Supporting Information

Xu et al. 10.1073/pnas.0908584106

SI Text

Supporting Information. While most of the genes in Table S3 have not been directly associated with schizophrenia (SCZ) in the literature, some of them have been implicated in psychiatric and other neurobehavioral disorders or in mechanisms that may underlie predisposition to such disorders. Two examples (*NRG3* and *RAPGEF2*) are discussed in the main text, but a few additional examples are provided below:

The Contactin Family. Burbach and van der Zwaag (1) recently provided a detailed account of the contactin system and its potential contribution to a number of different psychiatric and neurobehavioral disorders, including SCZ and autism. In an early study, Fernandez et al. (2, 3) showed that disruption of contactin 4 (*CNTN4*) results in developmental delay and other features of the 3p deletion syndrome. In the present study, we identified a CNV at ≈130 kb downstream of *CNTN6* (also named *NB3*), another member of the contactin protein family. Interestingly, in an independent study, we identified another CNV disrupting the contactin associated protein 2 (*CNTNAP2*) that co-segregates with psychosis in one family. Structural variants in the *CNTNAP2* gene have been reported in patients with psychosis in at least two previous genome-wide scans (4, 5). Thus, our results support the notion that abnormalities in the contactin family proteins may be involved in SCZ and other psychiatric conditions.

The CSMD1 (CUB and Sushi Multiple Domains 1) Gene. One CNV identified in the present study is a duplication at 8p23.1–8p23.2 involving the *CSMD1* gene. A recent study (6) showed that a transmitted duplication at 8p23.1–8p23.2 affecting the *CSMD1* gene may be associated with speech delay, autism, and learning difficulties.

The *ADARB1* (adenosine deaminase, RNA-specific, B1) gene: *ADARB1* (also named *ADAR2*) encodes for adenosine deaminase, a key RNA editing enzyme. Although *ADARB1* has not been directly associated with psychiatric conditions, mRNAs for several important neuronal receptors such as *5-HT2CR* and *GLUR2* undergo posttranscriptional editing. Abnormal editing of HT2CR has been implicated in depression and SCZ (7).

Other potentially interesting genes include the *RXFP2* (relaxin/insulin-like family peptide receptor 2) gene (also known as *LGR8*), a member of the relaxin family peptide receptor system whose members are distributed in both excitatory and inhibitory neurons in the brain and modulate diverse behaviors (8), as well as the *LRFN5* (leucine-rich repeat and fibronectin type III domain containing 5) gene (also named *SALM5*), which is expressed exclusively in neuronal tissues, especially in mature neurons. Several members of the LRFN protein family were shown to associate with PSD-95 (9, 10).

SI Methods

Patient Cohorts. We performed a genome-wide survey of rare inherited CNVs in a total of 182 individuals, consisting of 48 probands with familial SCZ (positive disease history in a first-degree ($n = 33$) or second-degree ($n = 15$) relative and both of their biological parents, as well as all additional affected relatives that were available for genotyping. Of the 33 families with a history of SCZ in a first-degree relative, 17 are parent-child pairs, 12 are sibling pairs, and four families are both. In eight of these 33 families, in addition to the first-degree relative, there is at least one second-degree relative who is also affected with SCZ (Fig. S1). Of the 15 families with a history of SCZ in a second-degree relative, 11 are avuncular pairs, three are grandparent-grandchild pairs, and one family is both. For our linkage studies, we genotyped 479 subjects from 130 families. Sixty-nine families are informative. In the 54 informative families with at least two affected members, there are 60 and 79 affected relative pairs for the narrow and broad affection categories, respectively. Of the broadly affected relative pairs, 30 are sibling, 24 are parent-child, 14 are avuncular, six are cousin, three are second cousin, and two are grandparent-grandchild.

All subjects were confirmed as Afrikaners by tracing ancestry back to the 1800s using state and church records and finally to the initial 2,000 founders through the Genealogies of Old South African Families. Diagnostic evaluations were conducted in-person by specially trained clinicians using the DIGS, which was translated and back-translated into Afrikaans. Upon completion of the DIGS, the interviewers assigned DSM-IV diagnoses and completed a detailed narrative report. All narratives and DIGS were independently evaluated by appropriately trained clinicians in New York.

Sample Characteristics. The familial cases sample consisted of 31 males (65%) and 17 females (35%) and both their biological parents. The sample enriched in sporadic cases consisted of 108 males (71%) and 44 females (29%) and both their unaffected biological parents. The control sample consisted of 93 females (58%) and 66 males (42%) and both their biological parents. It should be noted that both in this study and our previous Xu et al. (11) study we did not find any sex-specific differences in distribution of CNVs. We checked for population stratification between cases and controls using principal component analysis as implemented in EIGENSTRAT (12). Both groups overlap in the graphical representation of the top two principal components (eigenvectors), indicating no evidence of stratification.

The phenotypic variables analyzed in the CNV scan, were defined as follows: History of developmental delay was recorded as positive when there was clear history of maturation lag or milestone delays before the age of 6 (i.e., delayed crawling, walking, speaking, etc.). History of learning disabilities was recorded as positive if there had been a diagnosis of a learning disability, or clear history of being a "slow learner," requiring remediation at school, or placement in a special class. Age at onset was defined as the age at which full DSM criteria for schizophrenia or schizoaffective disorder were met.

Genotyping. For the linkage analysis, genotyping for 2005 di-, tri-, and tetra-nucleotide repeat microsatellite markers was performed by deCODE (Reykjavik, Iceland), through their fee-for-service genotyping facility. Among them, 1,904 autosomal markers and 76 chromosome X markers passed our internal quality checks. Of the 960,395 possible genotypes, 878,046 were successfully called, corresponding to an average (±SD) per marker genotyping rate of 91% (±7%).

For the CNV analysis, the families were genotyped using Human Genome-Wide SNP Array 5.0 (Affymetrix), which contains 500,568 SNPs, as well as 420,000 additional nonpolymorphic probes that can assess other genetic differences, such as CN variation. Samples were processed as previously described (11). Average call rate on arrays used in this study was 99.43%. All microarray experiments were performed in the Vanderbilt Microarray Shared Resource directed by Dr. Shawn Levy.

CNV Identification, Inherited CNV Detection, and Inherited CNV Verification. CEL files from all chips were analyzed using two software packages, DNA-Chip Analyzer (dChip) and Partek Genomics Suite (Partek software, version 6.3 Beta, build #6.07.1127; Partek), as previously described (11). The entire procedure included five steps: (*i*) quality control; (*ii*) CNV identification; (*iii*) inherited CNV detection; (*iv*) inherited CNV verification by intensity and genotyping filters; and (*v*) inherited CNV confirmation by MLPA.

(*i*) Quality control: A quality control step was conducted to exclude families with one or more hybridization failures.

(*ii*) CNV identification: All CEL files that passed the quality control step were imported and analyzed by dCHIP (13–15) and Partek according to the authors' instructions. For the dCHIP analysis, data were processed in batches, each batch containing arrays processed simultaneously in 96-well plate formats. An array with median overall intensity within each batch was chosen as the baseline to adjust the brightness of all arrays to a comparable level. Arrays were then normalized at probe intensity level using the Invariant Set Normalization procedure (14). Following normalization, a CNV identification step ensued. A reference signal value was first calculated for each probe set using a model-based method. To obtain reference signal distribution (blind to the affected status information), we assumed that for any locus, less than 10% of all of the samples show abnormal CN. Thus for a given probe set, 5% of samples with extreme signals were trimmed from each end, and the rest of the values were used to estimate the mean and standard deviation of the signal distribution of normal CN ($n = 2$). For probe sets located on the X chromosome, a "Gender" information file was used so that signal from male subjects was considered as one copy and signal from female subjects as two copies when computing mean signal intensity. Two pseudoautosomal regions (PARs) on chromosome X defined by Affymetrix annotation files [GenomeWideSNP_5 Annotations, CSV format (11/28/07)] were excluded from our analysis. The observed raw CN was then defined based on the reference signal for each probe set of each sample and a "Hidden Markov Model" (HMM) algorithm was used to infer CN and identify CNV regions. In determining CN, we also used the median smoothing method implemented in the dCHIP package. Compared to the HMM method, median smoothing has the advantage of providing results close to the raw CN but robust to outliers, and it does not require the stringent parameter specification required in HMM fitting. We used the CN inferred by this method for the inherited CNV confirmation step (see below). To reduce the false positive rate, we excluded the triads that have one or more individuals with more than 35 CNVs per chip for all follow-up analysis, as suggested in Szatmari et al. (15).

The Partek Genomics Suite (version 6.3 Beta, build #6.07.1127) uses both the SNP and nonpolymorphic CN probe sets on Affymetrix GWS 5.0 chip. During the data importation stage, all raw signal intensities were adjusted for fragment length and probe sequence, and quantile-normalized to generate normalized signal intensities for all genomic markers on the arrays. To identify CNVs, we used a reference sample set consisting of all unaffected mothers. Normalization was first performed on the reference sample set; the experimental samples were then normalized to the same distribution as the reference samples. A CN baseline created from the Reference set was used for CNV identification. The resulting data were scanned for regions of genomic gain or loss, with a minimum region size of three probe sets using a HMM algorithm. For every male sample, losses with a CN of 0, and gains with a CN of 2 for any X chromosome region were extracted.

CNV regions for each sample identified by the two software packages were then merged into a single list of nonredundant CNVs. CNVs were combined if they overlapped by 50% or more in length. The borders listed are the outermost borders defined by either analysis.

(*iii*) Rare inherited CNV detection: The CNV list generated above was first divided into two lists, a deletion list (mean CN $<$ 1.25 for autosomal or mean CN $<$ 0.25 for chromosome X) and a duplication list (mean CN $>$ 2.75 for autosomal or mean CN $>$ 1.75 for chromosome X). Deletion and duplication analyses were conducted separately. A candidate inherited CNV was considered for further analysis only if there was more than 50% overlap in size with a variant at the same locus in the biological parental chromosomes (size of overlap/total size of merged CNVs $>$ 0.5), but less than 50% overlap in size with any variant in the other parental chromosomes.

(*iv*) Inherited CNV verification using genotype information and median smoothing inferred CN: The vast majority of CNVs are inherited from parents (16) and since the false-positive or false-negative prediction rate is relatively high in all available CNV detection algorithms (17), it is important to use independent CNV inferring methods, as well as family information for validation of inherited CNV calling results. Genotyping calls and Median smoothing inferred CN calls for all samples were stored in MySQL database (version 5.1) along with the related pedigree information. For each identified candidate CNV, all genotypes and inferred CN of SNPs within the region were retrieved for each triad (the subject and his/her parents). Loss of heterozygosity (LOH), Mendelian inconsistencies and average median smoothing inferred CN of the region were analyzed in Microsoft Excel driven by a VBA script and further examined visually. A number of rules were applied to determine whether a candidate CNV in a child is a true inherited CNV:

1. Average inferred CN: Average inferred CN of the normal parent should be close to two for an autosomal region. The average inferred CN of the child and the mutation-carrying parent should be at least 20% greater (for duplication) or less (for deletion) than the normal parent.

2. Homozygosity/LOH: If a stretch of homozygosity was observed within a candidate deletion CNV, the latter was considered to be a true positive call. Moreover, observation of the same homozygosity in a child and mutation-carrying parent while the normal parent maintains proper heterozygosity was considered a strong indication of an inherited CNV.

3. Mendelian inconsistencies: Presence of Mendelian inconsistencies between the genotype of a child and the normal parent was considered a good indicator of the existence of an inherited CNV in the child.

All rules 1 through 3 were used to provide parental origin information.

(*v*) Inherited CNV confirmation by MLPA: We independently confirmed the in silico verified CNVs in a subset of families (see main text) using the MLPA approach. Two to three pairs of MLPA target-probes were designed based on the unique sequences within each CNV region. Additionally, three pairs of MLPA control probes from the unique sequences of the *VGEFA* locus were included in each MLPA reaction. All probes were synthetic oligonucleotides. MLPA reagents were prepared according to the instructions at http://www.mlpa.com/pages/support_pagepag.html. Final PCR products were analyzed on an ABI3730XL for peak identification and quantification. The peak profiles of all test samples were visualized as shown in Fig. S2, and parameters (height and area) were extracted using Peak Scanner Software v1.0 (Applied Biosystems). For copy number quantification, the peak areas and heights were exported to a Microsoft Excel worksheet. Peak area and height for each probe was normalized to the mean value for all control probes. The relative ratio of each peak was calculated by comparisons between proband sample and the samples of his/her relatives. Deletion was identified as relative ratio $<$0.8 and duplication as relative ratio $>$1.2.

Simulation Test. An excel VBA script was used to drive the randomization of the phenotypes within each family and the simulation process. For all of the pedigrees used, the pedigree structure, CNV inheritance pattern, and number of affected individuals were kept constant. For the individuals within a pedigree with known phenotype but no genotype (CNV) information, the program randomly assigned a CNV with a probability of 0.25 (the chance that a CNV is inherited from a heterozygous parent). The pattern of co-inheritance of CNVs and diagnosis within each family was determined based on two rules: (*i*) two or more affected individuals carry the CNV in the given family and (*ii*) all of the affected individuals in the family carry the CNV. When these two conditions were satisfied, the family was scored as showing co-segregation of CNV and clinical diagnosis. To estimate the empirical *P* value, we ran the simulation 10,000 times.

Linkage Analysis. DNA from all study participants was extracted from 24 mL EDTA-treated blood, according to standard procedures (18). The deCODE genotyping protocol involves PCR amplification followed by capillary electrophoresis and automated allele calling by deCODE's Allele Caller software. Among all of the markers genotyped, 1,904 autosomal markers and 76 chromosome X markers passed our internal quality checks, which are described in detail below.

Error Checking. Before linkage analysis, we verified reported relationships using genetic marker data. To do this, we used GRR to examine identity-by-state distributions for all pairs of individuals (19) (Fig. S3). We identified three instances of nonpaternity and one set of individuals who appeared to be switched. Using the X chromosome information, we identified one subject who had the appropriate levels of identity-by-state with their relatives, but was labeled as a male while appearing to be genetically female. The pedigree file was corrected to resolve these discrepancies.

Additionally, we identified problematic markers that either exhibited more than four Mendelian errors or in a subset of 210 unrelated subjects, deviated from Hardy-Weinberg equilibrium (HWE) at $P < 0.01$. These tests help identify markers with high rates of genotyping error (20). We excluded 12 markers with more than four Mendelian errors and another 12 markers with evidence for deviations from HWE. Thus, among the 2,005 genotyped microsatellites, 1,981 (99%) passed these initial quality filters. The error checks were conducted using Pedstats (21).

Genetic Map Validation. All of our multipoint analyses are based on the deCODE linkage map (22). For markers that did not have unique positions, we adjusted cM positions slightly according to the UniSTS National Center for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov). This small adjustment avoided problems with likelihood calculations when obligate recombinants are encountered in intervals of length zero. Since the results of multipoint linkage analysis can be sensitive to errors in genetic maps, we checked the agreement of our genotyped data and the published map by comparing the likelihood of the entire genotype set when: (*i*) all markers were analyzed in their original locations; (*ii*) the order of two consecutive markers was switched; and (*iii*) the position of a specific marker was changed so that it was unlinked to all others (this analysis should identify markers that are mismapped, perhaps due to an error in tracking primers during genotyping). By performing these checks, we identified one marker that appeared to be at the incorrect location on the chromosome (D14S56) (difference in likelihood is 116.3; average ± SD likelihood difference for all other markers is −14.12 ± 16.17). This marker was not included in subsequent analyses, resulting in a total of 1,980 analyzed markers.

MOD Score Parametric Linkage Analysis. Since the mode of inheritance for SCZ is unknown, we performed multipoint parametric MOD score linkage analysis using LAMP (23, 24), calculating the MOD scores at 1-cM intervals along the autosomes. In contrast to conventional parametric linkage analysis, MOD score analysis does not require the disease allele frequency and penetrance parameters to be specified a priori; they are estimated at each location using maximum likelihood. Thus, we expect MOD score analysis to be more powerful than both traditional parametric and nonparametric analysis in situations where the mode of inheritance is uncertain. We initially used the unconstrained (free) model, however since most of our families include a single affected relative pair, we used a multiplicative model that required estimation of only two parameters (disease allele frequency and effect size). We report the multiplicative results unless otherwise specified.

Nonparametric Linkage Analysis. We also tested for linkage to a SCZ locus with multipoint and singlepoint nonparametric linkage using the S_{all} statistic (25). Both multipoint and singlepoint analyses were performed, because multipoint analysis is more powerful, while singlepoint analysis may be more robust to genotyping errors (26–28). The S_{all} statistic tests for excess identity-by-descent allele sharing between all pairs of affected individuals within a family. We used the excess-sharing parameterization modified statistic (29) as implemented in Merlin (30). LOD scores were calculated at 1-cM intervals across all chromosomes. To evaluate evidence for parent-of-origin effects, we defined a nonparametric linkage statistic (25, 31) that measured allele sharing for maternally inherited alleles only. In each case, these statistics considered all pairs of affected individuals in each pedigree. For each inheritance vector, each pair was scored according to whether they shared their maternally inherited allele [1] or not [0]. An overall nonparametric statistic was calculated for each pedigree by summing all pair specific statistics and converted to a LOD score as previously described (25, 31). An analogous statistic was defined to evaluate sharing of paternally inherited alleles.

Empirical Significance Levels. To assess the significance of our results, we simulated genotypes for 1,000 data sets by gene dropping as implemented in Merlin. The data sets used the same family structures, included the same 1,980 markers and both phenotypes. To determine an empirical distribution of maximum genome-wide MOD (or LOD) scores, we repeated our analysis using the simulated data sets. We then counted the number of times a simulated dataset resulted in a genome-wide maximum MOD (or LOD) score greater than or equal to our experimental genome-wide maximum MOD (or LOD) score. To obtain our empirical *P* value, we divided this count by the number of simulations performed.

Detection of Copy Number Mutations at the Linkage Loci. To assess the contribution of copy number mutations to the linkage signal in the Afrikaner families, a whole genome CNVs annotation was conducted using dCHIP program, in all 241 cases included in our scan as well as in the 361 parents of our unaffected controls (control sample). The quality control and CNV identification steps were similar to the description for rare inherited CNV detection steps except all CNVs in cases and controls (224 and 361, respectively) were not filtered through the criteria used to identify rare CNVs. Following the CNV identification step, all CNVs within the cytoband corresponding to the linkage peak region were extracted. A copy number region (CNR) was identified if two or more CNVs overlapped 50% in length. All CNVs within a CNR were then separated according to diagnosis and the frequencies in cases and control were calculated.

1. Burbach JP, van der Zwaag B (2009) Contact in the genetics of autism and schizophrenia. *Trends Neurosci* 32:69–72.
2. Fernandez T, et al. (2004) Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet* 74:1286–1293.
3. Fernandez T, et al. (2008) Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet* 82:1385.
4. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
5. Friedman JI, et al. (2008) CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* 13:261–266.
6. Glancy M, et al. (2009) Transmitted duplication of 8p23.1–8p23.2 associated with speech delay, autism and learning difficulties. *Eur J Hum Genet* 17:37–43.
7. Maas S, Kawahara Y, Tamburro KM, Nishikura K (2006) A-to-I RNA editing and human disease. *RNA Biol* 3:1–9.
8. Gundlach AL, et al. (2009) Relaxin family peptides and receptors in mammalian brain. *Ann N Y Acad Sci* 1160:226–235.
9. Morimura N, Inoue T, Katayama K, Aruga J (2006) Comparative analysis of structure, expression and PSD95-binding capacity of Lrfn, a novel family of neuronal transmembrane proteins. *Gene* 380:72–83.
10. Homma S, Shimada T, Hikake T, Yaginuma H (2009) Expression pattern of LRR and Ig domain-containing protein (LRRIG protein) in the early mouse embryo. *Gene Expr Patterns* 9:1–26.
11. Xu B, et al. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885.
12. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909.
13. Li C, Hung Wong W (2001) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biol* 2:RESEARCH0032.
14. Lin M, et al. (2004) dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 20:1233–1240.
15. Szatmari P, et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39:319–328.
16. Locke DP, et al. (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* 79:275–290.
17. Redon R, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
18. Ciulla TA, Sklar RM, Hauser SL (1988) A simple method for DNA purification from peripheral blood. *Anal Biochem* 174:485–488.
19. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2001) GRR: Graphical representation of relationship errors. *Bioinformatics* 17:742–743.
20. Douglas JA, Boehnke M, Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am J Hum Genet* 66:1287–1297.
21. Wigginton JE, Abecasis GR (2005) PEDSTATS: Descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 21:3445–3447.
22. Kong A, et al. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247.
23. Li M, Boehnke M, Abecasis GR (2005) Joint modeling of linkage and association: Identifying SNPs responsible for a linkage signal. *Am J Hum Genet* 76:934–949.
24. Li M, Boehnke M, Abecasis GR (2006) Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am J Hum Genet* 78:778–792.
25. Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. *Biometrics* 50:118–127.
26. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am J Hum Genet* 58:1347–1363.
27. Abecasis GR, Cherny SS, Cardon LR (2001) The impact of genotyping error on family-based analysis of quantitative traits. *Eur J Hum Genet* 9:130–134.
28. Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60:1187–1192.
29. Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188.
30. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101.
31. Kong A, et al. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247.

Fig. S1. Familial cases cohort. Pedigree structures of the 48 familial cases used in the CNV study. All affected individuals are marked black. Red arrows indicate the 48 probands included in the screen. The 19 probands that carry at least one rare inherited CNV are indicated by an asterisk.

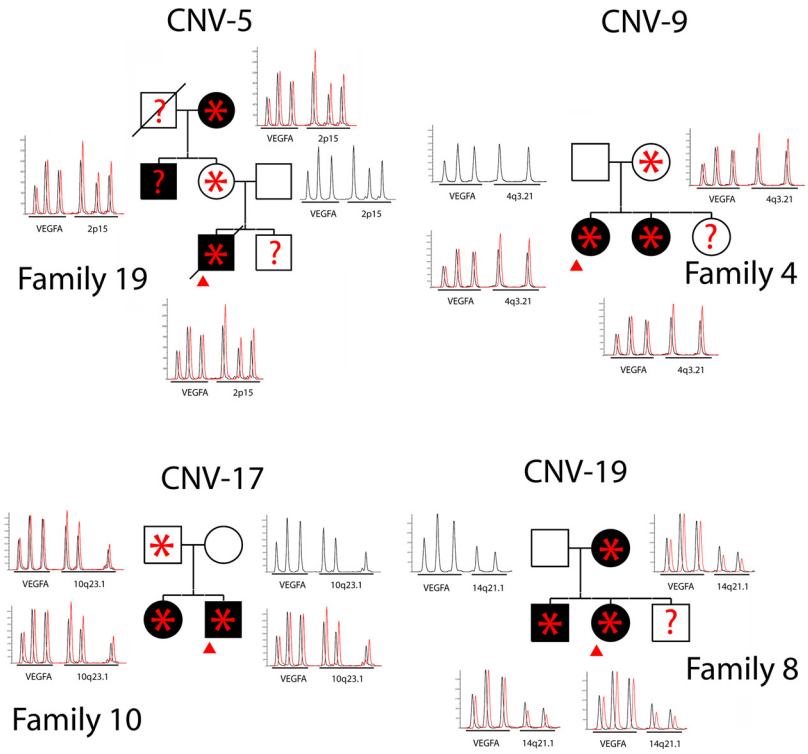

Fig. S2. Confirmation of the in silico verified CNVs in a subset of families using the MLPA approach. MLPA analysis confirms that the respective families segregate a duplication on chromosome 2p15, a deletion on chromosome 14q21.1, a duplication on chromosome 4q32.1 and a duplication on chromosome 10q23.1. For each individual in each pedigree, MLPA assay examines three genomic fragments within a control gene (VEGFA) (first three peaks along the x-axis) and two or three genomic fragments within the target CNV region (the rest of the peaks). Within each family, the MLPA tracing of an unaffected individual was used as reference (black tracing). Red tracings in other tracing graphs show a shifted overlaying of the corresponding individual's tracing compared to the reference tracing. When the relative copy number (the peak height) of the control genes is the same between the individual and the reference, the relative copy number of the target region of the individual can be significantly lower than that of the reference (indicative of a genomic loss) or significantly higher than that of the reference (indicative of a genomic gain).

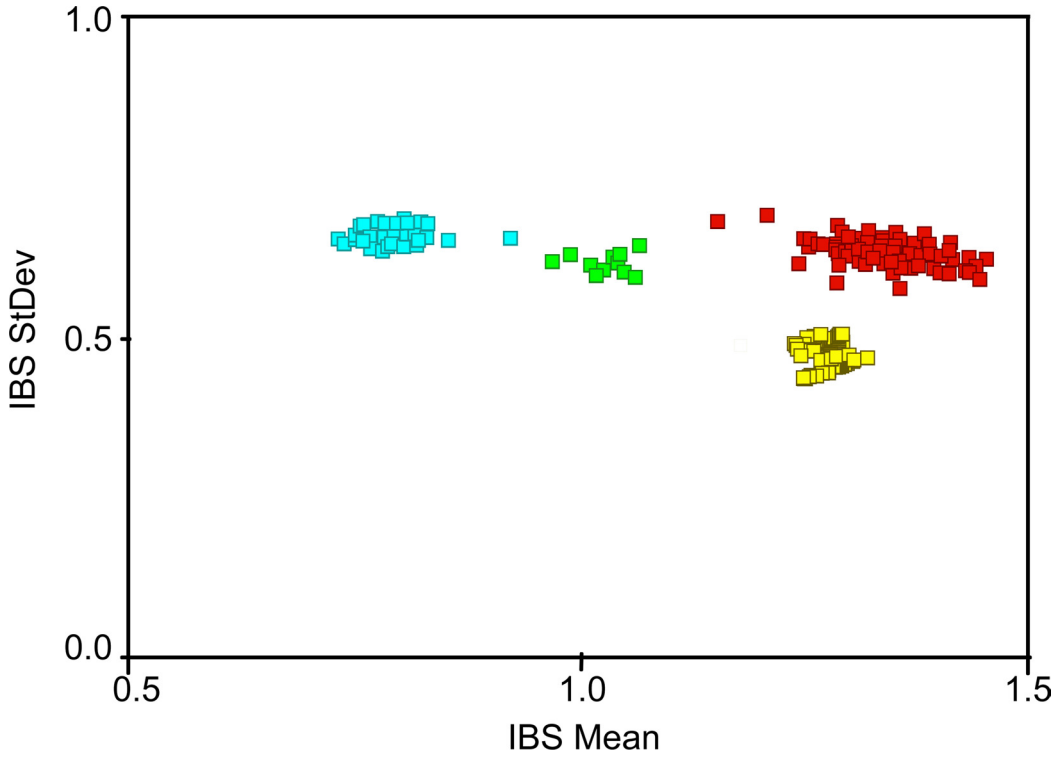

Fig. S3. Plot of mean identity-by-state by standard deviation identity-by-state for a pair of individuals from the same family. Each square represents a pair of individuals who have been genotyped for at least 900 of the same markers. Blue squares are unrelated individuals, green squares are half-siblings, red squares are full siblings, and yellow squares are parents-offsprings. The plot was created by Graphical Relationship Representation (19).

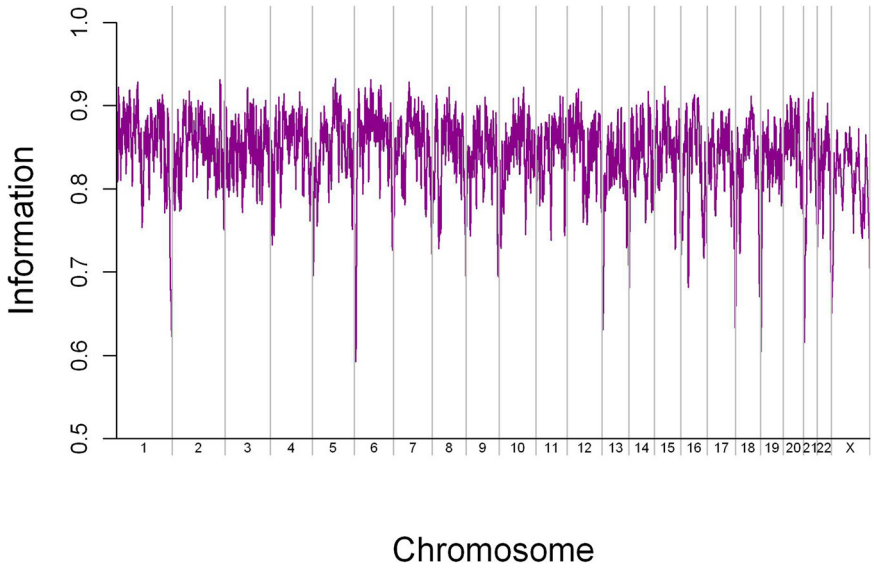

Fig. S4. Genome-wide information content.

Table S1. Summary of informative families

	Narrow	Broad
Affected individuals	110 (37%)	126 (42%)
Unaffected individuals	192 (63%)	176 (58%)
Affected females	44 (40%)	50 (40%)
Affected males	66 (60%)	76 (60%)
Families with:		
≥1 affected individual	65	69
≥2 affected individuals	44	53
≥3 affected individuals	8	12
Average affected individuals per family with ≥ 2 affected individuals	2.23	2.28
Number of affected relative pairs	67	87

Table S2. Identified rare inherited copy number mutations

Chr	art	End	Cytoband	CNV ID	Proband Sex	Size, kb	CN change	Genes involved	Overlap with ISC study [a]	Parental Origin
Chr1	144190576	144439040	1q21.1	CNV-1	M	248.46	Gain	10	1 case / 3 controls	Father
Chr1	150141044	150251582	1q21.3	CNV-2	M	110.54	Gain	2	None	Mother
Chr1	176050752	176201904	1q25.2	CNV-3	F	151.15	Gain	*SEC16B*	None	Mother
Chr1	246404508	246701712	1q44	CNV-4	M	297.2	Loss	11	None	Father
Chr2	61103180	61294012	2p15	CNV-5*	M	190.83	Gain	4	1 case / 0 controls	Mother
Chr3	1553749	1636120	3p26.3	CNV-6*	M	82.37	Gain	≈133 kb 3′ of *CNTN6*	4 cases / 6 controls	Mother
Chr3	2072813	2743794	3p26.3	CNV-7	M	670.98	Gain	*CNTN4*	0 cases / 2 controls	Mother
Chr3	113511195	113779889	3q13.2	CNV-8	M	268.69	Loss	4	None	Father
Chr4	160104841	160821269	4q32.1	CNV-9 *	F	716.43	Gain	2	None	Mother
Chr6	162768919	162902279	6q26	CNV-10	F	133.36	Loss	*PARK2*	6 cases / 1 control	Father
Chr7	16127873	16366347	7p21.1	CNV-11	M	238.47	Loss	*LOC729920*	None	Father
Chr7	75985157	76603404	7q11.23	CNV-12	M	618.25	Gain	5	None	Father
Chr7	157405851	157495760	7q36.3	CNV-13*	F	89.91	Loss	*PTPRN2*	None	Mother
Chr8	2326916	3462554	8p23.2	CNV-14*	F	1135.64	Gain	*CSMD1*	None	Mother
Chr8	87256056	87404313	8q21.3	CNV-15	M	148.26	Gain	*SLC7A13*	5 cases / 4 controls	Father
Chr10	53377917	53463422	10q21.1	CNV-16	F	85.51	Gain	*PRKG1*	0 cases / 1 control	Father
Chr10	83704352	83777944	10q23.1	CNV-17*	M	73.59	Gain	*NRG3*	1 case / 0 controls	Father
Chr13	31151027	31234534	13q13.1	CNV-18	M	83.51	Loss	*RXFP2*	None	Mother
Chr14	40761200	40826162	14q21.1	CNV-19*	F	64.96	Loss	≈320 kb 5′ of *LRFN5*	1 case / 0 control	Mother
Chr16	85842462	85897888	16q24.2	CNV-20	M	55.43	Loss	≈22 kb 3′ of *FBXO31*	None	Mother
Chr20	15000514	15092469	20p12.1	CNV-21*	M	91.96	Loss	*MACROD2*	1 case / 6 controls	Father
Chr21	13267540	14003100	21q11.2	CNV-22	M	735.56	Loss	2	17 cases / 19 controls	Father
Chr21	13687186	13991762	21q11.2	CNV-23*	M	304.58	Gain	2	13 cases / 7 controls	Mother
Chr21	45330407	45373040	21q22.3	CNV-24	M	42.63	Gain	*ADARB1*	1 case / 0 controls	Mother

* denotes CNVs showing co-segregation with disease in respective families (see main text). For non-genic CNVs we list the location of the closest flanking gene. ISC: International Schizophrenia Consortium study: *Nature* 455, 237–241 (2008); Total number of cases used in the ISC study = 3,391; total number of controls = 3,181. [a]A CNV is declared as present in both datasets when ≥ 50% overlap in length between the two CNVs was observed.

Table S3. Genes located within the inherited CN changes identified in familial cases

Gene Symbol	Gene Name	RefSeq ID	Chr	Cytoband	Strand	Start	End	CN change
ANKRD35	Ankyrin repeat domain 35	NM_144698	1	1q21.1	+	144260565	144279883	Gain
CD160	CD160 antigen	NM_007053	1	1q21.1	–	144407154	144426922	Gain
ITGA10	Integrin, alpha 10	NM_003637	1	1q21.1	+	144236346	144255225	Gain
LIX1L	LIX1 homolog (mouse) like	NM_153713	1	1q21.1	+	144188441	144210448	Gain
NUDT17	Nudix (nucleoside diphosphate linked moiety X)-type motif 17	NM_001012758	1	1q21.1	–	144297849	144300792	Gain
PEX11B	Peroxisomal biogenesis factor 11b	NM_003846	1	1q21.1	+	144227739	144235088	Gain
PIAS3	Protein inhibitor of activated STAT, 3	NM_006099	1	1q21.1	+	144287344	144297903	Gain
POLR3C	Polymerase (RNA) iii (DNA directed) polypeptide C (62 Kda)	NM_006468	1	1q21.1	–	144303961	144322241	Gain
RBM8A	RNA binding motif protein 8a	NM_005105	1	1q21.1	+	144218994	144222801	Gain
ZNF364	Zinc finger protein 364	NM_014455	1	1q21.1	+	144322392	144400084	Gain
S100A10	S100 calcium binding protein a10 (Annexin II ligand, Calpactin I, light polypeptide (p11))	NM_002966	1	1q21.3	–	150222009	150233338	Gain
THEM4	Thioesterase superfamily member 4	NM_053055	1	1q21.3	–	150112683	150148737	Gain
SEC16B	Leucine zipper transcription regulator 2	NM_033127	1	1q25.2	–	176164864	176205673	Gain
OR14C36	Olfactory receptor, family 5, subfamily BF, member 1	NM_001001918	1	1q44	+	246578699	246579638	Loss
OR2M2	Olfactory receptor, family 2, subfamily M, member 2	NM_001004688	1	1q44	+	246409910	246410954	Loss
OR2M3	Olfactory receptor, family 2, subfamily M, member 3	NM_001004689	1	1q44	+	246432992	246433931	Loss
OR2M4	Olfactory receptor, family 2, subfamily M, member 4	NM_017504	1	1q44	+	246468853	246469789	Loss
OR2M7	Olfactory receptor, family 2, subfamily M, member 7	NM_001004691	1	1q44	–	246553554	246554493	Loss
OR2T1	Olfactory receptor, family 2, subfamily T, member 1	NM_030904	1	1q44	+	246635918	246637028	Loss
OR2T12	Olfactory receptor, family 2, subfamily T, member 12	NM_001004692	1	1q44	–	246524540	246525503	Loss
OR2T2	Olfactory receptor, family 2, subfamily T, member 2	NM_001004136	1	1q44	+	246682721	246683696	Loss
OR2T33	Olfactory receptor, family 2, subfamily T, member 33	NM_001004695	1	1q44	–	246502776	246503739	Loss
OR2T4	Olfactory receptor, family 2, subfamily T, member 4	NM_001004696	1	1q44	+	246591505	246592552	Loss
OR2T6	Olfactory receptor, family 2, subfamily T, member 6	NM_001005471	1	1q44	+	246617532	246618459	Loss
AHSA2	Aha1, activator of heat shock 90Kda protein ATPase homolog 2 (yeast)	NM_152392	2	2p15	+	61258324	61267562	Gain
KIAA1841	KIAA 1841 protein	NM_032506	2	2p15	+	61146866	61218673	Gain
PEX13	Peroxisome biogenesis factor 13	NM_002618	2	2p15	+	61098373	61129896	Gain
USP34	Ubiquitin specific peptidase 34	NM_014709	2	2p15	–	61268093	61551353	Gain
CNTN4	Contactin 4	NM_175607	3	3p26.3	+	2117246	3074645	Gain
ATG3	ATG3 autophagy related 3 homolog (S. cerevisiae)	NM_022488	3	3q13.2	–	113734048	113763175	Loss
BTLA	B and T lymphocyte associated	NM_001085357	3	3q13.2	–	113665502	113701098	Loss
CD200	CD200 antigen	NM_001004196	3	3q13.2	+	113534605	113564348	Loss
SLC35A5	Solute carrier family 35, member A5	NM_017945	3	3q13.2	+	113763584	113785693	Loss
C4orf45	Hypothetical protein FLJ25371	NM_152543	4	4q32.1	–	160034133	160175783	Gain
RAPGEF2	Rap guanine nucleotide exchange factor 2	NM_014247	4	4q32.1	+	160408447	160500751	Gain
PARK2	Parkinson's disease (autosomal recessive, juvenile) 2, Parkin	NM_013988	6	6q26	–	161688579	163068824	Loss
LOC729920	Notch1-induced protein	NM_001101426	7	7p21.1	–	16097785	16427472	Loss
CCDC146	KIAA1505 protein	NM_020879	7	7q11.23	+	76589869	76762457	Gain
LOC100132832	Similar to cDNA sequence BC021523	NM_001013729	7	7q11.23	–	13973491	13975330	Loss
PMS2L11	Deltex homolog 2	NM_023383	7	7q11.23	+	76448074	76491012	Gain
POMZP3	POM (Pom121 homolog, rat) and ZP3 fusion	NM_152992	7	7q11.23	–	76077238	76094556	Gain
UPK3B	Uroplakin 3B	NM_182684	7	7q11.23	+	75977680	75995135	Gain
PTPRN2	Protein tyrosine phosphatase, receptor type, N polypeptide 2	NM_002847	7	7q36.3	–	157024515	158073179	Loss
CSMD1	Cub and sushi multiple domains 1	NM_033225	8	8p23.2	–	2780281	4839736	Gain
SLC7A13	Solute carrier family 7, (cationic amino acid transporter, y + system) member 13	NM_138817	8	8q21.3	–	87295403	87311720	Gain
PRKG1	Protein kinase cGMP-dependent type I isoform	NM_001098512	10	10q21.1	+	52420950	53725280	Gain
NRG3	Neuregulin 3	NM_001010848	10	10q23.1	+	83625076	84735341	Gain
RXFP2	Relaxin/insulin-like family peptide receptor 2	NM_130806	13	13q13.1	+	31211678	31275009	Loss
MACROD2	Chromosome 20 open reading frame 133	NM_080676	20	20p12.1	+	13924145	15981841	Loss
A26B3	Ankyrin repeat domain 21	NM_174981	21	21q11.2	+	13904368	13935777	Loss
LOC441956	Similar to cDNA sequence BC021523	NM_001013729	21	21q11.2	–	13973491	13975330	Loss
ADARB1	Adenosine deaminase, RNA-specific, B1 (Red1 homolog rat)	NM_001033049	21	21q22.3	+	45318942	45470902	Gain

Table S4. Parametric singlepoint MOD scores > 2.0 for either SCZ status (multiplicative model)

Chr	Marker	Position (cM)	Narrow MOD	Broad MOD
1p36	D1S1612	14	2.34	1.81
2p24	D2S320	41	1.19	2.09
2p23	D2S352	57	2.07	1.71
3q21	D3S1558	129	2.63	2.41
3q21	D3S3646	135	2.64	1.34
9p21	D9S270	54	1.91	2.99
9p13	D9S50	60	2.51	3.56
9q34	D9S2168	164	1.82	2.11
13q33	D13S1265	117	0.77	2.08
13q33–34	D13S285	127	3.30	3.67
17q25	DG17S14	138	2.28	2.17
20p13	D20S113	8	1.12	2.10
21q22	D21S1894	43	2.26	1.69
21q22	D21S1900	46	2.46	1.54
21q22	D21S1919	46	2.70	1.40
21q22	D21S270	46	3.00	1.88

Table S5. Nonparametric singlepoint LOD scores > 2.0 for either SCZ status

Chr	Marker	Position (cM)	Narrow LOD	Broad LOD
1p36	D1S2885	46	2.30	2.87
2p23	D2S352	57	2.06	1.86
2p13	D2S286	102	1.70	2.01
3q21	D3S3646	135	2.68	1.06
9p13	D9S50	60	2.03	2.44
13q33–34	D13S285	127	2.80	2.70
21q22	D21S1900	46	2.30	1.53
21q22	D21S1919	46	2.55	1.24
21q22	D21S270	46	2.31	1.20

Table S6. Nonparametric multipoint LOD scores > 1.5 for either SCZ status

Chr	Position (cM)	Nearest Marker	1-LOD interval	Narrow LOD	Broad LOD
3q21	137	D3S1589	129–147	1.61	0.81
8q11	65	D8S1831	58–78	1.52	1.62
13q33–34	125	D13S261	119-qter	2.65	2.66
21q22	46	D21S1900	38–48	2.16	0.83

1-MOD, region in which the LOD score is within 1 LOD score of the highest LOD score.

Table S7. CNVs identified within 13q34 and 1p36 linkage loci

Index	Chr	Start	End	Cytoband	Size (kb)	Gain/Loss	Freq. cases	Freq. controls
CNR1	13	110478211	110549062	13q34	70.9	Gain	2/224	3/361
CNR2	1	1429542	1464799	1p36.33	35.3	Gain	1/224	3/361
CNR3	1	2445723	2469754	1p36.32	24.0	Gain	2/224	2/361
CNR5	1	4801597	4851425	1p36.32	49.8	Gain	1/224	0/361
CNR6	1	10185627	10275699	1p36.22	90.1	Gain	1/224	1/361
CNR7	1	12791304	13057044	1p36.21	265.7	Gain	1/224	0/361
CNR8	1	16981856	17134270	1p36.13	152.4	Gain	4/224	3/361
CNR9	1	14361606	14374749	1p36.21	13.1	Loss	1/224	0/361