Human Molecular Genetics, 2013, Vol. 22, No. 4 **816–824**
doi:10.1093/hmg/dds476
Advance Access published on November 11, 2012

A genome-wide study shows a limited contribution of rare copy number variants to Alzheimer's disease risk

Jade Chapman†, Elliott Rees†, Denise Harold†, Dobril Ivanov, Amy Gerrish, Rebecca Sims, Paul Hollingworth, Alexandra Stretton, GERAD1 Consortium‡, Peter Holmans, Michael J. Owen, Michael C. O'Donovan, Julie Williams* and George Kirov

MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK

Received July 4, 2012; Revised September 25, 2012; Accepted October 31, 2012

We assessed the role of rare copy number variants (CNVs) in Alzheimer's disease (AD) using intensity data from 3260 AD cases and 1290 age-matched controls from the genome-wide association study (GWAS) conducted by the Genetic and Environmental Risk for Alzheimer's disease Consortium (GERAD). We did not observe a significant excess of rare CNVs in cases, although we did identify duplications overlapping *APP* and *CR1* which may be pathogenic. We looked for an excess of CNVs in loci which have been highlighted in previous AD CNV studies, but did not replicate previous findings. Through pathway analyses, we observed suggestive evidence for biological overlap between single nucleotide polymorphisms and CNVs in AD susceptibility. We also identified that our sample of elderly controls harbours significantly fewer deletions >1 Mb than younger control sets in previous CNV studies on schizophrenia and bipolar disorder ($P = 8.9 \times 10^{-4}$ and 0.024, respectively), raising the possibility that healthy elderly individuals have a reduced rate of large deletions. Thus, in contrast to diseases such as schizophrenia, autism and attention deficit/hyperactivity disorder, CNVs do not appear to make a significant contribution to the development of AD.

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia with a prevalence of ~1% in western populations at the age of 65, rising to 25–35% in those over 85 (1). While AD is genetically complex, it is also highly heritable, with recent estimates of heritability ranging from 58 to 79% (1). Neuropathologically, the disease is characterized by extracellular senile plaques containing β-amyloid (Aβ), intracellular neurofibrillary tangles containing hyperphosphorylated tau protein and loss of synapses (2).

Mutations of the genes *APP*, *PSEN1* and *PSEN2* cause rare Mendelian forms of the disease, usually with early onset. Until recently, the only unequivocal susceptibility gene for the late-onset form of the disease was *APOE* (3). Although several candidate gene studies had previously shown suggestive evidence of association [e.g. Bertram *et al*. (4)], within the past 3 years genome-wide association studies (GWAS) of AD have identified nine genome-wide significant susceptibility loci [see Hollingworth *et al*. for a review (5)]. These are *CLU*, *PICALM*, *CR1*, *BIN1*, *MS4A*, *ABCA7*, *CD33*, *EPHA1* and *CD2AP* (6–10). It has been estimated that common variants

*To whom correspondence should be addressed at: Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, CF14 4XN, UK. Tel: +44 02920687075; Fax: +44 02920687068; Email: williamsj@cardiff.ac.uk
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint first authors.
‡Data used in the preparation of this article were obtained from the Genetic and Environmental Risk for Alzheimer's disease (GERAD1) Consortium. As such, the investigators within the GERAD1 consortia contributed to the design and implementation of GERAD1 and/or provided data but did not participate in analysis or writing of this report. A full list of GERAD1 investigators and their affiliations can be found in the Supplementary material.

within GWAS conducted by the Genetic and Environmental Risk for Alzheimer's disease (GERAD) consortium (8) account for ~24% of the estimated heritability of AD (Lee *et al*., submitted for publication) and so other sources of genetic variation that contribute to the disease remain to be identified.

Structural variation, including copy number variants (CNVs), may account for some of the unexplained heritability. A number of rare CNV loci have been implicated in brain disorders (11) and several specific CNVs have been identified that increase the risk for neurodevelopmental disorders such as schizophrenia, autism and mental retardation (12–15). To date, there have been four published genome-wide case–control association studies to assess the contribution of CNVs in late-onset AD, three of which were conducted on Caucasian samples (16–18) and one on a sample of Caribbean Hispanic origin (19). None of these studies found a global excess of CNVs in AD cases; however, Heinzen *et al*. reported a rare duplication in the schizophrenia and epilepsy risk region at 15q13.3, affecting the *CHRNA7* gene, with 2% of their cases and 0.3% of their controls having the duplication ($P = 0.053$) (16). Swaminathan *et al*. carried out a CNV analysis in participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (17), and found a significant excess of CNVs overlapping the genes *CSMD1* and *HNRNPCL1* in AD cases, but these findings were not significant after correction for multiple testing. They also identified CNVs overlapping the genes *NRXN1, ERBB4, ATXN1, HLA-DPB1, RELN, CHRFAM7A, DOPEY2* and *GSTT1* in AD cases but not in controls, although the excess in these loci was not significant. These findings were subsequently confirmed by the authors through CNV analysis of participants in the National Institute of Aging-LOAD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study (18). This study also reported an excess of CNVs overlapping the gene *IMMP2L* in 1.6% of their AD cases with no CNVs identified in their controls, although this finding was not significant (uncorrected $P = 0.059$). Furthermore, Ghani *et al*. (19) CNV study on an AD data set of Caribbean Hispanic origin found nominal association with a duplication on chromosome 15q11.2 (chr15: 20.3 Mb–20.65 Mb). Although the loci highlighted by these studies are not significant, they warrant further investigation in larger data sets. These studies are relatively small ($n < 1200$) in comparison with the more recent collaborative AD GWAS which have been conducted (6–10), and are unlikely to have the power required to identify a significant association of CNVs in AD.

We aimed to identify CNVs contributing to AD development by undertaking a powerful analysis of association using CNV data on 3260 AD cases and 1290 age-matched controls. These samples had been genotyped on Illumina 610-quad chip arrays (Illumina, Inc.) as part of a GWAS conducted by the GERAD consortium (8).

RESULTS

Unless otherwise stated, analyses were restricted to the 2690 CNVs in 3260 AD cases and 1290 controls that passed QC, were <1% frequent in each sample set, were >100 kb in length, called with at least 20 probes and were validated using *z*-score analysis.

CNV burden analyses

The rates of rare CNVs of different sizes and the corresponding *P*-values when AD cases were compared with controls are shown in Supplementary Material, Table S1. When examining the rates of all CNVs >100 kb, we observed a significant excess of deletions in the controls rather than cases (case–control ratio = 0.90, $P = 0.0332$), but this excess does not remain significant (adjusted $P = 0.30$) after Bonferroni correction for multiple testing for different size ranges and types of CNVs (deletions/duplications). No statistically significant differences between cases and controls were observed in CNVs >500 kb. Very large deletions (>1 Mb) are likely to be the most pathogenic class of CNV (20–23) and an excess of deletions >1 Mbs was observed in cases (case–control ratio = 4.19, $P = 0.023$). This excess is not significant when corrected for multiple testing (adjusted $P = 0.21$). However, we did observe that the rate of these rare larger deletions is very low in our elderly control population compared with younger control sets studied previously. For example, the International Schizophrenia Consortium (ISC) study (20) identified rare deletions >1 Mb at a rate of 1.2% in 3181 controls and a study by Grozeva *et al*. (24) identified such CNVs at a rate of 0.7% in 2806 population controls. We observed such CNVs at a rate of only 0.16% in our 1290 controls, which is a significantly lower rate than the ISC study ($P = 8.9 \times 10^{-4}$, $\chi^2 = 11.04$, 1 df) and the study by Grozeva *et al*. ($P = 0.024$, $\chi^2 = 5.1$, 1 df).

Significant loci

As the global burden analysis highlighted a significant excess of deletions >100 kb in controls compared with cases, we carried out regional analysis of the whole genome to identify any specific regions which harboured an excess of deletions in controls. We identified four regions which showed an excess of deletions in controls, but these regions did not remain significant after correction for multiple testing (corrected $P > 0.4$).

We also carried out regional analysis to identify any regions which may harbour an excess of CNVs in cases, but no regions showed significant association (uncorrected $P > 0.05$).

CNVs in loci previously highlighted by AD CNV studies

We sought to replicate the findings of the four previous AD CNV studies (16–19). As the CNVs in these studies were not filtered for frequency, we carried out this analysis on the 7718 CNVs present before filtering for 1% frequency, and all CNVs identified in these regions were validated using *z*-score analysis, like we did for the rare CNVs. As shown in Table 1, we did not replicate any of the findings of the previous studies. Out of the five loci in which we also identified CNVs, four loci had a greater rate of CNVs in our controls (*ERBB4:* case–control ratio = 0.38; *CSMD1*: case–control ratio = 0.48; *CHRNA7*: case–control ratio = 0.95; *DOPEY2*: case–control ratio = 0.65). The only previously reported

Table 1. CNVs overlapping previously reported loci in AD CNV studies

Locus	Position (b36)	Cases Number	Rate (%)	Controls Number	Rate (%)	*P*-value	Rate in other AD cases (%)	Rate (%) in other controls	Publication
HNRNPCL1	chr1:12829847-12831165	N/A	N/A	N/A	N/A	N/A	2.7	0	(17)
NRXN1	chr2:51003297-51113178	0	0	0	0	1	0.96	0	(17,18)
ERBB4	chr2:211948687-213111597	1	0.03	1	0.08	1	0.86	0	(17,18)
ATXN1	chr6:16407321-16869700	0	0	0	0	1	4.3	0	(17,18)
HLA-DPB1	chr6:33151737-33162954	0	0	0	0	1	0.21	0	(17,18)
RELN	chr7:102899472-103417198	0	0	0	0	1	0.11	0	(17,18)
IMMP2L	chr7:110090342-110989583	0	0	0	0	1	1.8	0.32	(17,18)
CSMD1	chr8:2780281-4839736	5	0.15	4	0.31	0.47	4.1	0	(17)
15q11.2	chr15:20269300-20650620	17	0.52	3	0.23	0.22	2.6	0.8	(19)
CHRFAM7A	chr15:28440735-28473156	0	0	0	0	1	1.5	0.32	(17,18)
15q13.3 (*CHRNA7*)	chr15:29800000-30300000	18	0.55	12	0.93	0.214	2.17	0.3	(16)
DOPEY2	chr21:36458708-36588442	5	0.15	3	0.23	0.69	0.21	0	(17,18)
GSTT1	chr22:22706138-22714284	0	0	0	0	1	0.21	0	(17,18)

Loci are reported if they were significantly associated in a study, or if CNVs have been identified in AD cases but not in controls. Where more than one study has implicated a locus, the rates in both studies have been combined, The CNVs in the *HNRNPCL1* locus were not called due to the large amount of segmental duplications in this region.

Table 2. CNVs in AD susceptibility/causative genes

Gene	Chr	Position (bp)	Deletions Cases	Controls	*P*-value	Duplications Cases (%)	Controls	*P*-value
APP	21	26174732–26465009	0	0	1	1 (0.03)	0	1
CR1	1	205736096–205881733	0	0	1	2 (0.06)	0	0.596

No CNVs were identified in *APOE, PSEN1, PSEN2*, *BIN1, PICALM*, *CLU,* the *MS4A* gene cluster*, EPHA1, CD33, CD2AP* or *ABCA7.* Positions are based on NCBI b36.

region in which we observed an excess of CNVs in AD cases was the 15q11.2 duplication locus identified by Ghani *et al*. (19). We found CNVs overlapping this region in 0.52% of our cases and 0.23% of our controls (case–control ratio = 2.26); however, the excess in cases was not significant (uncorrected $P = 0.22$).

CNVs in genes known to contribute to AD

We sought to identify CNVs which overlapped the known AD risk genes, regardless of their frequency. We used the data set of 7718 CNVs present before filtering for <1% frequency was carried out and all CNVs were validated using *z*-score analysis. Table 2 shows the results of the analysis of CNVs that overlapped genes known to contribute to AD. *APOE, PSEN1, PSEN2*, *BIN1, PICALM*, *CLU,* the *MS4A* gene cluster (chr11: 59.5–60.35 Mb, NCBI b36), *EPHA1, CD33* and *CD2AP* were not overlapped by any CNVs and as such are not shown in the table. In total, three rare CNVs overlapping the genes *APP* (one duplication) and *CR1* (two duplications) were identified in cases. Although no CNVs overlapping these genes were identified in controls, the higher rates of CNVs in cases were not significant ($P > 0.596$). Figure 1a shows the position of the duplication that overlaps the *APP* locus. This duplication is over 5 Mb and overlaps another 13 genes as well as the entire length of *APP*. This duplication was identified in a post-mortem sample from an individual whose exact age of onset was unknown but who was a patient in a residential nursing home by 57 years of age and who had severe dementia (25). This individual is also part of the sample analysed by McNaughton *et al*. (25). That study aimed to identify duplications at the *APP* and *PRNP* loci using a three tiered screening assay which included exonic real-time quantitative PCR, fluorescent microsatellite quantitative PCR, and Illumina arrays, resulting in the identification of an *APP* duplication in the same individual, thus confirming our CNV call. The duplications overlapping the *CR1* gene are shown in Figure 1b, and both of the individuals with these duplications have late-onset forms of AD (onset ages of 66 and 67 years).

CNVs in pathogenic CNV Loci

We analysed CNVs that were present in regions of the genome that are regarded as harbouring potentially pathogenic CNVs that increase the risk of schizophrenia, autism, mental retardation and/or epilepsy (14), as shown in Table 3. None of these regions showed a significant excess of CNVs in our AD cases, although the relatively small number of cases in our study almost precludes the detection of statistically significant differences for the very rare CNVs, in the region of 1:5000 controls (1q21.1, *NRXN1*, 15q13.3, 16p11.2 and 17p12) or even lower rates (3q29 and 22q11.2). The most common of the pathogenic CNVs (deletions at 15q11.2 and duplications at 16p13.1) were found in our cases at rates slightly lower to those in previous control populations (26), and were lower than the rates in our own controls. We identified higher rates

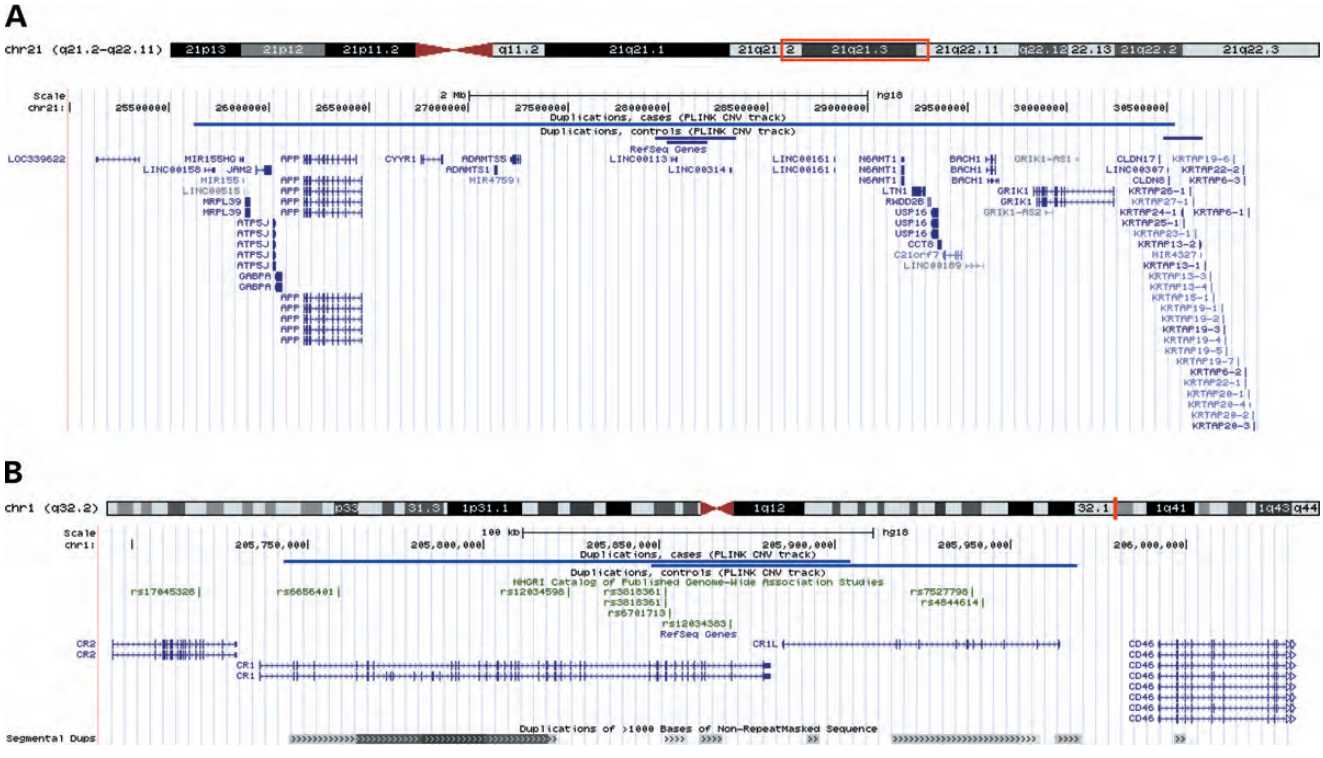

Figure 1. Positions of duplications overlapping *APP* (A) and *CR1* (B). The blue lines indicate the positions of the duplications identified which overlap *APP* (A) and *CR1* (B). No deletions or duplications were found to be overlapping this gene in the controls.

Table 3. Analysis of CNV loci regarded as pathogenic for several neurodevelopmental phenotypes

Locus	Type	Position (Mb)	Cases Number	Rate (%)	Controls Number	Rate (%)	Rate (%) in SZ cases	Rate (%) in other controls	Publication
1q21.1	Del	144.9–146.3	1	0.031	0	0	0.18	0.02	(26)
2p16.3 (*NRXN1*)	Del	50.0–51.2	0	0	0	0	0.16	0.021	(45)
3q29	Del	197.0–198.8	0	0	0	0	0.097	0.002	(26)
15q11.2	Del	20.2–20.8	7	0.215	5	0.388	0.57	0.27	(26)
15q13.3	Del	28.7–30.3	0	0	0	0	0.19	0.023	(26)
15q11-q13	Dup	21.2–26.2	0	0	0	0	0.05	0.012	(52)
16p11.2	Del	29.5–30.1	0	0	1	0.078	0.03	0.03	(13)
16p11.2	Dup	29.5–30.1	4	0.123	0	0	0.3	0.031	(26)
16p13.1	Del	15.0–16.4	0	0	0	0	0.12	0.04	(53)
16p13.1	Dup	15.0–16.4	3	0.092	2	0.155	0.28	0.13	(26)
17p12	Del	14.0–15.4	2	0.061	0	0	0.16	0.026	(26)
17q12	Del	31.8–33.3	0	0	0	0	0.06	0.003	(26)
22q11.2	Del	17.4–19.8	0	0	0	0	0.31	0	(26)

Del, deletions; dup, duplications. The columns 'Rate (%) in other controls' and 'Rate (%) in SZ cases' show the rate reported in controls and schizophrenia cases from previous reviews of these CNV loci. For simplicity, only the rates of these loci in schizophrenia cases are shown.

of deletions at the 1q21.1 locus, duplications at the 16p11.2 locus and deletions at the 17p12 locus in our cases than have been observed in previous control populations, but these rates are still far lower than those observed in schizophrenia case populations and are more similar to rates reported among controls (26).

Pathway analysis

We performed pathway analysis by testing 1833 gene sets for an excess of genes hit by CNVs in cases. We identified 113 pathways which were significantly enriched ($P < 0.05$). None of these pathways survived correction for multiple testing of pathways at a false discovery rate of less than 0.05 (all q-values were >0.15). We also tested these pathways for enrichment of gene hits in deletions and duplications separately. One hundred and seventy pathways were significantly enriched for case gene hits in deletions, but none survived multiple testing correction (q-values > 0.08). Likewise, of the 97 pathways significantly enriched for case hits in duplications, none survived multiple testing correction (q-values > 0.45).

We have previously conducted pathway analysis on the genome-wide single nucleotide polymorphism (SNP) data for these sample sets and found a number of pathways involved in cholesterol metabolism and the immune system to be significantly enriched (27). To test for overlap between the pathways identified through SNPs or CNVs, we ran the ALIGATOR algorithm (28) on the GERAD GWAS data, using more up-to-date pathway definitions than those used in (27). We restricted these analyses to pathways that were enriched for genes hit by all CNVs, deletions and duplications in cases. Of the pathways that were nominally significantly enriched ($P < 0.05$) for genes hit by deletions (>100 kb) in cases, 15 were enriched for SNP signal in GERAD at $P < 0.05$ and 6 at $P < 0.01$. These numbers are significantly larger than expected by chance ($P = 0.034$ and $P = 0.015$, respectively), suggesting some biological mechanisms for AD susceptibility acting through both SNPs and CNVs. The significant pathways are listed in Supplementary Material, Table S2, and include pathways related to lipid/cholesterol homoeostasis and cell signalling. No significant overlap was observed for pathways enriched for genes hit in cases by duplications or CNVs in general.

DISCUSSION

We conducted a large-scale study of rare CNVs in AD. Overall, there was no excess of CNVs in cases compared with controls, which is what has been observed in the other smaller genome-wide studies of CNVs in AD published so far (16–19). When the CNVs were divided into four size ranges and type (deletions and duplications), there was a significant excess of deletions >1 Mb in cases, but this finding did not remain significant after multiple testing.

This is in contrast with the findings of CNV studies of neurodevelopmental disorders. For example, the ISC (20) identified a significant excess of rare CNVs >100 kb in cases with schizophrenia (case–control ratio = 1.15) and this excess was more pronounced for deletions >500 kb (case–control ratio = 1.67). Williams *et al*. (29) also found a significant excess of CNVs >500 kb in cases with attention deficit/hyperactivity disorder (ADHD) (case–control ratio = 2.09). We observed an excess of deletions >100 kb in controls, but this was not significant after correction for multiple testing. This is a similar finding to that of Swaminathan *et al*. in their recent genome-wide study of CNVs in AD, where they observed a trend towards a reduced rate of both deletions and duplications in their AD cases (18). However, much higher CNV rates overall were observed in that study (CNVs per person = 9.3) as they did not filter the data for rare CNVs >100 kb when carrying out this analysis.

We investigated potentially interesting loci in which previous AD CNV studies identified an excess of CNVs in AD cases (16–19). Although we observed an excess of CNVs in our cases in the 15q11.2 region identified by Ghani *et al*. (19), this excess did not reach significance in our study. The rate of CNVs in this region in their AD cases of Caribbean Hispanic origin is five times what we have observed in our cases. This difference may be due to population differences in CNV rates at this locus. We identified CNVs in four other loci which had been highlighted in the two studies by Swaminathan *et al*. (17,18), but we observed a higher rate of CNVs in our controls than in our cases. This discrepancy may be due to the small control sample sizes of the previous two studies (combined $n = 339$). This present study consists of nearly four times as many controls and so has greater power to detect more CNVs in these regions.

We investigated if any known AD risk genes were intersected by CNVs. *APP* was overlapped by a duplication identified in an individual with early-onset AD, but no CNVs overlapped this gene in controls. This duplication was independently identified and validated in this same sample by another group that specifically focused on the *APP* gene (25). A number of previous studies have also identified duplications at the *APP* locus in early-onset AD cases (15,30–33). We also identified duplications of *CR1* in two individuals with late-onset AD. One of these duplications overlaps the low-copy repeat-associated CNV in *CR1*. This is particularly interesting as Brouwers *et al*. have shown that duplication of this intragenic CNV in the CR1-S isoform of the gene increases risk for AD, possibly by increasing the number of C3b/C4b-binding sites (34).

Our pathway analysis did not show a significant enrichment of any biological pathways after correction for multiple testing. However, we have shown that a number of pathways were found to be enriched in both the SNP data and in deletions, more so than would be expected by chance. This suggests that some biological mechanisms for AD susceptibility may act through both SNPs and CNVs (specifically, deletions). The signifcant pathways include lipid/cholesterol homoeostatsis, as well as cell signalling. Cholesterol homoeostatsis is of particular interest with AD as a number of AD risk genes are thought to be involved in lipid metabolism; *APOE* and *CLU* are involved in the formation and transport of lipoprotein particles, both systematically and in the brain (1,5) and *ABCA7* is involved in the release of cholesterol and phopholipids from cells to lipoprotein particles (35). *BIN1* and *PICALM* may also have roles in the internalization and transport of lipids through receptor-mediated endocytosis (1,5).

We observed a low rate of very large (>1 Mb) deletions in our elderly control population compared with younger control sets used in the CNV studies conducted by the ISC (20) and Grozeva *et al*. (24). Although different arrays were used in each of these analyses, deletions >1 Mb in size are the most reliable to call, allowing for fairly confident comparisons between studies. This raises the question of whether healthy elderly individuals have a reduced rate of large deletions, in other words, do deletions >1 Mb cause other general health problems that reduce life expectancy? This finding needs to be replicated in additional samples of elderly individuals before conclusions can be drawn.

In summary, we did not find a global excess of rare and large CNVs in AD cases and we did not replicate findings for an excess of CNVs in loci highlighted by previous AD CNV studies. Furthermore, we did not find an excess of CNVs overlapping AD candidate genes in cases, but did identify duplications overlapping *APP* and *CR1* which may be pathogenic. We have also shown potential biological overlap between the involvement of SNPs and CNVs in AD susceptibility, centred on lipid/cholesterol metabolism. We also find a

reduced rate of large, rare deletions in our elderly controls than has been observed in other control sets, raising the possibility that this class of CNVs not only increase the rate of various neurodevelopmental disorders, but might also reduce life expectancy in general. Thus, in contrast to diseases such as schizophrenia, autism and ADHD, CNVs do not appear to make a significant contribution to the development of AD.

MATERIALS AND METHODS

The sample used in this study consisted of 4112 cases and 1602 elderly screened controls. All samples were genotyped on Illumina 610-quad chip arrays as part of a GWAS of AD conducted by the GERAD consortium, as previously described by Harold *et al*. (8).

Sample ascertainment criteria

All AD cases met criteria for either probable [NINCDS-ADRDA (36), DSM-IV) or definite (CERAD (37)] AD. Controls were screened for dementia using the MMSE or ADAS-cog, and were determined to be free from dementia at neuropathological examination or had a Braak score of 2.5 or lower. Any controls with a known history of psychiatric illness were excluded. All individuals included in these analyses provided informed consent to take part in genetic association studies. Control samples were prepared in the same way as case samples at each collection centre to minimize sample differences. Participants were recruited from the UK and Ireland (2774 cases and 1165 controls) by the Medical Research Council (MRC) Genetic Resource for AD (Cardiff University; Institute of Psychiatry, London; Cambridge University; Trinity College Dublin), the Alzheimer's Research UK (ARUK) Collaboration [University of Nottingham; University of Manchester; University of Southampton; University of Bristol; Queen's University Belfast; the Oxford Project to Investigate Memory and Ageing (OPTIMA), Oxford University]; MRC PRION Unit, University College London and London and the South East Region AD project (LASER-AD), University College London. Individuals from Germany (680 cases and 137 controls) were recruited by the Competence Network of Dementia and Department of Psychiatry, University of Bonn. Individuals from the USA (658 cases and 300 controls) were recruited by the National Institute of Health (NIMH) AD Genetics Initiative and Washington University, St Louis, USA. Cases or controls described here have not been part of previous publications on CNVs.

Genotyping and CNV analysis

Genotyping was performed using Illumina 610-quad arrays at the Sanger Institute, UK. 200 ng of input DNA per sample were used and prepared for genotyping using the Illumina Infinium system following the manufacturer's protocols. The Log R Ratios (LRR) and B allele frequency (BAF) values for each sample were calculated from the signal intensity files by Illumina BeadStudio v3.2. Initially, we observed strong batch effects in the data set, manifested by large differences in the mean number of CNVs (both overall and >100 kb in size) produced on samples from different plates. These increased numbers of CNVs indicate the presence of false-positive calls in such plates. To correct the batch effects, we clustered SNPs in BeadStudio on a plate by plate basis, rather than for the whole sample before exporting the data. The exported data were then used to generate CNV calls using the PennCNV software (27 August 2009 version), applying the GC-model wave adjustment (38). This software detects CNVs using a hidden Markov model-based approach.

QC measures

Case and control samples were subject to the same stringent QC filters in order to prevent any DNA quality differences between case and controls samples affecting the results. After minimizing batch effects in the sample, four plates (consisting of 179 cases and 42 controls) appeared to be outliers from the distribution of the mean number of CNVs and were excluded. The remainder of the samples were excluded from the analyses if they had SNP call rates $<98\%$, Log R Ratio standard deviations >0.3, B allele frequency drift > 0.0002, wave factor out of the range -0.04 to 0.04 and a BAF median out of the range 0.45–0.55. We also filtered out individual samples that had >70 CNVs, which were outliers from the distribution. See Supplementary Material, Table S3 for a breakdown of the sample exclusions. This resulted in a final sample set of 3260 cases (average age at interview = 77.67, SD = 9.13; average age at onset = 72.91, SD = 8.49) and 1290 controls (average age at interview = 76.36, SD = 6.91).

The PennCNV algorithm identified a total of 165 361 CNVs. To reduce false positives, we set a relatively high CNV quality threshold: we excluded CNVs if they included fewer than 20 SNPs and if they had a low SNP density (>15 kb per SNP). We also chose to limit the analysis to CNVs >100 kb in size, as small CNVs are more difficult to call reliably. These cut-offs have been used in a number of papers (20,24,39). We excluded any CNVs if $>50\%$ of their length overlapped regions of segmental duplications, as defined by the 'Segmental Dups' track in UCSC (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=164967197&c=chr2&g=genomicSuperDups) (40,41).

We then joined any CNVs that appeared to be artificially split by the PennCNV algorithm. The rationale we used involved joining two CNVs if the length of the sequence between them was $<50\%$ of the length of the larger CNV. After QC, 7718 CNVs remained.

Consistent with previous CNV studies (20,24,39), we analysed only rare CNVs ($<1\%$) due to the problems of reliably calling more common CNVs with standard SNP arrays (42). Research on neurodevelopmental disorders is also justified in analysing only rare CNVs, as selection against pathogenic CNVs keeps them at low rates (43). Selection pressure might not be relevant in a disorder with such a late age at onset, like AD, but we decided that the technical problems associated with the calling of common CNVs still justified their exclusion. In addition, common CNVs have been shown to be in linkage disequilibrium with common SNPs, and therefore the signals would have been detected by GWAS (42). We filtered out the common CNVs present in

>1% of each sample using PLINK v1.06 (44), leaving 3593 CNVs.

To evaluate the remaining CNVs, we used a slight modification of the algorithm MeZOD reported by McCarthy *et al*. (13). This method is discussed in more detail in a recent article by Kirov *et al* (39). Briefly, the median *z*-score outlier method is a three-stage process: (i) the signal from each probe on an individual array is assigned a *z*-score based upon the distribution of all probe signals on that array (individual-wise standardization) (ii) each resulting *z*-score for each probe from (i) is assigned a new *z*-score based upon the distribution of all individual z-scores for that probe (probe-wise standardization) (iii) the median of the *z*-scores for all probes within a region of interest from (ii) is calculated and displayed as a histogram. Outlier detection is performed by visual inspection of the histogram, and outliers represent CNVs in the particular region of interest as shown in Supplementary Material, Figure S1. Software to perform this analysis and visualize the results can be obtained from: http://x004.psycm.uwcm.ac.uk/~dobril/z_scores_cnvs. We produced histograms of the *z*-scores produced by each CNV (deletions and duplications), and selected cut-offs in order to remove CNVs which did not appear to be true outliers, as shown in Supplementary Material, Figure S2. For deletions, CNVs were considered to be real if they had a *z*-score ≤ -5, and those with a *z*-score > -4 were excluded. Deletions with *z*-scores between -4 and -5 were manually inspected. Duplications were considered to be real if they had a *z*-score ≥ 4 and were excluded if they had a *z*-score <3. Duplications with *z*-scores between 3 and 4 were manually inspected. This filtering resulted in the exclusion of 317 deletions and 586 duplications, leaving 1220 deletions and 1470 duplications. The mean size of these CNVs in cases is 324.8 kb and in controls it was 308.9 kb.

Statistical analysis

To compare the global CNV burden between cases and controls, a Poisson regression was fitted on the number of CNVs per individual in cases versus controls. Ten principal components from Eigenstrat analysis of SNP data as well as country of origin (i.e. UK and Ireland, USA or Germany) were used as covariates in this analysis to control for systematic differences between centres as well as ethnic differences. Significant loci were identified using the 'segment group' function in PLINK v1.06 (44). Association analyses of these loci, and of CNVs overlapping AD associated genes and regions highlighted by previous studies of AD CNVs were also carried out using PLINK v1.06 (44). *P*-values are two-tailed, based on comparing the number of CNVs per individual cases and controls with the use of 10 000 permutations. The genomic coordinates used in this study are based on the March 2006 human genome sequence assembly (UCSC hg18, National Centre for Biotechnology Information build 36).

CNVs in previously implicated loci in other neuropsychiatric conditions were included if they spanned at least 50% of that region. For deletions in *NRXN1*, we used the criterion employed in previous reviews that found the strongest associations for deletions: >100 kb and disrupting exons (45,46).

Pathway analysis

The gene sets used in our pathway analyses came from three sources and have been previously described: (i) Gene Ontology (GO) (47), (ii) Kyoto Encyclopedia of Genes and Genomes pathways (KEGG) (ftp://ftp.genome.jp/pub/kegg/genes/organisms/hsa/hsa_pathway.list) and (iii) the canonical pathways set from the Molecular Signatures Database (MSigDB) (48). Pathways were required to contain between 3 and 1500 genes to be included in the analysis, giving a total of 10 086 pathways. We did not exclude CNVs which were only identified in controls in this analysis as we assumed incomplete penetrance of these CNVs. Gene sets were tested for an excess of genes hit by large, rare CNVs in cases by fitting the following logistic model to the combined set of case and control CNVs:

logit(pr(case))
= CNV size + Total number of annotated genes hit outside the gene set + number of genes hit in the gene set

The change in deviance was then compared between this and the model:

logit(pr(case))
= CNV size + Total number of annotated genes hit outside the gene set

A one-sided test for an excess of genes hit by case CNVs was performed. This overcomes biases relating to gene and CNV size (49). The comparison of case to control CNVs allows for the possibility of non-random CNV location unrelated to disease (i.e. CNVs being more likely to occur in certain specific locations of the genome in both cases and controls). The inclusion of the CNV size in the regression allows for the possibility of case CNVs being larger than control CNVs (and thus likely to hit more genes, regardless of function). Inclusion of the total number of genes hit outside the gene set in the regression corrects for case CNVs hitting more genes overall (regardless of function) than control CNVs. Analysis was restricted to gene sets containing at least 10 CNV hits in total (case and control combined), since pathways with a large number of gene hits are more likely to be biologically meaningful. This resulted in a total of 1833 gene sets being analysed. Correction of the enrichment *P*-values for each gene set for the multiple testing of gene sets was carried out by calculating *q*-values (50)—these are equivalent to the minimum value of the false discovery rate at which the gene set would be counted as significant (51).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. M.O. has received funding from GlaxoSmithKline plc and holds patents. J.W. has received lecture fees from Eli Lilly and company and Eisai Ltd and holds patents.

FUNDING

The work was made possible by the generous participation of the control subjects, the subjects with Alzheimer's disease, and their families. This work was supported by funding from the following organisations: the Wellcome Trust (grant number GR082604MA); Medical Research Council (grant number G0300429); Alzheimer's Research UK; Welsh Assembly Government; Alzheimer's Society; Ulster Garden Villages, Northern Ireland R&D Office; Royal College of Physicians/Dunhill Medical Trust; Mercer's Institute for Research on Ageing; Bristol Research into Alzheimer's and Care of the Elderly (BRACE); Charles Wolfson Charitable Trust; NIH (grant number PO1-AG026276, PO1-AG03991, RO1-AG16208, P50-AG05681); NIA; Barnes Jewish Foundation; Charles and Joanne Knight Alzheimer's Research Initiative of the Washington University Alzheimer's Disease Research Centre; the UCLH/UCL Biomedical Centre; Lundbeck SA; German Federal Ministry of Education and Research (BMBF): Kompetenznetz Demenzen (grant number 01GI0420); Bundesministerium für Bildung und Forschung and Competence Network Dementia (CND) Förderkennzeichen (grant number 01GI0102, 01GI0711). Funding to pay the Open Access publication charges for this article was provided by Medical Research Council – Identifying Genetic Risk for Late-onset Alzheimer's Disease: The GERAD Consortium.

REFERENCES

1. Jones, L., Harold, D. and Williams, J. (2010) Genetic evidence for the involvement of lipid metabolism in Alzheimer's disease. *Biochim. Biophys. Acta*, **1801**, 754–761.
2. Avramopoulos, D. (2009) Genetics of Alzheimer's disease: recent advances. *Genome Med.*, **1**, 34.
3. Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H., Pericak-Vance, M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., Alberts, M.J. *et al.* (1993) Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, **43**, 1467–1472.
4. Bertram, L., Lange, C., Mullin, K., Parkinson, M., Hsiao, M., Hogan, M.F., Schjeide, B.M., Hooli, B., Divito, J., Ionita, I. *et al.* (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am. J. Hum. Genet.*, **83**, 623–632.
5. Hollingworth, P., Harold, D., Jones, L., Owen, M.J. and Williams, J. (2010) Alzheimer's disease genetics: current knowledge and future challenges. *Int. J. Geriatr. Psychiatry*, **26**, 793–802.
6. Naj, A.C., Jun, G., Beecham, G.W., Wang, L.S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K. *et al.* (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.*, **43**, 436–441.
7. Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V. *et al.* (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.*, **43**, 429–435.
8. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A. *et al.* (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.*, **41**, 1088–1093.
9. Lambert, J.C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B. *et al.* (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.*, **41**, 1094–1099.
10. Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C. *et al.* (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, **303**, 1832–1840.
11. Lee, J.A. and Lupski, J.R. (2006) Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, **52**, 103–121.
12. O'Donovan, M.C., Kirov, G. and Owen, M.J. (2008) Phenotypic variations on the theme of CNVs. *Nat. Genet.*, **40**, 1392–1393.
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O. *et al.* (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.*, **41**, 1223–1227.
14. Kirov, G. (2010) The role of copy number variation in schizophrenia. *Expert Rev. Neurother.*, **10**, 25–32.
15. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M. *et al.* (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.*, **38**, 24–26.
16. Heinzen, E.L., Need, A.C., Hayden, K.M., Chiba-Falek, O., Roses, A.D., Strittmatter, W.J., Burke, J.R., Hulette, C.M., Welsh-Bohmer, K.A. and Goldstein, D.B. (2010) Genome-wide scan of copy number variation in late-onset Alzheimer's disease. *J. Alzheimers Dis.*, **19**, 69–77.
17. Swaminathan, S., Kim, S., Shen, L., Risacher, S.L., Foroud, T., Pankratz, N., Potkin, S.G., Huentelman, M.J., Craig, D.W., Weiner, M.W. *et al.* (2011) Genomic copy number analysis in Alzheimer's disease and mild cognitive impairment: an ADNI Study. *Int. J. Alzheimers Dis.*, **2011**, Article ID: 729478.
18. Swaminathan, S., Shen, L., Kim, S., Inlow, M., West, J.D., Faber, K.M., Foroud, T., Mayeux, R. and Saykin, A.J. (2012) Analysis of copy number variation in Alzheimer's disease: the NIA-LOAD/NCRAD Family Study. *Curr. Alzheimer Res.*, **9**, 801–814.
19. Ghani, M., Pinto, D., Lee, J.H., Grinberg, Y., Sato, C., Moreno, D., Scherer, S.W., Mayeux, R., St George-Hyslop, P. and Rogaeva, E. (2012) Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics. *G3 (Bethesda)*, **2**, 71–78.
20. ISC (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, **455**, 237–241.
21. Heinzen, E.L., Radtke, R.A., Urban, T.J., Cavalleri, G.L., Depondt, C., Need, A.C., Walley, N.M., Nicoletti, P., Ge, D., Catarino, C.B. *et al.* (2010) Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am. J. Hum. Genet.*, **86**, 707–718.
22. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., Craddock, N., Owen, M.J. and O'Donovan, M.C. (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.*, **18**, 1497–1503.
23. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V. *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.*, **43**, 838–846.
24. Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E. *et al.* (2010) Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch. Gen. Psychiatry*, **67**, 318–327.
25. McNaughton, D., Knight, W., Guerreiro, R., Ryan, N., Lowe, J., Poulter, M., Nicholl, D.J., Hardy, J., Revesz, T., Rossor, M. *et al.* (2010) Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series. *Neurobiol. Aging*, **44**, 426.e413–426.e421.
26. Grozeva, D., Conrad, D.F., Barnes, C.P., Hurles, M., Owen, M.J., O'Donovan, M.C., Craddock, N. and Kirov, G. (2012) Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. *Schizophr. Res.*, **135**, 1–7.
27. Jones, L., Holmans, P.A., Hamshere, M.L., Harold, D., Moskvina, V., Ivanov, D., Pocklington, A., Abraham, R., Hollingworth, P., Sims, R. *et al.* (2010) Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS ONE*, **5**, e13950.
28. Holmans, P., Green, E.K., Pahwa, J.S., Ferreira, M.A., Purcell, S.M., Sklar, P., Owen, M.J., O'Donovan, M.C. and Craddock, N. (2009) Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.*, **85**, 13–24.
29. Williams, N.M., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., Stefansson, H., Stefansson, K., Magnusson, P.,

Gudmundsson, O.O. *et al.* (2010) Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet*, **376**, 1401–1408.
30. Guyant-Marechal, L., Rovelet-Lecrux, A., Goumidi, L., Cousin, E., Hannequin, D., Raux, G., Penet, C., Ricard, S., Mace, S., Amouyel, P. *et al.* (2007) Variations in the APP gene promoter region and risk of Alzheimer disease. *Neurology*, **68**, 684–687.
31. Sleegers, K., Brouwers, N., Gijselinck, I., Theuns, J., Goossens, D., Wauters, J., Del-Favero, J., Cruts, M., van Duijn, C.M. and Van Broeckhoven, C. (2006) APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain*, **129**, 2977–2983.
32. Rovelet-Lecrux, A., Frebourg, T., Tuominen, H., Majamaa, K., Campion, D. and Remes, A.M. (2007) APP locus duplication in a Finnish family with dementia and intracerebral haemorrhage. *J. Neurol. Neurosurg. Psychiatry*, **78**, 1158–1159.
33. Kasuga, K., Shimohata, T., Nishimura, A., Shiga, A., Mizuguchi, T., Tokunaga, J., Ohno, T., Miyashita, A., Kuwano, R., Matsumoto, N. *et al.* (2009) Identification of independent APP locus duplication in Japanese patients with early-onset Alzheimer disease. *J. Neurol. Neurosurg. Psychiatry*, **80**, 1050–1052.
34. Brouwers, N., Van Cauwenberghe, C., Engelborghs, S., Lambert, J.C., Bettens, K., Le Bastard, N., Pasquier, F., Montoya, A.G., Peeters, K., Mattheijssens, M. *et al.* (2012) Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol. Psychiatry*, **17**, 223–233.
35. Abe-Dohmae, S., Ikeda, Y., Matsuo, M., Hayashi, M., Okuhira, K., Ueda, K. and Yokoyama, S. (2004) Human ABCA7 supports apolipoprotein-mediated release of cellular cholesterol and phospholipid to generate high density lipoprotein. *J. Biol. Chem.*, **279**, 604–611.
36. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. and Stadlan, E.M. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**, 939–944.
37. Mirra, S.S., Heyman, A., McKeel, D., Sumi, S.M., Crain, B.J., Brownlee, L.M., Vogel, F.S., Hughes, J.P., van Belle, G. and Berg, L. (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*, **41**, 479–486.
38. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. and Bucan, M. (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.
39. Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L. *et al.* (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry*, **17**, 142–153.
40. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. and Eichler, E.E. (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, **11**, 1005–1017.
41. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003–1007.
42. WTCCC (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **464**, 713–720.
43. Rees, E., Moskvina, V., Owen, M.J., O'Donovan, M.C. and Kirov, G. (2011) De novo rates and selection of schizophrenia-associated copy number variants. *Biol. Psychiatry*, **70**, 1109–1114.
44. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
45. Kirov, G., Rujescu, D., Ingason, A., Collier, D.A., O'Donovan, M.C. and Owen, M.J. (2009) Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr. Bull.*, **35**, 851–854.
46. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. *et al.* (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.*, **18**, 988–996.
47. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
48. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
49. Raychaudhuri, S., Korn, J.M., McCarroll, S.A., Altshuler, D., Sklar, P., Purcell, S. and Daly, M.J. (2010) Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet.*, **6**, e1001097.
50. Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
51. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57**, 289–300.
52. Ingason, A., Kirov, G., Giegling, I., Hansen, T., Isles, A.R., Jakobsen, K.D., Kristinsson, K.T., le Roux, L., Gustafsson, O., Craddock, N. *et al.* (2011) Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. *Am. J. Psychiatry*, **168**, 408–417.
53. Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O.P., Buizer-Voskamp, J.E., Strengman, E., Francks, C., Muglia, P. *et al.* (2011) Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol. Psychiatry*, **16**, 17–25.