

RESEARCH ARTICLE

Open Access



Sensitivity to gene dosage and gene expression affects genes with copy number variants observed among neuropsychiatric diseases

Maria Yamasaki^{1*}, Takashi Makino², Seik-Soon Khor³, Hiromi Toyoda^{3,4}, Taku Miyagawa⁵, Xiaoxi Liu⁶, Hitoshi Kuwabara⁷, Yukiko Kano^{8,9}, Takafumi Shimada¹⁰, Toshiro Sugiyama⁸, Hisami Nishida¹¹, Nagisa Sugaya¹², Mamoru Tochigi¹³, Takeshi Otowa¹⁴, Yuji Okazaki¹⁵, Hisanobu Kaiya¹⁶, Yoshiya Kawamura¹⁷, Akinori Miyashita¹⁸, Ryozo Kuwano^{18,19}, Kiyoto Kasai²⁰, Hisashi Tani²¹, Tsukasa Sasaki²², Makoto Honda⁵ and Katsushi Tokunaga^{3,4}

Abstract

Background: Copy number variants (CNVs) have been reported to be associated with diseases, traits, and evolution. However, it is hard to determine which gene should have priority as a target for further functional experiments if a CNV is rare or a singleton. In this study, we attempted to overcome this issue by using two approaches: by assessing the influences of gene dosage sensitivity and gene expression sensitivity. Dosage sensitive genes derived from two-round whole-genome duplication in previous studies. In addition, we proposed a cross-sectional omics approach that utilizes open data from GTEx to assess the effect of whole-genome CNVs on gene expression.

Methods: Affymetrix Genome-Wide SNP Array 6.0 was used to detect CNVs by PennCNV and CNV Workshop. After quality controls for population stratification, family relationship and CNV detection, 287 patients with narcolepsy, 133 patients with essential hypersomnia, 380 patients with panic disorders, 164 patients with autism, 784 patients with Alzheimer disease and 1280 healthy individuals remained for the enrichment analysis.

(Continued on next page)

* Correspondence: myamasaki-ky@umin.ac.jp

¹Department of Health Data Science Research, Healthy Aging Innovation Center, Tokyo Metropolitan Geriatric Medical Center, Tokyo, Japan
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Results: Overall, significant enrichment of dosage sensitive genes was found across patients with narcolepsy, panic disorders and autism. Particularly, significant enrichment of dosage-sensitive genes in duplications was observed across all diseases except for Alzheimer disease. For deletions, less or no enrichment of dosage-sensitive genes with deletions was seen in the patients when compared to the healthy individuals. Interestingly, significant enrichments of genes with expression sensitivity in brain were observed in patients with panic disorder and autism. While duplications presented a higher burden, deletions did not cause significant differences when compared to the healthy individuals. When we assess the effect of sensitivity to genome dosage and gene expression at the same time, the highest ratio of enrichment was observed in the group including dosage-sensitive genes and genes with expression sensitivity only in brain. In addition, shared CNV regions among the five neuropsychiatric diseases were also investigated.

Conclusions: This study contributed the evidence that dosage-sensitive genes are associated with CNVs among neuropsychiatric diseases. In addition, we utilized open data from GTEx to assess the effect of whole-genome CNVs on gene expression. We also investigated shared CNV region among neuropsychiatric diseases.

Keywords: Copy number variants, Ohnolog, Two-round whole-genome duplication, Gene dosage sensitivity, Gene expression sensitivity, Neuropsychiatric diseases

Background

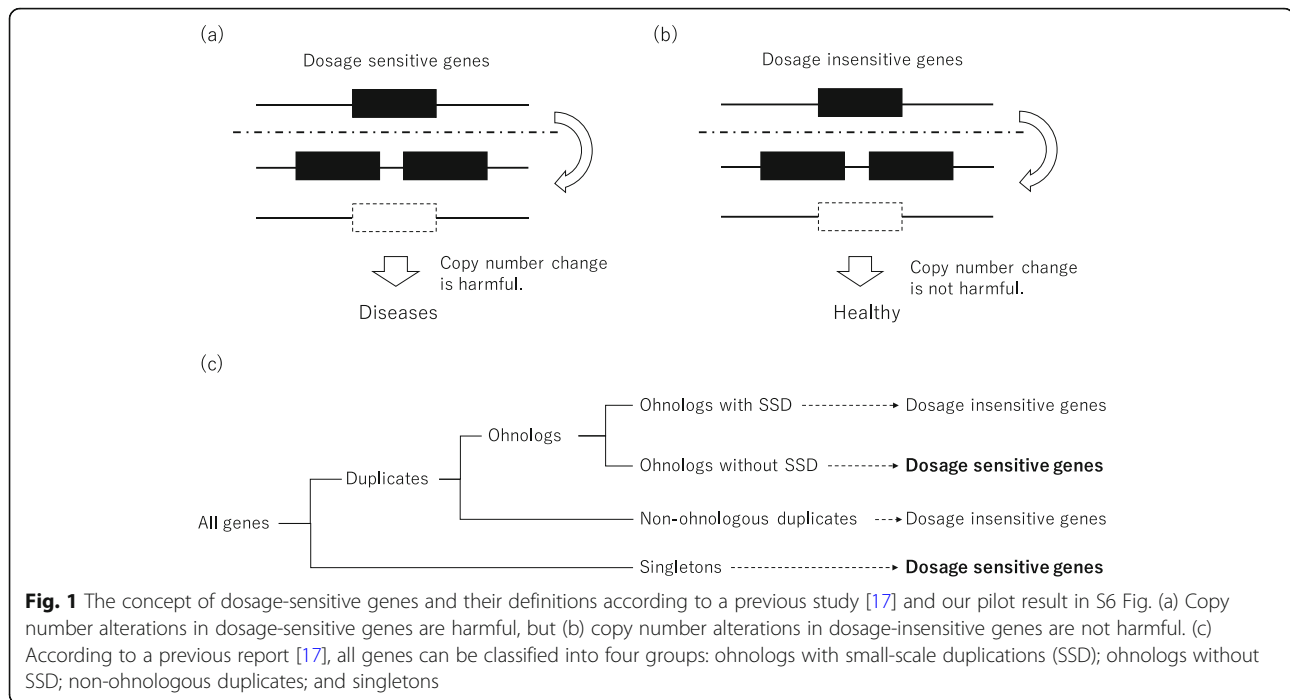
Copy number variants (CNVs) have been reported to be associated with diseases, traits, and evolution [1–5]. With new technologies for detecting CNVs, research in these fields have progressed rapidly [6–8]. However, when it comes to clinical applications, several issues still remain. One is that although rare CNVs have been reported to be associated with diseases [2, 3], the rarity of the CNVs makes them difficult to study for elucidating the pathogenicity of the diseases. This is similar to the situation of whole- or exome-sequencing analysis, from which a large number of rare mutations or singletons have been discovered [9]. Secondly, CNVs often span a few megabases and cover several genes [2, 3]; this makes it hard to determine which gene should have priority as a target for further functional experiments. Recent mega biobank projects [10–12] and international data-sharing consortia [13] have enabled the first issue to be overcome; however, the second issue remains unresolved. In this study, we attempted to overcome the second issue by using two approaches: by assessing the influences of gene dosage sensitivity and gene expression sensitivity.

Gene dosage sensitivity has been of increasing interest because it might provide a clue for elucidating the pathogenicity of diseases. As such, ClinGen Dosage Sensitivity Map (<https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/>) has started to accumulate data on dosage-sensitive genes. The notion of gene dosage sensitivity derived from the gene balance hypothesis, which was suggested in earlier studies [14–16]. In short, the gene balance hypothesis states that a stoichiometric balance is maintained among all of the complex gene products in a pathway, so a copy number change in a single gene of a pathway would be deleterious. Thus, genes under this hypothesis are thought to be dosage-sensitive

genes. Copy number alterations, for example, CNVs, in dosage-sensitive genes are harmful and might affect the onset of diseases (Fig. 1(a) and (b)). In contrast, copy number alterations in dosage-insensitive genes are not harmful and might have no effect on the onset of diseases. As such, dosage-sensitive genes may provide useful information in the research of diseases.

Recently, dosage-sensitive genes were effectively identified based on a particular hypothesis and applied to narrow down susceptible genes. In 1970, Susumu Ohno proposed a hypothesis that over the course of evolution, vertebrates experienced genome-wide duplication twice; this is otherwise known as two-round whole-genome duplication (2R-WGD) [18]. Genes that existed at the age of 2R-WGD are called ohnologs, named in honor of Ohno. Studies of ohnologs revealed that the pattern of retained ohnologs was not random and that the currently existing ohnologs are in gene balance as dosage-sensitive genes (Fig. 1(c)) [17, 19]. Several studies have demonstrated the importance and applicability of the dosage sensitivity of ohnologs using CNVs from databases or previously reported pathogenic CNVs [20–23]. Firstly, the susceptibility region for Down syndrome on 21q22.13 was reported to be enriched with ohnologs [19]. Human monogenic disease genes in the Online Mendelian Inheritance in Man (OMIM) (<https://www.omim.org>) database and previous literature were also found to be enriched in ohnologs [20, 22]. Similarly, previously reported pathogenic genes involved in neuropsychiatric diseases were frequently uncovered to be ohnologs [21, 23].

Previous studies about application of ohnologs limited their survey to target CNVs from databases or previously reported pathogenic CNVs. They did not assess the influence of dosage-sensitive ohnologs on small CNVs or



the CNVs detected in each patient. In this study, CNVs > 100 kb in size that were observed in individuals with neuropsychiatric diseases were investigated to assess the burden of the dosage-sensitive ohnologs on these diseases.

Expression quantitative trait locus (eQTL) analysis has been the focus of attention because it allows the assessment of whether single nucleotide polymorphisms (SNPs) or variants affect the expression of genes [24]. However, eQTL analysis does not examine whether an alteration in the gene expression level is deleterious or not. We propose a concept for genes with expression sensitivity (Fig. 2(a) and (b)): modification of the expression level of genes with expression sensitivity is deleterious, while modification of the expression level of genes without expression sensitivity is not deleterious. CNVs are one possible cause of expression level changes because CNVs themselves can change the gene dosage, so our assumption is that CNVs in genes with expression sensitivity might be deleterious. We utilized open data from the Genotype-Tissue Expression (GTEx) project (<https://www.gtexportal.org/home/datasets>) [24] in order to simulate this concept. Genes with expression sensitivity and, in other words, stable expression in a certain tissue were taken to be genes that are expressed in the tissue and that do not have any eQTL SNPs in the tissue. In contrast, genes without expression sensitivity and, in a different way, unstable expression in a certain tissue were taken to be genes that are expressed in the tissue and that have at least one eQTL SNP in the tissue.

Using this definition, we assessed the effect of genes with expression sensitivity in the CNVs observed among neuropsychiatric diseases.

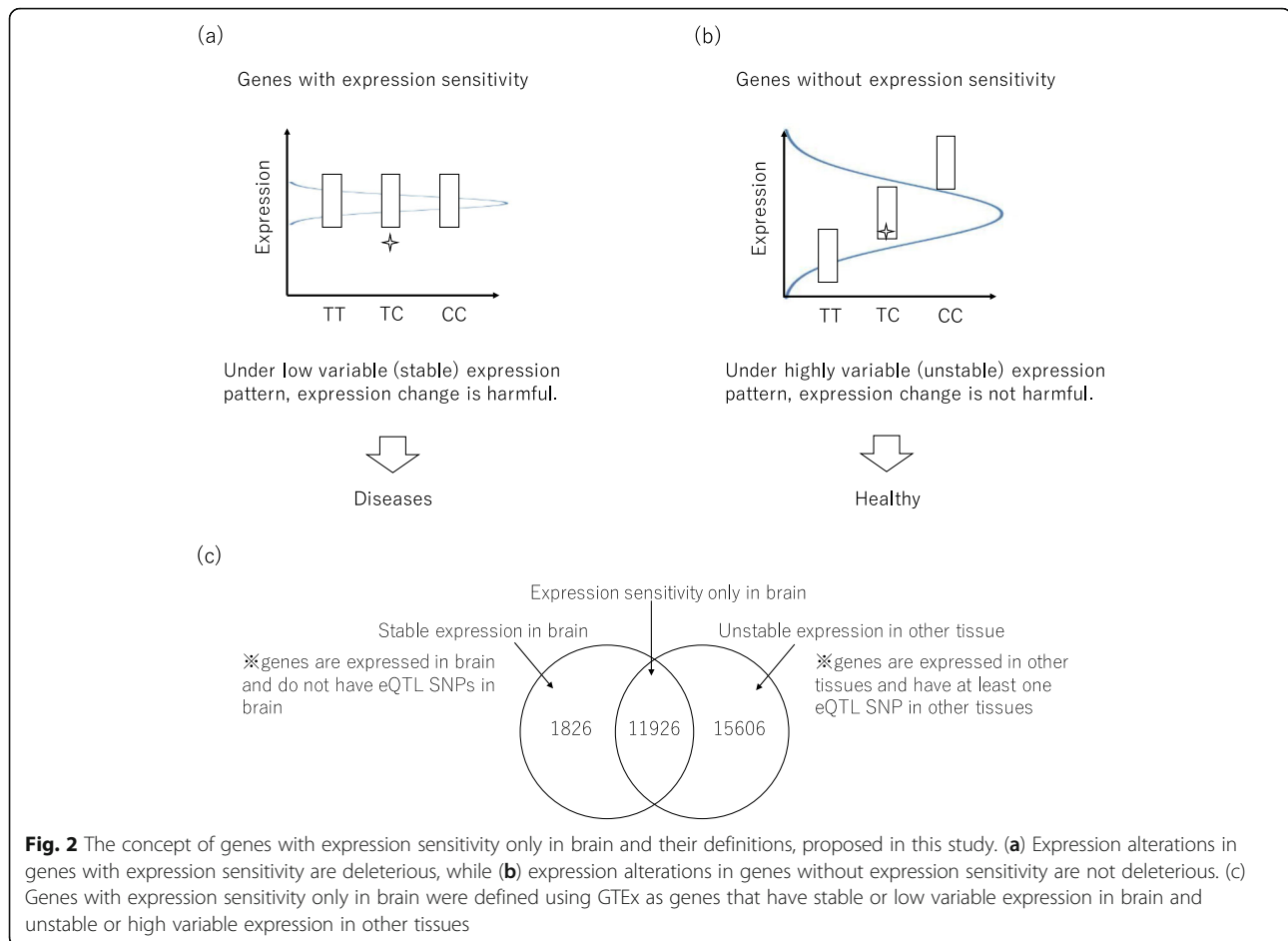
Methods

Subjects

The participants in this study were 425 patients with narcolepsy [25], 171 patients with essential hypersomnia (EHS) [26], 595 patients with panic disorders [27], 246 patients with autism [28], 1032 patients with Alzheimer disease [29] and 2135 healthy individuals. All subjects were genotyped in our previous studies, and their data were included for analysis in this study. Ethical approval was obtained from the local institutional review boards of all participating organizations. Age and gender were not matched between the participants with neuropsychiatric diseases and the healthy individuals. All individuals provided written informed consent for their inclusion in this study.

Genotyping and quality controls

Genomic DNA from all participants was genotyped for 906,622 SNPs using the Affymetrix Genome-Wide SNP Array 6.0 (Thermo Fisher Scientific, Waltham, MA) (S1 Fig). Genotype calling was done using the Birdseed algorithm in Affymetrix Power Tools software (Thermo Fisher Scientific, Waltham, MA). Quality control procedures were performed using PLINK v1.07 (<http://zzz.bwh.harvard.edu/plink/>). Samples with a call rate < 97% were excluded. For SNP quality control, SNPs with a



minor allele frequency < 0.05 , a Hardy-Weinberg equilibrium $p < 0.001$ for either the patient group or the healthy control group, and a SNP call rate $< 99\%$ were excluded. Samples with a reported family relationship with other participants or a mean probability of being identity-by-descent (PIHAT, calculated in PLINK) value > 0.185 were also excluded. Outliers in the principal component analysis using EIGENSOFT (<http://www.hsph.harvard.edu/alkes-price/software/>) were also excluded to eliminate population stratification. In the principal component analysis, data from 91 Japanese in Tokyo, Japan (JPT), 90 Han Chinese in Beijing, China (CHB), 180 Utah residents with Northern and Western European ancestry (CEU), and 180 Yoruba in Ibadan, Nigeria (YRI), obtained from the HapMap Project, were also included [30]. Data from the HapMap populations and the present sample sets were combined using common SNPs among all populations after the quality control steps described above.

CNV detection and quality controls

PennCNV (<http://www.openbioinformatics.org/>) [31] and CNV Workshop (<http://cnv.sourceforge.net>) were utilized

to detect CNVs (Supporting Information and S1 Fig) [32]. PennCNV employs the Hidden Markov Model. Briefly, PennCNV requires external reference values for the B allele frequency and log R ratio because its algorithm applying the Hidden Markov Model identifies CNV regions using the degree of deviation from these references. We used an in-house reference comprising some of the healthy individuals in our sample set.

After CNV detection by PennCNV, samples with a low detection signal were removed from the subsequent analyses. Samples with a log R ratio standard deviation $> |0.3|$, B allele frequency drift > 0.01 and CNV call count > 100 were excluded [33, 34]. CNVs with < 10 detection probes and a size < 30 kb were excluded. After the quality controls for population stratification, family relationships and CNV detection, 287 patients with narcolepsy, 133 patients with EHS, 380 patients with panic disorders, 164 patients with autism, 784 patients with Alzheimer disease and 1280 healthy individuals remained.

Gene coordinates were converted from hg18 to hg19 using Lifter (<https://genome.ucsc.edu/cgi-bin/hgLift-Over>). Artifact regions that tend to cause false-positive CNV detection, centromeric and telomeric regions (\pm

500 kb) and immunoglobulin regions (± 500 kb), were removed based on a previous study and software tutorial [3, 31].

Dosage-sensitive genes

According to a previous study, all genes can be classified into four groups based on evolutionary features and estimation methodology (Fig. 1(c)) [19]. Briefly, genes were divided into singletons and duplicates by means of a BLAST search. Next, among duplicates, genes were classified into ohnologs and non-ohnologous duplicates by comparison with other species. Ohnologs were separated based on whether or not they had undergone small-scale duplication (SSD) after 2R-WGD; these were labelled ohnologs with SSD or ohnologs without SSD. Of these four categories, ohnologs without SSD and singletons were defined as dosage-sensitive genes because genes under the gene balance hypothesis tend not to have undergone SSD after 2R-WGD, and because singletons were found to be less likely to have CNVs in our data (S2 Fig). The classification of human ohnologs was done based on a previous paper by Makino, T et al. using coordinates in Ensembl 73 [19]. According to previous reports, the numbers of dosage-sensitive genes and dosage-insensitive genes are 11,927 and 8360, respectively.

Another classification of ohnologs from OHNOLOGS (<http://ohnologs.curie.fr>) was utilized by Singh, PP et al. to identify ohnologs [35]. This previous study did not classify ohnologs based on past SSD, so the dosage-sensitive genes were not clearly defined. Here, we used a gene list from OHNOLOGS to validate the results. OHNOLOGS provides results from three different criteria for the identification of ohnologs, and a list of ohnologs defined using the strictest criteria was utilized in our analysis.

Genes with expression sensitivity only in brain

Genes with expression sensitivity *only in brain* were focused on in this study. We made the assumption that if genes with expression sensitivity in any tissue are disrupted, expression balance might be affected not only in brain, but also in other tissues, and it might contribute to neuropsychiatric diseases as well as diseases related to any of the other tissues. However, if genes with expression sensitivity *only in brain* are disrupted, it might only affect expression balance in brain and contribute to neuropsychiatric diseases. Therefore, in this study, genes with expression sensitivity *only in brain* were analyzed to evaluate the genetic background of neuropsychiatric diseases.

Genes with expression sensitivity *only in brain* were defined as those that had stable or low variable expression *in brain* and unstable or high variable expression *in other tissues* (Fig. 2(c)). The GTEx project (<https://www.gtexportal.org/home/datasets>) [24] was utilized to

simulate genes with expression sensitivity in brain. In the database, data from 26 different tissues, including 10 different brain tissues, were registered. (I) (Tissue)_Analysis.v6p.egenes.txt which is a list of genes expressed in each tissue and (II) (Tissue)_Analysis.v6p.signif_snpgene_pairs.txt which is a list of eQTL SNPs within each gene were utilized. Genes with stable expression in a certain tissue were taken to be those that are expressed in that tissue and are thus listed in (I), and those that did not have any eQTL SNPs in the tissue and are thus not listed in (II). Genes with expression sensitivity *in brain* were defined as genes with stable expression in the 10 different tissues from brain. In contrast, genes with unstable expression in a certain tissue were defined as genes that are expressed in that tissue and they are included in (I), and those that have at least one eQTL SNP in the tissue and they are included in (II). Genes with unstable expression *in other tissues* were defined as genes with unstable expression in the 16 different tissues other than brain. Finally, overlapping genes with expression sensitivity *in brain* and unstable expression *in other tissues* were defined as genes with expression sensitivity *only in brain*.

Statistical analysis

The frequency of CNVs was calculated among the patients of each disease and healthy individuals. CNVs with a size > 100 kb and a frequency < 1% were used in enrichment tests. In the tests, the average number of genes overlapped by CNVs were compared between the cases and the controls for the following categories: (1) dosage-sensitive genes (2); genes with expression sensitivity only in brain; and (3) a combined category of [1, 2]. In more detail, enrichment tests were conducted using the --cnv-count, --cnv-subset and --cnv-enrichment test options in PLINK to assess whether a subset of genes was enriched relative to all genes.

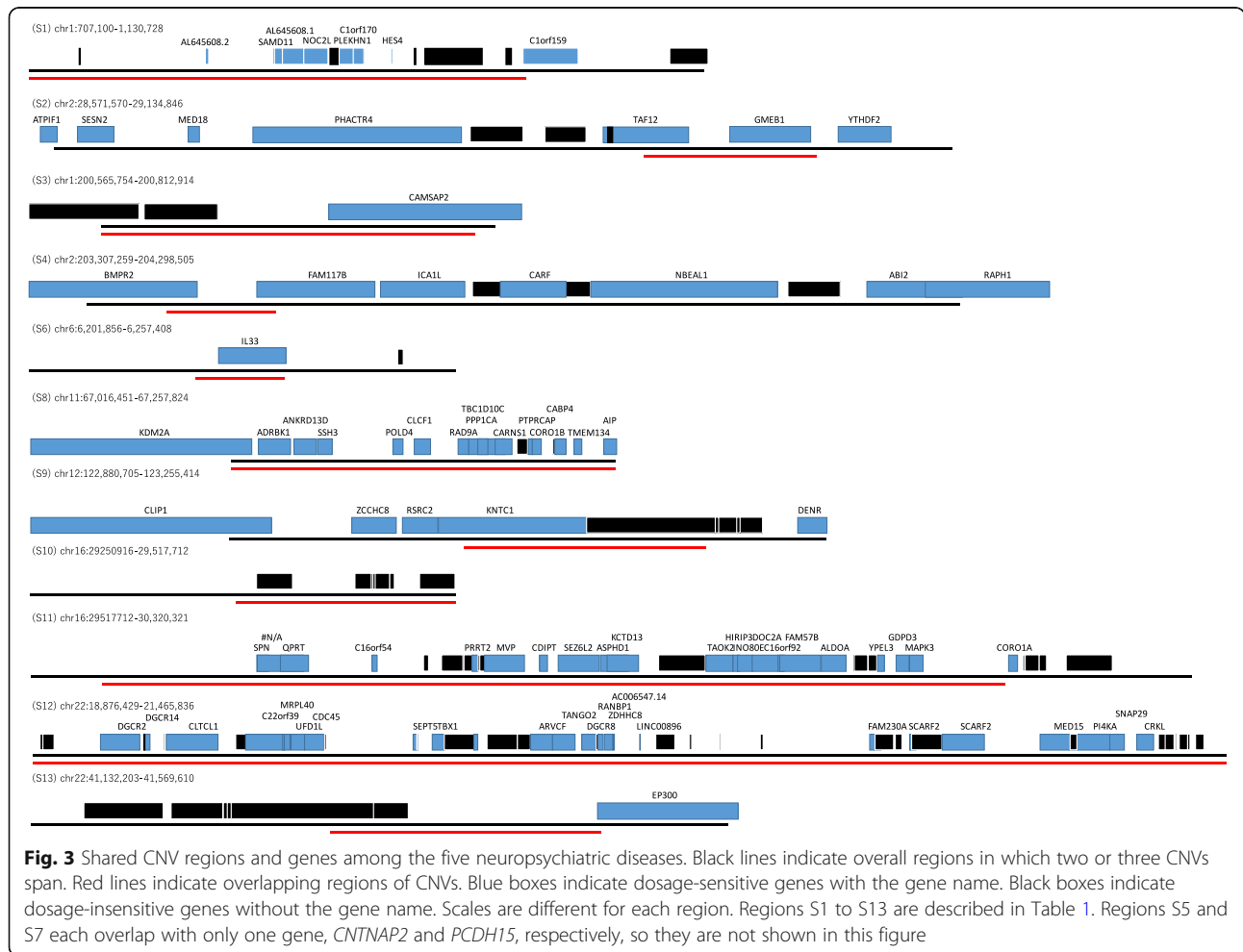
Inspection of regions detected only in the patients

Regions that were found only in the patients of the five neuropsychiatric diseases were examined. CNVs with a size of > 100 kb and a frequency of < 1% were examined. To narrow down the possible candidate regions, the following criteria were used: (i) previously reported regions; (ii) shared regions among the five neuropsychiatric diseases; or (iii) regions with more than six dosage-sensitive genes. Previously reported regions were taken from the following papers and database: a paper by Itsara, A et al. [36] and a list of candidate genes from the Autism Database (AutDB; <http://autism.mindspec.org/autdb/>) [37] for autism; a paper by Howe, A et al. [38] for panic disorders; a paper by Lane, J et al. [39] for sleep disorders; a paper by Ripke, S et al. [40] for schizophrenia; and a paper by Van Cauwenberghe, C et al. [41] for Alzheimer disease. Regions with (i) and (ii), or with (ii) and (iii)

Table 1 Shared regions among the five neuropsychiatric diseases

# Region	Panic disorder	Autism	Narcolepsy	EHS	Alzheimer	Sum of DS	Sum of genes with expression sensitivity	Total genes	Previously reported association	# of CNVs reported in ClinGen	Type
S1						8	6	14		> 50	Del & Dup
S2						7	6	10		3	Del & Dup
S3						1	0	3	Autism	> 5 <10	Del & Dup
S4						6	6	9	Autism	> 20 <30	Del & Dup
S5						1	1	1	Autism	> 30	Del & Dup
S6						1	0	2	Alzheimer	> 50	Del & Dup
S7						1	1	1	Autism, Sleep disorder	5	Del & Dup
S8						15	8	17		4	Del & Dup
S9						2	5	8	Autism	5	Del & Dup
S10						0	3	7	Schizophrenia	> 20	Del & Dup
S11						21	17	34	Autism, Schizophrenia	> 20	Del & Dup
S12						23	21	50	Autism	> 200	Del & Dup
S13						1	3	6	Autism, Schizophrenia	> 10	Del & Dup

Light and dark gray indicate deletion and duplication. Number of dosage sensitive genes, total genes within each region and CNVs reported in ClinGen are shown in this table. Regions S1~13 are used in Fig. 6. Regions S2 and S8 might be novel findings in this study because these regions were not reported previously and but had high density of dosage sensitive genes and few CNV reports in ClinGen



were listed as shared CNVs among the neuropsychiatric diseases in Table 1 and Fig. 3. The disease-specific regions, defined as those with (iii) and without (i) or with a $p < 0.05$ on Fisher's exact test, are listed in S7 Table.

Results

Enrichment of dosage-sensitive genes

The average number of dosage-sensitive genes was compared between the patients of each disease and the healthy individuals. According to the data from previous reports, 11,927 dosage-sensitive genes and 8360 dosage-insensitive genes were included in the analysis. Overall, in comparison to healthy individuals, a significant enrichment of dosage-sensitive genes was found among individuals with narcolepsy, panic disorders, or autism (Fig. 4 and S1–5 Tables). The similar enrichments in panic disorders and autism were also observed using ohnologs estimated by Singh PP et al. (S3 Fig). In addition, the weaker enrichment was also found using a different CNV detection algorithm (S4 and S5 Figs). In detail, reproducibility of significant enrichments of dosage sensitive genes is partly limited between software, however we could see tendency of enrichment across diseases. Overall, dosage-sensitive genes were significantly enriched in the patients with narcolepsy, panic disorders, or autism. Of note, significant enrichment of dosage-sensitive genes with duplications were observed in all diseases except for Alzheimer disease (Fig. 4). Among the five diseases, patients with panic disorders or autism showed higher enrichment of dosage-sensitive

genes with duplications when compared to individuals with narcolepsy or essential hypersomnia (EHS). For deletions, less or no enrichment of dosage-sensitive genes with deletions was seen in the patients when compared to the healthy individuals (Fig. 4).

Enrichment of genes with expression sensitivity only in brain

To demonstrate the effect of CNVs on gene expression, the average number of genes with expression sensitivity only in brain among the CNVs of the five neuropsychiatric diseases were compared between the cases and the controls. In our analysis, we proposed the idea of defining expression sensitivity by utilizing the GTEx database. We defined 11,926 genes as genes with expression sensitivity only in brain from the 39,769 expressed genes in any tissue in the GTEx database (Fig. 2 (c)). Significant enrichment of genes with expression sensitivity only in brain was observed among patients with panic disorders and autism; in contrast, patients with narcolepsy did not show significant enrichment of genes with expression sensitivity only in brain (Fig. 5). While duplications presented a higher burden, deletions did not cause significant differences when compared to the healthy individuals. An enrichment of genes with expression sensitivity only in brain was also seen using a different CNV detection algorithm (S6 Fig). Significant enrichments of genes with expression sensitivity is not partly concordant between software, however we could see slight tendency of enrichment across diseases.

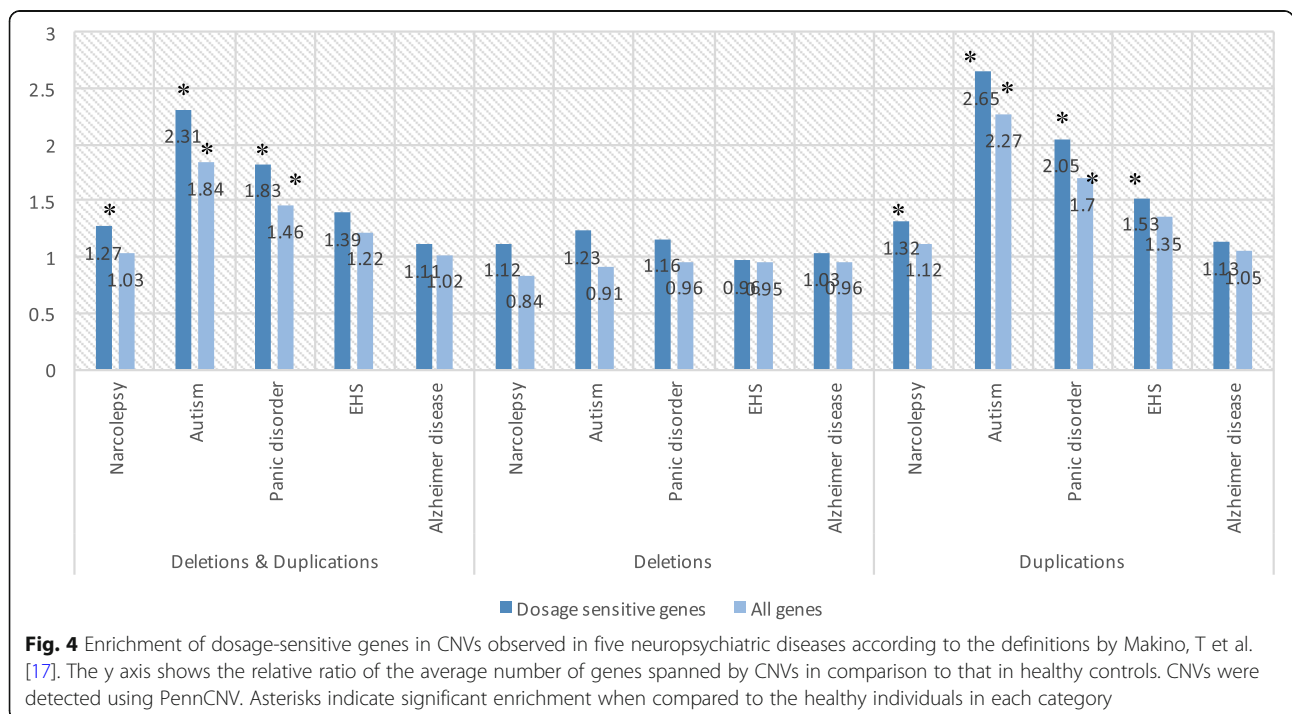
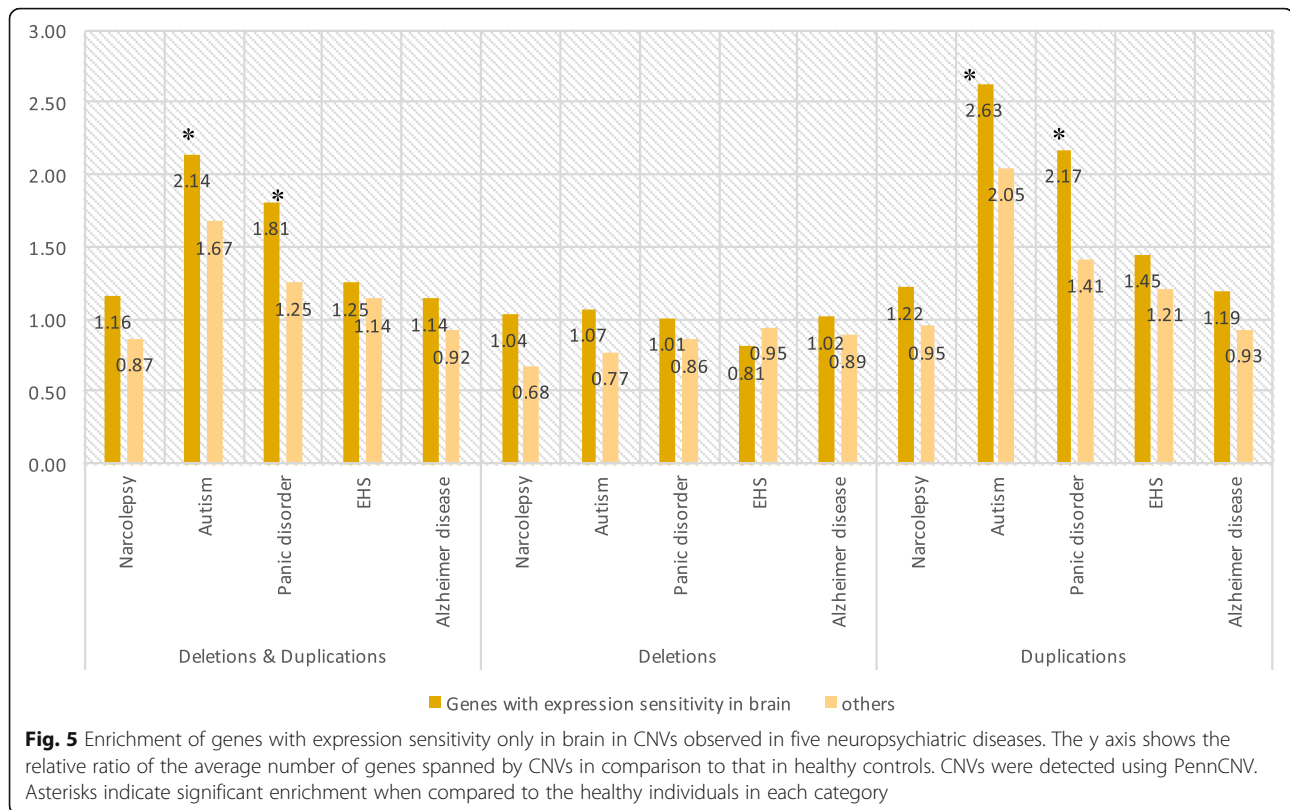


Fig. 4 Enrichment of dosage-sensitive genes in CNVs observed in five neuropsychiatric diseases according to the definitions by Makino, T et al. [17]. The y axis shows the relative ratio of the average number of genes spanned by CNVs in comparison to that in healthy controls. CNVs were detected using PennCNV. Asterisks indicate significant enrichment when compared to the healthy individuals in each category



Combined effect of dosage-sensitive genes and genes with expression sensitivity only in brain

To assess the effect of sensitivity to genome dosage and gene expression at the same time, enrichment tests were also performed. Genes were categorized into four groups based on the combinations of dosage-sensitive genes and genes with expression sensitivity only in brain: (a) dosage-sensitive genes and genes with expression sensitivity only in brain, including 5590 genes; (b) dosage-sensitive genes and genes without expression sensitivity only in brain, including 5527 genes; (c) dosage-insensitive genes and genes with expression sensitivity only in brain, including 3179 genes; and (d) dosage-insensitive genes and genes without expression sensitivity only in brain, including 4480 genes. After converting the genome coordinates to Ensembl 73, 9169 and 10,007 genes were mapped as genes with and without expression sensitivity only in brain, respectively.

Among the four categories of genes, the group including dosage-sensitive genes and genes with expression sensitivity only in brain showed the highest enrichment among the CNVs from the patients when compared to healthy individuals (Fig. 6). The group including dosage-sensitive genes and genes without expression sensitivity only in brain showed the second highest enrichment, followed by the group including dosage-insensitive genes and genes with expression sensitivity only in brain, then

the group including dosage-insensitive genes and genes without expression sensitivity only in brain. A similar tendency was observed in CNVs detected by another software (S6 Table). When the ratio of the average number of genes was compared between cases and controls among all tests for all diseases, the highest ratio of enrichment was observed in the group including dosage-sensitive genes and genes with expression sensitivity only in brain.

Inspection of regions detected only in the patients

Shared CNV regions among the five neuropsychiatric diseases were also investigated. Table 1 and Fig. 3 show the previously implicated regions or regions with more than six dosage-sensitive genes and regions associated with at least two different diseases. With the use of gene dosage sensitivity, regions S2 (chr1: 28,571,570-29,134,846) and S8 (chr11: 67,016,451-67,257,824) might represent novel findings because these regions have not been reported previously and have few reported CNVs in the ClinGen database (Table 1). In particular, region S8 was significantly enriched in dosage-sensitive genes when compared to the genome-wide ratio of dosage-sensitive genes and dosage-insensitive genes ($p = 0.013$ on Fisher's exact test), and while region S2 was not significantly enriched ($p = 0.540$ on Fisher's exact test), it had a higher density than the genome-wide average (density of

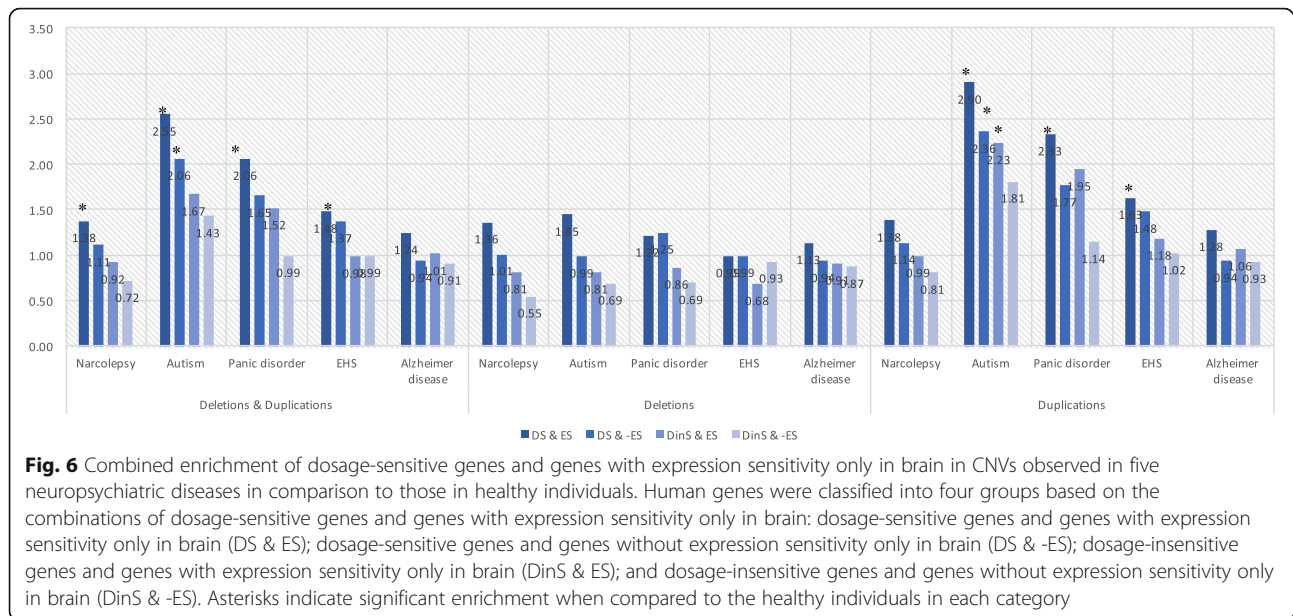


Fig. 6 Combined enrichment of dosage-sensitive genes and genes with expression sensitivity only in brain in CNVs observed in five neuropsychiatric diseases in comparison to those in healthy individuals. Human genes were classified into four groups based on the combinations of dosage-sensitive genes and genes with expression sensitivity only in brain: dosage-sensitive genes and genes with expression sensitivity only in brain (DS & ES); dosage-sensitive genes and genes without expression sensitivity only in brain (DS & -ES); dosage-insensitive genes and genes with expression sensitivity only in brain (DinS & ES); and dosage-insensitive genes and genes without expression sensitivity only in brain (DinS & -ES). Asterisks indicate significant enrichment when compared to the healthy individuals in each category

dosage-sensitive genes at region S2: 70.0%; genome-wide average of dosage-sensitive genes: 58.7%). Disease-specific regions were also examined, including regions with more than six dosage-sensitive genes; regions that have not been reported previously; or regions with a $p < 0.05$ on Fisher's exact test. All diseases except for narcolepsy had disease-specific regions that fulfilled the above criteria.

Discussion

In this paper, the phenotypic impact of sensitivity to gene dosage was demonstrated using CNVs detected in patients of five neuropsychiatric disorders. These results were concordant with a previous study using intolerance score [8]. The previous study utilized a ranking system that scored genes according to their tolerance to functional variation and demonstrated that CNVs among patients with schizophrenia had higher genic intolerance scores than that of healthy controls. In other words, CNVs in schizophrenia occurred within genes intolerance to functional variations. This is similar to our results, because it was known that genes with dosage sensitivity hardly experienced copy number alteration by CNV [19].

Duplications appeared to have a substantial impact on disease onset, as we saw from the significant enrichment of dosage-sensitive genes with duplications in all diseases except for Alzheimer disease. Until recently, duplications were often deprioritized and analyzed after deletions, perhaps because duplications are thought to be less harmful than deletions [3]. However, a recent study suggested that duplications exhibit more diversity, weaker selective constraint, and a four-fold greater chance of affecting genes than deletions, indicating that they possess

signatures of adaptive evolution [5]. Indeed, several studies have observed that high-copy CNVs were associated with human traits [4, 42–45]. It might be because once deletions occurred, they disappeared quickly or their frequencies were reduced in a population because they were lethal or extremely harmful and disadvantageous. However, when duplications occurred, they were retained within the diversity of a population because they were not lethal nor sufficiently harmful or disadvantageous to have had their frequencies reduced. From the viewpoint of the total number of dosage-sensitive genes affected by CNVs, the accumulative impact of duplications might be larger than that of deletions. This study provides another example that demonstrates the importance of duplications in the onset of diseases.

Genes with sensitive expression only in brain were identified utilizing publically available GTEx data, and the effects of CNVs on the expression of these genes were evaluated. Similar to the gene balance hypothesis, expression balance might also be maintained; thus, alterations in the gene expression level of genes with expression sensitivity are deleterious, while such alterations in genes without expression sensitivity are not. Indeed, it is already reported that dosage change of expressed genes in brain are less frequent than those of other genes and are controlled by tighter transcriptional regulation [46]. Also, according to an analysis using intolerance score, genes highly expressed in the brain showed the most intolerance to CNV [8]. CNVs could be one cause of change at the expression level because CNVs modify gene dosage, so it is possible that CNVs in genes with expression sensitivity might contribute to the onset of diseases.

Our finding of no significant enrichment of genes with expression sensitivity only in brain among individuals with narcolepsy suggests that the genetic background of narcolepsy may not be related to brain function, at least not in relation to CNVs. Previous studies of narcolepsy showed an autoimmune etiology for the disease and orexin (hypocretin) deficiency among patients with narcolepsy. *HLA-DQB1*06:02* was a major genetic factor for the onset of narcolepsy [47–51], and the T cell receptor alpha gene (*TRA*) and purinergic receptor P2Y, G-protein coupled, 11 gene (*P2RY11*) were also reported to be associated with narcolepsy [52, 53]. Pathway analysis of CNVs in patients with narcolepsy found enrichment of immune-related pathway [54]. Additionally, the hypocretin-1 level was found to be reduced or undetectable in the cerebrospinal fluid of narcoleptic patients [55], and postmortem examination showed a marked reduction of hypocretin-producing neurons in the hypothalamus [56]. Nevertheless, regarding CNVs, our result of no significant enrichment of genes with expression sensitivity only in brain demonstrated that overall, narcolepsy might be an autoimmune disease rather than a brain-related disease.

In this study, we proposed a method of assessing the influence of CNVs on gene expression. Analysis of eQTL enables the evaluation of whether SNPs or variants affect the gene expression level [24]; however, these analysis does not assess whether alterations of the gene expression level are deleterious or not. Here, genes with expression sensitivity were able to evaluate influence of change at expression-level. In addition to expression sensitivity, it is necessary to assess the impact of CNVs on gene expression [57] because CNVs contribute to 18 to 99% of the expression level of genes [58, 59]. Indeed, previous studies analyzed expression-level of CNV locus using human brain tissue from healthy individuals and patients with neuropsychiatric diseases [60, 61]. Yet, it has been difficult to know the effect of each CNVs on gene expression easily because there is no gene expression reference panel for CNVs like there is for SNPs. In this study, we proposed a cross-sectional omics approach using publicly available data. To the best of our knowledge, this is a novel method for assessing the influence of CNVs on gene expression.

The shared regions among five neuropsychiatric diseases were assessed. With the use of gene dosage sensitivity, regions S2 and S8 may be novel findings (Table 1 and Fig. 3). Region S8 was shared between autism and Alzheimer disease, and within region S8, the carnosine synthase 1 gene (*CARNS1*) was highly expressed in brain. This gene was known to catalyze the formation of carnosine and homocarnosine. Base on ARCHS4 database, top predicted biological process in Gene Ontology is fatty acid elongation [62]. Increased expression of fatty

acid synthesis in model of autism was demonstrated previously [63]. Correlation between deficient biosynthesis of fatty acid and cognitive impairment in Alzheimer's disease was reported [64]. This gene might be involved pathogenesis of these diseases. The protein tyrosine phosphatase receptor type C-associated protein gene (*PTPRCAP*) was reported to contain one of the top differentially methylated probes in autism [65]. A genetic link between autism and Alzheimer disease has previously been reported, and this provides additional evidence for a shared background between autism and Alzheimer disease [66].

Disease-specific regions of five neuropsychiatric diseases were evaluated. For panic disorders, region P3 spanned seven dosage-sensitive genes (S7 Table). One of them was the solute carrier family 17 member 7 gene (*SLC17A7*); this gene is highly expressed in brain, and it was reported that expression of this gene resulted in the uptake of glutamate as a vesicular glutamate transporter [67]. Reduced expression of this gene leads to reduced uptake of glutamate and an increased amount of glutamate in brain. Previously, a high amount of glutamate was reported to be associated with panic attacks [68]. Therefore, it seems that *SLC17A7* within this deletion may be a candidate causative gene in patients with panic disorders. Region P2 (chr14:50,043,390-50,311,552) spanned eight dosage-sensitive genes. Among these genes, the ribosomal protein S29 gene (*RPS29*) showed a marginal significant association in a genome-wide association study with posttraumatic stress disorder [69]. Another gene, the kelch domain containing 1 gene (*KLHDC1*), was reported to show a moderate significant association with bipolar disease [70]. Four individuals with panic disorders had duplications in region P1 (chr4:39,500,375-39,784,412; $p = 0.0027$ on Fisher's exact test). The ubiquitin-conjugating enzyme gene (*UBE2K*) and small integral membrane protein 14 gene (*SMIM14*) were overlapped in all duplications among these four individuals. *UBE2K* was reported to be correlated with positive symptoms of psychosis in schizophrenia and bipolar patients [71]. By thoroughly inspecting dosage-sensitive genes within CNVs, it seems possible to narrow down candidate genes.

We found several regions with more than six dosage-sensitive ohnologs in Alzheimer disease. No increase in the average number of CNVs per person was observed in Alzheimer disease. Nevertheless, when we further investigated each CNV that occurred only among patients, Alzheimer disease seemed to associate with dosage-sensitive genes. This result was concordant with the results of a recent paper [23]. In particular, a duplication in region AD6 (chr11: 104,756,445-107,834,208) was observed in one case in this study (S7 Table), and it overlapped with the contactin 5 gene (*CNTN5*) and ELMO

domain-containing 1 gene (*ELMOD1*), which were reported as candidate genes in the recent paper. In addition, when we investigated CNVs with a size < 100 kb, the amyloid beta precursor protein gene (*APP*) overlapped with a deletion from one patient with Alzheimer disease in this study. Previously, duplications of *APP* were reported to be causative variants for early onset familial Alzheimer disease [72–74]. Although a deletion was observed in this study, *APP* was a dosage-sensitive gene, so it is possible that not only a gain, but also a loss of gene dosage might have contributed to the onset of disease in this patient. Therefore, both the total burden of CNVs and inspection of dosage-sensitive genes around CNVs are useful strategies for identifying susceptibility genes.

The scope of our study was limited to ohnologs that underwent 2R-WGD and were very ancient. However, recent human-specific segmental duplications are also known to contribute to disease onset or phenotypic diversity [75]. It is important to inspect CNVs from both viewpoints. In addition, we only evaluated the effect of CNVs on neighboring genes. It was reported that among genes with expression influenced by CNVs, 53% of the expression is affected by CNVs that are distant from the target genes [1]. However, our analysis did not consider the distant effects of CNVs. Also the method we proposed in this study simplify expression among brain and did not consider different gene expression of region in brain and dynamics of gene expression through life span so on.

Conclusions

In this study, we demonstrated the impact of sensitivity to gene dosage and gene expression using CNVs identified in patients with neuropsychiatric diseases. A novel approach was proposed, and the effect of CNVs on gene expression was globally assessed. These results will help elucidate the pathogenicity of diseases more clearly than before.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12920-020-0699-9>.

Additional file 1: Figure S1. Overview of the study. CNVs identified from each detection method were filtered and analyzed independently. **Figure S2.** CNV frequency and proportion of dosage-sensitive genes overlapped by CNVs. The x axis shows the frequency of CNVs in each patient and healthy individual. The y axis shows the proportion of ohnologs with SSD, singletons, ohnologs without SSD, or non-ohnologous duplications. **Figure S3.** Enrichment of ohnologs in CNVs observed in five neuropsychiatric diseases according to the definitions by Singh, PP et al. **Figure S4.** Enrichment of dosage-sensitive genes in CNVs observed in five neuropsychiatric diseases with the use of another software, CNV Workshop. **Figure S5.** Enrichment of ohnologs in CNVs observed in five neuropsychiatric diseases with the use of another software, CNV Workshop. **Figure S6.** Enrichment of genes with expression sensitivity only in

brain in CNVs observed in five neuropsychiatric diseases with the use of another software, CNV Workshop. **Table S1.** Enrichment of dosage-sensitive ohnologs in individuals with narcolepsy using CNVs detected by PennCNV. **Table S2.** Enrichment of dosage-sensitive ohnologs in individuals with autism using CNVs detected by PennCNV. **Table S3.** Enrichment of dosage-sensitive ohnologs in individuals with panic disorders using CNVs detected by PennCNV. **Table S4.** Enrichment of dosage-sensitive ohnologs in individuals with essential hypersomnia using CNVs detected by PennCNV. **Table S5.** Enrichment of dosage-sensitive ohnologs in individuals with Alzheimer disease using CNVs detected by PennCNV. **Table S6.** Combined enrichment of dosage-sensitive genes and genes with expression sensitivity only in brain in CNVs observed in five neuropsychiatric diseases with the use of another software, CNV Workshop. **Table S7.** Disease-specific regions for five neuropsychiatric diseases.

Abbreviations

2R-WGD: Two-round whole-genome duplication; APP: Amyloid beta precursor protein gene; AutDB: Autism Database; CARN1: Carnosine synthase 1 gene; CEU: Utah residents with Northern and Western European ancestry; CHB: Han Chinese in Beijing, China; CNTN5: Contactin 5 gene; CNVs: Copy number variants; EHS: Essential hypersomnia; ELMOD1: ELMO domain-containing 1 gene; EQL: Expression quantitative trait locus; GTEx: Genotype-Tissue Expression; HLA-DQB1: Major histocompatibility complex, class II, DQ beta 1; JPT: Japanese in Tokyo, Japan; KLHDC1: Kelch domain containing 1 gene; OMIN: Online Mendelian Inheritance in Man; P2RY11: Purinergic receptor P2Y, G-protein coupled, 11 gene; PIHAT: Probability of being identity-by-descent; PTPRCAP: Protein tyrosine phosphatase receptor type C-associated protein gene; RPS29: Ribosomal protein S29 gene; SLC17A7: Solute carrier family 17 member 7 gene; SMIM14: Small integral membrane protein 14 gene; SNPs: Single nucleotide polymorphisms; SSD: Small-scale duplication; TRA: T cell receptor alpha gene; UBE2K: Ubiquitin-conjugating enzyme gene; YRI: Yoruba in Ibadan, Nigeria

Acknowledgements

This work was supported by the Japan Agency for Medical Research and Development (AMED) under grant number JP17km0405205h0002 and 18km0405205h0003 to Katsushi Tokunaga.

Authors' contributions

MY analyzed and interpreted the patient data regarding CNV and ohnologs and wrote the manuscript. TM1 provided list of ohnologs and has been involved in revising the manuscript critically for important intellectual content. SSK provided supportive advice regarding data analysis, interpretation and writing the manuscript. HT1, TM2 and MH were contributors in collection of individuals with sleep disorders. XL, HK1, YK1, TS1, TS2, HN and KK were contributors in acquisition of individuals with autism. NS, MT, TO, YO, HK2, HT2 and TS3 provided support to collect patients with panic disorder and healthy individuals. YK2, AM and RK contributed in acquisition of individuals with Alzheimer disease and healthy individuals. KT contributed this study in writing the manuscript and funding. All authors read and approved the final manuscript.

Funding

This study was supported by Japan Agency for Medical Research and Development, AMED-17km0405205h0002.

The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used in the current study are available from National Bioscience Database Center (<https://biosciencedbc.jp>) and from the authors with the reasonable request.

Ethics approval and consent to participate

This study was approved by the Human Genome, Gene Analysis Research Ethics Committee of the University of Tokyo, the National Center of Neurology and Psychiatry Ethics Committee and the Research Ethics Committee of TMIMS (Tokyo Metropolitan Institute of Medical Science).

All individuals provided written informed consent for their inclusion in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Health Data Science Research, Healthy Aging Innovation Center, Tokyo Metropolitan Geriatric Medical Center, Tokyo, Japan. ²Laboratory of Evolutionary Genomics, Graduate School of Life Sciences, Tohoku University, Sendai, Japan. ³Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Tokyo, Japan. ⁴Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ⁵Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan. ⁶RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. ⁷Department of Psychiatry, Hamamatsu University School of Medicine, Hamamatsu, Japan. ⁸Department of Child and Adolescent Psychiatry, Hamamatsu University School of Medicine, Shizuoka, Japan. ⁹Department of Child Psychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ¹⁰Division for Counseling and Support, The University of Tokyo, Tokyo, Japan. ¹¹Asunaro Hospital for Child and Adolescent Psychiatry, Mie, Japan. ¹²Unit of Public Health and Preventive Medicine, School of Medicine, Yokohama City University, Kanagawa, Japan. ¹³Department of Neuropsychiatry, Teikyo University Hospital, Tokyo, Japan. ¹⁴Department of Neuropsychiatry, NTT Medical Center Tokyo, Tokyo, Japan. ¹⁵Department of Psychiatry, Koseikai Michinoo Hospital, Nagasaki, Japan. ¹⁶Panic Disorder Research Center, Warakukai Med Corp, Tokyo, Japan. ¹⁷Department of Psychiatry, Shonan Kamakura General Hospital, Kanagawa, Japan. ¹⁸Department of Molecular Genetics, Bioresource Science Branch, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan. ¹⁹Asahigawaso Research Institute, Asahigawaso Medical-Welfare Center, Okayama, Japan. ²⁰Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ²¹Center for Physical and Mental Health, Mie University, Tsu, Mie, Japan. ²²Division of Physical and Health Education, Graduate School of Education, The University of Tokyo, Tokyo, Japan.

Received: 25 October 2018 Accepted: 24 February 2020

Published online: 29 March 2020

References

- 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(7118):444–54.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569–73.
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43(9):838–46.
- Falchi M, El-Sayed Moustafa JS, Takouis P, Pesce F, Bonfond A, Andersson-Assarsson JC, et al. Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet*. 2014;46(5):492–7.
- Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. *Science*. 2015;349(6253):aab3761.
- Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, et al. Large multiallelic copy number variations in humans. *Nat Genet*. 2015;47(3):296–303.
- Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16(3):172–83.
- Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet*. 2016;48(10):1107–11.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–24.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
- Yamaguchi-Kabata Y, Nariai N, Kawai Y, Sato Y, Kojima K, Tateno M, et al. iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. *Hum Genome Var*. 2015;2:15050.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–91.
- Terry SF. The global alliance for genomics & health. *Genet Test Mol Biomarkers*. 2014;18(6):375–6.
- Blakeslee AF, Belling J, Farnham ME. Chromosomal duplication and Mendelian phenomena in *Datura* mutants. *Science*. 1920;52(1347):388–90.
- Bridges CB. Sex in relation to chromosomes and genes. *Am Nat*. 1925;59(661):127–37.
- Papp B, Pal C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature*. 2003;424(6945):194–7.
- Makino T, McLysaght A, Kawata M. Genome-wide deserts for copy number variation in vertebrates. *Nat Commun*. 2013;4:2283.
- Ohno S. Evolution by gene duplication. Berlin: Springer; 1970.
- Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A*. 2010;107(20):9270–4.
- Chen WH, Zhao XM, van Noort V, Bork P. Human monogenic disease genes have frequently functionally redundant paralogs. *PLoS Comput Biol*. 2013;9(5):e1003073.
- McLysaght A, Makino T, Grayton HM, Tropeano M, Mitchell KJ, Vassos E, et al. Ohnologs are overrepresented in pathogenic copy number mutations. *Proc Natl Acad Sci U S A* 2014;111(1):361–366.
- Singh PP, Affeldt S, Malaguti G, Isambert H. Human dominant disease genes are enriched in paralogs originating from whole genome duplication. *PLoS Comput Biol*. 2014;10(7):e1003754.
- Sekine M, Makino T. Inference of causative genes for Alzheimer's disease due to dosage imbalance. *Mol Biol Evol*. 2017;34(9):2396–407.
- Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
- Toyoda H, Miyagawa T, Koike A, Kanbayashi T, Imanishi A, Sagawa Y, et al. A polymorphism in CCR1/CCR3 is associated with narcolepsy. *Brain Behavior and Immunity*. 2015;49:148–55.
- Khor SS, Miyagawa T, Toyoda H, Yamasaki M, Kawamura Y, Tanii H, et al. Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia. *PeerJ*. 2013;1:e66.
- Otowa T, Kawamura Y, Nishida N, Sugaya N, Koike A, Yoshida E, et al. Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. *Transl Psychiatry*. 2012;2:e186.
- Liu X, Kawamura Y, Shimada T, Otowa T, Koishi S, Sugiyama T, et al. Association of the oxytocin receptor (OXTR) gene polymorphisms with autism spectrum disorder (ASD) in the Japanese population. *J Hum Genet*. 2010;55(3):137–41.
- Miyashita A, Wen YN, Kitamura N, Matsubara E, Kawarabayashi T, Shoji M, et al. Lack of genetic association between TREM2 and late-onset Alzheimer's disease in a Japanese population. *J Alzheimers Dis*. 2014;41(4):1031–8.
- International HapMap C. The international HapMap project. *Nature*. 2003;426(6968):789–96.
- 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(11):1665–74.
- Gai X, Perin JC, Murphy K, O'Hara R, D'Arcy M, Wenocur A, et al. CNV workshop: an integrated platform for high-throughput copy number variation discovery and clinical diagnostics. *BMC Bioinformatics*. 2010;11:74.
- Zhang X, Du R, Li S, Zhang F, Jin L, Wang H. Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinformatics*. 2014;15:50.
- Mace A, Tuke MA, Beckmann JS, Lin L, Jacquemont S, Weedon MN, et al. New quality measure for SNP array based CNV detection. *Bioinformatics*. 2016;32(21):3298–305.
- Singh PP, Arora J, Isambert H. Identification of Ohnolog genes originating from whole genome duplication in early vertebrates, based on Synteny comparison across multiple genomes. *PLoS Comput Biol*. 2015;11(7):e1004394.

36. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet.* 2009;84(2):148–61.
37. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* 2009;37(Database issue):D832–6.
38. Howe AS, Buttenschon HN, Bani-Fatemi A, Maron E, Otowa T, Erhardt A, et al. Candidate genes in panic disorder: meta-analyses of 23 common variants in major anxiogenic pathways. *Mol Psychiatry.* 2016;21(5):665–79.
39. Lane JM, Liang JJ, Vlasac I, Anderson SG, Bechtold DA, Bowden J, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat Genet.* 2017;49(2):274–81.
40. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. 2014; 511(7510):Nature, 421–427.
41. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in Medicine.* 2016;18(5):421–30.
42. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science.* 2005;307(5714):1434–40.
43. McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis.* 2008;67(3):409–13.
44. Fellermann K, Stange DE, Schaeffeler E, Schmalz H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet.* 2006;79(3):439–48.
45. Hollox EJ, Huffmeier U, Zeeuwen PLJM, Palla R, Lascorz J, Rodijk-Olthuis D, et al. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet.* 2008;40(1):23–5.
46. Henrichsen CN, Vinckenbosch N, Zollner S, Chagnat E, Pradervand S, Schutz F, et al. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet.* 2009;41(4):424–9.
47. Juji T, Satake M, Honda Y, Doi Y. HLA antigens in Japanese patients with narcolepsy. All the patients were DR2 positive. *Tissue Antigens.* 1984;24(5): 316–9.
48. Langdon N, Welsh KI, van Dam M, Vaughan RW, Parkes D. Genetic markers in narcolepsy. *Lancet.* 1984;2(8413):1178–80.
49. Matsuki K, Juji T, Tokunaga K, Naohara T, Satake M, Honda Y. Human histocompatibility leukocyte antigen (HLA) haplotype frequencies estimated from the data on HLA class I, II, and III antigens in 111 Japanese narcoleptics. *J Clin Invest.* 1985;76(6):2078–83.
50. Miyagawa T, Hohjoh H, Honda Y, Juji T, Tokunaga K. Identification of a telomeric boundary of the HLA region with potential for predisposition to human narcolepsy. *Immunogenetics.* 2000;52(1–2):12–8.
51. Yamasaki M, Miyagawa T, Toyoda H, Khor SS, Liu X, Kuwabara H, et al. Evaluation of polygenic risks for narcolepsy and essential hypersomnia. *J Hum Genet.* 2016;61(10):873–8.
52. Hallmayer J, Faraco J, Lin L, Hesselson S, Winkelmann J, Kawashima M, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat Genet.* 2009;41(6):708–11.
53. Kornum BR, Kawashima M, Faraco J, Lin L, Rico TJ, Hesselson S, et al. Common variants in P2RY11 are associated with narcolepsy (vol 43, pg 66, 2011). *Nat Genet.* 2011;43(10):1040.
54. Yamasaki M, Miyagawa T, Toyoda H, Khor SS, Koike A, Nitta A, et al. Genome-wide analysis of CNV (copy number variation) and their associations with narcolepsy in a Japanese population. *J Hum Genet.* 2014; 59(5):235–40.
55. Mignot E, Lammers GJ, Ripley B, Okun M, Nevsimalova S, Overeem S, et al. The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. *Arch Neurol-Chicago.* 2002;59(10): 1553–62.
56. Nishino S, Ripley B, Overeem S, Lammers GJ, Mignot E. Hypocretin (orexin) deficiency in human narcolepsy. *Lancet.* 2000;355(9197):39–40.
57. Gamazon ER, Stranger BE. The impact of human copy number variation on gene expression. *Brief Funct Genomics.* 2015;14(5):352–7.
58. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* 2007;315(5813):848–53.
59. Fehrmann RS, Karjalainen JM, Krajewska M, Westra HJ, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet.* 2015;47(2):115–25.
60. Mehta D, Iwamoto K, Ueda J, Bundo M, Adati N, Kojima T, et al. Comprehensive survey of CNVs influencing gene expression in the human brain and its implications for pathophysiology. *Neurosci Res.* 2014;79:22–33.
61. Ye T, Lipska BK, Tao R, Hyde TM, Wang L, Li C, et al. Analysis of copy number variations in brain DNA from patients with schizophrenia and other psychiatric disorders. *Biol Psychiatry.* 2012;72(8):651–4.
62. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun.* 2018;9(1):1366.
63. Chen J, Wu W, Fu Y, Yu S, Cui D, Zhao M, et al. Increased expression of fatty acid synthase and acetyl-CoA carboxylase in the prefrontal cortex and cerebellum in the valproic acid model of autism. *Exp Ther Med.* 2016;12(3): 1293–8.
64. Astarita G, Jung KM, Berchtold NC, Nguyen VQ, Gillen DL, Head E, et al. Deficient liver biosynthesis of docosahexaenoic acid correlates with cognitive impairment in Alzheimer's disease. *PLoS One.* 2010;5(9):e12538.
65. Loke YJ, Hannan AJ, Craig JM. The role of epigenetic change in autism spectrum disorders. *Front Neurol.* 2015;6.
66. Malishkevich A, Amram N, Hacoheh-Kleiman G, Magen I, Giladi E, Gozes I. Activity-dependent neuroprotective protein (ADNP) exhibits striking sexual dichotomy impacting on autistic and Alzheimer's pathologies. *Transl Psychiatry.* 2015;5:e501.
67. Takamori S, Rhee JS, Rosenmund C, Jahn R. Identification of a vesicular glutamate transporter that defines a glutamatergic phenotype in neurons. *Nature.* 2000;407(6801):189–94.
68. Zwanzger P, Zavorotnyy M, Gencheva E, Diemer J, Kugel H, Heindel W, et al. Acute shift in glutamate concentrations following experimentally induced panic with cholecystokinin Tetrapeptide-a 3T-MRS study in healthy subjects. *Neuropsychopharmacol.* 2013;38(9):1648–54.
69. Xie PX, Kranzler HR, Yang C, Zhao HY, Farrer LA, Gelernter J. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biol Psychiatry.* 2013;74(9):656–63.
70. Feng T, Zhu XF. Genome-wide searching of rare genetic variants in WTCCC data. *Hum Genet.* 2010;128(3):269–80.
71. Bousman CA, Chana G, Glatt SJ, Chandler SD, May T, Lohr J, et al. Positive Symptoms of Psychosis Correlate With Expression of Ubiquitin Proteasome Genes in Peripheral Blood. *Am J Med Genet B.* 2010;153b(7):1336–41.
72. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet.* 2006;38(1): 24–6.
73. Sleegers K, Brouwers N, Gijssels I, Theuns J, Goossens D, Wauters J, et al. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain.* 2006;129:2977–83.
74. McNaughton D, Knight W, Guerreiro R, Ryan N, Lowe J, Poulter M, et al. Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series. *Neurobiol Aging.* 2012;33(2):426 e13–21.
75. Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, Sudmant PH, et al. Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature.* 2016;536(7615):205–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.