article

Characteristics of Highly Polymorphic Segmental Copy-Number Variations Observed in Japanese by BAC-Array-CGH

Norio Takahashi,¹ Yasunari Satoh,¹ Keiko Sasaki,¹ Yuko Shimoichi,¹ Keiko Sugita,² and Hiroaki Katayama¹

¹Department of Genetics, Radiation Effects Research Foundation, 5-2 Hijiya Park, Minami-ku, Hiroshima 732-0815, Japan

²Department of Information Technology, Radiation Effects Research Foundation, 5-2 Hijiya Park, Minami-ku, Hiroshima 732-0815, Japan

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

Received 2 July 2010; Revised 17 September 2010; Accepted 1 October 2010

Academic Editor: Xin-yuan Guan

Copyright © 2011 Norio Takahashi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Segmental copy-number variations (CNVs) may contribute to genetic variation in humans. Reports of the existence and characteristics of CNVs in a large Japanese cohort are quite limited. We report the data from a large Japanese population. We conducted population screening for 213 unrelated Japanese individuals using comparative genomic hybridization based on a bacterial artificial chromosome microarray (BAC-aCGH). We summarize the data by focusing on highly polymorphic CNVs in $\geq 5.0\%$ of the individual, since they may be informative for demonstrating the relationships between genotypes and their phenotypes. We found a total of 680 CNVs at 16 different BAC-regions in the genome. The majority of the polymorphic CNVs presented on BAC-clones that overlapped with regions of segmental duplication, and the majority of the polymorphic CNVs observed in this population had been previously reported in other publications. Some of the CNVs contained genes which might be related to phenotypic heterogeneity among individuals.

1. Introduction

Segmental copy-number variations (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. This heterogeneity may contribute to noted phenotypic variations and different susceptibilities to various diseases. Molecular genetics analyses and cytogenetic analyses have provided significant information about these variations in the human genome, specifically as they relate to disease, such as cancer, and to congenital malformation (see [1, 2], and review in [3]). Following the development of methodologies and the introduction of new research platforms [4–9], information regarding the nature and pattern of CNVs from representative populations have accumulated. 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 [10–13], oligo-arrays [14–16], and others [17]. The results are not always consistent and it is likely that different human populations bear different CNVs. The numbers of Japanese individuals examined to date are not so large compared to the studies for other ethnicities [18]. 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. In this report, we focus on CNVs which were observed at a high frequency ($\geq 5.0\%$ of the individuals) in the population residing in Hiroshima and Nagasaki, Japan by aCGH with BAC-clones as targets.

2. Materials and Methods

In the study, the population studies were conducted at two stages: Stage (1): 80 unrelated Japanese individuals were examined using BAC-aCGH with an array having 2,241 BAC

clones, and Stage (2): 133 unrelated Japanese individuals were examined using BAC-aCGH that contained 2,622 BAC-clones.

2.1. Target BAC-Clones Preparation. 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 [19] 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), in addition to BAC clones used in Stage (1), an additional 381 BAC clones were used, a majority of which were collaboratively obtained from Dr. N. Matsumoto of Yokohama City University. The 2,241 clones of chromosomal fragments from chromosome 1 to chromosome 22 were used in Stage (1) and 2,622 clones were used in Stage (2), respectively. That is, the additional 381 BAC clones were examined for only 133 unrelated individuals in Stage (2). Those clones are distributed every 1.2 Mb across all of the human autosomes in Stage (1), and 1.1 Mb in Stage (2), respectively. In addition to autosomal clones, four kinds of X-chromosomal clones were used as internal references. With respect to examination for Stage (1), three sets of arrays were constructed and imprinted: slide no.1, consisted of 698 clones on chromosomes 1 to 4; slide no.2 consisted of 718 clones on chromosomes 5 to 10, plus two BAC clones on chromosome 3, and six clones on chromosome 4; and slide no.3 consisted of 817 clones on chromosomes 11 to 22 [20]. For Stage (2), all of the clones were printed onto one glass slid.

2.2. Genomic DNA Samples. The genomic DNA samples used in this study were principally the same as those used in a previous study [20]. The DNA samples used for reference purposes were extracted from mononuclear cells of two physically and clinically normal volunteers (a 57-year-old Japanese male and a 54-year-old Japanese female). The DNA used for testing and analyses of this population was extracted from lymphoblastoid cell lines obtained from the offspring of atomic-bomb survivors. High molecular weight genomic DNA was isolated using conventional methods as described in detail elsewhere [21]. Lymphoblastoid cell lines were derived from a cryopreserved archive of approximately 1000 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 of peripheral B-lymphocytes. The composition of the families has been reported elsewhere [22]. Three hundred five offspring were initially screened. Since the offspring include some siblings, we selected one representative offspring to construct "unrelated individuals" to avoid double counting of polymorphic CNVs from families containing two or more siblings. We selected the offspring who first visited our institution for donating blood rather than other siblings. The 213 offspring selected as unrelated individuals included 124 offspring from Hiroshima and 89 from Nagasaki. The individuals gave their informed consent prior to the study.

The Institutional Review Board of our Foundation approved this study.

2.3. Array Preparations. Cloned DNAs for microarray targets were isolated from bacterial cultures using NucleoBond BAC 100 (NIPPON Genetics, Tokyo). With respect to Stage (1), DNA was digested by *NotI* followed by phenol-chloroform-isoamyl alcohol (25:24:1) extraction and ethanol-precipitation. On the other hand, in Stage (2), cloned DNA was digested with *MseI* followed by phenol-chloroform-isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The fragmented DNAs were amplified by ligation-mediated PCR carried out as described by Snijders et al. [23]. The target DNAs (0.5 $\mu\text{g}/\mu\text{L}$) were dissolved in 50%-dimethylsuloxide and printed in triplicate onto the glass slides (Matsunami Glass Co. Ltd.) using the Affymetrix 417 Arrayer (Affymetrix).

2.4. Array-CGH Analysis

2.4.1. Screening by the Array CGH Method. The screenings of both stages were conducted following the procedures described previously [20]. In brief, for labeling DNA, test and reference genomic DNA (1.25 μg each) was cut by *BamHI*, 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). The labeled probes were mixed and centrifuged with Microcon column (Millipore Co., Bedford, MA, USA) to purify the probes. Subsequently, human *CotI* DNA (120 μg ; Roche Diagnostic GmbH, Mannheim, Germany) was added to the column, and recentrifuged. After the volume of the mixture became less than 20 μL , it was transferred to microtubes with 100 μL of hybridization solution (50% formamide, 10% dextran sulfate, 1% Tween 20, 2 \times SSC, 10 mM Tris-HCl [pH 7.4] and 800 μg of yeast t-RNA [Invitrogen, Carlsbad, CA, USA]). The hybridization mixture was then denatured at 70°C for 10 minutes, and subsequently incubated at 37°C for at least five hours to block repetitive sequences of the labeled probes.

Prehybridization was conducted in order 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 (overnight at 37°C), the prehybridization solution was removed, and fresh hybridization solution with Cy-labeled DNA (prepared as described above) was added. Again, hybridization processes were carried out. The prehybridization and the hybridization were conducted with continuous mixing. After hybridization, the arrays were washed by the procedures reported previously [20]. All of the procedures were conducted using the GeneTAC Hybridization Station (Genomic Solutions Inc., Ann Arbor, MI, USA).

2.4.2. Imaging and Data Management. Fluorescent images of the hybridized arrays were obtained using a ScanArray 5000 confocal laser scanner (PerkinElmer Life Sciences). Array-Suite (Scanalytics Inc., Fairfax, VA, USA) in Stage (1) and Gene Pix (Axon Instruments, Sunnyvale, CA) in Stage (2),

respectively, were used to quantify the fluorescence of each spot on the array images. We then used software specifically developed to perform the following three measurements.

- (1) Fluorescent ratios of the total integrated Cy3 and Cy5 intensities for each target (triplicate spots for each target) were calculated, along with the mean ratios of the triplicate spots (if the raw ratio of one of the triplicate spots differed by more than 10% from the other two, the value was automatically excluded from the mean ratio calculation).
- (2) Normalized ratios were computed by dividing each raw ratio by the mean raw ratios of every autosomally mapped target (therefore, after normalization, the averaged ratio for all targets in one array was 1.0).
- (3) The spots whose ratios were 2.58 standard deviations (SD) below the mean (1.0), or 2.58 SD above the mean, were marked "CNV." To confirm the quality of each analysis, the ratio of X-linked clones was verified.

2.4.3. Pulse Field Gel Electrophoreses. EB-transformed lymphoblastoid cells established from the offspring were embedded in agarose plugs. DNA in the plugs was cleaved by restriction enzymes (*PacI*, or *Sse8387I*). The resulting fragments were then separated by pulse field gel electrophoreses (PFGE) on 1% Pulse Field Certified Agarose (BioRad, Hercules, CA, USA) with $\times 0.5$ TBE (Tris-Borate-Ethylenediaminetetraacetic acid). Using the CHEF-DR II system (BioRad) [24], electrophoresis was carried out using 6 V/cm at 14°C for 22 hrs. The angle of pulse was 120°. The switch time was used: ramped from 0.3 seconds to 15 seconds.

2.4.4. Southern Blotting and Hybridization. Southern blot analyses were carried out using conventional, well-described procedures [25]. In brief, after completion of PFGE, the DNA in the gel was cleaved by UV irradiation and blotted onto nitrocellulose filters (Schleicher & Schuell, Dassel, Germany). The filters were prehybridized with human *CotI* DNA (48 μ g/ml, Roche Diagnostic GmbH) and salmon testis DNA (14 μ g/ml, Sigma-Aldrich) to decrease the background due to repetitive sequences. Subsequently, the filters were hybridized with whole BAC-DNA as a probe. DNA probes were labeled with [α -³²P] dCTP (Amersham Biosciences, Piscataway, NJ, USA) and preannealed with human *CotI* DNA and salmon testis DNA (10.5 μ g/ml). Prehybridization and hybridization were performed overnight at 37°C in a solution containing 50% formamide, 10% dextran sulfate, 1% Tween 20, $2 \times$ SSC, and 10 mM Tris-HCl (pH 7.4). After hybridization, the filters were washed at 65°C with $1.0 \times$ SSC containing 0.1% SDS (Sodium dodecyl sulfate) and $0.5 \times$ SSC containing 0.1% SDS. Banding patterns were obtained by either exposure to X-ray film (Fuji Film, Tokyo, Japan) or through use of the Molecular Imager FX (BioRad).

2.4.5. The Procedure of Quantitative Polymerase Chain Reaction (qPCR). The qPCR was performed using SYBR

premix EX Taq (Takara-Bio) and the Light Cycler System (Roche Diagnostics), according to the manufacturers' protocols. Primers were designed with Primer3 software (<http://primer3.sourceforge.net>), and the size of PCR products was confirmed by the pattern of restriction enzyme digested fragments using the LabChip DNA 500 Kit on the 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany). Two, 0.5, and 0.125 ng of genome DNA from an individual with CNV were used and the quantification of each amplicon was carried out at 45 cycles of PCR. The results were analyzed with Light Cycler Data Analysis software using a second derivative maximum model. Relative DNA content of variants for every amplicon was calculated using "normals" (individuals who did not have CNVs) as a control.

3. Results

3.1. Population Study. We examined 213 unrelated individuals. The main purpose of this paper is to report the accumulation of the data about highly polymorphic CNVs found in $\geq 5.0\%$ of the individuals. (The number of CNVs was 11 or more in each BAC-spot.) As shown in Table 1, 680 polymorphic CNVs were observed on 16 BAC-regions. As described before, the results of two BAC (RP11-259N12 and RP11-121A8) were obtained from 133 unrelated individuals examined in Stage (2).

3.2. Southern Blot Analysis Followed by Pulse Field Gel Electrophoreses. Southern blot analyses followed by PFGE were carried out for the highly polymorphic CNVs. As shown by the typical cases in Figures 1 and 2, the highly polymorphic CNVs exhibited complicated patterns of alternation. The patterns of these two BAC clones (RP11-79F15 and RP11-88L18) are shown in Figures 1 (RP11-79F15) and 2 (RP11-88L18) as the typical examples. Since each individual contained a different number of core segmental duplication units, each individual showed bands having different motilities. The numbers of repeat units corresponded to the intensity of each spot. The data described in the figure legends in more detail.

3.3. The qPCR Analysis of CNVs Detected on Two BAC Clones. The results of qPCR conducted for two BAC clone regions are described in Figure 3 (RP11-89B15) and Figure 4 (RP11-79O18). For the former case, a part of a gene (*MEOX2*) was deleted. On the contrary, for the latter case, the copy number of a part of the gene (*NSF*) increased, but the copy number of the gene (*WNT3*) did not change.

3.4. Comparing Our Data with Published Data. There are many segmental duplications which have already been summarized in a public data base, such as Human Genome Segmental Duplication Database (TCAG Database; <http://projects.tcag.ca/cgi-bin/variation/gbrowse>), 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

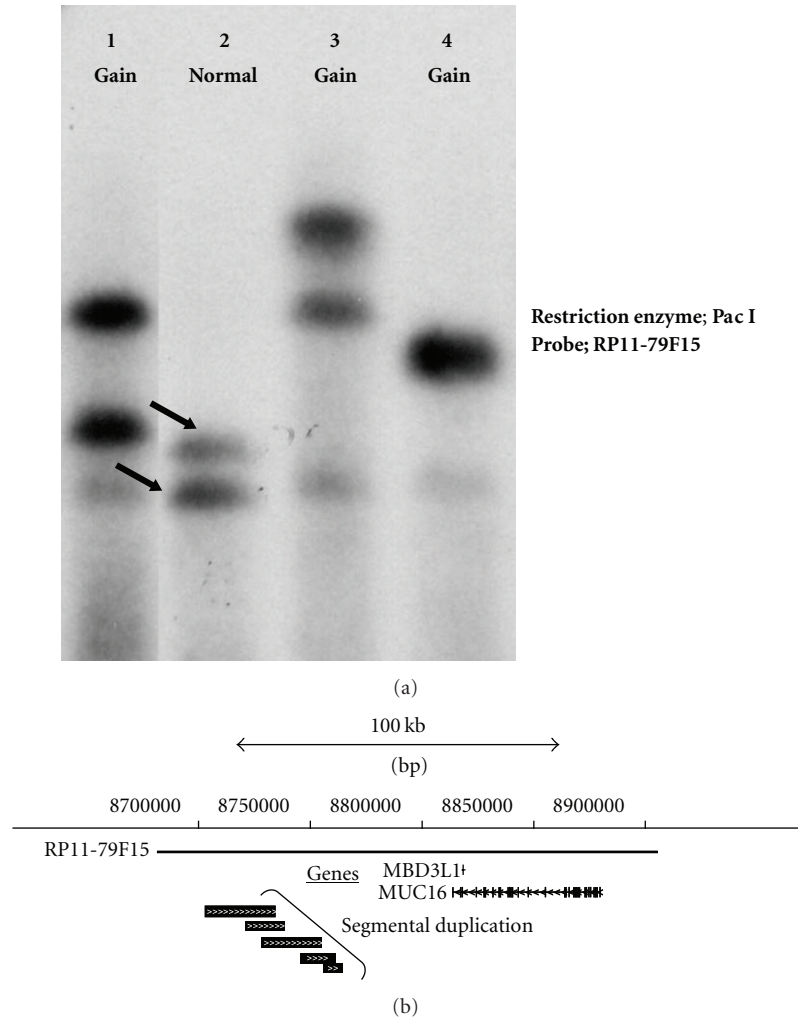


FIGURE 1: (a) Typical patterns of CNVs detected on the BAC clone (RP11-79F15) from Southern blot analyses after PFGE. (b) The Genome Map from the UCSC genome Browser. (a) The figure illustrated the pattern from three individuals showing gain on the array and one showing normal on the array. When the fragment length of the “Gain” (lanes; 1, 3, 4) was compared with “Normal” (lane 2), the segment sizes of the three (lanes; 1, 3, 4) individuals were larger than those (shown by arrows) of “Normal.” As shown in (b), five segmental duplications existed at the 5’-region of the BAC clone. That is, as these three (lanes; 1, 3, 4) contained a larger number of core segments (shown by blocks in (b)), the signals on the BAC spot were stronger than normal. So, those three were identified as “Gain.”

databases as above. The CNV data obtained in our studies are summarized in Table 1 in addition to the presence or absence of CNVs already reported in the databases. One BAC clone, named “RP11-115G22,” was mapped on two chromosomes, 6 and 15 in TCAG Database, so it is likely that this clone is present on two discrete chromosomes. On the contrary, however, that clone was mapped on only Chromosome no.15 in UCSC Database and NCBI Database. We accepted the reports from the latter two databases and described that this clone was mapped on only Chromosome no.15. With respect to segmental duplications, 10 out of 16 (about 63%) were present in the above databases (UCSC Database, and NCBI Database). It was noteworthy that the majority of BAC clones containing our CNVs were known to overlap to at least one CNV reported in the database. However, that does not mean that our highly polymorphic CNVs are exactly the same as those reported in previous

reports, since precise comparisons between our study and the other studies were not carried out.

3.5. *Comparing Our Data and the Data of Japanese Published by Redon et al. [18].* As mentioned before, there was very little information about the CNVs of 45 Japanese individuals for which relatively large sizes of population have been systematically screened. We compared our data with the data of Japanese including the HapMap project as reported by Redon et al. [18], in which CNVs were examined by a tiling BAC-array. The results are summarized in Table 1. Ten out of 16 CNVs identified in our study were reported in the data reported by Redon et al. [18], although we should emphasize again that the CNVs identified in our BAC region are not exactly the same as those reported by Redon et al. [18]. CNVs identified in our study with lower numbers tend to not be identified in Redon’s report [18]. On the other

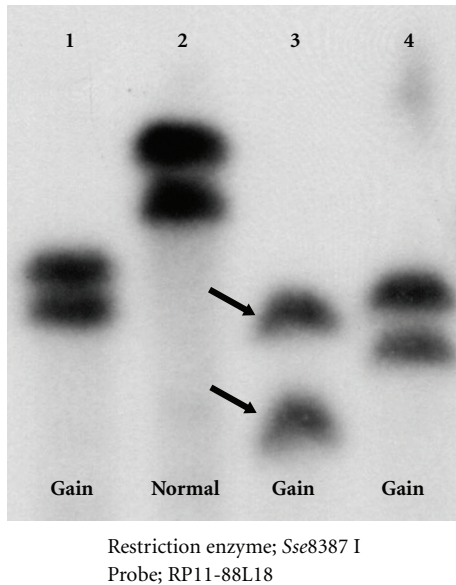


FIGURE 2: Typical patterns of CNVs detected on the BAC clone (RP11-88L18) from Southern blot analyses after PFGE. Very complicated segmental duplication patterns are present in the region of this BAC clone. The electrophoresis pattern showed the complicated patterns according to the segmental duplication. When the fragment length of the “Gain” (lanes; 1, 2, 4) was compared with “Normal” (lane 3), the segment sizes of the three (lanes; 1, 2, 4) individuals were larger than those (shown by arrows) of “Normal.” That is, as those three contained a larger number of core segments, the signals on the BAC spot were stronger than normal. So, those three were identified as “Gain.”

hand, when our data are compared with Redon’s oligo-data from Affymetrix 500EA array conducted for Japanese, only two CNVs were overlapped to our CNVs. As Redon et al. mentioned in their report [18], the reason appears to be that oligo arrays have some limitations for a complicated genome, such as segmental duplication areas.

3.6. Genes and Disease-Related Genes Listed in OMIM (Online Mendelian Inheritance in Man) (<http://www.ncbi.nlm.nih.gov/omim>). We summarized the genes and disease-related genes in OMIM which overlap to the BAC-clone region with our CNVs (Table 2). In addition to those two categories, mRNAs have been also reported in the database, but they are too many to describe here. All BAC-clone regions contained at least one mRNA, although the functions of a majority of those mRNAs are not known yet (data not shown).

4. Discussion

We examined 213 unrelated Japanese using BAC-aCGH and found a total of 680 CNVs on 16 BAC clones. A large fraction of the regions involved in the CNVs observed in our study (i.e., 625 out of 680 (92%), Table 1) have been reported previously in other studies listed in the database. That suggests that the structural rearrangements are evolutionarily

ancient. A majority ($\approx 63\%$) of the CNVs had been found on the BAC clones that overlapped with segmental duplication, suggesting the notion that segmental duplication might play a significant role in the creation of CNVs (Table 1). The observations are supported by previous data from Sharp et al. [11] in which they reported the sharing of CNVs among several populations, meaning those specific genomic imbalances either predated the dispersal of modern humans out of Africa or arose independently in different populations.

On the other hand, the data from Japanese are limited to about 45 individuals from the HapMap study [18]. We compare our CNVs data to the Japanese data from Redon et al. [18]. Our CNVs, especially those showing high frequencies, were also identified by Redon’s work [18]. On the contrary, our CNVs showing low frequency, such as less than about 10% of individuals, were not observed in their work (Table 1). That result suggests that those CNVs might be identified if the number of individuals examined by Redon et al. were increased to the level of our study (about 200 individuals). Moreover, as described before, when our data compared with Redon’s oligo-data conducted for Japanese, only two of their CNVs were overlapped to our CNVs. Redon et al. [18] suggested that BAC- and oligo-based methods complement each other. Although the oligo-based method tends to detect smaller CNVs (about a few kilobases), this approach is less effective in tracking CNVs in genomic regions of complex structure, like segmental duplications, that are not sufficiently tagged by oligo targets. On the contrary, the BAC platform can only identify larger CNVs (>40 kb), but this method has some advantage for detecting CNVs present in the regions of segmental duplications.

The CNVs in the human genome are often associated with developmental disorders and susceptibility to diseases. More importantly, CNVs may represent a major genetic component of our phenotypic diversity. Large duplications and deletions have been known to be present within the human genome based initially on cytogenetic observations in the course of etiological studies of congenital malformations (e.g., [1, 2]). The frequency of those duplications and deletions was presumed to be low and, for the most part, directly related to specific genetic disorders. A limited number of studies reported the presence of specific large duplications and deletions that were not apparently related to diseases (e.g., [26]).

Pinkel et al. [4] reported that CGH on a BAC-DNA-based microarray could reliably detect single-copy gene decreases or increases from normal diploidy. Following this, other array platforms, such as cDNA [6], and oligo-nucleotide [11], have been developed and many data from them have been reported.

As we mentioned before, many CNVs were reported to be closely related to disease phenotypes, and recent studies based on advanced molecular technologies, such as genomewide association studies [27–29] and next generation sequencing [30, 31], reported that many genes appear to play important roles in the etiology of common diseases. We report our highly polymorphic CNVs in BAC clones, which were reported to contain genes, expected to be related to phenotypic heterogeneity of each individual, based on the

TABLE 1: Summary of highly polymorphic CNVs detected.

Clone number	Chrom. Number	Position (Mb)	Variant [¶]		Known (K) or Novel (N) [§]	Segmental Duplication*	Data of Redon [18] [#]
			Gain	Loss			
RP11-259N12	1	103.8	10	48	K	P	P
RP11-89B17	2	132.1	24	0	K	P	P
RP11-11M9	4	19.1	0	15	N	A	A
RP11-231J7	4	158.7	0	23	N	A	A
RP11-88L18	5	17.5	28	33	K	P	P
RP11-90A9	5	79.9	57	3	K	P	A
RP1-299C21	6	103.8	4	79	K	A	P
RP11-121A8	7	38.0	13	0	K	P	A
RP11-89B15	7	15.4	0	17	N	A	A
RP11-90G23	8	86.7	0	51	K	P	P
RP11-95J4	9	112.9	11	0	K	A	P
RP11-100C24	13	56.6	38	22	K	P	P
RP11-115G22	15	30.3	72	0	K	P	P
RP11-79O18	17	42.1	26	0	K	P	P
CTD-2100F13	18	19.3	14	0	K	A	A
RP11-79F15	19	8.7	85	7	K	P	P

[¶] Number of individuals with CNVs.

[§] Known (K). The CNVs are already known by the published reports. Novel (N), The CNVs are not reported previously.

* P: Present, A: Absent. P means that the presence of segmental duplications are reported in the "Database of Genomic Variants" made by The Centre for Applied Genomics, or A means that those are not reported in the database.

[#] P means that our CNVs present in BAC-array data from Redon et al. [18]. A means that the CNVs absent in the data.

TABLE 2: Genes or disease gene (OMIM*) in BAC clones with highly polymorphic CNVs.

Clone number	Genes	OMIM
RP11-89B17	CCDC74A, AC093838, POTEK	
RP11-88L18	AC106774	
RP11-90A9	ANKRD34B, DHFR, SSU_rRNA_5, AC008434, MSH3	(i) DIHYDROFOLATE REDUCTASE; DHFR (ii) MutS
RP1-299C21		
RP11-89B15	MEOX2, AC005550	
RP11-90G23	REXO1L2P, AC232323	
RP11-95J4		
RP11-100C24	PRR20A, PRR20B, PRR20C, PRR20D, PRR20E	
RP11-115G22	CHRNA7	CHOLINERGIC RECEPTOR, NEURONAL NICOTINIC, ALPHA POLYPEPTIDE 7; CHRNA7
RP11-79O18	NSF, WNT3	WINGLESS-TYPE MMTV INTEGRATION SITE FAMILY, MEMBER 3; WNT3
RP11-79F15	ZNF558, MBD3L1, AC008734	
RP11-259N12	AMY1A, AMY1B, AMY1C, AMY2A	
RP11-121A8	STARD3NL, AC006033, TRGC2, TRGJ2, TRGJP2, TRGC1, TRGJ1, TRGJP, TRGJP1, TRGV11, TRGVB, TRGV10, TRGV9, TRGVA, TRGV8, TRGV7, TRGV6, TRGV5P, TRGV5	
RP11-11M9		
RP11-231J7	GRIA2	
CTD-2100E13	RIOK3, C18orf8, NPC1, ANKRD29	NPC1 GENE; NPC1

* OMIM: Online Mendelian Inheritance in Man.

TCAG Database. We focused on genes reported in the above database, although many mRNA were listed in the database in addition to genes (Table 2).

BAC clone (RP11-90A9) contains two genes: ankyrin repeat domain 34B (*ANKRD34B*) and dihydrofolate reductase (*DHFR*). A phosphoprotein encoded by *ANKRD34B* is induced during bone marrow commitment to dendritic cells which play an important role in vertebrate immunity [32]. *DHFR* genes were reported to be related to various malignancies including lymphoproliferative disorders such as systemic non-Hodgkin's lymphoma, primary central nervous system lymphoma (PCNSL) [33], and childhood acute lymphoblastic leukemia [34]. Those two genes are fully overlapped to the BAC clone (RP11-90A9). It is likely that the CNVs might affect the copy number of those genes.

One BAC clone (RP11-89B15) contains a gene mesenchyme homeobox 2 (*MEOX2*). The analyses by qPCR (Figure 3) demonstrated that the copy number of gene (*MEOX2*) is deleted in the individual having CNVs detected by this BAC clone. Loss of *MEOX2* gene is associated with Wilms tumor [35]. *MEOX2* induces senescence through controlling INK4a activity [36]. *MEOX2* suppressed epithelial cell proliferation in cooperation with TGF-beta1, and mediated induction of the cell-cycle inhibitor gene *p21* [37]. Finally, the data from genome wide association study (GWAS) reported that this is one of the candidate genes that might be associated with ischaemic stroke [38].

Another BAC clone (RP11-115G22) was mapped on the Chromosomes 15. The BAC clone contains one gene "cholinergic receptor, nicotinic, alpha 7" (*CHRNA7*). That gene was used as a candidate target for examining interactions on the severity of adult attention deficit hyperactivity disorder (ADHD) [39]. It is suggested that *CHRNA7* regulates airway epithelium differentiation by controlling basal cell proliferation [40]. Moreover, it was reported that the gene is one of the candidates for Alzheimer's disease [41]. The gene *CHRNA7* is overlapped to the segmental duplication region of BAC clone (RP11-115G22). Many CNVs were listed in the Databases mentioned as above, and those are overlapped to the gene. For those reasons, our CNV appears to be overlapped to the genes, and it might affect the copy number of gene *CHRNA7*.

A BAC clone (RP11-79O18) contains two genes which are N-ethylmaleimide-sensitive factor (*NSF*) and wingless-type MMTV integration site family, member 3 (*WNT3*). The qPCR results (Figure 4) demonstrated that the copy number of a gene (*NSF*) increased but no change was observed in the gene (*WNT3*). That result suggested that the CNV might affect the *NSF* gene, but not the *WNT3* gene. However, there may be an opportunity for the CNVs to become a surrogate marker of *WNT3* for the future association study between the CNV and some disease phenotype. The *NSF* gene is one of the essential components of membrane fusion machinery which is an important homeostatic process in eukaryotic cells [42]. The gene also plays an important role for assembly during human sperm exocytosis [43]. A recent study showed that the *NSF* gene is a good candidate marker for association studies for genetic risk underlying Parkinson's disease [44]. The *WNT3* gene's single nucleotide polymorphisms (SNPs)

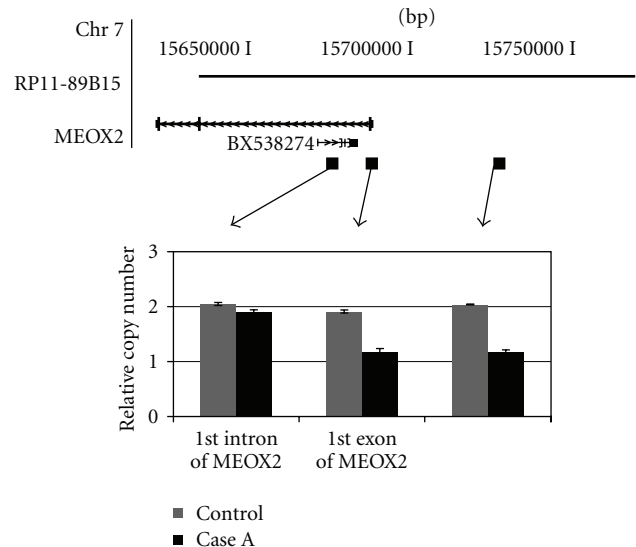


FIGURE 3: Characterization of a loss-type CNV detected on a BAC-clone RP11-89B15. A genome map of the RP11-89B15-containing region was obtained from the UCSC Genome Browser website. The distance (bp) from the p-terminal of chromosome 7 is shown in the uppermost part. That region contains the gene which is the gene mesenchyme homeobox 2 (*MEOX2*). 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 ($n = 3$). The individual (Case A) has variant fragments with a deletion that contains the 1st exon and 5'-end of the *MEOX2* gene.

were used as candidate markers for association studies of hemorrhagic stroke [45], hypertension [46], and chronic kidney disease [47]. Upregulation of the *WNT* gene family, including *WNT3*, suggested involvement of the *WNT*'s canonical and/or noncanonical signaling pathway in chronic lymphocytic leukemia [48]. *WNT3* signaling is also required in primary axis formation during vertebrate embryogenesis [49].

A BAC clone (RP11-79F15) contains two genes: methyl-CpG binding domain protein 3-like 1 (*MBD3L1*) and cell surface associated mucin 16 (*MUC16*). *MBD3L1* encodes a protein that is related to methyl-CpG-binding proteins. The protein is localized to discrete areas in the nucleus, and expression appears to be restricted to round spermatids, suggesting that the protein plays a role in the postmeiotic stages of male germ cell development [50]. On the other hand, *MUC16* is a member of the mucin gene family and encodes cancer antigen 125 (CA125) which is a blood biomarker routinely used to monitor the progression of human epithelial ovarian cancer (EOC), although its potential role in EOC is poorly understood [51]. Maintenance of an intact mucosal barrier, one of whose components is a gene product of *MUC16*, is critical to preventing damage and infection of wet-surfaced epithelia [52]. As we demonstrated by Southern blot analysis (Figure 1), the polymorphism was caused by segmental duplications in the BAC clone. The series of segmental duplications were on the 5'-region of two

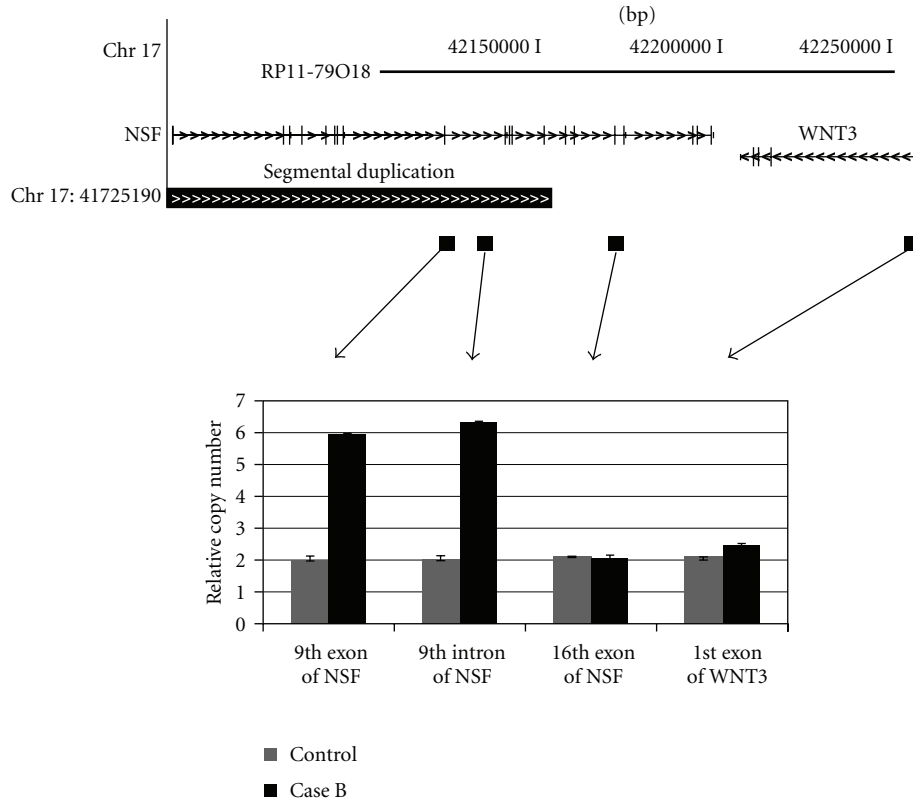


FIGURE 4: Characterization of a gain-type CNV detected on a BAC-clone RP11-79O18. A genome map of the RP11-79O18-containing region was obtained from the UCSC Genome Browser website. The distance (bp) from the p-terminal of chromosome 17 is shown in the uppermost part. That region contains two genes, that is, N-ethylmaleimide-sensitive factor (*NSF*) and wingless-type MMTV integration site family, member 3 (*WNT3*). The positions of PCR primers examined are shown by the solid boxes. The experimental conditions are described more precisely in the text. The individual (Case B) has variant fragments with gain type. For the individual showing a CNV, a part of region of the gene *NSF* was amplified ($\times 3$). However, no copy number change was observed in the region of *WNT3*.

genes (*MBD3L1* and *MUC16*). The CNVs do not affect the expression pattern of the genes, but these CNVs might be useful surrogate markers for future studies.

A BAC clone (RP11-231J7) contains one gene “glutamate receptor, ionotropic, (*AMPA 2*).” AMPAs are ligand-activated cation channels that mediate the fast component of excitatory postsynaptic currents in neurons of the central nervous system [53]. Since the size of the gene (*AMPA 2*) is larger than that of the BAC clone (RP11-231J7), it is likely that the CNVs affect the copy number of the gene.

A BAC clone (CTD-2100F13) contains two genes: RIO kinase 3 (yeast) (*RIOK3*) and Niemann-Pick disease, type C1 (*NPC1*). *RIOK3* promotes pancreas ductal cell motility and invasion in pancreatic cancer [54]. *NPC1* can have a function in the egress of certain membrane-impermeable lysosomal cargo [55]. The membrane-bound *NPC1* and soluble *NPC2* play an important role for the release of cholesterol from lysosomes. As a result of that mechanism, the gene is associated with obesity [56–58]. The BAC clone (CTD-2100F13) is fully overlapped to both of the two genes (*RIOK3* and *NPC1*). The size of CNVs summarized in the databases shows that the reported CNVs were overlapped to both genes. We assumed that our CNVs may reflect the change of copy number of the genes themselves.

Since the genes mentioned are good candidate markers for enabling us to examine the etiology of common diseases and phenotypical heterogeneities among individuals, our highly polymorphic CNVs should be able to become good markers in future studies.

5. Future Plans

We are currently planning to examine the same population using a high-density oligo-array platform to accumulate more CNV data for Japanese. The reason, as mentioned before, is that the BAC platform and oligo methods complement each other. The oligo platform is known to be more effective in detecting smaller CNVs (around a few kilobases), even though this approach is less effective in tracing CNVs in genomic regions of complex structure that are covered by the BAC approach conducted in this study. We expect to construct a more definite Japanese CNV database by the combination of BAC- and oligo-platform arrays.

6. Conclusion

We conducted population screening for 213 unrelated Japanese, and observed 680 highly polymorphic CNVs. The

majority of the polymorphic CNVs presented on BAC clones that overlapped with regions of segmental duplication, and had been previously reported in other publications. Some CNVs contained genes which might be related to phenotypic heterogeneity among individuals. Moreover, it is expected that the CNVs might be good surrogate markers for detecting etiological genes, even if CNVs did not directly affect the genes themselves.

Acknowledgments

The authors thank H. Omine, J. Kaneko, A. Miura, M. Imanaka, and E. Nishikori for their technical assistant. The authors are grateful to E. B. Douple for critical reading of the paper. This publication was supported by Research Protocols RP 1-01 and RP 2-07 of the Radiation Effects Research Foundation (RERF) and in part by Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology and the Japan Science and Technology Agency (Core Research for Evolutional Science and Technology). RERF, Hiroshima and Nagasaki, Japan is a private nonprofit foundation funded by the Japanese Ministry of Health, Labour and Welfare and the U.S. Department of Energy, the latter in part through the National Academy of Sciences. Norio Takahashi and Yasunari Satoh equally contributed to this work as the first authors.

References

- [1] K. Patau, D. Smith, E. Therman, S. Inhorn, and H. Wagner, "Multiple congenital anomaly caused by an extra autosome," *The Lancet*, vol. 275, no. 7128, pp. 790–793, 1960.
- [2] P. A. Jacobs, C. Browne, N. Gregson, C. Joyce, and H. White, "Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding," *Journal of Medical Genetics*, vol. 29, no. 2, pp. 103–108, 1992.
- [3] J. S. Beckmann, X. Estivill, and S. E. Antonarakis, "Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability," *Nature Reviews Genetics*, vol. 8, no. 8, pp. 639–646, 2007.
- [4] D. Pinkel, R. Seagraves, 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] A. J. Sharp, D. P. Locke, S. D. 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.
- [12] 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.
- [13] T. Díaz 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.
- [14] 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.
- [15] D. Pinto, C. Marshall, L. Feuk, and S. W. Scherer, "Copy-number variation in control population cohorts," *Human Molecular Genetics*, vol. 16, pp. R168–R173, 2007.
- [16] G. Zogopoulos, K. C. H. 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.
- [17] S. J. White, L. E. L. M. Vissers, A. Geurts Van Kessel et al., "Variation of CNV distribution in five different ethnic populations," *Cytogenetic and Genome Research*, vol. 118, no. 1, pp. 19–30, 2007.
- [18] 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.
- [19] V. G. Cheung, N. Nowak, W. Jang et al., "Integration of cytogenetic landmarks into the draft sequence of the human genome," *Nature*, vol. 409, no. 6822, pp. 953–958, 2001.
- [20] 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.
- [21] T. Maniatis, E. F. 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, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 1982.
- [22] C. Satoh, N. Takahashi, J.-I. Asakawa et al., "Genetic analysis of children of atomic bomb survivors," *Environmental Health Perspectives*, vol. 104, no. 3, pp. 511–519, 1996.
- [23] A. M. Snijders, N. Nowak, R. Seagraves et al., "Assembly of microarrays for genome-wide measurement of DNA copy number," *Nature Genetics*, vol. 29, no. 3, pp. 263–264, 2001.
- [24] G. Chu, D. Vollrath, and R. W. Davis, "Separation of large DNA molecules by contour-clamped homogeneous electric fields," *Science*, vol. 234, no. 4783, pp. 1582–1585, 1986.
- [25] J. Sambrook, E. F. Fritsch, and T. Maniatis, "Analysis of genomic DNA by Southern hybridization," in *Molecular Cloning. A Laboratory Manual*, pp. 9.31–9.62, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 1989.
- [26] J. J. M. Engelen, U. Moog, J. L. H. Evers, H. Dassen, J. C. M. Albrechts, and A. J. H. Hamers, "Duplication of chromosome region 8p23.1 → p23.3: a benign variant?" *American Journal of Medical Genetics*, vol. 91, no. 1, pp. 18–21, 2000.

- [27] 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.
- [28] S. A. McCarrroll, F. G. Kuruville, J. M. Korn et al., "Integrated detection and population-genetic analysis of SNPs and copy number variation," *Nature Genetics*, vol. 40, no. 10, pp. 1166–1174, 2008.
- [29] 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.
- [30] 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, pp. e151–e151, 2010.
- [31] 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.
- [32] N. Al-Shaibi and S. K. Ghosh, "A novel phosphoprotein is induced during bone marrow commitment to dendritic cells," *Biochemical and Biophysical Research Communications*, vol. 321, no. 1, pp. 26–30, 2004.
- [33] D. Kurzwelly, S. Knop, M. Guenther et al., "Genetic variants of folate and methionine metabolism and PCNSL incidence in a German patient population," *Journal of Neuro-Oncology*. In press.
- [34] F. Al-Shakfa, S. Dulucq, I. Brukner et al., "DNA variants in region for noncoding interfering transcript of Dihydrofolate reductase gene and outcome in childhood acute lymphoblastic leukemia," *Clinical Cancer Research*, vol. 15, no. 22, pp. 6931–6938, 2009.
- [35] J. Ohshima, M. Haruta, Y. Arai et al., "Two candidate tumor suppressor genes, MEOX2 and SOSTDC1, identified in a 7p21 homozygous deletion region in a Wilms tumor," *Genes Chromosomes and Cancer*, vol. 48, no. 12, pp. 1037–1050, 2009.
- [36] J. T. Ireland, A. G. del Arroyo, A. Gutierrez et al., "A functional screen for regulators of CKDN2A reveals MEOX2 as a transcriptional activator of INK4a," *PLoS ONE*, vol. 4, no. 4, article e5067, 2009.
- [37] U. Valcourt, S. Thuault, K. Pardali, C.-H. Heldin, and A. Moustakas, "Functional role of Meox2 during the epithelial cytotatic response to TGF- β ," *Molecular Oncology*, vol. 1, no. 1, pp. 55–71, 2007.
- [38] M. Matarín, W. M. Brown, S. Scholz et al., "A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release," *The Lancet Neurology*, vol. 6, no. 5, pp. 414–420, 2007.
- [39] D. J. Müller, A. Chiesa, L. Mandelli et al., "Correlation of a set of gene variants, life events and personality features on adult ADHD severity," *Journal of Psychiatric Research*, vol. 44, no. 9, pp. 598–604, 2010.
- [40] K. Maouche, M. Polette, T. Jolly et al., " α 7 nicotinic acetylcholine receptor regulates airway epithelium differentiation by controlling basal cell proliferation," *American Journal of Pathology*, vol. 175, no. 5, pp. 1868–1882, 2009.
- [41] E. L. Heinzen, A. C. Need, K. M. Hayden et al., "Genome-wide scan of copy number variation in late-onset Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 19, no. 1, pp. 69–77, 2010.
- [42] Y. Thielmann, O. H. Weiergra, P. Ma, M. Schwarten, J. Mohrlüder, and D. Willbold, "Comparative modeling of human NSF reveals a possible binding mode of GABARAP and GATE-16," *Proteins*, vol. 77, no. 3, pp. 637–646, 2009.
- [43] V. E. P. Zarelli, M. C. Ruete, C. M. Roggero, L. S. Mayorga, and C. N. Tomes, "PTP1B dephosphorylates N-ethylmaleimide-sensitive factor and elicits SNARE complex disassembly during human sperm exocytosis," *Journal of Biological Chemistry*, vol. 284, no. 16, pp. 10491–10503, 2009.
- [44] J. Simón-Sánchez, C. Schulte, J. M. Bras et al., "Genome-wide association study reveals genetic risk underlying Parkinson's disease," *Nature Genetics*, vol. 41, no. 12, pp. 1308–1312, 2009.
- [45] T. Yoshida, K. Kato, K. Yokoi et al., "Association of genetic variants with hemorrhagic stroke in Japanese individuals," *International Journal of Molecular Medicine*, vol. 25, no. 4, pp. 649–656, 2010.
- [46] M. Oguri, K. Kato, K. Yokoi et al., "Assessment of a polymorphism of SDK1 with hypertension in Japanese individuals," *American Journal of Hypertension*, vol. 23, no. 1, pp. 70–77, 2010.
- [47] T. Yoshida, K. Kato, K. Yokoi et al., "Association of gene polymorphisms with chronic kidney disease in Japanese individuals," *International Journal of Molecular Medicine*, vol. 24, no. 4, pp. 539–547, 2009.
- [48] A. Memarian, M. Hojjat-Farsangi, H. Asgarian-Omran et al., "Variation in WNT genes expression in different subtypes of chronic lymphocytic leukemia," *Leukemia and Lymphoma*, vol. 50, no. 12, pp. 2061–2070, 2009.
- [49] P. Liu, M. Wakamiya, M. J. Shea, U. Albrecht, R. R. Behringer, and A. Bradley, "Requirement for Wnt3 in vertebrate axis formation," *Nature Genetics*, vol. 22, no. 4, pp. 361–365, 1999.
- [50] C.-L. Jiang, S.-G. Jin, D.-H. Lee et al., "MBD3L1 and MBD3L2, two new proteins homologous to the methyl-CpG-binding proteins MBD2 and MBD3: characterization of MBD3L1 as a testis-specific transcriptional repressor," *Genomics*, vol. 80, no. 6, pp. 621–629, 2002.
- [51] J. A.A. Gubbels, M. Felder, S. Horibata et al., "MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells," *Molecular Cancer*, vol. 9, article 11, 2010.
- [52] P. Argüeso, A. Guzman-Aranguez, F. Mantelli, Z. Cao, J. Ricciuto, and N. Panjwani, "Association of cell surface mucins with galectin-3 contributes to the ocular surface epithelial barrier," *Journal of Biological Chemistry*, vol. 284, no. 34, pp. 23037–23045, 2009.
- [53] M. A. Sutton, E. F. Schmidt, K.-H. Choi et al., "Extinction-induced upregulation in AMPA receptors reduces cocaine-seeking behaviour," *Nature*, vol. 421, no. 6918, pp. 70–75, 2003.
- [54] A. C. Kimmelman, A. F. Hezel, A. J. Aguirre et al., "Genomic alterations link Rho family of GTPases to the highly invasive phenotype of pancreas cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19372–19377, 2008.
- [55] S. D. B. Goldman and J. P. Krise, "Niemann-Pick C1 functions independently of Niemann-Pick C2 in the initial stage of retrograde transport of membrane-impermeable lysosomal cargo," *Journal of Biological Chemistry*, vol. 285, no. 7, pp. 4983–4994, 2010.
- [56] K. Hotta, M. Nakamura, T. Nakamura et al., "Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population," *Journal of Human Genetics*, vol. 54, no. 12, pp. 727–731, 2009.
- [57] H. J. Kwon, L. Abi-Mosleh, M. L. Wang et al., "Structure of N-terminal domain of NPC1 reveals distinct subdomains for

binding and transfer of cholesterol," *Cell*, vol. 137, no. 7, pp. 1213–1224, 2009.

- [58] E. D. Carstea, J. A. Morris, K. G. Coleman et al., "Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis," *Science*, vol. 277, no. 5323, pp. 228–231, 1997.