# Serveur Académique Lausannois SERVAL serval.unil.ch

# Author Manuscript
## Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but dos not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

serval
serveur académique lausannois

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

# Genome Wide Association Analysis of Copy Number Variation in Recurrent Depressive Disorder

**James J.H. Rucker, M.B.B.S., B.Sc., M.R.C.Psych**[1], **Gerome Breen, Ph.D.**[1,5], **Dalila Pinto, Ph.D**[2], **Inti Pedroso, B.Sc.**[1,5], **Cathryn M. Lewis, Ph.D.**[1], **Sarah Cohen-Woods, Ph.D.**[1], **Rudolf Uher, M.D., Ph.D., M.R.C.Psych.**[1], **Alexandra Schosser, M.D., Ph.D.**[1,13], **Margarita Rivera, Ph.D.**[1,16], **Katherine J. Aitchison, M.D., M.R.C.Psych., Ph.D.**[1], **Nick Craddock, F.R.C.Psych., F.Med.Sci., Ph.D.**[3], **Michael J. Owen, F.R.C.Psych., F.Med.Sci., Ph.D.**[3], **Lisa Jones, Ph.D.**[15], **Ian Jones, M.R.C.Psych., Ph.D.**[3], **Ania Korszun, M.D., Ph.D., M.R.C.Psych.**[6], **Pierandrea Muglia, M.D.**[4], **Michael R. Barnes, Ph.D.**[4], **Martin Preisig, M.D., M.P.H.**[7], **Ole Mors, M.D.**[14], **Mike Gill, F.R.C.Psych.**[8], **Wolfgang Maier, M.D., Ph.D.**[9], **John Rice, Ph.D.**[10], **Marcella Rietschel, M.D.**[11], **Florian Holsboer, M.D.**[12], **Anne E. Farmer, M.D., F.R.C.Psych.**[1,5], **Ian W. Craig, Ph.D.**[1], **Stephen W. Scherer, Ph.D**[2], and **Peter McGuffin, F.R.C.P., F.R.C.Psych., F.Med.Sci., Ph.D.**[1,5]

[1]MRC SGDP Centre, Institute of Psychiatry, King's College London, Denmark Hill, London, UK
[2]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, MaRS Centre–East Tower, 101 College Street, Toronto, Ontario, Canada
[3]MRC Centre for Neuropsychiatric Genetics and Genomics, Neuroscience and Mental Health Research Institute, Cardiff University [4]Medical Genetics GlaxoSmithKline, Verona, Italy and Greenford, UK [5]National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust and Institute of Psychiatry, King's College London, London, UK [6]Barts and The London Medical School, Queen Mary University of London, London, UK [7]Department of Adult Psychiatry, University Hospital of Lausanne, Switzerland [8]Department of Psychiatry, Trinity Centre for Health Sciences, St James's Hospital, Dublin, Ireland [9]Department of Psychiatry, University of Bonn, Bonn, Germany [10]Department of Psychiatry, Washington University, St Louis, USA [11]Central Institute of Mental Health, Mannheim, Germany [12]Max Planck Institute of Psychiatry, Munich, Germany [13]Department of Psychiatry and Psychotherapy, Medical University Vienna, Austria [14]Centre of Psychiatric Research, Aarhus

University Hospital, Risskov, Denmark [15]Department of Psychiatry, University of Birmingham, Birmingham, UK [16]CIBERSAM, University of Granada, Section of Psychiatry, Institute of Neurosciences, Biomedical Research Centre (CIBM), Granada, Spain

## Abstract

Large, rare copy number variants (CNV) have been implicated in a variety of psychiatric disorders, but the role of CNVs in recurrent depression is unclear. We performed a genome-wide analysis of large, rare CNVs in 3,106 cases of recurrent depression, 459 controls screened for lifetime-absence of psychiatric disorder and 5,619 unscreened controls from phase 2 of the Wellcome Trust Case Control Consortium (WTCCC2). We compared the frequency of cases with CNVs against the frequency observed in each control group, analysing CNVs over the whole genome, genic, intergenic, intronic and exonic regions. We found that deletion CNVs were associated with recurrent depression while duplications were not. The effect was significant when comparing cases to WTCCC2 controls ($p=7.7 \times 10^{-6}$, OR $=1.25$ (95% CI 1.13 - 1.37)) and to screened controls ($p=5.6 \times 10^{-4}$, OR$=1.52$ (95% CI 1.20 - 1.93). Further analysis showed that CNVs deleting protein coding regions were largely responsible for the association. Within an analysis of regions previously implicated in schizophrenia, we found an overall enrichment of CNVs in our cases when compared to screened controls ($p=0.019$). We observe an ordered increase of samples with deletion CNVs, with the lowest proportion seen in screened controls, the next highest in unscreened controls and the highest in cases. This may suggest that the absence of deletion CNVs, especially in genes, is associated with resilience to recurrent depression.

## INTRODUCTION

Recurrent depressive disorder is a common psychiatric disorder associated with high morbidity, high economic burden and high rates of completed suicide (1-3). Depressive disorder is heritable, with recurrent and severe types being substantially so (4, 5), but the specific genes involved are not known. Since recurrent depression is associated with increased mortality at a young age (3) it is logical to expect that some of the genetic contribution to the disorder may be explained by rare genomic variants operating under negative selection pressure (6).

Genome-wide association studies of single nucleotide polymorphisms have been inconsistent in depression (7, 8). However there is increasingly robust evidence that copy number variants (CNV), defined as submicroscopic deletions or duplications of genomic DNA seen at a frequency of less than 1 in 100 in the general population (9), are associated with a range of psychiatric disorders (10-16). A genome wide analysis of CNVs in recurrent depression has not yet been undertaken. Thus we decided to analyse Illumina microarray data for evidence of copy number variation in a large cohort of cases of recurrent depression and compare this to two control samples; one consisting of participants screened for a lifetime-absence of psychiatric disorder and another larger, unscreened population control cohort from phase 2 of the Wellcome Trust Case Control Consortium (WTCCC2).

## METHODS

### SAMPLES

3,106 cases (2,188 female & 918 male) of depression were taken from two studies of recurrent depression and a pharmacogenetic study of antidepressant response. Details of these studies have been published elsewhere (17-19) and further details are also provided in the supplementary information. All studies were approved by local ethics committees and informed written consent was gained from all participants. All cases were ascertained by interview with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (20) and fulfilled criteria for depression of at least moderate severity (2 or more episodes) under ICD-10 or DSM IV criteria. Subjects were excluded if there was a history or family history of schizophrenia or bipolar disorder or if mood symptoms were secondary to alcohol or substance misuse. Subjects with mood congruent psychotic symptoms were not excluded.

459 controls (281 female & 178 male) comprehensively screened for a lifetime-absence of psychiatric disorder were recruited. Potential subjects for the screened control group were collected from students and staff at Kings College London by internal email advertisement and by local media advertisement. Subjects were interviewed with a modified version of the Past History Schedule (21) and were included only if they demonstrated no evidence of past or present psychiatric disorder. Subjects were also interviewed with the Beck Depression Inventory (22) and excluded if they scored more than 10.

As an additional control set we also used 5,619 control samples (2,611 female & 2,668 male) from phase 2 of the Wellcome Trust Case Control Consortium (WTCCC2) composed of the 1958 British birth cohort and the national blood service cohort. The 1958 British birth cohort is a sample of sequential live births in the UK during one week in 1958, whilst the national blood service cohort is drawn from subjects who have donated blood to the UK blood services collection (further details can be found in the supplementary information).

Our control samples are of exclusively UK ethnic origin. However some of our case samples have non-UK, European heritage. Within our analyses we have attempted to account for this potential confounding factor by performing an additional analysis restricted to samples with exclusively UK origins.

### GENOTYPING

All DNA samples were obtained from venous blood. We genotyped our case sample and screened control samples on the Illumina (San Diego, CA, USA) HumanHap 610-Quad Beadchip and all DNA samples were processed contemporaneously at the same laboratory with stringent quality control and full Laboratory Information Management Systems (LIMS) monitoring. Samples from the WTCCC2 cohort were genotyped on a modified Illumina 1M beadchip which is technically very similar to the 610-Quad beadchip. We were granted access to the chip intensity data (Illumina '.idat' files). All probe intensity data was normalised and processed using standard Illumina protocols with Illumina's GenomeStudio platform to obtain the log R ratio (LRR) and B allele frequency (BAF) at each marker. The LRR and BAF represent, for each marker in each sample, a summed probe intensity ratio

derived from comparison to a canonical value calculated from all samples, and, in the case of bi-allelic probes, an allelic intensity ratio, respectively.

## CNV CALLING

All our calls and analysis are based on build 18 of the human genome reference sequence. We processed LRR and BAF values for autosomal markers in common to both Illumina arrays (n=562,680) using PennCNV (version released August 2009), a popular, open-source package designed for Illumina array data that implements a hidden Markov model, a Viterbi algorithm, expectation maximisation and takes account of the distance between consecutive markers to make copy number calls (23).

## SAMPLE AND CNV QUALITY CONTROL

Samples with a genotype call rate of less than 98%, a B allele frequency standard deviation (BAFSD) of more than 0.045 or a log R ratio standard deviation (LRRSD) of more than 0.3 were excluded from analysis. We excluded all CNVs of less than 100kb and all CNVs made with less than 10 consecutive markers. We excluded samples with more than a total of 20 CNVs, as such samples are more likely to contain artefactual calls. We excluded calls that fell within 500kb of the centromere or telomere of each chromosome and any calls falling within immunoglobulin regions. We excluded calls occurring in more than 1% of our control sample, and calls made over regions of the genome where the marker density was low (<1 marker per 200,000bp (n=134)).

As an additional level of assurance that our results were not biased by poorly performing samples clustering in particular cohorts, we removed the worst performing 10% of samples as defined by the LRRSD and BAFSD across all cohorts and re-analysed our data. This tightened our thresholds for inclusion of a sample in this analysis to a LRRSD of 0.2241, a BAFSD of 0.0390 and a genotype call rate of greater than 99%.

## STATISTICAL ANALYSIS

To test for statistical significance and effect sizes we calculated odds ratios, Pearson's chi-square or Fisher's exact statistic for the frequency of samples with CNVs in cases and controls. We estimated the proportion of genetic risk for depression contributed to by our results by performing a logistic regression analysis and calculating pseudo $R^2$. All statistical calculations were performed in STATA IC v10.1 (24) or PLINK v1.07 (25). Graphs were created using STATA IC v10.1 and Microsoft Excel (2011).

Further information on methods and analysis can be found in the supplementary information.

# RESULTS

2,723 cases of recurrent depression, 348 screened controls and 4,828 unscreened controls passed quality control and were used in our main analysis. Overall, the proportion of samples with a CNV was increased in cases when compared to screened controls and WTCCC2 controls (p=0.019, OR=1.31 (95% CI 1.04 - 1.64) and p=0.025, OR=1.11 (95%

CI 1.01 - 1.23) respectively). After stratifying by type of CNV, deletions accounted for this change in both cohorts (p=$5.6 \times 10^{-4}$, OR=1.52 (95% CI 1.20 - 1.93) and p=$7.7 \times 10^{-6}$, OR =1.25 (95% CI 1.13 - 1.37) respectively). Table 1 illustrates the frequency of samples with a CNV, stratified into deletions and duplications, in screened controls, WTCCC2 controls, and cases. Fig 1 illustrates the proportion of samples with a CNV, stratified by type of CNV, across cohorts.

We then stratified our dataset into samples containing CNVs covering gene coding (genic) regions and non-gene coding (intergenic) regions, as defined by RefSeq gene annotation coordinates available from the UCSC genome browser (http://genome.ucsc.edu/). The frequency of samples with deletion CNVs in gene coding regions was significantly increased in cases when compared to screened controls and WTCCC2 controls (p=$4.2 \times 10^{-5}$, OR 1.78 (95% CI 1.35 - 2.35) and p=$7.2 \times 10^{-6}$, OR 1.27 (95% CI 1.14 - 1.41) respectively). We further divided the samples with gene coding deletion CNVs into those with deletions interrupting exons and those with only intronic deletions. We found an increased frequency of cases with an exonic deletion CNV when compared to screened controls and WTCCC2 controls (p=$1.70 \times 10^{-5}$, OR 1.87 (95% CI 1.40 - 2.49) and p=$1.04 \times 10^{-5}$, OR 1.27 (95% CI 1.14 - 1.41) respectively). Intronic deletions were not associated (p=1.00, OR 1.00 (95% CI 0.38 - 2.60) and p=0.58, OR 0.91 (95% CI 0.66 - 1.24) respectively) however low numbers of samples with only intronic CNVs reduced our power to detect an effect. Table 2 details results from this analysis and Fig. 2 illustrates the frequency of samples with exonic and intronic deletions, across cohorts. Tables S1a and S1b in the supplementary information detail all results for all regions, whilst Fig S1a contrasts the results for deletion CNVs shown in Fig 2 with the results for duplication CNVs. Further details on absolute numbers of CNVs, number of CNVs per sample, and mean and median CNV size can be found in table S1k in the supplementary information. Within an analysis for CNVs that occur only once in the dataset (singleton CNVs) we found a broadly similar enrichment within our cases when compared to the WTCCC2 control cohort (1 sided empirical p value = 0.0054). Further details of this analysis can be found in the supplementary information.

Analyses such as ours can be affected by differences in sample quality between groups, and also by population stratification. We attempted to account for this by performing analyses, post hoc, excluding the worst performing samples across our data set and restricted to a UK-only sample set. The association between deletion CNVs occurring across the whole genome, gene coding regions and exons remains significant in both analyses. Full details of these analyses are contained in pages 6-15 the supplementary information.

Given the differences indicated in Figs 1 & 2 between the screened control cohort and the WTCCC2 control cohort we further analysed these two cohorts for significant differences, again stratifying by type of CNV made over A) genic and intergenic areas and B) exonic and intronic areas. There were no significant differences for the frequency of samples with CNVs throughout the genome between screened controls and WTCCC2 controls, although there was a trend for deletion CNVs (p=0.096, OR 1.22 (95% CI 0.97 - 1.54)). However, in the analysis stratified into genic/intergenic and exonic/intronic CNVs, significantly more deletion CNVs were seen in the WTCCC2 group than the screened control group in genic regions and exons (p=0.016, OR 1.40 (95% CI 1.06 - 1.84); p=0.0074, OR 1.47 (95% CI

1.11 - 1.95) respectively) but not in intergenic and intronic regions (p=0.79, OR 1.04 (95% CI 0.78 - 1.39) and p=1.00, OR 1.06 (95% CI 0.52 - 2.15) respectively. Again, low numbers of samples with only intronic CNVs reduced our power to detect an effect. No significant differences were seen in duplication CNV frequency. Results are presented in table S1g in the online supplementary information.

The WTCCC2 cohort is comprised of the 1958 British birth cohort and the national blood service cohort. As these represent samples drawn from different populations we performed an additional analysis comparing the frequency of samples with CNVs between these two subsets. More detail on this, details on stratification of our results by gender, and other information can be found in pages 18-22 of the supplementary information.

CNVs in various regions of the genome have previously been associated with psychiatric disorders such as schizophrenia (26). We performed an analysis across regions previously implicated in schizophrenia and a more detailed analysis of regions 1q21.1, 15q13.3 and 22q11.2 (13, 14, 16). We found that our cases were significantly enriched with CNVs in regions previously implicated in schizophrenia when compared to screened controls and when both deletions and duplications were considered together (p=0.0190, OR=3.21 (95%CI 1.07 - 9.68)). However no significant difference was observed when cases were compared to the WTCCC2 controls. We observed no significant differences in CNV frequency in the 1q21.1 region. We saw a nominally reduced frequency of cases with deletion or duplication CNVs in 15q13.3 compared to the national blood service subset of the WTCCC2 sample (p=0.044). We saw a nominally increased frequency of samples with deletion CNVs in 22q11.2 in our case sample when compared to the national blood service subset of the WTCCC2 control sample (p=0.045). Full details, including tables of frequencies for all regions and UCSC browser illustrations and individual CNV plots for regions 1q21.1, 15q13.3 and 22q11.2, can be found in pages 25-33 of the supplementary information.

## DISCUSSION

This is the largest analysis of CNVs in recurrent depression performed to date. The results of our analysis suggest that rare deletion CNVs over 100Kb in size are significantly enriched in cases of recurrent depression. Furthermore, genic and exonic deletion CNVs show the greatest enrichment in our sample. We see a continuum of increasing deletion CNV frequency throughout our cohorts with the lowest frequency in screened controls, raised in unscreened controls and highest in cases. The only other study of CNVs in depressive disorder thus far focussed on individual variants rather than results from across the genome (27).

There is some weight to the hypothesis that cases of recurrent depression will be associated with CNVs. Other studies have found that deletion CNVs, and to a lesser extent duplication CNVs, are also associated with psychiatric disorders such as autism, schizophrenia and ADHD (see, for example (11, 12, 15)). Furthermore, other research suggests that similar CNVs can predispose to a variety of neurological and psychiatric disorders (10). Other studies of CNVs in affective disorders have concentrated on bipolar disorder (28, 29). These

studies do not find an association of CNVs with cases. Recurrent depression and bipolar disorder are overlapping but distinct disorders, so both results may be correct. In general the literature on CNVs in psychiatric disorders (especially autism and schizophrenia) supports the involvement of CNVs in specific areas of the genome, although the absolute numbers at each locus are low, and large sample sizes are needed to provide statistically significant evidence of association. Within table S2a in the supplementary information we present counts of CNVs in areas of the genome previously associated with schizophrenia. Whilst we find a significant difference when the total number of cases with CNVs in these areas is compared with the total number of screened controls with CNVs in these areas, no comparisons within individual areas are significant. However some areas warrant further research. For example, we see two and three cases respectively with the microdeletion and microduplication of 16p11.2 that has previously been associated with autism and schizophrenia respectively (30, 31). 2 out of 3 cases with the 16p11.2 duplication, but neither of the cases with the deletion, also had mood congruent psychotic symptoms. One case has the large 22q11.2 microdeletion associated with numerous clinical syndromes (as well as schizophrenia) (32). Exonic deletions in neurexin 1 have also been suggested to play a role in schizophrenia (33, 34). We see two cases with microdeletions affecting neurexin 1 in our cohort and 1 case with a deletion immediately downstream of the gene. 4 WTCCC2 controls have neurexin 1 deletions, and a further 2 have deletions immediately downstream of the gene. No screened controls have deletions of neurexin 1 within or downstream of the gene. Our data provides evidence for the presence of specific CNVs in some regions of the genome previously implicated in other psychiatric disorders in a minority of our cases.

In 1967 Gottesman and Shields commented that psychiatric disorders may represent the extremes of more general continuums of normal psychosocial functioning (35). They highlighted that this implied that the genetic aetiology of psychiatric disorders was likely to be highly complex. Modern research in psychiatric genetics indicates that this model may be applicable, and its renaissance is gathering momentum (25, 36). Perhaps those with a low number of deleterious CNVs, particularly deletions, are more likely to undergo development in a manner that fosters emotionally and cognitively adaptive behaviours associated with higher resilience to developing psychiatric disorders, such as recurrent depression. Such a theory may be applicable to other complex disorders in medicine, however it does suggest that the genetic aetiology of complex diseases will involve multiple genes, each with multiple biological effects, each subject to gene-gene interactions, and the entire milieu subject to environmental modification over the lifespan of the organism. Particular patterns of genetic variation and environmental interaction may be more strongly associated with complex disease than specific genes per se.

We performed a logistic regression analysis on our data (details in methods and supplementary information) which indicated that 0.87% (pseudo $R^2$, Nagelkerke method, table S3g in the supplementary information) of the variance between RDD cases and screened controls is explained by our results. Although not explaining a large proportion, this is nonetheless a significant finding in a study of the genetic architecture of a complex disorder and underlines the genetic complexity of disorders such as recurrent depression. Only a limited number of large CNV events can be reliably detected using microarrays such as the Illumina Beadarray used in our experiment, and what we observe may extend to rare

variants in general. In the near future more detailed studies, such as those involving large scale whole genome sequencing efforts, may uncover more rare variants than are currently detectable. This information will then allow us to make a better assessment of the role of rare variants in recurrent depression and psychiatric disorders in general.

Recurrent depression represents an increasingly prominent health problem, especially in developed countries. It conspicuously lacks any clinically useful biomarkers to guide diagnosis or treatment, which often proceeds on a trial-and-error basis and frequently does not lead to sustained remission. Elucidating the genetic architecture of diseases like recurrent depression has the potential to, in collaboration with other avenues of research, inform both diagnosis and treatment and our results thus represent a significant step towards identifying the biological underpinnings of a complex and hitherto genetically obscure psychiatric disorder.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Greenberg P, Stiglin L, Finkelstein S, Berndt E. The economic burden of depression in 1990. Journal of Clinical Psychiatry. 1993; 54:405–418. [PubMed: 8270583]

2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003; 289:3095–3105. [PubMed: 12813115]

3. Harris EC, Barraclough B. Excess mortality of mental disorder. Br J Psychiatry. 1998; 173:11–53. [PubMed: 9850203]

4. Farmer A, Harris T, Redman K, Sadler S, Mahmood A, McGuffin P. Cardiff depression study. A sib-pair study of life events and familiality in major depression. Br J Psychiatry. 2000; 176:150–155. [PubMed: 10755052]

5. McGuffin P, Katz R, Watkins S, Rutherford J. A hospital-based twin register of the heritability of DSM-IV unipolar depression. Arch Gen Psychiatry. 1996; 53:129–136. [PubMed: 8629888]

6. Uher R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. Mol Psychiatry. 2009; 14:1072–1082. [PubMed: 19704409]

7. Bosker FJ, Hartman CA, Nolte IM, et al. Poor replication of candidate genes for major depressive disorder using genome-wide association data. Mol Psychiatry. 2010

8. Lewis CM, Ng MY, Butler AW, et al. Genome-wide association study of major recurrent depression in the U.K. population. Am J Psychiatry. 2010; 167:949–957. [PubMed: 20516156]

9. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006; 7:85–97. [PubMed: 16418744]

10. Cook EHJ, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. Nature. 2008; 455:919–923. [PubMed: 18923514]

11. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008; 455:232–236. [PubMed: 18668039]

12. Williams NM, Zaharieva I, Martin A, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet. 2010

13. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008; 455:237–241. [PubMed: 18668038]

14. Kirov G, Grozeva D, Norton N, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet. 2009; 18:1497–1503. [PubMed: 19181681]

15. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. Science. 2007; 316:445–449. [PubMed: 17363630]

16. Shinawi M, Schaaf CP, Bhatt SS, et al. A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. Nat Genet. 2009

17. Farmer A, Breen G, Brewster S, et al. The Depression Network (DeNT) Study: methodology and sociodemographic characteristics of the first 470 affected sibling pairs from a large multi-site linkage genetic study. BMC Psychiatry. 2004; 4:42. [PubMed: 15588307]

18. Gaysina D, Cohen S, Craddock N, et al. No association with the 5,10-methylenetetrahydrofolate reductase gene and major depressive disorder: results of the depression case control (DeCC) study and a meta-analysis. Am J Med Genet B Neuropsychiatr Genet. 2008; 147B:699–706. [PubMed: 18165972]

19. Uher R, Maier W, Hauser J, et al. Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. Br J Psychiatry. 2009; 194:252–259. [PubMed: 19252156]

20. Wing JK, Babor T, Brugha T, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry. 1990; 47:589–593. [PubMed: 2190539]

21. McGuffin P, Katz R, Aldrich J. Past and present state examination: the assessment of 'lifetime ever' psychopathology. Psychol Med. 1986; 16:461–465. [PubMed: 3726017]

22. Beck AT, Steer RA. Internal consistencies of the original and revised Beck Depression Inventory. J Clin Psychol. 1984; 40:1365–1367. [PubMed: 6511949]

23. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17:1665–1674. [PubMed: 17921354]

24. StataCorp. Stata Statistical Software (Release 10). 2007

25. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]

26. Kirov G. The role of copy number variation in schizophrenia. Expert review of neurotherapeutics. 2010; 10:25–32. [PubMed: 20021318]

27. Glessner JT, Wang K, Sleiman PM, et al. Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. PLoS One. 2010; 5:e15463. [PubMed: 21152026]

28. Grozeva D, Kirov G, Ivanov D, et al. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. Arch Gen Psychiatry. 2010; 67:318–327. [PubMed: 20368508]

29. McQuillin A, Bass N, Anjorin A, et al. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. Eur J Hum Genet. 2011

30. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009

31. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008; 358:667–675. [PubMed: 18184952]

32. Robin NH, Shprintzen RJ. Defining the clinical spectrum of deletion 22q11.2. J Pediatr. 2005; 147:90–96. [PubMed: 16027702]

33. Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophrenia bulletin. 2009; 35:851. [PubMed: 19675094]

34. Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. Human molecular genetics. 2009; 18:988. [PubMed: 18945720]

35. Gottesman II, Shields J. A polygenic theory of schizophrenia. Proc Natl Acad Sci U S A. 1967; 58:199–205. [PubMed: 5231600]

36. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. Nat Rev Genet. 2009; 10:872–878. [PubMed: 19859063]
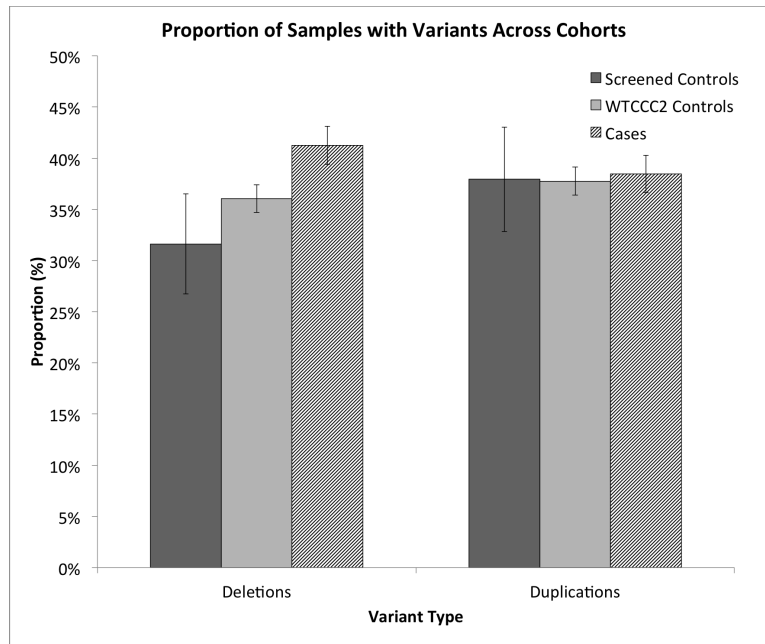
**Fig 1. Proportion of samples with deletion and duplication CNVs across cohorts.**
Error bars represent 95% confidence intervals.
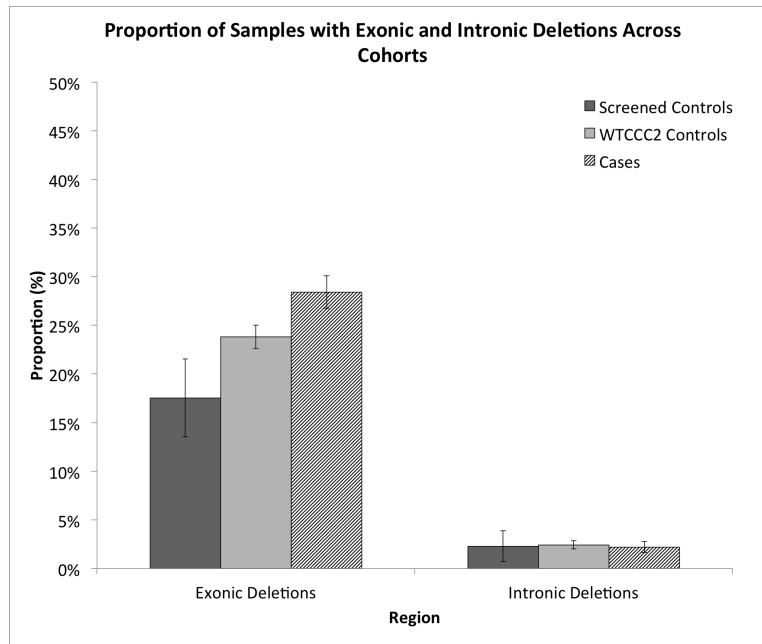
**Fig 2. Proportion of samples with exonic and intronic deletion CNVs across cohorts.**
Error bars represent 95% confidence intervals.

**Table 1**

**Frequency of samples with CNVs stratified by type, in screened controls, WTCCC2 controls and cases.**

|  | Screened Controls (N=348) | WTCCC2 Controls (N=4828) | Cases (N=2723) |
|---|---|---|---|
| **No. samples with a deletion** | 110 31.6% | 1740 36.0% | 1123 41.2% |
| **No. samples with a duplication** | 132 37.9% | 1823 37.8% | 1047 38.5% |
| **No. samples with a deletion or a duplication** | 198 56.9% | 2932 60.7% | 1725 63.3% |

**Table 2**

**Frequency of samples with gene coding deletion CNVs stratified into exonic, intronic, genic and intergenic regions, in screened controls, WTCCC2 controls and cases.**

| | Screened Controls (N=348) | WTCCC2 Controls (N=4828) | Cases (N=2723) |
|---|---|---|---|
| No. samples with an exonic deletion [*] | 61 17.5% | 1150 23.8% | 774 28.4% |
| No. samples with an intronic deletion [**] | 8 2.3% | 117 2.4% | 60 2.2% |
| No. samples with a genic deletion | 67 19.2% | 1208 25.0% | 811 29.8% |
| No. samples with an intergenic deletion | 59 17.0% | 846 17.5% | 523 19.2% |

[*] Refers to samples with deletions with any exonic coverage

[**] Refers to samples with deletions with only intronic coverage