

## Genome-wide Burden of Rare Short Deletions Is Enriched in Major Depressive Disorder in Four Cohorts

Xianglong Zhang, Abdel Abdellaoui, James Rucker, Simone de Jong, James B. Potash, Myrna M. Weissman, Jianxin Shi, James A. Knowles, Carlos Pato, Michele Pato, Janet Sobell, Johannes H. Smit, Jouke-Jan Hottenga, Eco J.C. de Geus, Cathryn M. Lewis, Henriette N. Buttenschøn, Nick Craddock, Ian Jones, Lisa Jones, Peter McGuffin, Ole Mors, Michael J. Owen, Martin Preisig, Marcella Rietschel, John P. Rice, Margarita Rivera, Rudolf Uher, Pablo V. Gejman, Alan R. Sanders, Dorret Boomsma, Brenda W.J.H. Penninx, Gerome Breen, and Douglas F. Levinson

### ABSTRACT

**BACKGROUND:** Major depressive disorder (MDD) is moderately heritable, with a high prevalence and a presumed high heterogeneity. Copy number variants (CNVs) could contribute to the heritable component of risk, but the two previous genome-wide association studies of rare CNVs did not report significant findings.

**METHODS:** In this meta-analysis of four cohorts (5780 patients and 6626 control subjects), we analyzed the association of MDD to 1) genome-wide burden of rare deletions and duplications, partitioned by length (<100 kb or >100 kb) and other characteristics, and 2) individual rare exonic CNVs and CNV regions.

**RESULTS:** Patients with MDD carried significantly more short deletions than control subjects ( $p = .0059$ ) but not long deletions or short or long duplications. The confidence interval for long deletions overlapped with that for short deletions, but long deletions were 70% less frequent genome-wide, reducing the power to detect increased burden. The increased burden of short deletions was primarily in intergenic regions. Short deletions in cases were also modestly enriched for high-confidence enhancer regions. No individual CNV achieved thresholds for suggestive or significant association after genome-wide correction.  $p$  values < .01 were observed for 15q11.2 duplications (*TUBGCP5*, *CYFIP1*, *NIPA1*, and *NIPA2*), deletions in or near *PRKN* or *MSR1*, and exonic duplications of *ATG5*.

**CONCLUSIONS:** The increased burden of short deletions in patients with MDD suggests that rare CNVs increase the risk of MDD by disrupting regulatory regions. Results for longer deletions were less clear, but no large effects were observed for long multigenic CNVs (as seen in schizophrenia and autism). Further studies with larger sample sizes are warranted.

**Keywords:** Copy number variation, Genetics, Genome-wide association study, Major depressive disorder, Meta-analysis, Neuroscience

<https://doi.org/10.1016/j.biopsych.2019.02.022>

Major depressive disorder (MDD) is a common psychiatric disorder with a lifetime prevalence of 10% to 20% (1). It was the third leading cause of global disability in 2015 (2). Heritability is approximately 37%, lower than that of several other psychiatric disorders (3). The genome-wide contribution of common single nucleotide polymorphisms (SNPs) to MDD risk is approximately 20% (4). Consistent with the moderate heritability and high population prevalence, it has required more than 100,000 MDD cases to detect large numbers of genome-wide significant SNP associations, e.g., 15 loci in 121,380 cases plus 338,101 control subjects (5) and 44 loci in a partially

overlapping sample (135,458 cases plus 344,901 control subjects) (6).

Rare copy number variants (CNVs) could be contributing to the unexplained portion of genetic risk and provide information about disease mechanisms. Two previous MDD studies of longer CNVs reported no significant genome-wide burden in patients with MDD (7,8). Here, to achieve a larger sample size, we performed a meta-analysis of the association of MDD to rare CNVs in 5780 patients and 6626 control subjects from four cohorts. A significant increase of rare, shorter deletions (<100,000 base pairs) was observed in

patients with MDD, and this was driven by CNVs in intergenic regions.

## METHODS AND MATERIALS

### Samples

We studied four European ancestry cohorts as shown in Table 1. All participants gave signed informed consent under protocols approved by the relevant institutional review boards.

**RADIANT Cohort.** The RADIANT cohort (7) included patients from three studies of recurrent MDD and two control subject cohorts (458 control subjects who were screened for lifetime absence of psychiatric disorder and 2699 control subjects from phase 2 of the National Blood Service Wellcome Trust Case Control Consortium subcohort). Patients were interviewed with the Schedules for Clinical Assessment in Neuropsychiatry (9) and diagnosed using ICD-10 or DSM-IV criteria. Exclusion criteria were any history or family history of schizophrenia or bipolar disorder or any history of mood disorder secondary to alcohol/substance misuse or of mood-incongruent psychosis (7).

**Netherlands Study of Depression and Anxiety/Netherlands Twin Register.** Patients with MDD and control subjects were drawn from the Netherlands Twin Register (NTR) (10) and the Netherlands Study of Depression and Anxiety (NESDA) (11). Cases had DSM-IV MDD diagnoses as assessed with the Composite Interview Diagnostic Instrument (12).

**Genetics of Recurrent Early-Onset Depression.** Patients with MDD and control subjects were drawn from the Genetics of Recurrent Early-Onset Depression (GenRED) cohort (13,14). Patients had a consensus DSM-IV MDD diagnosis based on a Diagnostic Interview for Genetic Studies interview and other information, with recurrence or chronicity (an episode lasting  $\geq 3$  years), onset before 31 years of age, one or more siblings or parents with recurrent MDD and onset before 41 years of age, MDD independent of substance dependence, no bipolar, schizoaffective disorder or schizophrenia diagnosis, and no parent or sibling with suspected bipolar disorder I. The control subjects ( $n = 1345$ ) from the Molecular Genetics of Schizophrenia cohort (15) denied (by online screen) ever meeting DSM-IV MDD gate criteria (no

2-week period of depressed mood or anhedonia most of the day, nearly every day), whereas the published GenRED genome-wide association study (GWAS) (13) included control subjects who never met full MDD criteria by online screen (16).

**GenRED II.** Patients with MDD were from the second GenRED GWAS wave (same criteria as GenRED). Control subjects were drawn from the Genomic Psychiatry Cohort (17), Depression Genes and Networks (18), and the Mayo Clinic (19). Control subjects were drawn from the Genomic Psychiatry Cohort (17) (screened for lifetime depression with a self-report questionnaire), Depression Genes and Networks (18) (screened with a Structured Clinical Interview for DSM-IV), and the Mayo Clinic (19) (screened based on diagnoses in the electronic medical record over an extended period).

### Genotyping

Patients with MDD in the RADIANT cohort and screened control subjects were genotyped with the HumanHap 610-Quad beadchip (Illumina, Inc., San Diego, CA), and the un-screened National Blood Service samples were genotyped with Illumina Infinium 1M beadchips (hg18 for both) (7). The NTR/NESDA (20) and GenRED cohorts were genotyped with the Affymetrix Human Genome-Wide SNP 6.0 Array (Affymetrix, Santa Clara, CA) (hg18) (14), and GenRED II patients and control subjects were genotyped with the Illumina Omni1-Quad beadchip (hg19) (21).

### Selection of CNV Calling Algorithms

CNVs were called with PennCNV (22), QuantiSNP (23) and iPattern (24) in the RADIANT dataset (using 562,329 probes common to the two platforms), with Birdsuite (25) and PennCNV (22) in the NTR/NESDA dataset, with Birdsuite (25) in the GenRED dataset, and with QuantiSNP (23) and PennCNV (22) in the GenRED II dataset. There is no consensus “optimal” calling algorithm for each platform. Various authors use a single calling method (8,26), agreement between two methods (20,27), or more complex approaches (28,29).

We conducted a preliminary analysis of CNV call concordance for duplicate genotypes for 115 Affymetrix 6.0 samples and 20 Illumina Human610-Quad samples. For Affymetrix, we compared CNVision (28), QuantiSNP (23), PennCNV (22), and Birdsuite (25) and each pair of algorithms, plus the addition of CNVision’s pCNV parameter (estimating the probability of a true CNV, based on per-SNP variability of log R ratio [LRR] and the number of SNPs consistent with a CNV based on B allele

**Table 1. Cohorts and Sample Sizes Before and After Quality Control Filtering**

Cohort	Pre-Quality Control Sample Size		Post-Quality Control Sample Size	
	Patients With MDD (Male/Female)	Control Subjects (Male/Female)	Patients With MDD (Male/Female)	Control Subjects (Male/Female)
RADIANT	3087 (908/2179)	3157 (1522/1635)	2460 (724/1736)	2587 (1240/1347)
NESDA/NTR	1637 (509/1128)	2030 (765/1265)	1568 (488/1080)	1913 (719/1194)
GenRED I	1089 (319/770)	1345 (784/561)	941 (271/670)	1264 (743/521)
GenRED II	831 (144/687)	944 (418/526)	811 (139/672)	862 (384/478)
Total	6644 (1880/4764)	7476 (3489/3987)	5780 (1622/4158)	6626 (3086/3540)

GenRED, Genetics of Recurrent Early-Onset Depression; MDD, major depressive disorder; NESDA, Netherlands Study of Depression and Anxiety; NTR, Netherlands Twin Register.

## Rare Deletions in Major Depressive Disorder

frequency [BAF]). For Illumina we compared all algorithms (except Birdsuite) and pairs plus the addition of pCNV. We also conducted the analyses for short (<100 kb) and long (>100 kb) CNVs separately.

For Affymetrix, Birdsuite had the highest concordance rate (deletions and duplications), whereas combining it with any other method slightly increased concordance but excluded >40% of calls (Supplemental Table S1). We therefore used Birdsuite alone for Affymetrix data. For Illumina data, QuantiSNP alone had the best concordance for deletions (Supplemental Table S2). For duplications, concordance was highest for QuantiSNP alone; calls made by both PennCNV and QuantiSNP showed improved concordance but excluded >30% of calls. We used QuantiSNP for primary analyses, plus a secondary “narrow” QuantiSNP + PennCNV analysis. For both platforms, concordance was similar for shorter and longer CNVs (Supplemental Tables S3–S6).

### Quality Control of Samples and CNV Calls

Exclusion criteria for samples were applied to each cohort separately. For NTR/NESDA (20) and GenRED (14), exclusions were applied to samples retained by the original studies, using the previous Birdsuite calls: 1) probe intensity variances >4 SDs above the cohort mean; 2) total number or length of deletions or duplications >3 SDs above the mean; 3) any chromosome with number or length of deletions or duplications >7 SDs above the mean; and 4) only autosomal CNVs were called. For Illumina data (RADIANT and GenRED II cohorts), we recalled CNVs with QuantiSNP and PennCNV from raw LRR and BAF data. Exclusion criteria for unfiltered calls were 1) genotype call rate <99%; 2) >5% of SNPs with LRR < -0.5 or > 0.5; 3) >1% of SNPs with LRR < -1; 4) BAF drift > 0.01; 5) LRR SD > 0.28; 6) waviness factor < -0.05 or > 0.05; or 7) total CNV number or length >3 SDs above the cohort mean (deletions or duplications).

For both platforms, we removed CNVs with fewer than 10 probes and of Birdsuite calls with a logarithm of odds score <10 (duplications) or <6 (deletions) and QuantiSNP calls with a maximum log(Bayes factor) <10. We then merged adjacent deletions (copy numbers 0 or 1) or adjacent duplications (copy numbers 3 or 4) if the number of probes separating them was <30% of probes in the merged region (iterating through each chromosome until all eligible segments were merged, using an in-house script). We removed CNVs with ≥50% overlap with centromeres, telomeres, segmental duplications, or immunoglobulin genes, or length <10 kb (too few probes to call reliably) or >4 Mb [in previous work (30), CNVs >4 Mb were disproportionately detected in DNA from lymphoblastic cell lines], or with a frequency of >1% in any of four large-sample cohorts included in the Database of Genomic Variants (31–34), or with a frequency of >1% (based on 50% overlap) in any of our control cohorts.

### Statistical Analysis Overview

All analyses were conducted for post-quality control (QC) deletions and duplications using PLINK and R software. Genomic locations with hg18 coordinates were converted to hg19 (University of California Santa Cruz LiftOver tool). We first determined (as described below) that effects of cohort and sex

had to be controlled appropriately. We chose primary analyses that directly compute an odds ratio (OR) and were equivalent to meta-analysis: logistic regression with sex and cohort covariates (for burden tests) or Cochrane-Mantel-Haenszel (CMH) tests stratified for sex and cohort (for single CNVs), plus meta-analysis and/or permutation tests to check results. We tested two main hypotheses, correcting for multiple tests within each hypothesis.

Our first hypothesis was that the global burden of rare CNVs is greater in patients with MDD than in control subjects. The four primary analyses were for deletions and duplications, each subdivided by size (<100 kb, >100 kb); the threshold of significance was  $p < .0125$  (.05/4).

Our second hypothesis was that patients with MDD are more likely to carry specific CNVs. Primary analyses tested association by 1) gene (CNVs impacting exon[s] of the gene) and 2) CNV region defined by pools of overlapping CNVs (PLINK). We established thresholds for significant suggestive association as described below (30). Genic tests considered only exonic CNVs because of the stronger mechanistic hypothesis and because exonic and “genic” CNVs were largely overlapping (Supplemental Table S7)—93.2% (deletions) and 99.4% (duplications) of long genic CNVs and 62.6% and 83.6% of short genic CNVs were exonic.

### Effects of Cohort and Sex

We evaluated two potential confounding variables: cohort and sex (the female proportion was higher in patients with MDD and was variable across cohorts). Multiple linear regressions were performed for total rare deletions or duplications per subject or summed length (Supplemental Table S8), with case-control status, cohort, and sex as independent variables. There were significant effects for cohort (deletions and duplications) and sex (deletions).

Genome-wide burden analyses were thus performed for short and long deletions and duplications, using logistic regression with sex and cohort as covariates to test for case-control difference. Secondary analyses considered intergenic and genic CNVs, separate analyses of exonic and intronic-only CNVs, singletons, CNVs >500 kb and >1000 kb, and short deletions by 10-kb length bins (10–20, 20–30, etc.). Results were checked against logistic regression for each cohort (with sex as a covariate) followed by meta-analysis of the beta coefficients and standard errors [R function “metagen” (35)], and permutation tests stratified for cohort and sex (randomly swapping case-control status within the same sex and cohort 100,000 times using the PLINK “-within” option).

### Down-sampled Analysis

As a check on the effects of uneven numbers of patients/control subjects and males/females per cohort, we repeated burden analyses using a down-sampled dataset: 1622 male and female patients and control subjects (6488 total) drawn from each cohort proportional to its size (Supplemental Table S9).

### Analyses of Single CNVs

We performed one-sided CMH tests (stratified by sex and cohort) of a case excess of exonic CNVs impacting each

RefSeq gene and of CNVs in each “CNV region,” and we checked results with a stratified permutation tests (results were almost identical). To define regions, we used the PLINK “-segment-group” command to identify 994 CNV “pools” of overlapping post-QC CNVs (from all cohorts) and termed the union a CNV region.

For any CNV with nominally increased case frequency (CMH  $p < .01$ ), we carried out additional filtering because calling artifacts often produce “significant” results for rare events. We visualized regional LRR and BAF plots for all carriers and a threefold number of noncarriers and superimposed on LRR a plot of estimated probe-by-probe copy number using a different algorithm (36). We also plotted all CNVs in the region. We excluded CNVs for which the probewise algorithm showed no copy number change. After excluding genes/regions where most calls were considered artifacts or were the edges of a common CNV region, we recomputed the CMH tests. We computed a proportion test across the four cohorts for each gene/region and excluded those with significant heterogeneity ( $p < 3.53 \times 10^{-5}$  to correct for multiple tests, see below). [Supplemental Table S10](#) lists the inspected regions and reasons for all exclusions.

Additional exploratory analyses (permutation tests) considered each transcript (<http://genome.ucsc.edu/>), Encyclopedia of DNA Elements regulatory region, Roadmap Project putative enhancer, promoter and dyadic region, and in aggregate for lists of CNVs with reported associations to psychiatric disorders (29,37) or developmental delay (32).

We used a previously described method (30) to estimate thresholds for significant association (expected by chance once in 20 genome-wide studies) and suggestive association (expected once per study). For all 994 CNV regions, the 329 deletion regions intersected with 487 genes, and 665 duplication regions intersected with 1475 genes (a total of 1962 genic tests). However, tests of genes within a region are correlated, and each region contained 4.64 genes on average. The 1962 genic tests represented 423 independent tests ( $\sim 1962/4.64$ ). We corrected for 1417 tests (994 regions and 423 genes)—a conservative estimate, because some regions were partially overlapping, and many genes were in more than one region, resulting in a  $p$  value threshold for significant association of  $3.53 \times 10^{-5}$  (.05/1417) and for suggestive association of  $7.06 \times 10^{-4}$  (1.0/1417).

### Power Analysis

Power analyses were conducted for detection of specific CNVs ([Supplemental Figure S1](#)). For the ranges of allele frequencies and genotypic relative risks that were observed in this study, power was good to excellent to detect associations at  $p = .01$ , but the detection of suggestive or significant association would have required larger relative risks than were observed here.

### Enrichment Analysis of Functional Pathways

To detect gene sets associated with MDD, we downloaded pathways from the Kyoto Encyclopedia of Genes and Genomes (<http://rest.kegg.jp/list/pathway>) and Gene Ontology (<http://geneontology.org/page/download-annotations>). Gene-set enrichment methods (38) were used to test for the

enrichment of CNVs (separately for all or exonic CNVs) in all the genes of each pathway relative to all genic CNVs using “-cnv-enrichment-test” in PLINK. Permutation tests of enrichment in cases were also performed by adding “-mperm 10000” in PLINK, with batch and sex as covariates. A set of schizophrenia-associated genes (39) was also tested.

We also evaluated whether case CNVs were enriched in high-confidence DNaseI regions ( $-\log_{10} p \geq 10$ ) from the Encyclopedia of DNA Elements (40) or the Roadmap Epigenomics Project (41) ([https://personal.broadinstitute.org/meuleman/reg2map/HoneyBadger2\\_release/](https://personal.broadinstitute.org/meuleman/reg2map/HoneyBadger2_release/)). Separately for promoter, enhancer, and dyadic regions, we analyzed all tissues together (i.e., whether more case short deletions intersected with at least one high-confidence regulatory sequence from any tissue) and then each tissue separately (counting high-confidence sequences for that tissue). For intergenic short deletions, averaged across tissues, the proportion of CNVs that overlap high-confidence regulatory regions was 1.3% for promoter regions, 2.0% for dyadic regions, and 14.9% for enhancer regions.

## RESULTS

Of 14,429 samples, 12,406 passed QC (5780 patients with MDD and 6626 control subjects) ([Table 1](#)). The total numbers of rare deletion and duplication calls are shown in [Supplemental Table S11](#).

### Genome-wide Burden

Cases had more CNVs per subject for rare, short (<100 kb) deletions ( $p = .00592$ , OR = 1.0483), driven by intergenic deletions ( $p = .00714$ , OR = 1.0716) ([Table 2](#)) and by cohort in [Supplemental Table S12](#). Similar results were observed by the primary logistic regression tests ([Supplemental Table S13](#)), meta-analysis of cohort-specific logistic regressions ([Supplemental Table S14](#)), stratified permutation tests ([Supplemental Table S7](#)), and the down-sampled dataset ([Supplemental Table S9](#)). Short deletions across the 10- to 100-kb range contributed to the case-control difference ([Supplemental Table S15](#) and [Supplemental Figures S2–S3](#)). No significant differences were observed for duplications or long deletions, but the OR for long deletions was positive (1.03), the confidence interval overlapped with that for short deletions ([Table 2](#)), and a secondary analysis of all rare deletions was significant (OR = 1.044; 95% confidence interval = 1.013–1.075;  $p = .0046$ ) ([Supplemental Table S16](#)). No significant effect was observed for singleton or very long (>500 kb, >1000 kb) deletions or duplications. There was no evidence of strong heterogeneity by cohort for short deletions (Cochran’s Q test;  $p = .31$ ) or short intergenic deletions ( $p = .14$ ) ([Supplemental Table S14](#) and [Supplemental Table S12](#) and [Supplemental Figure S4](#) for results by cohort). The excess of short deletions in cases became more significant when CNVs with frequency >1% in each cohort separately were excluded (rather than >1% in any cohort) ([Supplemental Table S17](#)) or when QuantiSNP + PennCNV calls were required for Illumina data ([Supplemental Table S18](#)). Burden results did not change after excluding nominally significant CNV regions that failed manual checks ([Supplemental Table S19](#)).

**Table 2. Genome-wide Burden Analyses of Long and Short Deletions and Duplications (CNVs/Subject)**

CNV Type	CNVs/Subject		OR (95% CI)	p Value
	Patients With MDD	Control Subjects		
<b>Deletions</b>				
>100 kb—All	0.324	0.318	1.0296 (0.9658–1.0975)	$3.71 \times 10^{-1}$
Intergenic	0.134	0.138	0.9881 (0.8956–1.0899)	$8.11 \times 10^{-1}$
Genic	0.191	0.181	1.0606 (0.9754–1.1531)	$1.68 \times 10^{-1}$
Exonic	0.175	0.168	1.0521 (0.9646–1.1475)	$2.51 \times 10^{-1}$
Intronic	0.015	0.012	1.1672 (0.8591–1.5876)	$3.23 \times 10^{-1}$
<100 kb—All	1.015	0.978	1.0483 (1.0139–1.0843)	$5.92 \times 10^{-3a}$
Intergenic	0.506	0.483	1.0716 (1.0190–1.1270)	$7.14 \times 10^{-3a}$
Genic	0.509	0.495	1.0343 (0.9877–1.0842)	$1.56 \times 10^{-1}$
Exonic	0.330	0.310	1.0552 (0.9965–1.1192)	$6.95 \times 10^{-2}$
Intronic	0.179	0.185	0.9952 (0.9149–1.0825)	$9.11 \times 10^{-1}$
<b>Duplications</b>				
>100 kb—All	0.496	0.476	1.0268 (0.9837–1.0725)	$2.29 \times 10^{-1}$
Intergenic	0.087	0.079	1.0912 (0.9654–1.2333)	$1.62 \times 10^{-1}$
Genic	0.409	0.397	1.0187 (0.9723–1.0677)	$4.37 \times 10^{-1}$
Exonic	0.406	0.395	1.0166 (0.9702–1.0657)	$4.89 \times 10^{-1}$
Intronic	0.004	0.002	1.5254 (0.7880–3.0095)	$2.13 \times 10^{-1}$
<100 kb—All	0.670	0.702	0.9850 (0.9512–1.0194)	$3.90 \times 10^{-1}$
Intergenic	0.252	0.266	0.9788 (0.9166–1.0449)	$5.21 \times 10^{-1}$
Genic	0.418	0.436	0.9845 (0.9410–1.0296)	$4.96 \times 10^{-1}$
Exonic	0.345	0.365	0.9730 (0.9253–1.0225)	$2.83 \times 10^{-1}$
Intronic	0.073	0.072	1.0586 (0.9285–1.2066)	$3.94 \times 10^{-1}$

For rare CNVs (carried by <1% of control subjects in each cohort), we defined four primary case-control tests of CNV subsets: deletions and duplications, and within each type, long (>100 kb) and short (<100 kb). For each subset, the case-control difference in CNVs per subject was tested by logistic regression, stratified for cohort and sex (Bonferroni-corrected threshold of significance  $p = .05/4$ , or .0125). Further exploration then considered genomic location: only intergenic, genic (exonic and/or intronic impact), exonic (subset of genic), and only intronic (subset of genic). See [Supplemental Table S8](#) for complete results.

CI, confidence interval; CNV, copy number variant; MDD, major depressive disorder; OR, odds ratio.

<sup>a</sup>Significant result.

We considered two possible within-cohort confounding factors: DNA source and genotyping platforms. In GenRED II, there were two DNA sources: blood (137 patients and all control subjects) or lymphoblastic cell lines (674 patients) ([Supplemental Table S20](#)). CNV burden did not significantly differ between blood and lymphoblastoid cell line case DNAs for any category, with a trend for more long deletions in lymphoblastoid cell line DNA ([Supplemental Table S21](#)). RADIANT CNV calls used probes common to Illumina 610-Quad (assayed in patients and screened control subjects) and Illumina 1M (unscreened control subjects). Burden results were similar for patients versus screened or unscreened control subjects, except that patients had more short deletions than screened control subjects (assayed with the same array) ([Supplemental Table S22](#)). Thus, neither factor accounted for the main finding.

### Exonic CNVs and CNV Regions

After all QC, no gene or region met the criteria for significant or suggestive association ([Supplemental Table S10](#)). Results with  $p < .01$  are shown in [Table 3](#). These represent four independent loci. Duplications in 15q11.2 achieved  $p = .00076$  (OR = 3.88). These duplications are reciprocal to a well-known deletion region (see Discussion), consistently impacting four genes. Less consistent results are observed in surrounding genes in segmental duplication regions ([Supplemental](#)

[Table S10](#)). Exonic deletions in *MSR1* achieve  $p = .0019$  (OR = 1.96); the region test includes several intronic deletions, with a similar result ( $p = .00075$ ; OR = 2.05). A CNV region containing exonic and intronic deletions in *PRKN* (formerly *PARK2*) produced  $p = .00097$  (OR = 1.92); the exonic test for *PRKN* had  $p > .01$ . Finally, there were six duplications, all in patients with MDD, in 6q21 ( $p = .0059$ ; OR =  $\infty$ ), including five exonic duplications in *ATG5* that overlapped with one upstream duplication. LRR/BAF plots of CNVs shown in [Table 3](#) are provided in [Supplemental Figure S5](#).

### Pathway Enrichment Analysis

After correction for multiple testing, no Kyoto Encyclopedia of Genes and Genomes or Gene Ontology pathway was enriched with short deletions in patients with MDD.

### Regulatory Regions

Enhancer regions were modestly enriched in patients with MDD for all tissues combined as defined above ( $p = .024$ ), and in 5 of 127 specific tissues ( $p < .05$ ) ([Supplemental Table S23](#)).

### Known Loci Associated With Psychiatric Disorders or Developmental Delay

Permutation tests did not demonstrate case enrichment of CNVs in loci associated with psychiatric disorders

**Table 3. Copy Number Variant Genes and Regions<sup>a</sup>**

Gene or Region	Chromosome	Start	End	All		CMH Test		RAD		GR2		GR1		Neth		Annotation
				Ca	Co	OR	<i>p</i> Value	Ca	Co	Ca	Co	Ca	Co	Ca	Co	
Genes (Exonic)																
Deletion																
<i>MSR1</i>	8	15,965,386	16,050,300	55	32	1.96	$1.9 \times 10^{-3}$	23	10	6	3	15	10	11	9	
Duplication																
<i>TUBGCP5</i>	15	22,833,394	22,873,891	24	7	3.88	$7.6 \times 10^{-4}$	7	3	11	0	2	1	4	3	15q11.2 (reciprocal to well-known deletion region)
<i>CYFIP1</i>	15	22,892,648	23,003,603													
<i>NIPA2</i>	15	23,004,683	23,034,427													
<i>NIPA1</i>	15	23,043,278	23,086,843													
Regions (Genic and/or Intergenic)																
Deletion																
6q26	6	162,136,159	163,489,668	65	40	1.92	$9.7 \times 10^{-4}$	33	17	10	11	9	5	13	7	<i>PRKN</i>
8p22	8	15,817,196	16,092,656	59	33	2.05	$7.5 \times 10^{-4}$	24	10	18	11	6	3	11	9	<i>MSR1</i>
Duplication																
6q21	6	106,549,398	107,026,323	6	0	Inf	$5.9 \times 10^{-3}$	2	0	1	0	0	0	3	0	<i>ATG5</i>
15q11.2	15	22,652,330	23,309,294	24	7	3.88	$7.6 \times 10^{-4}$	7	3	2	1	11	0	4	3	<i>TUBGCP5</i> , <i>CYFIP1</i> , <i>NIPA1</i> , and <i>NIPA2</i>

Shown are the numbers of cases (patients with major depressive disorder) (out of 5780) and of control subjects (out of 6626) carrying each CNV with post-quality control  $p < .01$ . Start and end are genomic positions in base pairs (build hg19) either for the gene for which one or more exons was impacted by each CNV or for the region within which CNVs were counted.

Ca, cases; CMH, Cochrane-Mantel-Haenszel; CNV, copy number variant; Co, control subjects; GR1, Genetics of Recurrent Early-Onset Depression (genotyped with Affymetrix); GR2, Genetics of Recurrent Early-Onset Depression II (genotyped with Illumina); Neth, Netherlands Study of Depression and Anxiety and Netherlands Twin Register (genotyped with Affymetrix); OR, CMH odds ratio; Inf, infinite; RAD, RADIANT (genotyped with Illumina).

<sup>a</sup> $p < .01$  case-control difference.

(Supplemental Table S24) or developmental delay (Supplemental Table S25). There was no overlap between the CNVs reported in Table 3 and significant MDD GWAS loci (6,42).

## DISCUSSION

This is the largest genome-wide study to date of the association of MDD with rare CNVs. An excess of long CNVs (>100 kb) was initially reported in an analysis of the RADIANT cohort that included additional controls (43), but a subsequent reanalysis (without the extra controls and with stricter QC, producing a substantial reduction in number of CNVs per subject similar to that reported here) detected no significant excess (7). Another study of longer CNVs in 452 patients with treatment-resistant depression and 811 control subjects also reported no significant differences (8). For schizophrenia, evidence for association of several long CNVs with large effects on risk could be detected with samples comparable in size to RADIANT (44). There were no such findings for single CNVs in the present, larger study. It appears that long, multigenic CNVs are less likely to have large effects on the risk of MDD.

### Global Burden of Short Deletions

We observed enrichment of short deletions (<100 kb) in patients with MDD, and particularly intergenic deletions. This suggests that the effect on MDD risk is due to the deletion of regulatory elements, consistent with the (modest) enrichment of high-confidence enhancer regions in short deletions in patients with MDD. This is consistent with the extensive analyses

of the Psychiatric Genomics Consortium's meta-analysis of depression GWAS data (6) that detected 44 significant associations primarily in nonexonic SNPs, including several in genes that are involved with extensive regulatory networks (*RBFOX1*, *RBFOX2*, *RBFOX3*, and *CELF4*), as well as genome-wide enrichment of highly conserved regions, open chromatin in human brain and an epigenetic mark of active enhancers (H3K4me1).

One might expect an increased burden of longer CNVs as well, because they contain more genes and regulatory elements. We analyzed short and long deletions separately because longer CNVs have been more frequently implicated in disease risk. Similar ORs were observed for burden of short and of long deletions in patients with MDD, and their confidence intervals overlapped, but we had less power to detect an excess of long deletions because they were 70% less frequent than short deletions. Thus, an increased burden of longer deletions might be observed in larger meta-analyses. We also suspect that the ascertainment methods of most MDD studies are biased against individuals with long multigenic CNVs, whose carriers are at higher risk of disorders such as schizophrenia, autism, and intellectual disability. Individuals with these phenotypes have an increased risk of depression (45,46), but they are often excluded from MDD cohorts and are often not specifically diagnosed with, or treated for, depression—resulting in exclusion even from registry-based cohorts. Thus, both short and long rare deletions could impact the risk of MDD, but the current results are significant only for shorter deletions (10–100 kb), and larger cohorts will be needed to resolve the issue.

### Individual Genes and Regions

No significant or suggestive associations were detected for individual exonic CNVs or for CNV regions, after conservative correction for genome-wide testing. Larger datasets will be needed to identify true positive findings. Nominal association was observed in several regions ( $p < .01$  but not achieving suggestive or significant thresholds). The first region was 15q11.2 duplications encompassing the small, nonimprinted BP1/BP2 segment of the Prader-Willi/Angelman region. Deletions of this segment are weakly associated with risk of schizophrenia (29,37) and have been reported to be associated with dyslexia and dyscalculia (with deletions and duplications associated with reductions or increases, respectively, in the size and activity of the left fusiform gyrus) (47). Second, deletions in exons of *MSR1* (or all deletions in that region) have been implicated in atherosclerosis, Alzheimer's disease, and host defense. Third, deletions in 6q26 impact introns or exons of *PRKN*, where recessive mutations cause early-onset Parkinson's disease (type 2), but heterozygous variants are not associated with Parkinson's disease (48), although Parkinson's disease is associated with increased depressive symptoms (49). The final region included duplications in exons of, or upstream sequence near, *ATG5*, which has multiple immune functions, including negative regulation of the type I interferon production pathway—this is of note because reduced white blood cell expression of interferon I response genes was reported (18) but not replicated (50) in studies of MDD.

### Limitations

The sample size is larger than previous CNV studies of MDD but remains underpowered. Combining CNV cohorts presents challenges including differences in clinical methods (inclusion criteria, ascertainment, and assessments) and genotyping (platforms that differ in genome coverage and signal to noise ratio). Also, the present cohorts are not ideal for testing whether the long, multi-genic “neuropsychiatric” CNVs are also predisposing for depression: the psychiatric and neurological features of these CNVs may be considered exclusion criteria from MDD studies; and the associated cognitive impairments reduce the probability of being recruited into MDD cohorts because individual carriers are less likely to volunteer or to be treated in the targeted clinical settings. On the other hand, the cohorts are broadly representative of the current concept of clinically significant MDD.

In conclusion, we found significant evidence for an increased global burden of shorter rare deletions that was mainly driven by intergenic deletions in patients with MDD from four cohorts. The evidence regarding longer deletions was inconclusive: They were not significantly increased in patients with MDD, but the confidence intervals overlapped with the case-control ORs for shorter and longer deletions, and there was less power to detect a difference because longer deletions are less frequent. Overall, the results suggest that the effects of CNVs on regulatory elements, primarily in intergenic regions, play a role in predisposition to MDD.

### ACKNOWLEDGMENTS AND DISCLOSURES

GenRED and GenRED II: These projects were supported by National Institute of Mental Health (NIMH) R01 Grants MH061686 (to DFL),

MH059542 (to WH Coryell), MH075131 (to WB Lawson), MH059552 (to JBP), MH059541 (to WA Scheftner) and MH060912 (to MMW). The NIMH Cell Repository at Rutgers University and the NIMH Center for Collaborative Genetic Studies on Mental Disorders made essential contributions to this project. Genotyping was carried out by the Broad Institute Center for Genotyping and Analysis with support from Grant U54 RR020278 (which partially subsidized the genotyping of the GenRED cases). Collection and quality control analyses of the control data set were supported by grants from NIMH and the National Alliance for Research on Schizophrenia and Depression.

For the Molecular Genetics of Schizophrenia (MGS) control cohort from which GenRED-I controls were drawn: This work was supported primarily by the National Institutes of Health (Grant Nos. R01MH067257 [to NG Buccola], R01MH 059588 [to BJ Mowry], R01MH059571 [to PVG], R01MH059565 to [R Freedman], R01MH059587 [to F Amin], R01MH060870 [to WF Byerley], R01M H059566 [to DW Black], R01MH059586 [to JM Silverman], R01MH061675 [to DFL], R01MH060879 to [CR Cloninger], R01MH081800 [to PVG], U01MH046276 to [CR Cloninger], U01MH046289 [to C Kaufmann], U01MH046318 [to MT Tsuang], U01MH079469 to [PVG], and U01MH079470 to [DFL]), the Genetic Association Information Network (GAIN, for genotyping of half of the EA sample), and The Paul Michael Donovan Charitable Foundation. Genotyping was carried out by the Center for Genotyping and Analysis at the Broad Institute of Harvard and MIT (to S Gabriel and DB Mirel), supported by NIH Grant No. U54RR020278. We are grateful to Knowledge Networks (Menlo Park, CA) for assistance in collecting the control data set.

NESDA/NTR: The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) contributed to GAIN-MDD and to MDD2000. Funding was from: the Netherlands Organization for Scientific Research (MagW/ZonMW Grants 904-61-090, 985-10002, 904-61-193, 480-04-004, 400-05-717, 912-100-20; Spinozapremie 56-464-14192; Geestkracht program Grant 10-000-1002); the Center for Medical Systems Biology (NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure, VU University's Institutes for Health and Care Research and Neuroscience Campus Amsterdam, NBIC/BioAssist/ RK (2008.024); the European Science Foundation (EU/QLRT-2001-01254); the European Community's Seventh Framework Program (FP7/2007-2013); ENGAGE (HEALTH-F4-2007-201413); and the European Science Council (ERC, 230374). Genotyping was funded in part by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). CM Middeldorp was supported by the Netherlands Organization for Scientific Research (NOW-VENI grant 916-76-125).

RADIANT: This work was supported by a joint grant from the United Kingdom Medical Research Council and GlaxoSmithKline (Grant No. G0701420) and the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley National Health Service (NHS) Foundation Trust and Institute of Psychiatry, King's College London. This work presents independent research in part funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. This work was also supported by the Wellcome Trust Grant No. 086635 (JJHR); NIHR Specialist Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, King's College London (SC-W); a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme; European Commission Grant Agreement No. 115008; and Canada Research Chairs program (<http://www.chairs-chaires.gc.ca/>). The Genome Based Therapeutic Drugs for Depression study was funded by a European Commission Framework 6 grant, European Commission Contract Reference LSHB-CT-2 003-503428, and GlaxoSmithKline. Genotyping was performed at the Centre Nationale De Genotypage, Evry, Paris. We acknowledge the contribution of phase 2 of the Wellcome Trust Case Control Consortium in providing access to control datasets from the 1958 British birth cohort and the National Blood Service cohort.

We thank Stephan Sanders from the University of California at San Francisco for his assistance using CNVision. We also thank the individuals who participated in these projects and to the many clinicians who facilitated or contributed to them.

Availability of Data and Biomaterials: Biomaterials and clinical data are available from the NIMH repository (<https://nimhgenetics.org>) for the GenRED cases (the GenRED1 cohort includes the family-based linkage cohort and part of the subsequent case collection; the GenRED2 cohort includes the remainder of the case collection); for the MGS controls; and for Genomic Psychiatry Cohort controls, including the Mayo Clinic controls.

The authors report no biomedical financial interests or potential conflicts of interest.

## ARTICLE INFORMATION

From the Departments of Psychiatry and Behavioral Sciences (XZ) and Genetics (XZ, DFL), Stanford University School of Medicine, Stanford, California; Department of Biological Psychology (AA, J-JH, EJCdG, DB), Amsterdam Public Health Research Institute; Department of Psychiatry (AA), Amsterdam Universitair Medische Centra (JHS, BWJHP), Vrije Universiteit Amsterdam; and the Department of Psychiatry, Amsterdam Public Health and GGz inGeest, Amsterdam, the Netherlands; The Institute of Psychiatry (JR, SdJ), MRC Social Genetic and Developmental Psychiatry Centre (CML, PM, MRiv, GB), Department of Medical and Molecular Genetics (CML), and the National Institute for Health Research Biomedical Research Center for Mental Health (GB), King's College London, London; Department of Psychological Medicine (NC, IJ) and the Medical Research Council Centre for Neuropsychiatric Genetics and Genomics (MJO), Cardiff University, Cardiff; and the Institute of Health and Society (LJ), University of Worcester, Worcester, United Kingdom; Department of Psychiatry (JBP), Johns Hopkins University School of Medicine, Baltimore, and the Division of Cancer Epidemiology and Genetics (JSh), Biostatistics Branch, National Cancer Institute, Bethesda, Maryland; Department of Psychiatry (MMW), Columbia University, and New York State Psychiatric Institute, New York, and the Departments of Cell Biology (JAK) and Psychiatry (MPa), State University of New York Downstate Medical Center College of Medicine (CP), Brooklyn, New York; Department of Psychiatry and Behavioral Sciences (JSO), University of Southern California, Los Angeles, California; Department of Clinical Medicine (HNB), Translational Neuropsychiatry Unit and the Centre for Integrative Sequencing (HNB), Aarhus University, and the Psychosis Research Unit (OM), Aarhus University Hospital Risskov, Aarhus, Denmark; Department of Psychiatry (MPn), University Hospital of Lausanne, Prilly, Switzerland; Department of Genetic Epidemiology in Psychiatry (MRie), Central Institute of Mental Health, Mannheim, Germany; Department of Psychiatry (JPR), Washington University in Saint Louis School of Medicine, Saint Louis, Missouri; Department of Biochemistry and Molecular Biology II (MRiv), Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain; Department of Psychiatry (RU), Dalhousie University, Halifax, Nova Scotia, Canada; and the Center for Psychiatric Genetics (PVG, ARS), NorthShore University HealthSystem, and the Department of Psychiatry and Behavioral Neuroscience (PVG, ARS), University of Chicago, Chicago, Illinois.

Address correspondence to Douglas F. Levinson, M.D., Department of Psychiatry and Behavioral Sciences, Stanford University, 401 Quarry Rd., Rm. 3322, Stanford, CA 94305-5797; E-mail: [dflev@stanford.edu](mailto:dflev@stanford.edu).

Received Sep 17, 2018; revised Jan 29, 2019; accepted Feb 19, 2019.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.biopsych.2019.02.022>.

## REFERENCES

- Kessler RC, Bromet EJ (2013): The epidemiology of depression across cultures. *Annu Rev Public Health* 34:119–138.
- GBD 2015 Disease and Injury Incidence and Prevalence Collaborators (2016): Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388:1545–1602.
- Sullivan PF, Daly MJ, O'Donovan M (2012): Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nat Rev Genet* 13:537–551.
- Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, *et al.* (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45:984–994.
- Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, *et al.* (2016): Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* 48:1031–1036.
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, *et al.* (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50:668–681.
- Rucker JJ, Tansey KE, Rivera M, Pinto D, Cohen-Woods S, Uher R, *et al.* (2016): Phenotypic association analyses with copy number variation in recurrent depressive disorder. *Biol Psychiatry* 79:329–336.
- O'Dushlaine C, Ripke S, Ruderfer DM, Hamilton SP, Fava M, Iosifescu DV, *et al.* (2014): Rare copy number variation in treatment-resistant major depressive disorder. *Biol Psychiatry* 76:536–541.
- Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, *et al.* (1990): SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 47:589–593.
- Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JH, Draisma HH, *et al.* (2013): The Adult Netherlands Twin Register: Twenty-five years of survey and biological data collection. *Twin Res Hum Genet* 16:271–281.
- Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, *et al.* (2008): The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *Int J Methods Psychiatr Res* 17:121–140.
- Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, *et al.* (2008): 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* 16:335–342.
- Levinson DF, Zubenko GS, Crowe RR, DePaulo RJ, Scheftner WS, Weissman MM, *et al.* (2003): Genetics of recurrent early-onset depression (GenRED): Design and preliminary clinical characteristics of a repository sample for genetic linkage studies. *Am J Med Genet B Neuropsychiatr Genet* 119B:118–130.
- Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, *et al.* (2011): Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry* 16:193–201.
- Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, *et al.* (2008): No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: Implications for psychiatric genetics. *Am J Psychiatry* 165:497–506.
- Sanders AR, Levinson DF, Duan J, Dennis JM, Li R, Kendler KS, *et al.* (2010): The Internet-based MGS2 control sample: Self report of mental illness. *Am J Psychiatry* 167:854–865.
- Pato MT, Sobell JL, Medeiros H, Abbott C, Sklar BM, Buckley PF, *et al.* (2013): The genomic psychiatry cohort: Partners in discovery. *Am J Med Genet B Neuropsychiatr Genet* 162B:306–312.
- Mostafavi S, Battle A, Zhu X, Potash JB, Weissman MM, Shi J, *et al.* (2014): Type I interferon signaling genes in recurrent major depression: Increased expression detected by whole-blood RNA sequencing. *Mol Psychiatry* 19:1267–1274.
- Sobell JL, Heston LL, Sommer SS (1993): Novel association approach for determining the genetic predisposition to schizophrenia: Case-control resource and testing of a candidate gene. *Am J Med Genet* 48:28–35.
- Abdellaoui A, Ehli EA, Hottenga JJ, Weber Z, Mbarek H, Willemsen G, *et al.* (2015): CNV concordance in 1,097 MZ twin pairs. *Twin Res Hum Genet* 18:1–12.
- Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, *et al.* (2013): A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18:497–511.
- Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al.* (2007): PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665–1674.
- Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, *et al.* (2007): QuantiSNP: An objective Bayes hidden-Markov model to



## Rare Deletions in Major Depressive Disorder

- detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 35:2013–2025.
24. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, *et al.* (2010): Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466:368–372.
  25. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, *et al.* (2008): Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40:1253–1260.
  26. Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, *et al.* (2014): Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 19:762–773.
  27. Buizer-Voskamp JE, Muntjewerff JW, Strengman E, Sabatti C, Stefansson H, Vorstman JA, *et al.* (2011): Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. *Biol Psychiatry* 70:655–662.
  28. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, *et al.* (2011): Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70:863–885.
  29. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, *et al.* (2017): Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 49:27–35.
  30. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, *et al.* (2011): Copy number variants in schizophrenia: Confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* 168:302–316.
  31. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, *et al.* (2014): Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 46:1063–1071.
  32. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, *et al.* (2011): A copy number variation morbidity map of developmental delay. *Nat Genet* 43:838–846.
  33. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, *et al.* (2012): An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
  34. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, *et al.* (2011): Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59–65.
  35. Schwarzer G, Carpenter JR, Rücker G (2015): *Meta-Analysis with R*. Basel, Switzerland: Springer International Publishing.
  36. Lai TL, Xing H, Zhang N (2008): Stochastic segmentation models for array-based comparative genomic hybridization data analysis. *Biostatistics* 9:290–307.
  37. Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, *et al.* (2014): Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry* 204:108–114.
  38. Raychaudhuri S, Korn JM, McCarroll SA, Altshuler D, Sklar P, Purcell S, *et al.* (2010): Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet* 6:e1001097.
  39. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427.
  40. ENCODE Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.
  41. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, *et al.* (2015): Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–330.
  42. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shiralil M, *et al.* (2019): Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* 22:343–352.
  43. Rucker JJ, Breen G, Pinto D, Pedroso I, Lewis CM, Cohen-Woods S, *et al.* (2013): Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Mol Psychiatry* 18:183–189.
  44. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, *et al.* (2008): Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
  45. Uptegrove R, Marwaha S, Birchwood M (2017): Depression and schizophrenia: Cause, consequence, or trans-diagnostic issue? *Schizophr Bull* 43:240–244.
  46. Baudewijns L, Ronsse E, Verstraete V, Sabbe B, Morrens M, Bertelli MO (2018): Problem behaviours and major depressive disorder in adults with intellectual disability and autism. *Psychiatry Res* 270:769–774.
  47. Ulfarsson MO, Walters GB, Gustafsson O, Steinberg S, Silva A, Doyle OM, *et al.* (2017): 15q11.2 CNV affects cognitive, structural and functional correlates of dyslexia and dyscalculia. *Transl Psychiatry* 7:e1109.
  48. Hattori N, Mizuno Y (2017): Twenty years since the discovery of the parkin gene. *J Neural Transm (Vienna)* 124:1037–1054.
  49. Larsen JP, Dalen I, Pedersen KF, Tysnes OB (2017): The natural history of depressive symptoms in patients with incident Parkinson's disease: A prospective cohort study. *J Neurol* 264:2401–2408.
  50. Jansen R, Penninx BW, Madar V, Xia K, Milaneschi Y, Hottenga JJ, *et al.* (2016): Gene expression in major depressive disorder. *Mol Psychiatry* 21:339–347.