

RESEARCH

Open Access



Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders

Joseph T. Glessner^{1†}, Jin Li^{1,2†}, Dai Wang³, Michael March¹, Leandro Lima¹, Akshatha Desai¹, Dexter Hadley¹, Charly Kao¹, Raquel E. Gur⁴, Nadine Cohen³, Patrick M. A. Sleiman^{1,5,6}, Qingqin Li⁷, Hakon Hakonarson^{1,5,6*} and the Janssen-CHOP Neuropsychiatric Genomics Working Group

Abstract

Background: Neurodevelopmental and neuropsychiatric disorders represent a wide spectrum of heterogeneous yet inter-related disease conditions. The overlapping clinical presentations of these diseases suggest a shared genetic etiology. We aim to identify shared structural variants spanning the spectrum of five neuropsychiatric disorders.

Methods: We investigated copy number variations (CNVs) in five cohorts, including schizophrenia (SCZ), bipolar disease (BD), autism spectrum disorders (ASD), attention deficit hyperactivity disorder (ADHD), and depression, from 7849 cases and 10,799 controls. CNVs were called based on intensity data from genome-wide SNP arrays and CNV frequency was compared between cases and controls in each disease cohort separately. Meta-analysis was performed via a gene-based approach. Quantitative PCR (qPCR) was employed to validate novel significant loci.

Results: In our meta-analysis, two genes containing CNVs with exonic overlap reached genome-wide significance threshold of meta P value $< 9.4 \times 10^{-6}$ for deletions and 7.5×10^{-6} for duplications. We observed significant overlap between risk CNV loci across cohorts. In addition, we identified novel significant associations of *DOCK8/KANK1* duplications (meta P value = 7.5×10^{-7}) across all cohorts, and further validated the CNV region with qPCR.

Conclusions: In the first large scale meta-analysis of CNVs across multiple neurodevelopmental/psychiatric diseases, we uncovered novel significant associations of structural variants in the locus of *DOCK8/KANK1* shared by five diseases, suggesting common etiology of these clinically distinct neurodevelopmental conditions.

Keywords: Copy number variation, *DOCK8*, Gene-based analysis, Meta-analysis, Neuropsychiatric disorders, Quantitative PCR

* Correspondence: hakonarson@chop.edu

†Equal contributors

¹The Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

⁵Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Full list of author information is available at the end of the article



Background

Neurodevelopmental and neuropsychiatric disorders represent a wide spectrum of heterogeneous yet inter-related disease conditions with significant overlap in phenotype expression. These diseases in children and young adults pose a major health burden and are growing in prevalence. As the clinical presentations of these diseases are not entirely distinct from each other, the clinical diagnostic boundaries are often hard to define [1].

Genome-wide association studies (GWAS) of single nucleotide polymorphisms (SNPs) in psychiatric diseases have begun to bear intriguing results [1–3]. Copy number variations (CNVs) have more direct gene dosage impacts and have been implicated in psychiatric diseases with larger effect size compared to SNPs [4–7].

Both GWAS and CNV studies have suggested different neuropsychiatric disorders share genetic determinants underlying disease development. It has been estimated that five major psychiatric disorders (schizophrenia (SCZ), bipolar disease (BD), autism spectrum disorders (ASD), attention deficit hyperactivity disorder (ADHD), and depression) share a degree of common genetic etiology. The Psychiatric Genomics Consortium (PGC) study exhibited a certain degree of genetic correlation which is relatively high between SCZ and BD, moderate between SCZ and depression, moderate between BD and depression, moderate between ADHD and depression, and non-zero between the other pairs of diseases, based on genome-wide SNP genotype data [3]. Common genetic loci have emerged from GWAS focusing on individual neuropsychiatric diseases [2, 8, 9]. Meta-analysis further suggested a shared genetic etiology and the need for molecular diagnostic technology development [1, 10]. Four loci located close to genes *ITIH3*, *AS3MT*, *CACNA1C*, and *CACNB2* reached genome-wide significance in meta-analysis of five major psychiatric disorders (SCZ, BD, ASD, ADHD, and depression) in the PGC study with the same direction of effects for these diseases [1]. For three out of the four loci, the five disorders did not show significant difference between each other in a meta-analysis homogeneity test and the model of best fit includes the contribution of all five disorders [1]. An additional seven loci were found approaching genome-wide significance (P value $< 1 \times 10^{-6}$) and 20 genes in the calcium channel activity pathway were significantly enriched in the dataset of each of the five psychiatric disorders [1]. With the rapid development of high-throughput sequencing technology, a growing catalog of de novo loss-of-function (LoF) mutations have been identified for each of the neuropsychiatric disorders. Statistical analyses have also revealed significant overlap for LoF mutations between neuropsychiatric disorders [11]. For example, LoF mutations in ten genes are

shared between ASD and SCZ (*CHD8*, *ZMYND11*, *CRYBG3*, *YTHDC1*, *HIVEP3*, *TNRC18*, *MOV10*, *ST3GAL6*, *PHF7*, *SMARCC2*) [11–14]. Compared to the other neuropsychiatric disorders, depression has been more enigmatic. To date, 23 genome-wide significant loci have been identified from five GWAS on major depressive disorder [15–19]. Among these 23 loci, ten (*PAX5*, *RERE*, *VRK2*, *MEF2C*, *L3MBTL2*, *DCC*, *SORCS3*, *NEGR1*, *VRK2*, *LIN28B*) were shared with other neuropsychiatric disorders (SCZ, BD, ASD, ADHD) reported in the GWAS catalog [20]. Less progress has been made with respect to CNVs. We have observed significant sharing of CNVs across different neurodevelopmental/psychiatric diseases, impacting genes that belong to the metabotropic glutamate receptor gene networks [6, 7, 21]; *CACNA1B* was identified as significant in a schizophrenia case-control CNV study [4], and subsequently a close homolog, *CACNA1C*, was identified as significant by the PGC GWAS [3]. Deletions upstream of *CNTN4* were identified as significant in an autism CNV study [5] as well as significant in an ADHD study [7]. Others have also found common CNV loci contributing to more than one neuropsychiatric disease. For example, the CNV at the 16p11.2 locus is associated with both SCZ and BD [22], the duplication at 17q12 is shared between ASD, intellectual disability, and SCZ [23], and the CNVs at 15q13.3, 22q11.2, and in the *NRXN1* gene are each associated with a wide spectrum of neurodevelopmental disorders [24–26]. However, no study has systematically examined which rare recurrent shared genetic loci bearing CNVs impact the development of multiple neuropsychiatric disorders.

Here, we analyze five major psychiatric disease cohorts, including 7849 cases and 10,799 controls, in a systematic manner to promote comparability of results, and more importantly to understand the degree to which the shared CNV loci may similarly or differently impact the development of neuropsychiatric disorders.

Methods

Study subjects and genotyping

Our study is composed of cases of neurodevelopmental and neuropsychiatric diseases and healthy controls from five independent cohorts (Table 1).

Janssen SCZ and BD cohort

The samples, including 3251 schizophrenia, 377 schizoaffective disorder, and 1344 bipolar cases, were collected from 28 clinical trials conducted by Janssen Research & Development, LLC, as described previously [10, 27]. These samples were genotyped on the Illumina 1MDuoV3 array. They were matched to controls, from the biorepository at the Center for Applied Genomics (CAG) of the Children's Hospital of Philadelphia

Table 1 The neurodevelopmental and neuropsychiatric disease cohorts analyzed after quality control filtering

Disease cohort	Cases	Controls	Array
Janssen SCZ and BD	2917	1113	Illumina 1MDv3
CHOP SCZ	965	1467	Affymetrix 6.0
CHOP ASD	2079	2519	Illumina 550v3
CHOP ADHD	1241	4110	Illumina 550v1
Depression	647	1590	Perlegen 660 k

SCZ schizophrenia, BD bipolar disorder, ASD autism spectrum disorders, ADHD attention-deficit hyperactivity disorder

(CHOP), which were also genotyped on the Illumina 1MDuoV3 arrays to ensure consistency in CNV discovery biases. All controls were recruited at CHOP and had no diagnosis or family history of psychiatric disease based on their medical record. [10]

CHOP SCZ cohort

The cases and controls were from CAG at CHOP and the Department of Psychiatry at the University of Pennsylvania, School of Medicine. All cases meet DSM-IV-TR criteria for schizophrenia or schizoaffective disorder [4]. A subject is excluded if he/she is unable to give informed consent to all aspects of the study, or is unable to speak and be interviewed in English, or has severe mental retardation. All samples were genotyped on the Affymetrix 6.0 array at CHOP, as described previously [4].

CHOP ASD cohort

The autism cohort included 3360 cases from Autism Genetics Resource exchange (AGRE), Autism Genome Project (AGP), and Autism Cases recruited and genotyped at CAG [5]. The control group included children of self-reported Caucasian ancestry, recruited at CHOP. All controls had no history of ASD, or any other central nervous system disorder, chromosomal disorder, syndrome or genetic disorder. All samples were genotyped on the Illumina HumanHap550 chip [5].

CHOP ADHD cohort

The 1013 ADHD cases of European descent were recruited and genotyped at CHOP; additional cases were from NIMH and The University of Utah [7]. The control group included healthy children aged 6–18 years old, with no serious underlying medical disorder, including but not limited to neurodevelopmental disorders, cancer, chromosomal abnormalities, and known metabolic or genetic disorders [7]. Samples were genotyped on the Illumina HumanHap 550 chip [7].

Depression cohort

The depression cohort included cases and controls from the Genetic Association Information Network (GAIN) major depressive disorder (MDD)/Netherlands Study of Depression and Anxiety (NESDA) project (phs000020.v2.p1) [28]. The depression cohort cases and controls were genotyped on the Perlegen 660 k array (Perlegen Sciences Mountain View, CA, USA).

Principal component analysis

PCA was conducted on the SNP genotype using the Eigenstrat [29] package to infer population structure. The first two principal components were plotted to exclude those outliers of non-European ancestry. Only subjects of European ancestry were kept for further association analysis.

CNV detection

For the Affymetrix 6.0 array, the CEL files were first converted to raw intensity data using our PennCNV [30] Affy workflow (<http://penncnv.openbioinformatics.org/en/latest/user-guide/affy/>). Then for all array types, CNVs from all samples were generated using PennCNV [30], a hidden Markov model (HMM)-based algorithm which combines multiple sources of information, including log R ratio (LRR), the B allele frequency (BAF) of each SNP, SNP spacing, and population frequency of B allele, to generate CNVs. Only CNVs containing more than three SNPs were generated. As large CNVs tend to be split into small fragments during the CNV calling procedure, adjacent CNV calls were merged via the `clean_cnv.pl` program implemented in PennCNV with the default parameters.

Quality control filtering

Sample quality control (QC) steps were performed to remove the related and/or problematic samples. First, gender discrepancies were examined using both the heterozygosity rate of the X-chromosome SNPs and the call rate of the Y-chromosome SNPs. Samples with discrepant and ambiguous gender information were excluded. Second, the relatedness of the genotyped samples was examined using pairwise Identity-by-State analysis via PLINK [31]. Duplicated samples with discrepant phenotype data were excluded from subsequent analyses. For each pair of samples that were duplicates with consistent phenotype data, or samples of related individuals (PI_HAT value exceeding 0.3), the sample with the smaller standard deviation of LRR (LRR SD) was retained.

QC was also conducted on samples based on CNV metrics: genotyping rate, LRR SD indication of intensity noise, |GC base pair wave factor (GCWF)| indication of

intensity waviness, and CNV count per sample suggestive of DNA quality.

Because of differences between array types, we reviewed data on each array type separately and plotted the distribution of each CNV metric; these metrics typically show a linear phase (the majority of the samples) but can also show an exponential phase (including in samples with outlier values). We excluded subjects with any CNV metric in the exponential phase. The QC criteria for each cohort based on the distributions of the CNV metrics are shown in (Additional file 1: Table S1).

Next, we performed QC on called CNVs, excluding CNVs < 20 kb in length and those encompassing fewer than ten probes.

All QC steps were performed for cases and controls together in each cohort. About 20% of samples in each cohort were excluded during QC filtering.

Fisher exact test

Fisher exact test implemented in software ParseCNV [32] was used to conduct the CNV association analysis on each of the individual cohorts CHOP SCZ, CHOP ASD, CHOP ADHD, and depression. It was also used to compare CNV frequency between cases and controls, as described in previous publications [5, 32]. Deletions and duplications were analyzed separately.

Linear mixed model association analysis

The “-includedeped” option in ParseCNV [32] was used to generate the ped files for additional CNV analysis using CNV “genotype” status. CNV was converted to “genotype” status in the following way: 1 1 for CN = 0, 1 2 for CN = 1, and 2 2 for others in the deletion ped file; 1 1 for CN = 4, 1 2 for CN = 3, and 2 2 for others in the duplication ped file. Then the CNV ped files were imported into GEMMA version 0.94 [33], which could correct for residual sample structure and population stratification. The LMM association testing was performed on the cohort of Janssen SCZ and BD, the samples of which came from various clinical trials. The relatedness matrix for genotype was calculated using the -gk 1 option. The matrix file was then imported for univariate linear mixed model (LMM) association, and the -lmm 4 option was used. We calculated Wald test, likelihood ratio test, and score test statistics. Then we used InsertPlinkPvalue program from the ParseCNV [32] package to insert the SNP *P* value generated by GEMMA back into ParseCNV to collapse neighboring SNPs into CNV regions.

CNV annotation

Because there is no strong consensus in defining a regulatory region and its targeted gene, we focused our analysis on genes with exonic CNVs which are most likely to directly affect the protein product

encoded by the genes. The gene(s) that each exonic CNV region resides in was used to annotate each CNV region. The CNV association *P* value from the Fisher exact test or LMM association analysis was assigned to the gene(s). For genes that contain more than one CNV, multiple-testing correction was conducted by taking the lowest *P* value of the CNVs multiplied by the number of CNVs in this gene.

Meta-analysis

Fixed effect meta-analysis was then carried out on a gene basis using the software METAL [34]. The logarithm of the odds ratio was taken to ensure consistency with Beta for the direction of association considerations.

Statistical analysis

We tested a total of 5347 genes harboring exonic CNVs in the meta-analysis of deletions and 6684 genes containing exonic CNVs in the meta-analysis of duplications. Therefore, the multiple-testing adjusted significance threshold is 9.4×10^{-6} for deletions and 7.5×10^{-6} for duplications. The significance of intersection of risk CNV loci between cohorts was computed using the R package SuperExactTest [35].

CNV quality review

We conducted manual visual review of the BAF and LRR plots of the significant CNV loci. As we are interested in shared risk genetic loci across different neuropsychiatric diseases, we focused on significant genes harboring case-enriched exonic CNVs in at least two cohorts. BAF and LRR plots of each CNV-containing sample for each significant CNV region were generated with the visualize_cnv.pl program implemented in the software Penncnv [30].

Quantitative PCR assay

Quantitative PCR (qPCR) was performed with the Universal Probe Library (UPL). UPL probes (Roche, Indianapolis, IN, USA) and corresponding primers for five assays across the targeted *DOCK8/KANK1* region were selected using the ProbeFinder v2.49 software (Roche, Indianapolis, IN, USA). qPCR was conducted on an ABI Prism™ 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA). For all samples, qPCR reactions were performed in triplicate, each in 10 µl of reaction mixture containing 10 ng genomic DNA, 100 nM of the UPL probe, 400 nM of each PCR primer, and 1× TaqMan Gene Expression Master Mix containing UDG and ROX (Life Technologies, Carlsbad, CA, USA), according to the manufacturer's protocol. Male and female genomic DNA (Promega, Madison, WI, USA) were included in the analysis as controls with expected normal copy number. Results were evaluated

using the Sequence Detection Software v2.4 (Applied Biosystems, Foster City, CA, USA) and further analyzed by the $\Delta\Delta C_T$ method. The GAPDH and SNCA genes were used as internal controls and the geometric mean of their C_T values was calculated and used as the reference value for ΔC_T calculations. The average of values from Promega male and Promega female genomic DNA (Promega, Madison, WI, USA) was considered the reference 2 N sample for $\Delta\Delta C_T$ calculations. Duplications were determined when the relative copy number value for a specific sample normalized to the reference sample was greater than 1.5.

Results

To identify shared structural variants underlying the development of neuropsychiatric disorders, we took an unbiased approach based on genome-wide SNP array intensity data from five psychiatric disease cohorts, including SCZ, BD, ASD, ADHD, and depression. After QC filtering based on sample quality and CNV metrics ("Methods"; Additional file 1: Table S1), 7849 cases and 10,799 controls were left for analysis (Table 1). Then we performed case-control association testing on CNVs from each cohort.

Next we performed gene-based meta-analysis. The SNP coverage differs between arrays and one necessary component for CNV calling, the SNP intensity data, cannot be imputed, so there may not be extensive direct overlap between CNVs across different cohorts. We annotated each CNV overlapping a gene exon(s) with the gene(s) in which the CNV region resides. Similar to other types of rare variants, CNVs may reside in different regions of a gene, which is important for brain development. Thus, we took a gene-based approach to allow for more dynamic matching between CNVs. The gene-based association testing method has been frequently used in common-variant and rare-variant analyses [36–42].

In our study, we focused on genes that contain exonic CNVs, which are case-enriched in at least two cohorts. This is because exonic CNVs are most likely to directly affect the protein product encoded by the genes and there is no strong consensus in defining a regulatory region and its targeted gene. Another consideration is the

direction of effects that each CNV gives rise to. Though control-enriched CNVs may have potential protective effects, e.g. those of the 22q11.2 CNV region against schizophrenia as reported by Rees et al. [43], this is still actively debated in the CNV study field, and the mechanism by which 22q11.2 CNVs affect the risk of disease development is under investigation. We favor the deleterious model of CNVs, consistent with the field's standard and the focus of our study is risk CNVs shared by neuropsychiatric disorders. Therefore, we kept only significant exonic CNVs that are case-enriched in at least two cohorts. We observed significant overlap of such CNV-containing genes between cohorts (Additional file 2: Figure S1). Two genes (*DOCK8* and *LOC100131257*) were found in all five cohorts (P value = 8.29×10^{-6}).

In the meta-analysis, the *ZNF280A* and *DOCK8* genes reached a genome-wide significance threshold of 9.4×10^{-6} for deletions and 7.5×10^{-6} for duplications (Table 2, Fig. 1) and passed the manual review of their BAF and LRR plots for each CNV carrier sample (Additional file 2: Figure S2 and S3). *KANK1*, which is next to *DOCK8* and contains exonic duplications, is of suggestive significance ($P = 3.45 \times 10^{-5}$). The locus of 22q11.22, encompassing the *ZNF280A* gene, is a known risk CNV locus for SCZ and BD [44]. Deletions overlapping the *ZNF280A* gene locus were observed in four cohorts. It is significantly case-enriched for the CHOP ADHD and ASD cohorts, in which 22q11.22 deletion was only observed among cases and not among controls, but this locus is control-enriched in the depression and Janssen SCZ and BD cohorts (Table 2, Fig. 1)

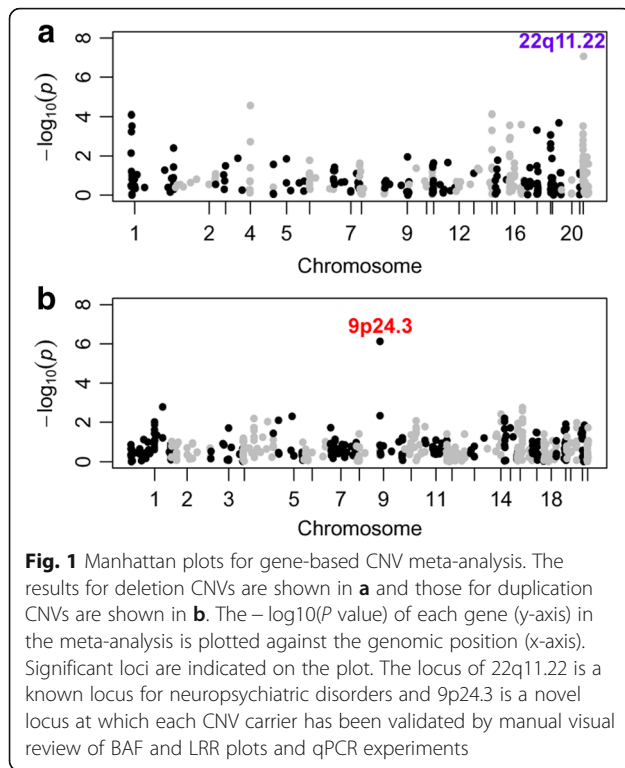
We identified a novel CNV locus demonstrating significant association across the neuropsychiatric disorders under study. The *DOCK8* gene at cytoband 9p24.3 showed significant association with neuropsychiatric disorders in meta-analysis. *DOCK8* and the neighboring gene *KANK1* exhibited significant or marginally significant case enrichment in all five cohorts (Tables 2 and 3, Fig. 2), especially in the Janssen SCZ and BD cohort as well as the CHOP ASD cohort, in which more than 5 cases but no controls carry CNVs in this region (Table 3). As the samples in the Janssen SCZ and BD cohort were from diverse resources and were not genotyped at CHOP, to ensure CNV quality at this region, we

Table 2 Significant loci in gene-based meta-analysis of the five neurodevelopmental/neuropsychiatric cohorts that contain case-enriched exonic CNVs in two or more cohorts

Marker name	Cytoband	CNV type	Number of cohorts observed	Meta P value	Direction of effect
<i>ZNF280A</i>	22q11.22*	Del	4	8.63E-08	-7++-
<i>DOCK8</i>	9p24.3	Dup	5	7.50E-07	+++++

The *Number of cohorts* is the number of cohorts in which CNVs were observed overlapping exons of the gene. The *Direction of effect* is reported in the order of Janssen SCZ and BD, CHOP SCZ, CHOP ASD, CHOP ADHD, and depression cohorts: a plus sign means CNVs are enriched in cases, a dash means CNVs are enriched in controls, a question mark means no CNV is observed overlapping the exons of this gene. *Del* deletion, *Dup* duplication.

* Known neuropsychiatric-disorder-associated regions



specifically conducted validation for CNV carriers in this cohort by two approaches. First, we manually examined the CNV calls by visually reviewing the raw intensity and genotype values for probes in this region and flanking regions (Additional file 2: Figure S3). We have previously reported that visually validated CNVs yield experimental validation success rates above 95% [32]. For the CNVs contributing calls in this region, we observed a moderate gain in LRR intensity and alternative banding of BAF at 0.33 and 0.66 instead of 0.5, which are indicative of clear duplication CNV signals. Second, we also tested the CNV status of this region using an independent experimental approach, the qPCR assay, on the above samples (Table 4). We included one sample from our cohort without a CNV in this region as a negative control. Each sample containing a duplication in the *DOCK8*–*KANK1* region was tested by five qPCR assays. The results showed that the CNV in each of the ten samples was

also detected by the corresponding qPCR assays, and the CN = 2 control sample did not show any CNV in all qPCR assays across this region. There were a few regions flanking the CNVs detected by arrays where duplications were observed with qPCR, refining the CNV boundaries. Thus, the CNV data from the arrays were validated by both an in silico approach and an independent experimental assay. An inconsistency between the array and qPCR results was seen for sample S9 with probe 141, and the array result for this sample is likely due to ambiguity in CNV boundary determination rather than presence vs. absence based on CNV calling from SNP arrays.

Each of the contributing CNVs overlapping regions from the five cohorts overlaps with *DOCK8* exons, which is likely to have an impact on *DOCK8* expression. In addition, assessing the annotations in ENCODE [45] and ROADMAP [46] databases, we found that the *DOCK8* overlapping CNV region from each cohort contains many histone marks and DNase sites (Additional file 1: Table S2). Further interrogating these regions in two eQTL databases—GTEx [47] and GRASP [48]—we observed significant brain tissue eQTL SNPs in *DOCK8* overlapping the CNV region in four out of the five cohorts (Additional file 1: Table S2). For the significant eQTLs, the genes regulated include *CBWD1*, *DMRT3*, *DOCK8*, *FOXD4*, and *KANK1*, all of which are located in the same topologically associating domains (TADs; Additional file 2: Figure S4), suggesting that the expression of these genes may also be affected.

Discussion

Cumulative evidence indicates a shared genetic etiology of neurodevelopmental and neuropsychiatric diseases. We performed CNV meta-analysis in five major neurodevelopmental/psychiatric diseases. Using gene-based association statistics, we robustly meta-analyzed different psychiatric conditions across distinct microarrays. This is the first large-scale CNV meta-analysis across a spectrum of neuropsychiatric disorders. We identified the *DOCK8*/*KANK1* locus as containing exonic CNVs with genome-wide significant meta *P* values and consistent direction of effects across all five cohorts. The highly significant associations of *DOCK8*/*KANK1* duplications

Table 3 Contributing signals for the *DOCK8* gene from each psychiatric disease cohort

Disease cohort	Overlapping CNV region (hg18)	CNV type	Odds ratio	<i>P</i> value	Number of case CNVs (%)	Number of control CNVs (%)
Janssen SCZ and BD	Chr9:396118-474850	Dup	Infinity	0.00693	6 (0.21)	0 (0)
CHOP SCZ	Chr9:372245-389052	Dup	10.71	0.008	7 (0.73)	1 (0.07)
CHOP ASD	Chr9:407918-474786	Dup	Infinity	0.00384	7 (0.34)	0 (0)
CHOP ADHD	Chr9: 293639-352917	Dup	3.32	0.0899	4 (0.32)	4 (0.10)
Depression	Chr9: 283360-294957	Dup	4.96	0.00731	8 (1.24)	4 (0.25)

Number of case CNVs is the number of cases having a *DOCK8* duplication and the frequency among all cases. *Number of control CNVs* is the number of controls having a *DOCK8* duplication CNV and the frequency among all controls

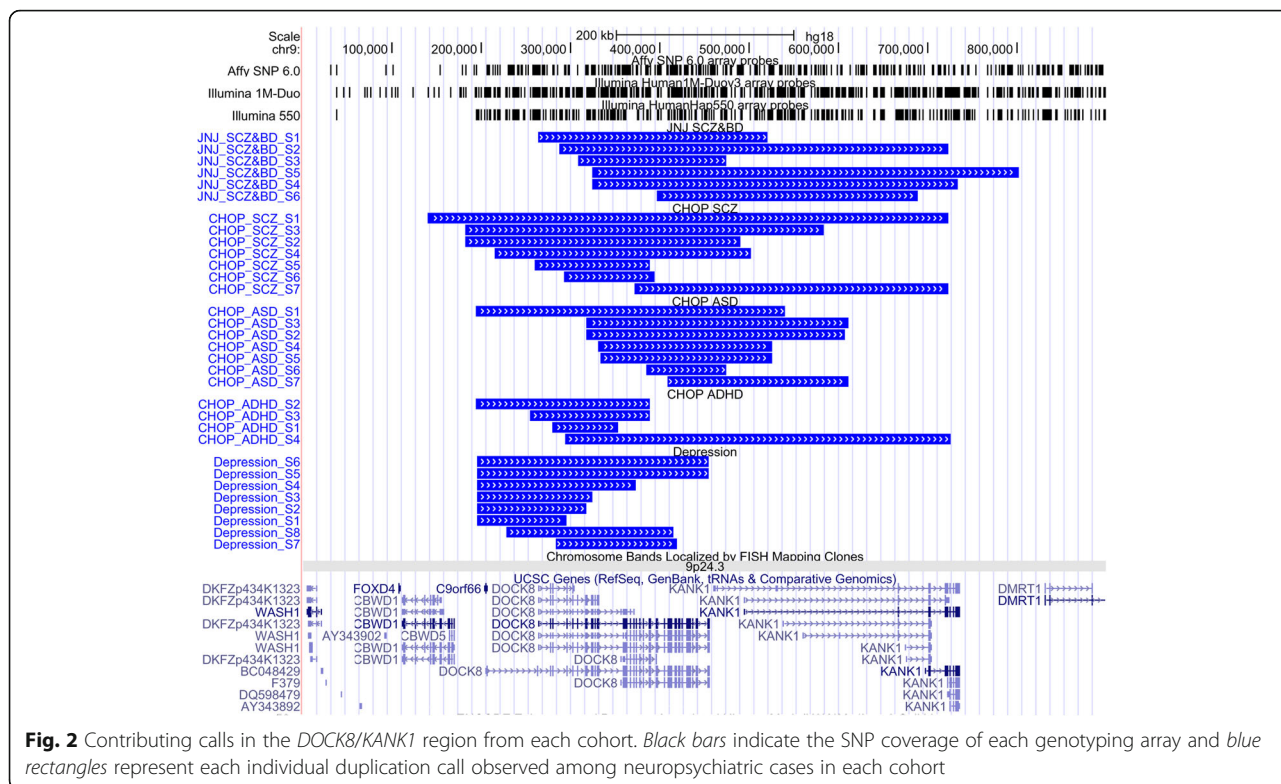


Fig. 2 Contributing calls in the *DOCK8/KANK1* region from each cohort. Black bars indicate the SNP coverage of each genotyping array and blue rectangles represent each individual duplication call observed among neuropsychiatric cases in each cohort

were further validated by an independent experimental approach. The identification of shared structural variants underlying the five neuropsychiatric disorders helps to refine the genetic basis for co-morbidity and co-occurrence of neuropsychiatric diseases among individuals or families has the potential to help in the

development of common therapeutics of shared genetic targets across different diseases.

The duplications of *DOCK8* and *KANK1* at 9p24.3 are intriguing given that these genes have been shown to be involved in neurodevelopment and neurological functions. *DOCK8* is the dedicator of cytokinesis 8, a

Table 4 qPCR validation of duplications in the *DOCK8–KANK1* region

Sample	CNV	Assay132 chr9:279,035-279138	Assay134 chr9:326,544-326608	Assay135 chr9:416,891-417000	Assay137 chr9:460,349-460443	Assay141 chr9:537,138-537211
S1	chr9:263161-520703	3	3	3	3	2
S2	chr9:287395-723374	2	3	3	3	3
S3	chr9:308154-474850	2	3	3	3	2
S4	chr9:323820-733353	2	3	3	3	3
S5	chr9:323820-801972	2	3	3	3	4
S6	chr9:396118-689065	2	2	3	3	3
S7	chr9:468154-697859	2	2	3	3	3
S8	chr9:474299-702599	2	2	3	3	3
S9	chr9:490811-534956	2	2	3	3	2
S10	chr9:516773-801972	2	2	2	2	3
Con		2	2	2	2	2

Each sample was tested by five qPCR assays which cover much of the duplication regions. Each result represents copy number calculated from triplicate runs. The table has been shaded gray for assays that were within the duplication call for that subject detected by arrays, and the CNVs validated by qPCR are highlighted in red

member of the DOCK180 family of guanine nucleotide exchange factors (GEF), which includes 11 DOCK genes [49, 50]. GEF proteins are important components of intracellular signaling networks, activating small GTPases by exchanging bound GDP for free GTP [51]. *DOCK8* is expressed in adult and fetal brain tissues and deletion or translocation breakpoints that disrupt its function have been found in individual patients with intellectual disability [52]. This is the first time that *DOCK8* duplications were found to be significantly associated with a spectrum of neuropsychiatric disorders, suggesting that a tightly regulated *DOCK8* expression level may be required for normal cellular function. The neighboring gene of *DOCK8* is *KANK1* (KN motif and ankyrin repeat domains 1), which has also been demonstrated to play a role in neuronal functions. *KANK1* functions in actin cytoskeleton formation by competing for 14-3-3 binding upon phosphorylation by Akt and inhibiting RhoA activity [53, 54]. It inhibits neurite outgrowth, actin fiber formation, and cell migration, depending on the competitive interaction with BAIAP2 to block its association with activated RAC1 [55]. *KANK1* functions in regulating microtubule dynamics at the cell cortex by recruiting KIF21A, which is important in neuronal development [56, 57]. Mutations in this gene cause cerebral palsy spastic quadriplegic type 2, a central nervous system development disorder [58]. In the DECIPHER dataset, 16 subjects with intellectual disability carry *DOCK8* duplications/gains and ten carry *KANK1* duplications/gains [59].

The novelty of our finding lies in the illustration of the significant association of *DOCK8/KANK1* with multiple neuropsychiatric diseases. In the CNV study of Coe et al. [60] for intellectual disability, developmental delay, and/or ASD, *DOCK8* is nominally significant for deletions ($P = 0.000281$) but not duplications. In the recent CNV study of schizophrenia by Marshall et al. [42], *DMRT1*, which is at the same cytoband of 9p24.3, was reported to be a novel CNV-containing gene significantly associated with schizophrenia in the test for a combined (deletion + duplication) CNV analysis. The reported *DMRT1* CNV locus (chr9:831690–959090, hg18) is more than 400 kb downstream of the *DOCK8/KANK1* region found in our study. In addition, the *DOCK8/KANK1* region is not highly prone to CNV in the general population. In our study, 0.25% or less of the control subjects carry *DOCK8/KANK1* duplications in each cohort. The Database of Genomic Variants (DGV) [61] shows 65 duplications in the *DOCK8* gene. Among them, 33 were reported in CNV studies of developmental delay [60, 62], and 32 duplications in *DOCK8* were reported from other CNV population studies. This is also similar to another gene, *NRXN1*,

the deletion of which is significantly associated with schizophrenia [63]. A total of 144 deletions/loss were reported in DGV for the *NRXN1* gene, among which 45 were reported in the CNV studies of developmental delay and the remaining 99 were from other CNV population studies. Therefore, the significant association in the meta-analysis reflects its potential contribution to the pathology of neurodevelopmental/psychiatric disorders. In addition, the novelty of our finding lies in the identification of the significant association of *DOCK8* duplication with multiple neurodevelopmental/psychiatric disorders across cohorts (association $P < 0.01$ for four out of five cohorts), implicating its common role as a risk locus for these diseases.

To compensate for the incomplete overlapping in SNP coverage between study cohorts, we conducted gene-based meta-analysis, which has been widely adopted in both common and rare variant analyses. One caveat is that CNV overlaps with different gene regions might not have the same effect on the expression of the gene. Some may result in frame shifts and the complete loss of function of the gene and some may have minor effects on the gene. Nevertheless, the biological function of each gene is maintained in a fine balanced state for cellular activities. Even minor perturbation of its expression could lead to pathological consequences. The discovery of CNVs in *DOCK8* and *KANK1* across all five diseases not only adds to the growing catalog of neurodevelopmental variants but also paves the way for new diagnostics opportunities and interventions which could be applied across multiple clinical indications. However, functional studies are needed to better understand the biological effect of these variations.

Conclusions

With the growing awareness of the high impact of childhood psychiatric conditions comes the important need for large-scale genetic studies and a unified picture of the catalog of rare variants underlying these conditions. We have undertaken the unprecedented step to meta-analyze CNVs across five neurodevelopmental/psychiatric diseases and have uncovered significant structural variation at the *DOCK8/KANK1* locus shared by these diseases, emphasizing the common genetic component involved in the pathogenesis of neuropsychiatric disorders.

Additional files

Additional file 1: Table S1. The quality control criteria based on CNV metrics applied to each cohort. **Table S2.** The number of histone marks, DNase, and eQTLs in each *DOCK8* CNV. (PDF 127 kb)

Additional file 2: Figure S1. Circular plot showing the intersections of genes harboring case-enriched CNVs between the five neuropsychiatric cohorts. **Figure S2.** The BAF and LRR plots of *ZNF280A* deletions. **Figure S3.** The BAF and LRR plots of *DOCK8/KANK1* duplications. **Figure S4.** The topologically associating domains (TAD) at the chromosome 9 *DOCK8/KANK1* region. (PDF 2076 kb)

Abbreviations

ADHD: Attention deficit hyperactivity disorder; AGP: Autism Genome Project; AGRE: Autism Genetics Resource Exchange; ASD: Autism spectrum disorders; BAF: B allele frequency; BD: Bipolar disease; CAG: Center for Applied Genomics; CHOP: the Children's Hospital of Philadelphia; CNV: Copy number variation; GAIN: Genetic Association Information Network; GCWF: GC base pair wave factor; GEF: Guanine nucleotide exchange factor; GWAS: Genome-wide association studies; KANK1: KN motif and ankyrin repeat domains 1; LMM: Linear mixed model; LRR: Log R ratio; MDD: Major depressive disorder; NESDA: Netherlands Study of Depression and Anxiety; PCA: Principal component analysis; QC: Quality control; qPCR: Quantitative polymerase chain reaction; SCZ: Schizophrenia; SNP: Single nucleotide polymorphism; TAD: Topologically associating domains; UPL: Universal Probe Library

Acknowledgments

We thank all the study volunteers for participating in the research studies and the clinicians and support staff for enabling patient recruitment and blood sample collection. We also thank the staff in the Neuroscience Biomarkers Genomic Lab at Janssen & the Center for Applied Genomics at CHOP for sample processing and the staff at Illumina for genotyping Janssen DNA samples. We thank Anthony Santos, Nicole Bottrel, Monique-Andree Franc, and William Cafferty of Janssen Research & Development for operational support, and Stefanie Rasnack of Janssen Research & Development for input in setting up the collaboration and for discussion.

Members of the Janssen-CHOP Neuropsychiatric Genomics Working Group: CHOP, Patrick Sleiman^{1,2}, Joseph Glessner¹, Dexter Hadley¹, Charly Kao¹, Zhi-liang Wu¹, Cecilia Kim¹, Kelly Thomas¹, Hakon Hakonarson^{1,2}; Janssen, Dai Wang³, Reyna Favis³, Dong-Jing Fu⁴, Hedy Chung⁴, Adam Savitz⁵, Srihari Gopal⁵, Nadine Cohen⁵, Qingqin Li⁵.

¹The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; ²Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; ³Janssen Research & Development, LLC, Raritan, NJ, 08869, USA; ⁴Janssen Scientific Affairs, LLC, Titusville, NJ, 08540, USA; ⁵Janssen Research & Development, LLC, Titusville, NJ, 08869, USA.

Funding

Funding for this study was provided by an Institutional Development Award to the Center for Applied Genomics from The Children's Hospital of Philadelphia; Adele and Daniel Kubert donation, University of Pennsylvania Biomedical Graduate Studies training grant; the Cotswold foundation; by 1R01MH097284-01 from the National Institute of Mental Health (NIMH), and sponsored research to The Children's Hospital of Philadelphia by Janssen Research & Development, LLC.

Availability of data and materials

The dbGaP study accessions: CHOP SCZ, phs000167.v1.p1 and phs000021.v3.p2; CHOP ASD, phs000267.v5.p2; CHOP ADHD, phs000016.v2.p2; depression, phs000020.v2.p1. The other datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

HH and NC conceived the project and HH supervised the study. QL managed the Janssen samples and genetic data. JG and JL performed the analysis with the help of JG, AD, LL, and DH. DW performed Janssen genotype data QC. REG assisted with phenotyping of the PENN cohort. MM conducted qPCR validation. JG, JL, AD, and HH wrote the manuscript. The Janssen-CHOP Neuropsychiatric Genomics Working Group collected samples and performed phenotype classification. All authors contributed to the data analysis review, discussions, and contributed to the final manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All cohorts in our study were from published datasets. Study protocols used in each cohort have been approved by corresponding Institutional Review Board or equivalent committees, and written informed consent was given from each participant or parent. Our research complies with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

Dr. Nadine Cohen is a former employee of Janssen Research & Development. Drs. Dai Wang and Qingqin Li are current employees of Janssen Research & Development. The remaining authors declare that they have no competing interests. Funding for this study included sponsored research to The Children's Hospital of Philadelphia by Janssen Research & Development, LLC.

Publisher's Note

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

Author details

¹The Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. ²Department of Cell Biology, Tianjin Medical University, Tianjin, China. ³Janssen Research & Development, LLC, Raritan, NJ 08869, USA. ⁴Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ⁵Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ⁶Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. ⁷Janssen Research & Development, LLC, Titusville, NJ 08560, USA.

Received: 13 July 2017 Accepted: 14 November 2017

Published online: 30 November 2017

References

- Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013;381(9875):1371–9.
- International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748–52.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984–94.
- Glessner JT, Reilly MP, Kim CE, Takahashi N, Albano A, Hou C, Bradfield JP, Zhang H, Sleiman PM, Flory JH, et al. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc Natl Acad Sci U S A*. 2010; 107(23):10584–9.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569–73.
- Glessner JT, Wang K, Sleiman PM, Zhang H, Kim CE, Flory JH, Bradfield JP, Imielinski M, Frackelton EC, Qiu H, et al. Duplication of the *SLIT3* locus on 5q35.1 predisposes to major depressive disorder. *PLoS One*. 2010;5(12): e15463.
- Elija J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D, Sleiman PM, Zhang H, Kim CE, Robison R, et al. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat Genet*. 2012;44(1):78–84.
- Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43(10):969–76.
- Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat Genet*. 2011;43(10):977–83.
- Sleiman P, Wang D, Glessner J, Hadley D, Gur RE, Cohen N, Li Q, Hakonarson H, Janssen CNGWG. GWAS meta analysis identifies *TSNARE1* as a novel Schizophrenia/Bipolar susceptibility locus. *Sci Rep*. 2013;3:3075.

11. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 2016;17(1):9–18.
12. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet.* 2011;43(9):864–8.
13. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, Rippey C, Shahin H, Consortium on the Genetics of S, Group PS, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell.* 2013;154(3):518–29.
14. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature.* 2014; 506(7487):179–84.
15. CONVERGE consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature.* 2015;523(7562):588–91.
16. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, Tung JY, Hinds DA, Perlis RH, Winslow AR. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet.* 2016;48(9):1031–6.
17. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, Meddens SF, Linner RK, Rietveld CA, Derringer J, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet.* 2016; 48(6):624–33.
18. Direk N, Williams S, Smith JA, Ripke S, Air T, Amare AT, Amin N, Baune BT, Bennett DA, Blackwood DHR, et al. An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biol Psychiatry.* 2017;82(5):322–9.
19. Mullins N, Lewis CM. Genetics of depression: progress at last. *Curr Psychiatry Rep.* 2017;19(8):43.
20. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45(D1):D896–901.
21. Hadley D, Wu ZL, Kao C, Kini A, Mohamed-Hadley A, Thomas K, Vazquez L, Qiu H, Mentch F, Pellegrino R, et al. The impact of the metabotropic glutamate receptor and other gene family interaction networks on autism. *Nat Commun.* 2014;5:4074.
22. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D, Moran JL, Sklar P, Ripke S, Chambert KD, et al. Copy number variation in bipolar disorder. *Mol Psychiatry.* 2016;21(1):89–93.
23. Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, Fromer M, Ruderfer D, Akterin S, Bergen SE, et al. Copy number variation in schizophrenia in Sweden. *Mol Psychiatry.* 2014;19(7):762–73.
24. Schneider M, Debbané M, Bassett AS, Chow EW, Fung WL, van den Bree M, Owen M, Murphy KC, Niarchou M, Kates WR, et al. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry.* 2014;171(6):627–39.
25. Hassfurther A, Komini E, Fischer J, Leipoldt M. Clinical and genetic heterogeneity of the 15q13.3 microdeletion syndrome. *Mol Syndromol.* 2016;6(5):222–8.
26. Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R, et al. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet.* 2010;153B(4):937–47.
27. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014; 511(7510):421–7.
28. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, Kluff C, Smit G, Nolen WA, Zitman FG, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* 2008; 16(3):335–42.
29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904–9.
30. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. 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.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
32. Glessner JT, Li J, Hakonarson H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* 2013;41(5):e64.
33. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods.* 2014; 11(4):407–9.
34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–1.
35. Wang M, Zhao Y, Zhang B. Efficient test and visualization of multi-set intersections. *Sci Rep.* 2015;5:16923.
36. Li YR, Li J, Zhao SD, Bradfield JP, Mentch FD, Maggadottir SM, Hou C, Abrams DJ, Chang D, Gao F, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med.* 2015; 21(9):1018–27.
37. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyster AE, Denny JC, Consortium GT, Nicolae DL, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47(9):1091–8.
38. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Investigators A, Hayward NK, Montgomery GW, Visscher PM, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet.* 2010;87(1):139–45.
39. Moutsianis L, Agarwala V, Fuchsberger C, Flannick J, Rivas MA, Gaulton KJ, Albers PK, Go TDC, McVean G, Boehnke M, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet.* 2015;11(4): e1005165.
40. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet.* 2010;11(12):843–54.
41. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Magi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012;44(9):991–1005.
42. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, Antaki D, Shetty A, Holmans PA, Pinto D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet.* 2017;49(1):27–35.
43. Rees E, Kirov G, Sanders A, Walters JT, Chambert KD, Shi J, Szatkiewicz J, O'Dushlaine C, Richards AL, Green EK, et al. Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol Psychiatry.* 2014;19(1):37–40.
44. Chen J, Calhoun VD, Perrone-Bizzozero NI, Pearlson GD, Sui J, Du Y, Liu J. A pilot study on commonality and specificity of copy number variants in schizophrenia and bipolar disorder. *Transl Psychiatry.* 2016;6(5):e824.
45. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
46. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenyi M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
47. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015; 348(6235):648–60.
48. Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics.* 2014;30(12): i185–94.
49. Ruusala A, Aspenstrom P. Isolation and characterisation of DOCK8, a member of the DOCK180-related regulators of cell morphology. *FEBS Lett.* 2004;572(1–3):159–66.
50. Su HC. Deducator of cytokinesis 8 (DOCK8) deficiency. *Curr Opin Allergy Clin Immunol.* 2010;10(6):515–20.
51. Rossman KL, Der CJ, Sondek J. GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol.* 2005; 6(2):167–80.
52. Griggs BL, Ladd S, Saul RA, DuPont BR, Srivastava AK. Deducator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. *Genomics.* 2008;91(2):195–202.

53. Kakinuma N, Roy BC, Zhu Y, Wang Y, Kiyama R. Kank regulates RhoA-dependent formation of actin stress fibers and cell migration via 14-3-3 in PI3K-Akt signaling. *J Cell Biol.* 2008;181(3):537–49.
54. Kakinuma N, Zhu Y, Wang Y, Roy BC, Kiyama R. Kank proteins: structure, functions and diseases. *Cell Mol Life Sci.* 2009;66(16):2651–9.
55. Roy BC, Kakinuma N, Kiyama R. Kank attenuates actin remodeling by preventing interaction between IRSp53 and Rac1. *J Cell Biol.* 2009;184(2):253–67.
56. Kakinuma N, Kiyama R. A major mutation of KIF21A associated with congenital fibrosis of the extraocular muscles type 1 (CFEOM1) enhances translocation of Kank1 to the membrane. *Biochem Biophys Res Commun.* 2009;386(4):639–44.
57. van der Vaart B, van Riel WE, Doodhi H, Kevenaer JT, Katrukha EA, Gumy L, Bouchet BP, Grigoriev I, Spangler SA, Yu KL, et al. CFEOM1-associated kinesin KIF21A is a cortical microtubule growth inhibitor. *Dev Cell.* 2013;27(2):145–60.
58. Lerer I, Sagi M, Meiner V, Cohen T, Zlotogora J, Abeliovich D. Deletion of the ANKRD15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy. *Hum Mol Genet.* 2005;14(24):3911–20.
59. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet.* 2009;84(4):524–33.
60. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46(10):1063–71.
61. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986–92.
62. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. A copy number variation morbidity map of developmental delay. *Nat Genet.* 2011;43(9):838–46.
63. Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr Bull.* 2009;35(5):851–4.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

