Supplementary Material for

Schizophrenia risk from complex variation of complement component 4

Aswin Sekar, Allison R. Bialas, Heather de Rivera, Avery Davis, Timothy R. Hammond, Nolan Kamitaki, Katherine Tooley, Jessy Presumey, Matthew Baum, Vanessa Van Doren, Giulio Genovese, Samuel A. Rose, Robert E. Handsaker, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Mark J. Daly, Michael C. Carroll, Beth Stevens,
Steven A. McCarroll

Contents:

Extended Methods

Sources of DNA samples

Genomic DNA samples for the HapMap CEU population sample were obtained from Coriel Repositories (HapMap CEU plates 1 and 2). DNA samples for two groups of brain tissue donors were obtained from the Stanley Brain Resource of the Stanley Medical Research Institute (SMRI) and corresponded to the SMRI Array (SMRI-A) and SMRI Neuropathology (SMRI-N) collections. DNA samples for a third group of brain tissue donors, comprising 90 tissue donors for the NHGRI Gene and Tissue Expression Project (GTEx), were obtained from GTEx under an approved analysis proposal.

Molecular analysis of C4 structural elements (A, B, L, S)

We first measured copy number of each individual *C4* structural element (*C4A*, *C4B*, *C4L*, and *C4S*) using droplet digital PCR (ddPCR)[1]. We used the following protocol for each genomic DNA sample in the study (including the HapMap CEU samples and the brain tissue donors). First, genomic DNA was digested with *AluI* so that multiple tandem copies of *C4* would then be on separate pieces of genomic DNA. (*AluI* cuts between structural features of *C4* but not within any of the amplicons used for detection of them below.) For each genomic DNA sample, 50 ng of genomic DNA was digested in *AluI* (1 unit of enzyme in 10 µl of 1x reaction buffer, New England Biolabs) at 37°C for 1 hour. The digested DNA was then diluted two-fold with water for subsequent analyses.

To measure the precise copy number of each structural element in each genomic DNA sample, we performed digital PCR using nanoliter droplets (ddPCR), in which individual DNA molecules are dispersed into separate droplets, amplified with fluorescence detection probes (that detect with separate fluorescence colors the sequence of interest and a control, two-copy locus), and fluorescence-positive and -negative droplets of each color are then digitally counted[1]. 6.25 µl of the digested, diluted DNA from the above reaction was mixed with 1 µl of a 20x primer-probe mix (containing 18 µM of forward and reverse primers each and 5 µM of fluorescent probe) for *C4* and a reference locus

(*RPP30*) each, and 2x ddPCR Supermix for Probes (Bio-Rad Laboratories). The oligonucleotide sequences for the primers and probes used for assaying copy number of *C4A*, *C4B*, *C4L*, and *C4S* were from Wu *et al*.[2] and are listed in **Supplementary Table 1** (at the end of this document). For each sample, this reaction mixture was then emulsified into approximately 20,000 droplets in an oil/aqueous emulsion, using a microfluidic droplet generator (Bio-Rad). The droplets containing this reaction mixture were subjected to PCR using the following cycling conditions: 95°C for 10 minutes, 40 cycles of 94°C for 30 seconds and 60°C (for *C4A* and *C4L*) or 59°C (for *C4B* and *C4S*) for 1 minute, followed by 98°C for 10 minutes. After PCR, the fluorescence (both colors) in each droplet was read using a QX100 droplet reader (Bio-Rad). Data were analyzed using the QuantaSoft software (Bio-Rad), which estimates absolute concentration of DNA templates by Poisson-correcting the fraction of droplets that are positive for each amplicon (*C4* or *RPP30*). Since there are two copies of *RPP30* (the control locus) in each diploid genome, the ratio of the concentration of the *C4* amplicon to that of the reference (*RPP30*) amplicon is multiplied by two to yield the measurement of copy number of the *C4* sequence per diploid genome (**Extended Data Fig. 2b**). A key feature of these data is that the resulting measurements show a multi-modal distribution in which individual measurements are very close to integers rather than mid-integer (**Extended Data Fig. 2b**), allowing a precise integer measurement (rather than a rough estimate) of the copy number of each structural element in each genome.

The accuracy of copy number measurements from the above approach was evaluated in two ways. First, in every genome analyzed, the following relationship between the copy number of *C4* structural elements is expected to hold because any given *C4* gene is defined by its length (long or short) and its paralogous form (*A* or *B*):

$$C4A + C4B = C4L + C4S$$

Any deviation from this equality (for any sample) could flag a genotyping error for *C4A, C4B, C4L,* or *C4S*. Copy number measurements for all HapMap DNA samples and all brain donor DNA samples in this study satisfied this test in every case. In addition, copy number measurements for *C4A* and *C4B*

from ddPCR were compared to those for 89 HapMap samples previously evaluated by Fernando et al.[3] using Southern blot analysis of the same samples; our measurements agreed with those of Fernando *et al.* for 89/89 samples.

Determining copy number of the compound C4 *structural forms (AL, AS, BL, BS)*

The above analysis determines copy number of individual structural elements (*A, B, L, S*) but not of compound structural forms (AL, AS, BL, BS). Given that we know (for example) the numbers of copies of *C4S*, determining the ratio of the number of copies of *C4*AS and *C4*BS allows the copy number of these compound structural features to be readily calculated.

To determine how the known number of *C4S* copies (measured above) was composed of *C4AS* and *C4BS* copies, we first performed PCR to amplify 5.2-kilobase DNA molecules derived from *C4S* and spanning to the *C4 A/B*-defining molecular features (**Extended Data Fig. 2c**); this PCR involved a forward primer specific to *C4S* and reverse primer designed to the right of the *C4 A/B*-defining molecular features in exon 26. The reaction was performed in 50 µl and consisted of 20 ng of input genomic DNA, 10 µl of 5X Long Range Buffer (Mg2+ free) (Kapa Biosystems), 1.75 mM $MgCl_2$, 0.3 mM of each dNTP, 0.5 μ M each of forward and reverse primers, and 1.25 units of Kapa LongRange DNA Polymerase. Cycling conditions were as follows: 94°C for 2 minutes; 35 cycles of 94°C for 25 seconds, 61.2°C for 15 seconds, and 68°C for 5 minutes and 12 seconds; and 72°C for 5 minutes and 12 seconds.

The PCR product from the long-range PCR was used as input into a ddPCR assay with which we could precisely measure the ratio of *C4*AS to *C4*BS gene copies. PCR products were diluted and 1 µl of this diluted DNA was added to a ddPCR mixture containing 1 µl of a 20x primer-probe mixture of the *C4A* assay (FAM), 1 µl of a 20x primer-probe mixture of the *C4B* assay (HEX), and 10 µl of 2x ddPCR Supermix for Probes (Bio-Rad). The generation of droplets and the PCR cycling conditions were as described above for the ddPCR assays of *C4* copy number, with an annealing temperature of

60°C. After droplets were read, the ratio of *C4*AS to *C4*BS was calculated from the relative estimated concentrations of *C4A*-defining and *C4B*-defining sequences among the *C4S* amplicons. The combination of this ratio with the earlier determination of *C4S* copy number (above) allowed determination of integer copy number of *C4*AS and *C4*BS.

Once *C4A, C4B, C4L, C4S, C4*AS, and *C4*BS copy numbers are calculated by the above methods, copy number of the remaining compound structural features (*C4*BL and *C4*AL) is easily calculated by the following formulas:

Copy number (CN) of *C4*BL = (CN of *C4B*) - (CN of *C4*BS)

Copy number (CN) of *C4*AL = (CN of *C4A*) - (CN of *C4*AS)

= (CN of *C4L*) - (CN of *C4*BL)

with the redundant calculation of *C4*AL copy number (by these two formulas) providing an additional checksum on the accuracy of measurements of copy number state.

Inference of allelic contribution to copy number in diploid genomes

For a multi-allelic CNV, multiple combinations of alleles can give rise to the same diploid copy number. For example, if a sample has 4 copies of the *C4*AL gene in a diploid genome, this could be a result of any of the following potential allelic combinations: 0+4, 1+3, or 2+2. To distinguish among these possibilities, we exploited allele frequency information that is implicit in the relative frequencies of the different diploid copy-number genotypes, together with additional constraints placed by inheritance in trios, as described below. An expectation-maximization (EM) algorithm that incorporated this information was applied to each *C4* structural form (AL, AS, BL, and BS) separately. In this approach, each allelic configuration that could potentially give rise to each diploid copy number was enumerated. In certain trios only one configuration was possible under Mendelian inheritance (e.g., a trio in which father, mother, and offspring had a copy number of 0, 2, and 1, respectively). In the rest of the trios, allelic contributions were inferred using an EM algorithm with the following steps. First, probabilistic inferences of haploid copy number were made in each sample

(with an "initial condition" that all possible combinations were equally likely). These inferences were then used to estimate frequencies of each copy-number allele in the population. The likelihood of each allelic combination in each trio was then re-calculated given these allele-frequency estimates. This allowed new estimates of allele frequency, which were then used to refine likelihoods of observing each allelic combination in each trio. This EM loop was repeated until the allele frequency estimates converged. In practice, these estimates converged very quickly to estimates that had low uncertainty in 45-55 of the 55 trios in the analysis (51 for AL, 55 for AS, 45 for BL, 49 for BS). In the remaining trios, the following further approach was used. First, a reference set of haplotypes was created from the trios in which inference of copy-number alleles had been unambiguous. This core set of haplotypes was then used as a reference to phase the remaining copy number alleles onto SNP haplotypes using Beagle[4].

Imputation of C4 alleles; leave-one-out trials to estimate imputation accuracy

We imputed *C4* alleles from SNP genotypes using Beagle[5]. To estimate the accuracy of inferences using our imputation approach, we performed leave-one-out trials. A different individual was removed from the reference panel in each trial, and the rest of the reference haplotypes were used to impute, using Beagle[5], the *C4* structural form and haplogroup, with different subsets of SNPs in the extended MHC locus (chr6: 25-34 Mb): Illumina OmniExpress, Affymetrix 6.0, and Illumina Immunochip. The correlation (r^2) between the probabilistic dosage from imputation and the experimentally-determined genotypes was calculated as a metric of imputation accuracy (**Supplementary Table 2**). Note that our estimates of imputation efficacy will in many cases be lower bounds: (i) they will be exceeded by what it should be possible to do in the future (with larger reference panels derived from whole genome sequencing of many hundreds of families); and (ii) even in the current analysis, we frequently observed that SNP haplotypes that were rare or unique in our reference panel (for example, the haplotypes grouped into the "-other" categories) were more common

in the PGC cohorts and were presumably imputed with greater accuracy than a leave-one-out analysis would predict.

Post mortem human brain tissue RNA samples

Expression of *C4A* and *C4B* was measured in eight panels of *post mortem* human brain RNA samples derived from three sets of donors. The first set (five brain-region-specific panels from one set of donors) was the Stanley Medical Research Institute Array Collection. This collection consists of 525 samples from 105 individuals. Five brain regions were sampled from each donor: anterior cingulate cortex, orbital frontal cortex, parietal cortex, cerebellum, and corpus callosum. The median age of the donors was 44 (range 19-64). Of the 105 individuals, 102 were of European ancestry and used in the analysis. The median post mortem interval (PMI) was 30 hours (range 9-84). 69 donors were male and 38 were female. Age, sex and PMI were evaluated as potential covariates in all analyses but were found to have insignificant regression coefficients in all analyses described. The second set (two tissue-specific panels) was obtained from the Stanley Medical Research Institute Neuropathology Consortium and contained 120 samples from 60 individuals. Two regions were sampled from each donor: anterior cingulate cortex and cerebellum. 36 donors were male and 24 were female. The median age was 47 (range: 30-68). The median PMI was 27 hours (range: 11-62). Age, sex and PMI were evaluated as potential covariates in all analyses but were found to have insignificant regression coefficients in all analyses described. The third set consisted of 93 samples (frontal cortex) from 93 individuals sampled by the Genotype-Tissue Expression (GTEx) Consortium. 67 donors were male and 26 were female. The median age was 53 (range: 22-59). Age, sex and BMI were evaluated as potential covariates in all analyses but were found to have insignificant regression coefficients in all analyses described. Copy number of *C4* structural elements was measured using ddPCR in blood-derived genomic DNA samples from all individuals as described above.

Molecular analysis of C4A and C4B expression levels

Expression measurements were made using reverse-transcription ddPCR, in which total RNA is dispersed into thousands of nanodroplets; reverse transcription, PCR amplification, and fluorescence detection are then performed in droplets. Gene-expression measurements were normalized to the expression of a control gene (*ZNF394*) to account for variation in the amount of input RNA across samples; this gene was selected as a normalization control because in earlier brain transcriptomics data it showed uniform (low-variance) expression level across brain tissues sampled from many different individuals. In each reaction, the number of *C4A*-positive (or *C4B*-positive) and -negative droplets was counted, as well as the number of *ZNF394*-positive and -negative droplets. These numbers were then Poisson-corrected to yield an estimate of the underlying expression level, using the QuantaSoft software (Bio-Rad). We use *ZNF394* as a normalization control and therefore calculate the ratio of *C4A* (or *C4B*) to *ZNF394* expression.

For each brain donor in the two SMRI Brain Collection cohorts (each of which sampled multiple brain regions from each donor), a composite measure of expression across multiple brain regions was calculated in the following way. We started with an *i* x *j* matrix (*i* individuals and *j* brain regions) of gene-expression measurements. We then performed a median normalization of the data for each region (more formally, the expression for i^{th} individual in region *j* was re-calculated as a percentage of the median expression value across all the individuals for region *j*). To then obtain an overall summary value (across multiple brain regions) for an individual, we then calculated the median (across regions) of these median-normalized values (more formally, a median value across the *j* columns was calculated for each row). Donors for whom measurements were available for at least 3 (of the 5) brain regions were carried into downstream analysis.

Association between *C4A* (or *C4B*) expression and *C4A* (or *C4B*) copy number (**Fig. 3a, b**) was tested using a (non-parametric) Spearman correlation test. In order to evaluate the relationship of *C4*-HERV (*C4L*) copy number to *C4* expression (**Fig. 3c**), we sought to neutralize the effects of gene copy number, linkage disequilibrium, and trans-acting influences by calculating the ratio of *C4A* expression

per copy (*C4A* expression divided by *C4A* copy number) to *C4B* expression per copy (*C4B* expression divided by *C4B* copy number). Normalizing for genomic copy number of *C4A* and *C4B* allowed us to ask about effects separate from the effect (or in LD with the effect) of increased gene copy number. Normalizing expression of *C4A* to expression of *C4B* allowed cleaner analysis of cis-acting effects by controlling for trans-acting effects. (This is analogous to what is done in studies that utilize allele-specific expression, only here with two paralogous genes rather than two alleles of the same gene.). Of course, this normalization leaves open the question of whether the observed positive relationship to *C4*-HERV copy number (**Fig. 3c**) is due to increased expression of *C4A* or reduced expression of *C4B*; regression of *C4A* and *C4B* expression against copy number of these structural features (see section below) indicated that it was mostly if not entirely due to increased expression of *C4A*.

In the SMRI samples, the availability of genome-wide SNP data (together with our measurements of *C4*A, L, B, S copy number) allowed us to infer (by imputation) the complex *C4* structures present on each chromosome. To calculate the effect of each of the four common *C4* structures on expression of *C4A* (**Fig. 5b**), *C4A* expression was fit to the dosage of that structure across the SMRI *post mortem* brain samples:

$$(C4A\ \text{expression})_i = \sum_j \beta_j \times (\text{dose})_{ij} + \theta$$

where $(\text{dose})_{i,j}$ is the number of chromosomes in each diploid genome i that carry the structure j and θ is a constant (intercept).

To determine the *C4* structural genotype for each individual in the SMRI array collection, copy number data for each *C4* structural element (*C4A*, *C4B, C4L*, and *C4S*) from ddPCR were integrated together with SNP genotypes for these samples (from the Illumina Omni 2.5 SNP microarray). For each individual, the list of structural genotypes consistent with the set of copy numbers of *C4* structural elements were enumerated, based on the 15 *C4* structures that were identified in the HapMap CEU population sample (**Fig. 1c**). For example, if the copy number of *C4A*, *C4B*, *C4L*, and *C4S* were 2, 1, 2, and 1, respectively, then two structural genotypes were possible: AL/ AL-BS and AL-AL / BS. Given the large number of structural genotypes theoretically

possible (120 possible genotypes based on 15 structural haplotypes), more than 5 structural genotypes were consistent with a set of copy number data for *C4* structural elements for many individuals. In order to identify the most likely structural genotype, the backbone SNP genotype data were used to estimate the likelihood of observing each structural genotype given a set of copy number as well as SNP genotype data. A vector of genotype likelihoods (of length 120) was provided as input for phasing in Beagle (version 4). Each structural genotype that was consistent with the copy number data was encoded as equally likely, and those that were inconsistent were assigned a $\log_{10}$ likelihood of -1000 (i.e., to indicate that they are extremely unlikely). These likelihoods were then phased together with SNP genotypes to obtain posterior genotype probabilities for each possible structural genotype, for every individual. These probability estimates readily identified the most likely genotype for each individual (with a mean probability of 0.99).

To test association between gene expression and clinical diagnosis, the Mann-Whitney (non-parametric) test was used. The alternative hypothesis was specified based on the direction of effect of *C4* structural variation on gene expression and on the risk of schizophrenia – given that *C4* structural variants associating to increased risk of schizophrenia also associated to higher expression, it was hypothesized that the expression of *C4* would be higher in patients with schizophrenia compared to unaffected controls. A Mann-Whitney test was performed to assess for differences in median normalized *C4A* expression values between patients with schizophrenia and unaffected controls. In order to test whether the expression of *C4A* associated with clinical diagnosis independently of structural variation in *C4*, the *C4A* expression-per-copy values were used and a Mann-Whitney test was again performed.

Expression of *C4A* and *C4B* was also tested for association to potential confounders, including age, sex, *post mortem* interval, preservation technique, and smoking. Parametric (Pearson) as well as non-parametric (Spearman) tests of correlation were used to evaluate correlation to continuous variables (age and *post mortem* interval), and association of expression to categorical variables (sex, preservation technique, and smoking) was tested using the Mann-Whitney test.

Model for genetically predicting C4A and C4B expression

To derive a model for genetically predicting *C4A* and *C4B* expression to be used in association analysis of schizophrenia (in which we expect that numerous genomes will have lower-frequency *C4* structural haplotypes that are sparsely represented among the samples with measured expression values), we sought to predict *C4A* and *C4B* expression levels as a function of the dosage of each structural element (*C4*AL, *C4*BL, *C4*AS, *C4*BS). All median-normalized expression data from samples across the SMRI array, SMRI Neuropathology, and GTEx cohorts was used to fit

$$(C4A \text{ or } C4B \text{ expression})_i = \sum_j \beta_j \times (\text{dose})_{ij} + \theta$$

where $(\text{dose})_{i,j}$ is the number of structural elements j in sample i. From this model, samples with lower-frequency *C4* haplotypes can have expected expression values computed by summing their structural element dosages multiplied by the corresponding coefficients. Regression coefficients that were significantly different from zero were included in the prediction models. The following prediction models were generated:

C4Aexpression = (0.47 * AL) + (0.47 * AS) + (0.20 * BL)

C4Bexpression = (1.03 * BL) + (0.88 * BS)

Note that these are parameterized in internally normalized "expression units" that are not comparable between *C4A* and *C4B*, but are comparable across individuals for the same gene. These models explained 71% and 42% of inter-individual variation in measured *C4A* and *C4B* expression levels (respectively) – far more than explained by most known cis-eQTLs, but still consistent with a role for additional factors (beyond cis-acting variation at *C4*) in shaping *C4* expression levels.

Case-control genotype data from the Psychiatric Genomics Consortium (PGC)

We used data from all 40 of the European-ancestry case-control cohorts for which individual-level data could be made available by the PGC for such analyses (individual-level data from some cohorts could not be made available due to restricted level of patient consent). As described in the

PGC manuscript[6], all subjects provided written informed consent (or legal guardian consent and subject assent) with the exception of the CLOZUK sample, which obtained anonymous samples via a drug monitoring service under ethical approval and in accordance with the UK Human Tissue Act. The cohorts and array platforms used are listed in **Supplementary Table 3**. These samples are further described in ref[6] and in the individual studies referenced in **Supplementary Table 3**.

Relatedness among samples and population structure was previously analyzed by the PGC Statistical Analysis Working Group, using a set of 19,551 autosomal SNPs across all cohorts, removing one member of each pair with $\pi \hat{} > 0.2$. The first ten principal components were included as covariates in all of the association analyses (as described below). All analyses were pursued in concordance with an analysis proposal approved by the PGC Schizophrenia Working Group. All analyses of individual-level genotype data were conducted on the PGC's computer server in the Netherlands.

Quality control for SNP data

The SNPs and individuals retained for association analysis were subject to the following quality control (QC) parameters previously applied by the PGC Statistical Analysis Group and including: (i) SNP missingness < 0.05 (before sample removal); (ii) subject missingness < 0.02; (iii) autosomal heterozygosity deviation ($|F_{het}| < 0.2$); (iv) SNP missingness < 0.02 (after sample removal); difference in SNP missingness between cases and controls < 0.02; and SNP Hardy-Weinberg equilibrium ($p > 10^{-6}$ in controls or $p > 10^{-10}$ in cases).

In addition to the above parameters that were analyzed on a genome-wide scale, additional QC filters were applied to the SNP genotype data from the extended MHC locus in each of the 40 cohorts analyzed. SNPs that met the following criteria were removed: (i) those that were within the duplicated *C4* locus (chromosome 6:31939608-32014384, hg 19); (ii) SNPs whose allele frequency differed by more than 0.15 from their frequency in our HapMap CEU reference panel for imputation; and (iii)

transversion SNPs (A/T and G/C) whose minor allele frequency was greater than 0.35 (as it can be problematic to determine whether they have the same strand assignment as SNPs in the reference panel for imputation).

Imputation of C4 *structural variation, genetically predicted* C4A *expression*, *and HLA classical alleles*

Imputation of *C4* structural variation into the PGC data set was done with Beagle[5], using the HapMap CEU reference panel that we had supplemented with *C4* structural alleles. *C4* structural variation was imputed into each of the 40 cohorts in the PGC data set separately. We performed imputation using two approaches, with highly similar results: (i) a "best guess" approach in which each genome is assigned the most likely pair of *C4* structural alleles given the SNP data; and (ii) a "dosages" approach in which imputation uncertainty is advanced into subsequent stages of analysis by performing association analysis on the probabilistic "dosages" of each allele in each genome.

Our reference panel consisted of 222 haplotypes from 111 unrelated individuals, with *C4* structural variants on haplotypes with HapMap phase III SNPs (see **Fig. 2**) in the extended MHC locus (chromosome 6: 25 - 34 Mb). The encoding of *C4* structural variation in this reference panel was based on both the *C4* structure as well as its MHC haplotype background (**Fig. 2**). *C4* structures that segregated on multiple MHC SNP haplotypes were encoded as separate alleles in the reference panel – AL-AL structures were divided into two alleles, AL-AL-1 and AL-AL-2, based on which of the two MHC SNP haplotypes they segregated on; AL-BL structures into three alleles that were based on the three well-defined haplotype backgrounds and a fourth allele to represent the remaining ("other") set of rarer haplotypes; and AL-BS structures into six alleles (five of which had common haplotype backgrounds, and the sixth of which collected the other, rarer haplotypes together).

This strategy enabled independent testing of association of each common combination of *C4* structure and MHC SNP haplotype background. This strategy also allowed (i) inference of copy number of *C4* structural elements (*C4A, C4B*, *C4L*, and *C4S*) based on the *C4* alleles imputed in each

individual (e.g., an individual with *C4* alleles AL-AL-1 and AL-BL-2 has a diploid copy number of 3 for *C4A*, 1 for *C4B*, 4 for *C4L* and 0 for *C4S*); and (ii) inference of expected expression of *C4A* and *C4B* in the brain based on calculated copy number of *C4* structural elements in each individual, using the linear model (described above) that was fit to the expression data from *post mortem* brain samples.

A reference panel consisting of 9,956 haplotypes based on data collected by the Type 1 Diabetes Genetics Consortium (T1DGC)[7] was used for imputation of *HLA* classical alleles from both class I and class II genes: *HLA-A, B, C, DRB1, DQA1, DQB1, DPB1, DPA1*. This reference panel enabled imputation of *HLA* classical alleles at four-digit resolution, *HLA* amino acids, intragenic SNPs in the MHC locus, and insertions/deletions.

Testing association of C4, *SNPs*, *and* HLA *classical alleles to schizophrenia*

A mega-analysis was performed that utilized individual-level genotype data from all 40 cohorts that were analyzed from the PGC data set. Association analysis was performed in a logistic regression framework that included study indicator variables to account for cohort-specific effects and principal components to control for population stratification:

$$\log(\text{odds}_i) = \beta_j \times (\text{dose}_{i,j}) + \sum_{c=1}^{39} \beta_c \times (\text{cohort}_{i,c}) + \sum_{p=1}^{10} \beta_p \times (\text{PC}_{i,p}) + \theta$$

where $\text{dose}_{i,j}$ is the number of chromosomes in each individual, *i*, that carried a *C4* structural allele, *j*, and β_j is the additive effect per copy of the *C4* allele. 39 study indicator variables (the number of cohorts minus 1) were included, with $\text{cohort}_{i,c}$ equal to 1 if the i^{th} individual belonged to the c^{th} cohort and equal to 0 otherwise. In addition, ten principal components that associated to phenotype were included as covariates, with $\text{PC}_{i,p}$ being the p^{th} principal component for the i^{th} individual. The same framework was used for testing association to (i) individual SNPs and *HLA* classical alleles, where $\text{dose}_{i,j}$ was the dosage of the minor allele, *j*, of the SNP or *HLA* classical allele in individual *i*; (ii) copy number of *C4* structural features, where $\text{dose}_{i,j}$ was the diploid copy number of the *C4* feature in individual *i*; (iii) genetically predicted expression of *C4A* and *C4B*, where $\text{dose}_{i,j}$ was calculated from

the imputed *C4* structures according to the above formulas (see the above section, "*Model for genetically predicting C4A and C4B expression*"). To test association to *C4* conditional on rs13194504 and rs210133 (representing the other two genome-wide significant associations within the extended MHC locus), the dosages of the minor alleles of those SNPs were used as additional covariates in the model.

We tested the association of *C4* alleles to schizophrenia in multiple ways. The first test used aggregate genetic predictors (of *C4A* and *C4B* expression levels) as a composite genetic variable that combined information across many different alleles into an omnibus test; we started with this omnibus test (**Fig. 4**) in order to avoid over-fitting the genetic data to *ad hoc* combinations of *C4* alleles.

We further measured the schizophrenia association of specific C4 structures (structural forms of the C4 locus) (**Fig. 5a**). An estimate of effect size for a *C4* structure (e.g., AL-AL) was obtained across all alleles that contained that given structure (e.g., AL-AL-1 and AL-AL-2), by performing an inverse variance meta-analysis based on the effect size and standard error associated with each *C4* allele that contained the given *C4* structure. These effect size estimates were then normalized to a reference value of 1.0 for the C4 BS allele.

Immunohistochemistry (human tissue)

Fresh frozen hippocampus and frontal cortex sections were obtained from the Stanley Medical Research Institute. Stained tissues were from schizophrenia patients aged 31-43. Sections were thawed on ice and then post-fixed for one hour at 4^{o}C in 4% paraformaldehyde in PBS. Sections were then washed three times in PBS and then permeabilized in 0.2% Triton X-100 in PBS on a shaker for one hour at room temperature. Sections were then blocked in 10% BSA with 0.2% Triton X-100 in PBS for one hour at room temperature on a shaker and then transferred into a carrier solution of 5% BSA in 0.2% Triton X-100 in PBS containing the primary antibody and were left to incubate overnight at 4^{o}C. For pre-adsorption experiments, purified human C4 protein (Quidel) was pre-incubated with the C4c antibody at double the antibody concentration for 30 minutes at room temperature before being added

to the slides for overnight incubation at 4°C. The following day sections were washed three times in PBS and incubated in carrier solution with Alexa-Flour conjugated secondary antibodies (1:500) and Hoechst (1:10,000) for one hour at room temperature on a shaker. The sections were then washed three times in PBS and then incubated in 0.5% Sudan Black dissolved in 70% ethanol to eliminate autoflourescence from lipofuscin vesicles. Sections were then washed 5-7 times in PBS to remove the excess Sudan Black. Coverslips were then added to the slides using 90% glycerol in PBS as the mounting media. Slides were imaged on an Ultraview Vox Spinning Disk Confocal for images of cellular colocalization or Zeiss ELYRA PS1 structured illumination microscope (SIM) for synapse analysis. The following antibodies were used for staining; anti-C4c (Quidel, A211, 1:1000), anti-NeuN (Abcam, AB104225, 1:500), anti-Vglut1 (Millipore, AB5905, 1:1000), anti-Vglut2 (Millipore, AB2251, 1:2000), and anti-PSD95 (Invitrogen, 51-6900, 1:200).

We performed IHC in brain tissue slices from 5 individuals affected with schizophrenia and 2 unaffected individuals. These were selected from the same brains as the RNA experiments (SMRI Neuropathology Consortium). Across different donors we observed variable intensity of staining (down to almost no staining) but did not observe qualitatively different patterns. The level of RNA expression of *C4* (in the corresponding RNA sample from the same donor) predicted the level of IHC staining – in tissue from donors with higher *C4* RNA expression, the IHC staining was also stronger; in tissue from donors with little-to-no *C4* RNA detected, we also observed little-to-no IHC staining. The images in **Fig. 6** are from tissue from one of the individuals affected with schizophrenia.

Immunocytochemistry

Primary human cortical neurons were obtained from Sciencell Research Laboratories (catalog no. 1520). The neurons were characterized by Sciencell to be immunopositive for MAP2, neurafilament, and beta-tubulin III; are guaranteed to be negative for HIV-1, HBV, HCV, mycoplasma, bacteria, yeast, and fungi; and are not listed as a commonly misidentified cell line by ICLAC. Human cortical neurons were cultured *in vitro* on PLL-coated coverslips in neuronal media

for up to 48 days. Coverslips were fixed with 4% paraformaldehyde at room temperature for 7 minutes. Non-specific binding sites were blocked with 5% BSA for 1 hour in PBST (0.1% Tween 20) followed by 4°C overnight incubation with primary antibodies anti-MAP2 (EMD-Millipore, rabbit polyclonal, 1:10,000), anti-200 kD Neurofilament (Abcam, chicken polyclonal, 1:100,000), anti-Synaptotagmin (Synaptic Systems, rabbit polyclonal, 1:500), anti-PSD95 (Abcam, goat polyclonal, 1:500), and/or anti-C4c (Quidel, mouse monoclonal, 1:200). Coverslips were then washed with PBST and incubated for 1 hour at room temperature with secondary antibodies (Abcam, donkey or goat, 1:1000 in 5% BSA-PBST). Coverslips were mounted on slides using Vectashield with DAPI and visualized by fluorescent microscope (Zeiss Confocal).

Western blot analysis

Conditioned media was collected from *in-vitro* cultured human neurons at days 7 and 30 and frozen at -80°C until quantification of C4 by western blot. Equal amounts of proteins (20 ug as determined by BCA Protein Assay) were diluted 1:1 with Native Sample Buffer (BioRad 161-0738) and separated on a 4-15% TGX precast polyacrylamide gel. Purified human C4 protein from Quidel (A402) was used as a positive control. Unconditioned neuronal media (Sciencell 1521) provided an appropriate negative control. Electrophoresis was performed using the Mini-PROTEAN Tetra Cell (BioRad). Proteins were then transferred onto polyvinylidene difluoride membranes (Immun-Blot PVDF, BioRad 162-0177) for Western Blot analysis. Membranes were blocked in a 5% milk solution in TBST (0.1% Tween 20) for 1 hour at room temperature and then incubated with anti-C4c (Dako, F016902-2, 1:1000) primary antibody overnight at 4°C. Following washes in TBST, secondary antibody goat-anti-rabbit HRP (Abcam, preadsorbed, 1:10,000) was hybridized for 1 hour at room temperature. Membranes were washed in TBST again and then reactivity was revealed by chemiluminescence reaction performed with ECL detection reagents (BioRad Clarity) and film exposure.

Mice

The generation of the *C4*-/- mice that were used to investigate synapse elimination in the retinogeniculate system is described in detail in earlier work[8]. In these mice, the sequence spanning part of exon 23 through exon 29 has been replaced with a PGK-Neo gene. Experiments involved litters created by crossing *C4*+/- heterozygous parents, so that all comparisons were among littermates of different *C4* genotypes. Sample sizes were determined based on power calculations for each data set (to obtain > 80% statistical power) and based on recommendations from IACUC to conserve animals. Mice from both sexes were analyzed in these experiments. Experiments were approved by the institutional animal use and care committee in accordance with NIH guidelines for the humane treatment of animals.

Analysis of dorsal lateral geniculate nucleus (dLGN)

Visualization and analysis of RGC synaptic inputs in the mouse dLGN was performed as described[9]. Cholera toxin- β subunit (CTB) conjugated to Alexa 488 (green label) and CTB conjugated to Alexa 594 (red label) were intraocularly injected into the left and right eyes, respectively, of P9 mice, which were sacrificed the following day. Images were acquired using the Zeiss Axiocam and quantified blind to experimental conditions and compared to age-matched littermate controls. The degree of left and right eye axon overlap in dLGN was quantified using an *R*-value analysis as described[9] and by quantifying the percent overlap as previously described[10]. Pseudocolored images representing the *R*-value distribution were generated in ImageJ.

For measurement of *C4* expression in the retinal ganglion cells (RGCs) and LGN, RNA was isolated from tissue with the Qiagen RNeasy Lipid mini kit (cat. No 74804) with optional DNase digestion according to the manufacturer's protocol. RGCs were isolated, lysed, and DNase digested with Ambion Cells to Ct kit[10]. 15ng of RNA was used as the input for the RT-ddPCR reaction with the primer-probe sets listed in **Supplementary Table 1**.

Measurement of C4 expression in mouse tissues and cell populations

Retinal ganglion cells were purified from p5 and p15 C57BL/6 mice through serial immunopanning as previously described[11]. To specifically isolate the lateral geniculate nucleus (LGN) from P5 C57BL/6 mice, LGN was first fluorescently labeled through bi-lateral intraorbital injection of flourophore-conjugated cholera toxin at P4 and then microdissected at P5 during visualization with a florescence dissecting microscope. Retinal tissue was harvested from separate P5 C57Bl6 mice. RNA was isolated from LGN and retinal tissue with the Qiagen RNeasy Lipid mini kit (cat. No 74804) with optional DNase digestion according to the manufacturer's protocol. RGCs were lysed, DNase digested with Ambion Cells to Ct kit, and RNA from the cell-free solution used in subsequent reactions. Mouse *C4* expression was calculated as the average of two *C4*-specific reverse transcription-ddPCR assays, one with the primer-probe set spanning the junction of exons 23 and 24 and the other, the junction of exons 25 and 26, each normalized to the housekeeping mRNA, *Eif4h*.

Immunohistochemistry (mouse tissue)

Brains were harvested from mice after transcardial perfusion with 4% paraformaldehyde (PFA). Tissue was then immersed in 4% PFA for 2 hours following perfusion, cryoprotected in 30% sucrose, and embedded in a 2:1 mixture of OCT:20% sucrose PBS. Tissue was cryosectioned (12-14 microns), sections were dried, washed three times in PBS, and blocked with 2% BSA+ 0.2% Triton X in PBS for 1 hr. Primary antibodies were diluted in antibody buffer (+ 0.05% triton + 0.5% BSA) as follows: anti-C3 (Cappel,1:300), anti-vglut2 (Millipore,1:2000) and incubated overnight at 4°C. Secondary Alexa-conjugated antibodies (Invitrogen) were added at 1:200 in antibody buffer for 2 hours at room temperature. Slides were mounted in Vectashield (+DAPI) and imaged using the Zeiss Axiocam, Zeiss LSM700. In addition to the analysis of C3 localization, we also tested several commercial antibodies for mouse C4 and found that none were sufficiently specific.

Retinal cell counts

Retinal flat mounts were prepared by dissecting out retinas whole from the eyecup and placing four cuts along the major axis, radial to the optic nerve. Each retina was stained with DAPI (Vector Laboratories, Burlingame, CA) to reveal cell nuclei. Measurements of RGC density based on Brn3a (goat anti-Brn3a, 1:200, Santa Cruz) immunohistochemistry were carried out blind to genotype from matched locations in the central and peripheral retina for all four retinal quadrants of each retina. Quantification was done on P10 retinas, which is the age at which eye specific segregation analysis was completed. For each retina (1 retina per animal; N = 4 mice per treatment condition or genotype), 12 images of peripheral retina and 8 images of central retina were collected. For each field of view collected (20 per retina), Macbiophotonics ImageJ software (NIH) was used to quantify the total number of Brn3a-positive cells using the cell counter plugin. All analyses were performed blind to genotype.

Supplementary Table 1

Primer and probe sequences used.

All sequences are provided in the 5' to 3' orientation. Assays identified with an asterisk (*) were based on Wu *et al.*[2]

Assay	Forward Primer	Reverse Primer	Probe
Copy number of human *C4A**	CCTTTGTGTTGAAGGTCCTGAGTT	TCCTGTCTAACACTGGACAGGGGT	VIC-CCAGGAGCAGGTAGGAGGCTCGC-MGB
Copy number of human *C4B**	TGCAGGAGACATCTAACTGGCTTCT	CATGCTCCTATGTATCACTGGAGAGA	VIC-AGCAGGCTGACGGC-MGB
Copy number of human *C4L**	TTGCTCGTTCTGCTCATTCCTT	GTTGAGGCTGGTCCCCAACA	VIC-CTCCTCCAGTGGACATG-MGB
Copy number of human *C4S**	TTGCTCGTTCTGCTCATTCCTT	GGCGCAGGCTGCTGTATT	VIC-CTCCTCCAGTGGACATG-MGB
Control for copy number assays of human DNA (*RPP30*)	GATTTGGACCTGCGAGCG	GCGGCTGTCTCCACAAGT	FAM-CTGACCTGAAGGCTCT-MGB
Expression of human *C4A*	CCTGAGAAACTGCAGGAGACAT	GTGAGTGCCACAGTCTCATCAT	FAM-CAGGACCCCTGTCCAGTGTTAGAC
Expression of human *C4B*	CCTGAGAAACTGCAGGAGACAT	GTGAGTGCCACAGTCTCATCAT	FAM-CTATGTATCACTGGAGAGAGGTCCTGGAAC
Expression of mouse *C4* (junction of exons 23 and 24)	GACAAGGCACCTTCAGAACC	CAGCAGCTTAGTCAGGGTTACA	FAM-CATAGTCTCCAAGGGTTCCGAGGC
Expression of mouse *C4* (junction of exons 25 and 26)	AAACTCAGAATCTTCAGCACGA	GGAAGAATGATGGCTCCTTTG	FAM-GCTGCTGTCCCGGTGTAACCAA
Control for expression assays of human RNA (*ZNF394*)	CATGTGGAAACTTTGCTTGC	CCTTGTTCTATGTCAGCACATCC	HEX-TTGTTCCCGTGTTCCTCACTGTCA
Control for expression assays of mouse RNA (*Eif4h*)	TGCAGCTTGCTTGGTAGC	GTAAATTGCCGAGACCTTGC	VIC-AGCCTACCCCTTGGCTCGGG
Control for expression assays of mouse RNA (*Hs2st1*)	CCCCTGATAGTCACACAGTCC	TGGAGTTTTGAGGGTTTTGG	Hex-TCCGCTGCTGCTCTGGCCTCCT
Amplifying human *C4S* Copies	TCAGCATGTACAGACAGGAATACA	GAGTGCCACAGTCTCATCATTG	

Supplementary Table 2

Imputation of *C4* structural alleles from SNP data.

The correlation (r^2) between experimentally derived genotypes of *C4* structural alleles and imputed probabilistic dosages from leave-one-out trials within the reference panel are shown, together with a 95% confidence interval for each estimate (see **Methods**). Imputation of *C4* structural alleles was tested using SNPs within the extended MHC locus (chr 6: 25-34 Mb) from the indicated SNP microarrays. 95% confidence intervals around the Pearson r^2 value are shown in parentheses. Our HapMap-based reference panel included 7,751 SNPs, of which 2,259 to 5,523 were present on the SNP arrays evaluated.

	SNP array platform (SNPs in common with MHC reference panel)		
***C4* allele**	**Illumina Omni Express** (5,523 SNPs)	**Illumina Immunochip** (3,703 SNPs)	**Affymetrix SNP 6.0** (2,259 SNPs)
BS	0.85 (0.80-0.90)	0.86 (0.81-0.91)	0.92 (0.89-0.95)
AL-BS-1	0.55 (0.43-0.67)	0.78 (0.71-0.85)	0.55 (0.43-0.67)
AL-BS-2	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.88 (0.84-0.92)
AL-BS-3	0.84 (0.79-0.89)	0.74 (0.66-0.82)	0.67 (0.57-0.77)
AL-BS-4	0.88 (0.84-0.92)	0.83 (0.77-0.89)	0.90 (0.87-0.93)
AL-BS-5	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.98 (0.97-0.99)
AL-BL-1	0.71 (0.62-0.8)	0.71 (0.62-0.8)	0.57 (0.45-0.69)
AL-BL-2	0.63 (0.52-0.74)	0.50 (0.37-0.63)	0.63 (0.52-0.74)
AL-BL-3	0.77 (0.7-0.84)	0.72 (0.63-0.81)	0.67 (0.57-0.77)
AL-AL-1	0.54 (0.42-0.66)	0.58 (0.46-0.70)	0.65 (0.55-0.75)
AL-AL-2	0.8 (0.73-0.87)	0.8 (0.73-0.87)	0.69 (0.60-0.78)

Supplementary Table 3

Psychiatric Genomics Consortium cohorts contributing to association analysis in this study.

Cohort name	PMID	Site	Genotyping array	Cases	Controls
scz_aarh_eur	19571808	Denmark	Illumina 650K	876	871
scz_aber_eur	19571811	Aberdeen, UK	Affymetrix 6.0	719	697
scz_ajsz_eur	24253340	Israel	Illumina 1M	894	1594
scz_asrb_eur	21034186	Australia	Illumina 650K	456	287
scz_boco_eur	19571808	Bonn/Mannheim, Germany	Illumina 550K	1773	2161
scz_buls_eur		Bulgaria	Affymetrix 6.0	195	608
scz_cati_eur	18347602	US (CATIE)	Affymetrix 500K	397	203
scz_caws_eur	19571811	Cardiff, UK	Affymetrix 500K	396	284
scz_cims_eur		Boston, US (CIDAR)	Illumina OmniExpress	67	65
scz_clm2_eur	22614287	UK (CLOZUK)	Illumina 1M	3426	4085
scz_clo3_eur	22614287	UK (CLOZUK)	Illumina OmniExpress	2105	1975
scz_cou3_eur	21850710	Cardiff, UK (CogUK)	Illumina OmniExpress	530	678
scz_denm_eur	19571808	Denmark	Illumina 650K	471	456
scz_dubl_eur	19571811	Ireland	Affymetrix 6.0	264	839
scz_edin_eur	19571811	Edinburgh, UK	Affymetrix 6.0	367	284
scz_egcu_eur	15133739	Estonia (EGCUT)	Illumina OmniExpress	234	1152
scz_ersw_eur	19571808	Sweden (Hubin)	Illumina OmniExpress	265	319
scz_fi3m_eur	19571808	Finland	Illumina 317K	186	929
scz_fii6_eur		Finnish	Illumina 550K	360	1082
scz_gras_eur	20819981	Germany (GRAS)	Affymetrix Axiom	1067	1169
scz_irwt_eur	22883433	Ireland (WTCCC2)	Affymetrix 6.0	1291	1006
scz_lacw_eur	22885689	Six countries, WTCCC controls	Illumina 550K	157	245
scz_lie2_eur	11381111	NIMH CBDB	Illumina Omni 2.5M	133	269
scz_lie5_eur	11381111	NIMH CBDB	Illumina 550K	497	389
scz_mgs2_eur	19571809	US, Australia (MGS)	Affymetrix 6.0	2638	2482
scz_msaf_eur	20489179	New York, US & Israel	Affymetrix 6.0	325	139
scz_munc_eur	19571808	Munich, Germany	Illumina 317K	421	312
scz_pewb_eur	23871474	Seven countries (PEIC, WTCCC2)	Illumina 1M	574	1812
scz_pews_eur	23871474	Spain (PEIC, WTCCC2)	Illumina 1M	150	236
scz_port_eur	19571811	Portugal	Affymetrix 6.0	346	215
scz_s234_eur	23974872	Sweden (sw234)	Affymetrix 6.0	1980	2274
scz_swe1_eur	23974872	Sweden (sw1)	Affymetrix 5.0	215	210
scz_swe5_eur	23974872	Sweden (sw5)	Illumina OmniExpress	1764	2581

scz_swe6_eur	23974872	Sweden (sw6)	Illumina OmniExpress	975	1145
scz_top8_eur	19571808	Norway (TOP)	Affymetrix 6.0	377	403
scz_ucla_eur	19571808	Netherlands	Illumina 550K	700	607
scz_uclo_eur	19571811	London, UK	Affymetrix 6.0	509	485
scz_umeb_eur		Umeå, Sweden	Illumina OmniExpress	341	577
scz_umes_eur		Umeå, Sweden	Illumina OmniExpress	193	704
scz_zhh1_eur	17522711	New York, US	Affymetrix 500K	190	190

Schizophrenia Working Group of the Psychiatric Genomics Consortium

Stephan Ripke[1,2], Benjamin M. Neale[1,2,3,4], Aiden Corvin[5], James T. R. Walters[6], Kai-How Farh[1], Peter A. Holmans[6,7], Phil Lee[1,2,4], Brendan Bulik-Sullivan[1,2], David A. Collier[8,9], Hailiang Huang[1,3], Tune H. Pers[3,10,11], Ingrid Agartz[12,13,14], Esben Agerbo[15,16,17], Margot Albus[18], Madeline Alexander[19], Farooq Amin[20,21], Silviu A. Bacanu[22], Martin Begemann[23], Richard A Belliveau Jr[2], Judit Bene[24,25], Sarah E. Bergen [2,26], Elizabeth Bevilacqua[2], Tim B Bigdeli [22], Donald W. Black[27], Richard Bruggeman[28], Nancy G. Buccola[29], Randy L. Buckner[30,31,32], William Byerley[33], Wiepke Cahn[34], Guiqing Cai[35,36], Murray J. Cairns[39,120,170], Dominique Campion[37], Rita M. Cantor[38], Vaughan J. Carr[39,40], Noa Carrera[6], Stanley V. Catts[39,41], Kimberly D. Chambert[2], Raymond C. K. Chan[42], Ronald Y. L. Chen[43], Eric Y. H. Chen[43,44], Wei Cheng[45], Eric F. C. Cheung[46], Siow Ann Chong[47], C. Robert Cloninger[48], David Cohen[49], Nadine Cohen[50], Paul Cormican[5], Nick Craddock[6,7], Benedicto Crespo-Facorro[210], James J. Crowley[51], David Curtis[52,53], Michael Davidson[54], Kenneth L. Davis[36], Franziska Degenhardt[55,56], Jurgen Del Favero[57], Lynn E. DeLisi[128,129] , Ditte Demontis[17,58,59], Dimitris Dikeos[60], Timothy Dinan[61], Srdjan Djurovic[14,62], Gary Donohoe[5,63], Elodie Drapeau[36], Jubao Duan[64,65], Frank Dudbridge[66], Naser Durmishi[67], Peter Eichhammer[68], Johan Eriksson[69,70,71], Valentina Escott-Price[6], Laurent Essioux[72], Ayman H. Fanous[73,74,75,76], Martilias S. Farrell[51], Josef Frank[77], Lude Franke[78], Robert Freedman[79], Nelson B. Freimer[80], Marion Friedl[81], Joseph I. Friedman[36], Menachem Fromer[1,2,4,82], Giulio Genovese[2], Lyudmila Georgieva[6], Elliot S. Gershon[209], Ina Giegling[81,83], Paola Giusti-Rodríguez[51], Stephanie Godard[84], Jacqueline I. Goldstein[1,3], Vera Golimbet[85], Srihari Gopal[86], Jacob Gratten[87], Lieuwe de Haan[88], Christian Hammer[23], Marian L. Hamshere[6], Mark Hansen[89], Thomas Hansen[17,90], Vahram Haroutunian[36,91,92], Annette M. Hartmann[81], Frans A. Henskens[39,93,94], Stefan Herms[55,56,95], Joel N. Hirschhorn[3,11,96], Per Hoffmann[55,56,95], Andrea Hofman[55,56], Mads V. Hollegaard[97], David M. Hougaard[97], Masashi Ikeda[98], Inge Joa[99], Antonio Julià[100], René S. Kahn[34], Luba Kalaydjieva[101,102], Sena Karachanak-Yankova[103], Juha Karjalainen[78], David Kavanagh[6], Matthew C. Keller[104], Brian J. Kelly[120], James L. Kennedy[105,106,107], Andrey Khrunin[108], Yunjung Kim[51], Janis Klovins[109], James A. Knowles[110], Bettina Konte[81], Vaidutis Kucinskas[111], Zita Ausrele Kucinskiene[111], Hana Kuzelova-Ptackova[112], Anna K. Kähler[26], Claudine Laurent[19,113], Jimmy Lee Chee Keong[47,114], S. Hong Lee[87], Sophie E. Legge[6], Bernard Lerer[115], Miaoxin Li[43,44,116] Tao Li[117], Kung-Yee Liang[118], Jeffrey Lieberman[119], Svetlana Limborska[108], Carmel M. Loughland[39,120], Jan Lubinski[121], Jouko Lönnqvist[122], Milan Macek Jr[112], Patrik K. E. Magnusson[26], Brion S. Maher[123], Wolfgang Maier[124], Jacques Mallet[125], Sara Marsal[100], Manuel Mattheisen[17,58,59,126], Morten Mattingsdal[14,127], Robert W. McCarley[128,129], Colm McDonald[130], Andrew M. McIntosh[131,132], Sandra Meier[77], Carin J. Meijer[88], Bela Melegh[24,25], Ingrid Melle[14,133], Raquelle I. Mesholam-Gately[128,134], Andres Metspalu[135], Patricia T. Michie[39,136], Lili Milani[135], Vihra Milanova[137], Younes Mokrab[8], Derek W. Morris[5,63], Ole Mors[17,58,138], Kieran C. Murphy[139], Robin M. Murray[140], Inez Myin-Germeys[141], Bertram Müller-Myhsok[142,143,144], Mari Nelis[135], Igor Nenadic[145], Deborah A. Nertney[146], Gerald Nestadt[147], Kristin K. Nicodemus[148], Liene Nikitina-Zake[109], Laura Nisenbaum[149], Annelie Nordin[150], Eadbhard O'Callaghan[151], Colm O'Dushlaine[2], F. Anthony O'Neill[152], Sang-Yun Oh[153], Ann Olincy[79], Line Olsen[17,90], Jim Van Os[141,154], Psychosis Endophenotypes International Consortium[155], Christos Pantelis[39,156], George N. Papadimitriou[60], Sergi Papiol[23], Elena Parkhomenko[36], Michele T. Pato[110], Tiina Paunio[157,158], Milica Pejovic-Milovancevic[159], Diana O. Perkins[160], Olli Pietiläinen[158,161], Jonathan Pimm[53], Andrew J. Pocklington[6], John Powell[140], Alkes Price[3,162], Ann E. Pulver[147], Shaun M. Purcell[82], Digby Quested[163], Henrik B. Rasmussen[17,90], Abraham Reichenberg[36], Mark A. Reimers[164], Alexander L. Richards[6], Joshua L. Roffman[30,32], Panos Roussos[82,165], Douglas M. Ruderfer[6,82], Veikko Salomaa[71], Alan R. Sanders[64,65], Ulrich Schall[39,120], Christian R. Schubert[166], Thomas G. Schulze[77,167], Sibylle G. Schwab[168], Edward M. Scolnick[2], Rodney J. Scott[39,169,170], Larry J. Seidman[128,134], Jianxin Shi[171], Engilbert Sigurdsson[172], Teimuraz Silagadze[173], Jeremy M. Silverman[36,174], Kang Sim[47], Petr Slominsky[108], Jordan W. Smoller[2,4], Hon-Cheong So[43], Chris C. A. Spencer[175], Eli A. Stahl[3,82], Hreinn Stefansson[176], Stacy Steinberg[176], Elisabeth Stogmann[177], Richard E. Straub[178], Eric Strengman[179,34], Jana Strohmaier[77], T. Scott Stroup[119], Mythily Subramaniam[47], Jaana

Suvisaari[122], Dragan M. Svrakic[48], Jin P. Szatkiewicz[51], Erik Söderman[12], Srinivas Thirumalai[180], Draga Toncheva[103], Paul A. Tooney[39,120,170], Sarah Tosato[181], Juha Veijola[182,183], John Waddington[184], Dermot Walsh[185], Dai Wang[86], Qiang Wang[117], Bradley T. Webb[22], Mark Weiser[54], Dieter B. Wildenauer[186], Nigel M. Williams[6], Stephanie Williams[51], Stephanie H. Witt[77], Aaron R. Wolen[164], Emily H. M. Wong[43], Brandon K. Wormley[22], Jing Qin Wu[39,170], Hualin Simon Xi[187], Clement C. Zai[105,106], Xuebin Zheng[188], Fritz Zimprich[177], Naomi R. Wray[87], Kari Stefansson[176], Peter M. Visscher[87], Wellcome Trust Case-Control Consortium 2[189], Rolf Adolfsson[150], Ole A. Andreassen[14,133], Douglas H. R. Blackwood[132], Elvira Bramon[190], Joseph D. Buxbaum[35,36,91,191], Anders D. Børglum[17,58,59,138], Sven Cichon[55,56,95,192], Ariel Darvasi[193], Enrico Domenici[194], Hannelore Ehrenreich[23], Tõnu Esko[3,11,96,135], Pablo V. Gejman[64,65], Michael Gill[5], Hugh Gurling[53], Christina M. Hultman[26], Nakao Iwata[98], Assen V. Jablensky[39,102,186,195], Erik G. Jönsson[12,14], Kenneth S. Kendler[196], George Kirov[6], Jo Knight[105,106,107], Todd Lencz[197,198,199], Douglas F. Levinson[19], Qingqin S. Li[86], Jianjun Liu[188,200], Anil K. Malhotra[197,198,199], Steven A. McCarroll[2,96], Andrew McQuillin[53], Jennifer L. Moran[2], Preben B. Mortensen[15,16,17], Bryan J. Mowry[87,201], Markus M. Nöthen[55,56], Roel A. Ophoff[38,80,34], Michael J. Owen[6,7], Aarno Palotie[2,4,161], Carlos N. Pato[110], Tracey L. Petryshen[2,128,202], Danielle Posthuma[203,204,205], Marcella Rietschel[77], Brien P. Riley[196], Dan Rujescu[81,83], Pak C. Sham[43,44,116] Pamela Sklar[82,91,165], David St Clair[206], Daniel R. Weinberger[178,207], Jens R. Wendland[166], Thomas Werge[17,90,208], Mark J. Daly[1,2,3], Patrick F. Sullivan[26,51,160] & Michael C. O'Donovan[6,7]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

[3]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

[4]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[5]Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin 8, Ireland.

[6]MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, CF24 4HQ, UK.

[7]National Centre for Mental Health, Cardiff University, Cardiff, CF24 4HQ, UK.

[8]Eli Lilly and Company Limited, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, GU20 6PH, UK. [9]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, SE5 8AF, UK.

[10]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, DK-2800, Denmark.

[11]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, 02115USA.

[12]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, SE-17176 Stockholm, Sweden. [13]Department of Psychiatry, Diakonhjemmet Hospital, 0319 Oslo, Norway. [14]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, 0424 Oslo, Norway.

[15]Centre for Integrative Register-based Research, CIRRAU, Aarhus University, DK-8210 Aarhus, Denmark.

[16]National Centre for Register-based Research, Aarhus University, DK-8210 Aarhus, Denmark.

[17]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark.

[18]State Mental Hospital, 85540 Haar, Germany.

[19]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA.

[20]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia 30033, USA.

[21]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta Georgia 30322, USA.

[22]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[23]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen 37075, Germany.

[24]Department of Medical Genetics, University of Pécs, Pécs H-7624, Hungary.

[25]Szentagothai Research Center, University of Pécs, Pécs H-7624, Hungary.

[26]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden.

[27]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242, USA.

[28]University Medical Center Groningen, Department of Psychiatry, University of Groningen NL-9700 RB, The Netherlands.

[29]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana 70112, USA.

[30]Athinoula A. Martinos Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA.

[31]Center for Brain Science, Harvard University, Cambridge, Massachusetts, 02138 USA.

[32]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, 02114 USA.

[33]Department of Psychiatry, University of California at San Francisco, San Francisco, California, 94143 USA.

[34]University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, 3584 Utrecht, The Netherlands.

[35]Department of Human Genetics, Icahn School of Medicine at Mount Sinai, New York, New York 10029 USA.

[36]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029 USA.

[37]Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine, 76301 Rouen, France.

[38]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

[39]Schizophrenia Research Institute, Sydney NSW 2010, Australia.

[40]School of Psychiatry, University of New South Wales, Sydney NSW 2031, Australia.

[41]Royal Brisbane and Women's Hospital, University of Queensland, Brisbane, St Lucia QLD 4072, Australia.

[42]Institute of Psychology, Chinese Academy of Science, Beijing 100101, China.

[43]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

[44]State Key Laboratory for Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

[45]Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27514, USA.

[46]Castle Peak Hospital, Hong Kong, China.

[47]Institute of Mental Health, Singapore 539747, Singapore.

[48]Department of Psychiatry, Washington University, St. Louis, Missouri 63110, USA.

[49]Department of Child and Adolescent Psychiatry, Assistance Publique Hopitaux de Paris, Pierre and Marie Curie Faculty of Medicine and Institute for Intelligent Systems and Robotics, Paris, 75013, France.

[50] Blue Note Biosciences, Princeton, New Jersey 08540, USA

[51]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA.

[52]Department of Psychological Medicine, Queen Mary University of London, London E1 1BB, UK.

[53]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6JJ, UK.

[54]Sheba Medical Center, Tel Hashomer 52621, Israel.

[55]Department of Genomics, Life and Brain Center, D-53127 Bonn, Germany.

[56]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany.

[57]Applied Molecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, B-2610 Antwerp, Belgium.

[58]Centre for Integrative Sequencing, iSEQ, Aarhus University, DK-8000 Aarhus C, Denmark.

[59]Department of Biomedicine, Aarhus University, DK-8000 Aarhus C, Denmark.

[60]First Department of Psychiatry, University of Athens Medical School, Athens 11528, Greece.

[61]Department of Psychiatry, University College Cork, Co. Cork, Ireland.

[62]Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway.

[63]Cognitive Genetics and Therapy Group, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Co. Galway, Ireland.

[64]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois 60637, USA.

[65]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois 60201, USA.

[66]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

[67]Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje 1000, Republic of Macedonia.

[68]Department of Psychiatry, University of Regensburg, 93053 Regensburg, Germany.

[69]Department of General Practice, Helsinki University Central Hospital, University of Helsinki P.O. Box 20, Tukholmankatu 8 B, FI-00014, Helsinki, Finland

[70]Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki 1, Haartmaninkatu 8, FI-00290, Helsinki, Finland.

[71]National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland.

[72]Translational Technologies and Bioinformatics, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.

[73]Department of Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA.

[74]Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA.

[75]Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298, USA.

[76]Mental Health Service Line, Washington VA Medical Center, Washington DC 20422, USA.

[77]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg , D-68159 Mannheim, Germany.

[78]Department of Genetics, University of Groningen, University Medical Centre Groningen, 9700 RB Groningen, The Netherlands.

[79]Department of Psychiatry, University of Colorado Denver, Aurora, Colorado 80045, USA.

[80]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA.

[81]Department of Psychiatry, University of Halle, 06112 Halle, Germany.

[82]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[83]Department of Psychiatry, University of Munich, 80336, Munich, Germany.

[84]Departments of Psychiatry and Human and Molecular Genetics, INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris, 75013, France.

[85]Mental Health Research Centre, Russian Academy of Medical Sciences, 115522 Moscow, Russia.

[86]Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, New Jersey 08869, USA.

[87]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, QLD 4072, Australia.

[88]Academic Medical Centre University of Amsterdam, Department of Psychiatry, 1105 AZ Amsterdam, The Netherlands.

[89]Illumina, La Jolla, California, California 92122, USA.

[90]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, DK-4000, Denmark.

[91]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[92]J. J. Peters VA Medical Center, Bronx, New York, New York 10468, USA.

[93]Priority Research Centre for Health Behaviour, University of Newcastle, Newcastle NSW 2308, Australia.

[94]School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle NSW 2308, Australia.

[95]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, CH-4058, Switzerland.

[96]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

[97]Section of Neonatal Screening and Hormones, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, DK-2300, Denmark.

[98]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Aichi, 470-1192, Japan.

[99]Regional Centre for Clinical Research in Psychosis, Department of Psychiatry, Stavanger University Hospital, 4011 Stavanger, Norway.

[100]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, 08035, Spain.

[101]Centre for Medical Research, The University of Western Australia, Perth, WA 6009, Australia.

[102]The Perkins Institute for Medical Research, The University of Western Australia, Perth, WA 6009, Australia.

[103]Department of Medical Genetics, Medical University, Sofia1431, Bulgaria.

[104]Department of Psychology, University of Colorado Boulder, Boulder, Colorado 80309, USA.

[105]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, M5T 1R8, Canada.

[106]Department of Psychiatry, University of Toronto, Toronto, Ontario, M5T 1R8, Canada.

[107]Institute of Medical Science, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

[108]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow123182, Russia.

[109]Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia.

[110]Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of Southern California, Los Angeles, California 90089, USA.

[111]Faculty of Medicine, Vilnius University, LT-01513 Vilnius, Lithuania.

[112] Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic.

[113] Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris 75013, France.

[114]Duke-NUS Graduate Medical School, Singapore 169857, Singapore.

[115]Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel.

[116]Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China.

[117]Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, 610041, Sichuan, China.

[118]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA.

[119]Department of Psychiatry, Columbia University, New York, New York 10032, USA.

[120]Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle NSW 2300, Australia.

[121]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, 70-453 Szczecin, Poland.

[122]Department of Mental Health and Substance Abuse Services; National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland

[123]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.

[124]Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany.

[125]Centre National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital de la Pitié Salpêtrière, 75013, Paris, France.

[126]Department of Genomics Mathematics, University of Bonn, D-53127 Bonn, Germany.

[127]Research Unit, Sørlandet Hospital, 4604 Kristiansand, Norway.

[128]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA.

[129]VA Boston Health Care System, Brockton, Massachusetts 02301, USA.

[130]Department of Psychiatry, National University of Ireland Galway, Co. Galway, Ireland.

[131]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH16 4SB, UK.

[132]Division of Psychiatry, University of Edinburgh, Edinburgh EH16 4SB, UK.

[133]Division of Mental Health and Addiction, Oslo University Hospital, 0424 Oslo, Norway.

[134]Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, Massachusetts 02114, USA.

[135]Estonian Genome Center, University of Tartu, Tartu 50090, Estonia.

[136]School of Psychology, University of Newcastle, Newcastle NSW 2308, Australia.

[137]First Psychiatric Clinic, Medical University, Sofia 1431, Bulgaria.

[138]Department P, Aarhus University Hospital, DK-8240 Risskov, Denmark.

[139]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

[140]King's College London, London SE5 8AF, UK.

[141]Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, EURON, 6229 HX Maastricht, The Netherlands.

[142]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK.

[143]Max Planck Institute of Psychiatry, 80336 Munich, Germany.

[144]Munich Cluster for Systems Neurology (SyNergy), 80336 Munich, Germany.

[145]Department of Psychiatry and Psychotherapy, Jena University Hospital, 07743 Jena, Germany.

[146]Department of Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, St Lucia QLD 4072, Australia.

[147]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

[148]Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland.

[149]Eli Lilly and Company, Lilly Corporate Center, Indianapolis, 46285 Indiana, USA.

[150]Department of Clinical Sciences, Psychiatry, Umeå University, SE-901 87 Umeå, Sweden.

[151]DETECT Early Intervention Service for Psychosis, Blackrock, Co. Dublin, Ireland.

[152]Centre for Public Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast BT12 6AB, UK.

[153]Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley, California 94720, USA.

[154]Institute of Psychiatry, King's College London, London SE5 8AF, UK.

[155]A list of authors and affiliations appear in the Supplementary Information.

[156]Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne, Vic 3053, Australia.

[157]Department of Psychiatry, University of Helsinki, P.O. Box 590, FI-00029 HUS, Helsinki, Finland.

[158]Public Health Genomics Unit, National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland.

[159]Medical Faculty, University of Belgrade, 11000 Belgrade, Serbia.

[160]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7160, USA.

[161]Institute for Molecular Medicine Finland, FIMM, University of Helsinki, P.O. Box 20

FI-00014, Helsinki, Finland.

[162]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.

[163]Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK.

[164]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[165]Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[166]PharmaTherapeutics Clinical Research, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.

[167]Department of Psychiatry and Psychotherapy, University of Gottingen, 37073 Göttingen, Germany.

[168]Psychiatry and Psychotherapy Clinic, University of Erlangen, 91054 Erlangen, Germany.

[169]Hunter New England Health Service, Newcastle NSW 2308, Australia.

[170]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan NSW 2308, Australia.

[171]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA.

[172]University of Iceland, Landspitali, National University Hospital, 101 Reykjavik, Iceland.

[173]Department of Psychiatry and Drug Addiction, Tbilisi State Medical University (TSMU), **N33, 0177** Tbilisi, Georgia.

[174]Research and Development, Bronx Veterans Affairs Medical Center, New York, New York 10468, USA.

[175]Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK.

[176]deCODE Genetics, 101 Reykjavik, Iceland.

[177]Department of Clinical Neurology, Medical University of Vienna, 1090 Wien, Austria.

[178]Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA.

[179]Department of Medical Genetics, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands.

[180]Berkshire Healthcare NHS Foundation Trust, Bracknell RG12 1BQ, UK.

[181]Section of Psychiatry, University of Verona, 37134 Verona, Italy.

[182]Department of Psychiatry, University of Oulu, P.O. BOX 5000, 90014, Finland

[183]University Hospital of Oulu, P.O.BOX 20, 90029 OYS, Finland.

[184]Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

[185]Health Research Board, Dublin 2, Ireland.

[186]School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth WA6009, Australia.

[187]Computational Sciences CoE, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.

[188]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore 138672, Singapore.

[190]University College London, London WC1E 6BT, UK.

[191]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[192]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, 52428 Juelich, Germany.

[193]Department of Genetics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel.

[194]Neuroscience Discovery and Translational Area, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.

[195]Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Medical Research Foundation Building, Perth WA 6000, Australia.

[196]Virginia Institute for Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[197]The Feinstein Institute for Medical Research, Manhasset, New York, 11030 USA.

[198]The Hofstra NS-LIJ School of Medicine, Hempstead, New York, 11549 USA.

[199]The Zucker Hillside Hospital, Glen Oaks, New York,11004 USA.

[200]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore.

[201]Queensland Centre for Mental Health Research, University of Queensland, Brisbane 4076, Queensland, Australia.

[202]Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[203]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam 3000, The Netherlands.

[204]Department of Complex Trait Genetics, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam 1081, The Netherlands.

[205]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam 1081, The Netherlands.

[206]University of Aberdeen, Institute of Medical Sciences, Aberdeen, AB25 2ZD, UK.

[207]Departments of Psychiatry, Neurology, Neuroscience and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA.

[208]Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200, Denmark.

[209]Departments of Psychiatry and Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.

[210]University Hospital Marqués de Valdecilla, Instituto de Formación e Investigación Marqués de Valdecilla, University of Cantabria, E□39008 Santander, Spain.

References

1 Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical chemistry* **83**, 8604-8610, doi:10.1021/ac202028g (2011).

2 Wu, Y. L. *et al.* Sensitive and specific real-time polymerase chain reaction assays to accurately determine copy number variations (CNVs) of human complement C4A, C4B, C4-long, C4-short, and RCCX modules: elucidation of C4 CNVs in 50 consanguineous subjects with defined HLA genotypes. *Journal of immunology (Baltimore, Md. : 1950)* **179**, 3012-3025 (2007).

3 Fernando, M. M. *et al.* Assessment of complement C4 gene copy number using the paralog ratio test. *Human mutation* **31**, 866-874, doi:10.1002/humu.21259 (2010).

4 Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics* **84**, 210-223, doi:10.1016/j.ajhg.2009.01.005 (2009).

5 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics* **81**, 1084-1097, doi:10.1086/521987 (2007).

6 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

7 Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PloS one* **8**, e64683, doi:10.1371/journal.pone.0064683 (2013).

8 Fischer, M. B. *et al.* Regulation of the B cell response to T-dependent antigens by classical pathway complement. *Journal of immunology (Baltimore, Md. : 1950)* **157**, 549-556 (1996).

9 Torborg, C. L. & Feller, M. B. Unbiased analysis of bulk axonal segregation patterns. *Journal of neuroscience methods* **135**, 17-26, doi:10.1016/j.jneumeth.2003.11.019 (2004).

10 Bialas, A. R. & Stevens, B. TGF-beta signaling regulates neuronal C1q expression and developmental synaptic refinement. *Nature neuroscience* **16**, 1773-1782, doi:10.1038/nn.3560 (2013).

11 Barres, B.A., Silverstein, B.E., Corey, D.R. & Chun, L.L.Y. Immunological, morphological, and electrophysiological variation among retinal ganglion cells purified by panning. *Neuron* **1**,791–803 (1988).