Molecular Psychiatry (2009) 14, 786–795

www.nature.com/mp

ORIGINAL ARTICLE

Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms

PA Holmans[1], B Riley[2], AE Pulver[3], MJ Owen[1], DB Wildenauer[4,5], PV Gejman[6], BJ Mowry[7], C Laurent[8], KS Kendler[2], G Nestadt[3], NM Williams[1], SG Schwab[4,9,10], AR Sanders[6], D Nertney[7], J Mallet[11], B Wormley[2], VK Lasseter[3], MC O'Donovan[1], J Duan[6], M Albus[12], M Alexander[13], S Godard[14], R Ribble[2], KY Liang[15], N Norton[1], W Maier[16], G Papadimitriou[17], D Walsh[18], M Jay[8], A O'Neill[19], FB Lerer[20], D Dikeos[17], RR Crowe[21], JM Silverman[22] and DF Levinson[13]

[1]*Department of Psychological Medicine, School of Medicine, Cardiff University, Cardiff, UK;* [2]*Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA;* [3]*Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA;* [4]*School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, WA, Australia;* [5]*Centre for Clinical Research in Neuropsychiatry, Graylands Hospital, Mt Claremont, WA, Australia;* [6]*Center for Psychiatric Genetics, Evanston Northwestern Healthcare Research Institute, Northwestern University, Evanston, IL, USA;* [7]*Queensland Centre for Mental Health Research, University of Queensland, Brisbane, QLD, Australia;* [8]*Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris, France;* [9]*School of Medicine and Pharmacology, University of Western Australia, Perth, WA, Australia;* [10]*Centre for Medical Research, Western Australian Institute for Medical Research, University of Western Australia, Perth, WA, Australia;* [11]*Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Centre National de la Recherche Scientifique, Hôpital de la Pitié Salpêtrière, Paris, France;* [12]*State Mental Hospital, Haar, Germany;* [13]*Department of Psychiatry, Stanford University, Stanford, CA, USA;* [14]*INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris, France;* [15]*Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA;* [16]*Department of Psychiatry, University of Bonn, Bonn, Germany;* [17]*Department of Psychiatry, University of Athens Medical School, Athens, Greece;* [18]*The Health Research Board, Dublin, Ireland;* [19]*Department of Psychiatry, Queens University, Belfast, Ireland;* [20]*Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel;* [21]*Department of Psychiatry, The University of Iowa College of Medicine, Iowa City, IW, USA and* [22]*Department of Psychiatry, Mt Sinai School of Medicine, New York, NY, USA*

A genomewide linkage scan was carried out in eight clinical samples of informative schizophrenia families. After all quality control checks, the analysis of 707 European-ancestry families included 1615 affected and 1602 unaffected genotyped individuals, and the analysis of all 807 families included 1900 affected and 1839 unaffected individuals. Multipoint linkage analysis with correction for marker–marker linkage disequilibrium was carried out with 5861 single nucleotide polymorphisms (SNPs; Illumina version 4.0 linkage map). Suggestive evidence for linkage (European families) was observed on chromosomes 8p21, 8q24.1, 9q34 and 12q24.1 in nonparametric and/or parametric analyses. In a logistic regression allele-sharing analysis of linkage allowing for intersite heterogeneity, genomewide significant evidence for linkage was observed on chromosome 10p12. Significant heterogeneity was also observed on chromosome 22q11.1. Evidence for linkage across family sets and analyses was most consistent on chromosome 8p21, with a one-LOD support interval that does not include the candidate gene *NRG1*, suggesting that one or more other susceptibility loci might exist in the region. In this era of genomewide association and deep resequencing studies, consensus linkage regions deserve continued attention, given that linkage signals can be produced by many types of genomic variation, including any combination of multiple common or rare SNPs or copy number variants in a region.
Molecular Psychiatry (2009) **14,** 786–795; doi:10.1038/mp.2009.11; published online 17 February 2009

Keywords: genome; human; genotype; humans; schizophrenia/*genetics; genetic predisposition to disease; *linkage (genetics)

Correspondence: Dr DF Levinson, Department of Psychiatry, Stanford University, Suite A-3325, 701 Welch Road, Palo Alto, CA 94305, USA.
E-mail: dflev@stanford.edu
Received 9 October 2008; revised 20 November 2008; accepted 25 November 2008; published online 17 February 2009

Introduction

Two major methods are currently available to scan the genome to detect disease susceptibility loci: genomewide linkage studies (GWLS) and genome-

wide association studies (GWAS). We report here on a GWLS of eight samples of families with multiple cases of schizophrenia (SCZ) using a dense map of single nucleotide polymorphism (SNP) markers, and the companion paper[1] reports on a meta-analysis of 32 SCZ GWLS including the eight samples studied here.

GWLS use hundreds or thousands of DNA markers to detect the broad regions (millions of base pairs) within which there are most likely to be disease susceptibility loci, based on the pattern of within-family correlations between marker alleles and disease. GWAS use hundreds of thousands of SNPs that tag (serve as proxies for) most of the common SNPs in the genome, to identify small regions (tens of thousands of base pairs) likely to harbor susceptibility variants. GWAS can detect loci with much weaker genetic effects if they are due to common SNPs. For common, genetically complex disorders, GWAS have proven more successful than GWLS in producing robust and well-replicated associations.[2] However, there are genetic effects for which GWLS can be more powerful, including loci with multiple rare pathogenic mutations in different families, or several different susceptibility loci in the same region.

The present study is a collaboration of seven research groups using pedigree samples collected by each group[3–11] plus a publicly available sample,[12] totaling over 800 pedigrees with ill individuals in constellations that are informative for linkage analysis. We previously carried out a set of studies of candidate linkage regions.[13–16] We now report on a new genomewide linkage scan of the entire sample. Whereas previously around 70% of these families had been included in published linkage scans using microsatellite markers,[3,5,6,10,11,17–19] we have now scanned all available families using a set of almost 6000 SNP markers genotyped with high accuracy, extracting on average around 90% of the possible linkage information from these pedigrees. In analyses of 707 European-ancestry pedigrees, significant linkage accounting for cross-site heterogeneity was observed on chromosome 10p, and suggestive evidence for linkage on chromosomes 8p, 8q and 12q, as well 9q when non-European families were included.

Materials and Methods

Subjects

The sample is described in Tables 1 and 2. Recruitment by each research group has been previously described.[3–12] Here, affected cases included probands with *Diagnostic and Statistical Manual of Mental Disorders*, third edition, revised diagnoses of SCZ and relatives with SCZ or schizoaffective disorder, which co-segregates with SCZ in families[20] and is often not differentiated reliably from SCZ.[21] Consensus diagnoses were based on information from semi-structured interviews, psychiatric records and family informants.

Genotyping

Genotyping was carried out at the Center for Inherited Disease Research (CIDR) using the Illumina GoldenGate assay[22] to analyze the Illumina version 4.0 linkage marker set of 6008 SNPs. SNPs were excluded by CIDR ($N=53$) based on internal quality control (QC) criteria, and by the investigators ($N=36$) for more than three parent–child inheritance errors or deviation from Hardy–Weinberg equilibrium at $P<0.001$, leaving 5861 autosomal or X chromosome SNPs for analyses. deCODE[23] map locations were provided by Illumina. There were 0.09% missing genotypes, 0.12% Mendelian inconsistencies before QC checks (0.00138% in the analyzed SNPs) and 0.002% discordant genotypes in 224 blind duplicate specimens. There were 132 DNAs excluded for poor performance in genotyping of a preliminary forensic panel or the full SNP panel; and 60 for inconsistency with reported sex, Mendelian inconsistencies greater than 0.5% or sample call rates less than 98%.

Pedigree relationship analyses

Pairwise identity-by-descent (IBD) proportions were analyzed for all pairs of subjects using PLINK,[24] and differences between specified and actual relationships within families were analyzed using PREST.[25] As a result, 50 DNAs were excluded to resolve pairs of identical specimens, three families were excluded because genotypic relationships did not fit the family and eight because the same family was found in two different samples (JHU-NIMH, JHU-ENH, ENH-NIMH or Cardiff-VCU). Pedigree structures were also corrected (for example, half-sib vs full-sib relationships) as required.

Because of the high accuracy of genotyping with a dense SNP map that facilitated analysis of relationships, and the enlarged samples, the present data replace previous analyses of candidate regions by this collaboration for this narrow phenotypic model.[13–16]

Assignment of families to ancestry subsamples

Families were assigned to predominantly European (EUR), African-American or African-European (AFR) or 'other' (OTH) groups based on STRUCTURE[26] analysis of 49 independent autosomal SNPs that had large allele frequency differences between ancestries (0.5–0.69 EUR vs AFR, 0.3–0.47 for EUR vs OTH) in this sample based on investigator-reported ancestries, or based on public databases. EUR or AFR families had an estimated 70% or more ancestry from that group, otherwise the family was considered OTH. Members of these groups had a mean 98 or 96% ancestry, from that group. Analyses were carried out for EUR families and then for ALL families (using allele frequencies estimated separately for EUR, AFR and OTH groups[27]).

Statistical analyses

The planned primary multipoint linkage analysis of EUR families used the S_{PAIRS} statistic ($Z_{\mathrm{LikelihoodRatio}}$ and its Kong–Cox equivalent LOD score[28] under the

Table 1 Clinical sample

Site	*European ancestry*				*African ancestry*				*Other ancestry*				*All families*			
	Fam	*Aff*	*UA*	*All*	*Fam*	*Aff*	*UA*	*All*	*Fam*	*Aff*	*UA*	*All*	*Fam*	*Aff*	*UA*	*All*
Australia–USA	49	114	126	240	8	24	12	36	2	5	10	15	59	143	148	291
Bonn/Perth	96	212	189	401	0	0	0	0	0	0	0	0	96	212	189	401
Cardiff	113	239	88	327	0	0	0	0	0	0	0	0	113	239	88	327
ENH/Northwestern	51	121	115	236	2	4	4	8	0	0	0	0	53	125	119	244
Paris/CNRS	29	75	68	143	8	25	29	54	26	86	91	177	63	186	188	374
Johns Hopkins	124	295	429	724	8	20	21	41	1	2	2	4	133	317	452	769
NIMH-SGI	61	142	120	262	39	107	62	169	6	12	6	18	106	261	188	449
VCU/Irish	184	417	467	884	0	0	0	0	0	0	0	0	184	417	467	884
Total	707	1615	1602	3217	65	180	128	308	35	105	109	214	807	1900	1839	3739

Sample sizes are shown for European ancestry (abbreviated as EUR in the text), African-American or African-European (AFR) and Other ancestry (OTH) pedigrees included in the combined analysis (ALL families). Shown are the numbers of families, genotyped affected and unaffected individuals in these families, and total numbers of DNA specimens.
Approximately 70% of these subjects and pedigrees had been included in previous published genome scan analysis (see text), using less informative marker sets.

exponential model) computed with MERLIN[29] using MERLIN's correction for linkage disequilibrium (LD) within clusters of markers based on a threshold of r^2 greater than 0.05 for consecutive pairs of markers.[30] Before analyses, unlikely genotypes were detected and excluded using MERLIN.[31]

Additional analyses included: (1) S_{PAIRS} analysis of ALL families using ALLEGRO 2.0[32] (analyzing each ancestry subset with it own allele frequencies) with a 'no-LD map' of 4365 autosomal and X chromosome SNPs with no marker–marker r^2 greater than 0.05 (because ALLEGRO cannot correct for LD); (2) the Kong–Cox exponential S_{ALL} statistic that gives more weight to larger families (results were similar and are not shown here, but are included in online Supplementary Files); (3) parametric heterogeneity LOD score (hLOD) analysis under dominant (risk allele frequency = 0.05; penetrances = 0, 0.001 and 0.001) and recessive (risk allele frequency = 0.1, penetrances = 0, 0 and 0.001) models, using MERLIN for EUR and ALLEGRO (and the no-LD map) for ALL families; (4) logistic regression analysis of IBD sharing[33,34] to assess heterogeneity across sites, linkage while accounting for heterogeneity, effects of parent-of-origin of each allele and of sex of the affected pair (M-M, M-F, F-F), and interactions between linkage regions (see online Supplementary Table 5 for description).

Thresholds for significant (0.05 or fewer peaks expected genomewide per genome scan) and suggestive (less than 1 peak per scan) evidence for linkage were determined by simulation for nonparametric and parametric analyses using data generated under the assumption of no linkage. 'Peaks' were defined as local maxima at least 30 cM from another peak. The empirical threshold for parametric analysis was corrected for two tests by taking the maximum result of the dominant and recessive analyses of simulated replicates at each point.

Data sharing
All genotypic data for this study will be made available to qualified scientists by the NIMH Center for Genetic Studies (nimhgenetics.org).

Results

Nonparametric and parametric linkage analyses
The empirical LOD or hLOD thresholds for suggestive linkage were 1.94 for nonparametric and 2.21 for parametric tests, or 3.26 and 3.66 for significant linkage. Mean information content was 0.88 (s.d. 0.026) using MERLIN's entropy measure (reflecting potential information with fully informative markers) and 0.908 (s.d. 0.028) using ALLEGRO's exponential measure (measuring potential information given the constellation of genotyped relatives).

Figure 1 shows Kong–Cox LODs and dominant and recessive hLOD scores for EUR and ALL families. Table 2 lists the maximum LOD and hLOD scores in each analysis on each chromosome. The nonparametric analysis of EUR families, considered the primary analysis here, produced suggestive evidence for linkage on chromosome 8p21 (in EUR families, LOD = 2.00, 45.9 cM; in ALL families, LOD = 2.51, 46.4 cM, with the latter, larger score at 26.61 Mb). The dominant and recessive analyses were considered an alternative approach, and suggestive evidence for linkage (taking both tests into account as noted above) was observed on chromosomes 8p21, 8q24.1, 9q34 and 12q24.1 in nonparametric and/or parametric analyses (see Table 1 for details). Evidence for linkage was most consistent for chromosome 8p21 (five of the six analyses).

Heterogeneity and linkage allowing for heterogeneity
Table 3 shows the results of the logistic regression analysis of linkage while allowing for intersite heterogeneity, in EUR families. Highly significant

Table 2 Maximum nonparametric and parametric LOD scores on each chromosome in European-ancestry and ALL families

	European-ancestry families									*ALL families*								
Chr	*Kong–Cox (nonpar)*			*Dominant*			*Recessive*			*Kong–Cox (nonpar)*			*Dominant*			*Recessive*		
	cM	*LOD*	*SNP*	*cM*	*hLOD*	*SNP*	*cM*	*hLOD*	*SNP*	*cM*	*LOD*	*SNP*	*cM*	*hLOD*	*SNP*	*cM*	*hLOD*	*SNP*
1	124.1	1.24	rs1517432	254.7	1.13	rs528011	124.1	0.91	rs1517432	122.4	0.91	rs508020	254.7	0.77	rs528011	47.1	0.83	rs8559
2	206.6	1.88	rs1396828	206.6	1.48	rs1396828	233.3	1.75	rs1435850	118.1	1.52	rs1026220	118.1	1.13	rs1026220	182.1	1.81	rs920557
3	64.6	0.37	rs1405796	18.3	0.42	rs1504034	9.2	0.13	rs2290610	64.6	0.33	rs1405793	28.3	0.47	rs749477	9.3	0.23	rs1153459
4	56.3	0.84	rs12142	66.1	0.67	rs1866989	180.6	1.12	rs724659	202.6	0.68	rs996026	65.1	0.71	rs969992	181.8	1.09	rs335077
5	166.2	0.96	rs1299048	161.5	1.69	rs728693	168.8	0.93	rs878953	161.9	1.15	rs949602	161.0	2.03	rs1432812	164.4	1.13	rs9216
6	33.0	0.67	rs1891284	21.7	0.47	rs767022	167.4	0.76	rs1866896	77.9	0.73	rs1409104	77.9	0.66	rs1409104	100.7	1.07	rs1488318
7	12.9	0.26	rs758263	8.7	0.33	rs1470539	78.8	0.61	rs517258	12.3	0.32	rs558030	8.6	0.36	rs1470539	78.4	0.46	rs2009526
8	**45.9**	**2.00**	**rs1561817**	**46.3**	**2.65**	**rs1561817**	**133.9**	**2.37**	**rs901592**	**46.4**	**2.51**	**rs9797**	**46.8**	**2.76**	**rs9797**	133.9	1.98	rs901592
9	146.5	1.69	rs886017	146.8	1.85	rs12335	143.3	1.81	rs10901140	146.5	1.91	rs886017	**146.5**	**2.75**	**rs886017**	**145.6**	1.77	rs456396
10	169.3	0.85	rs1536087	43.9	1.61	rs1339048	68.6	1.85	rs1822861	52.9	0.74	rs332188	42.8	1.26	rs729245	68.0	1.41	rs1822861
11	126.4	0.17	rs668183	90.0	0.13	rs1278402	124.9	0.16	rs596437	126.1	0.14	rs668183	90.0	0.20	rs948142	124.5	0.33	rs665035
12	126.7	1.39	rs233722	**126.7**	**2.25**	**rs233722**	130.0	1.14	rs1920586	119.0	1.55	rs1862032	126.2	1.87	rs737280	130.4	1.17	rs1920568
13	58.1	0.53	rs301653	12.5	0.85	rs6490970	127.8	0.78	rs755992	122.0	0.36	rs1894758	51.2	0.60	rs1853987	127.8	0.51	rs912007
14	14.8	0.52	rs4982599	14.8	0.63	rs4982599	58.0	1.06	rs999881	72.1	0.82	rs7155380	14.8	0.58	rs4982599	57.5	1.62	rs999881
15	58.6	1.39	rs383902	68.3	1.21	rs2439378	68.7	0.86	rs745103	68.3	1.63	rs2439378	68.3	1.63	rs2439378	68.7	1.68	rs745103
16	84.7	1.51	rs149156	85.1	0.84	rs1177648	111.9	1.49	rs1387370	85.6	1.47	rs1541979	116.4	1.23	rs2052904	111.1	1.49	rs723919
17	136.9	0.48	rs599314	1.0	0.50	rs7813	132.4	0.82	rs1062935	120.8	0.68	rs454138	17.4	0.46	rs1443417	124.0	1.21	rs1552173
18	97.5	0.18	rs1539964	97.9	0.21	rs1539964	77.6	0.52	rs732982	97.8	0.36	rs1539964	97.8	0.51	rs1539964	77.5	0.48	rs732982
19	72.5	0.62	rs1603	14.3	0.58	rs352500	44.0	0.61	rs273265	66.5	0.39	rs268666	14.3	0.51	rs352500	42.9	0.54	rs4808095
20	114.6	0.46	rs379042	101.6	0.69	rs1570160	101.6	0.47	rs1570160	108.7	0.65	rs6587239	101.4	0.73	rs1570160	101.4	0.41	rs1570160
21	61.0	0.65	rs875060	61.4	0.99	rs875060	61.4	0.46	rs875060	50.8	0.62	rs2837121	50.8	0.75	rs2837121	41.0	0.15	rs1892687
22	63.1	1.55	rs2399153	74.5	1.93	rs6520165	71.4	1.00	rs739240	73.3	1.52	rs137930	74.5	1.43	rs6520165	71.4	0.97	rs739240
X	19.3	1.08	rs1656651	14.5	1.51	rs1852456	15.4	1.37	rs1869588	18.9	1.22	rs768567	14.2	1.67	rs1852456	15.4	1.79	rs1869588

Shown are multipoint linkage scores computed with MERLIN for EUR families, correcting for marker–marker linkage disequilibrium; and with ALLEGRO for ALL families using SNPs with minimal LD ($r^2 > 0.05$) and separate allele frequency estimates for EUR, AFR and OTH families. Bold scores exceed empirical suggestive evidence for linkage (1.94 for Kong–Cox; 2.21 for dominant and recessive analyses taking the two tests into account). Chr, chromosome; cM, location of the peak score (centimorgans); LOD, LOD score; hLOD, heterogeneity LOD; SNP, assayed single nucleotide polymorphism closest to each score; Kong—Cox, equivalent LOD for a *Z*-likelihood ratio score computed with the Kong–Cox exponential model. For the four regions with suggestive evidence for linkage, the physical position of the largest LOD and 1-LOD support interval in cM and Mb (megabases) were: 8p: 27.61 Mb, 37.1–49.8 cM, 21.37–29.36 Mb; 8q: 128.06 Mb, 130.1–141.6 cM, 126.84–132.63 Mb; 9q: 133.00 Mb, 143.0–151.7 cM, 131.96–134.55 Mb; 12q: 111.08 Mb, 117.3–133.6 cM, 104.07–115.64 Mb.

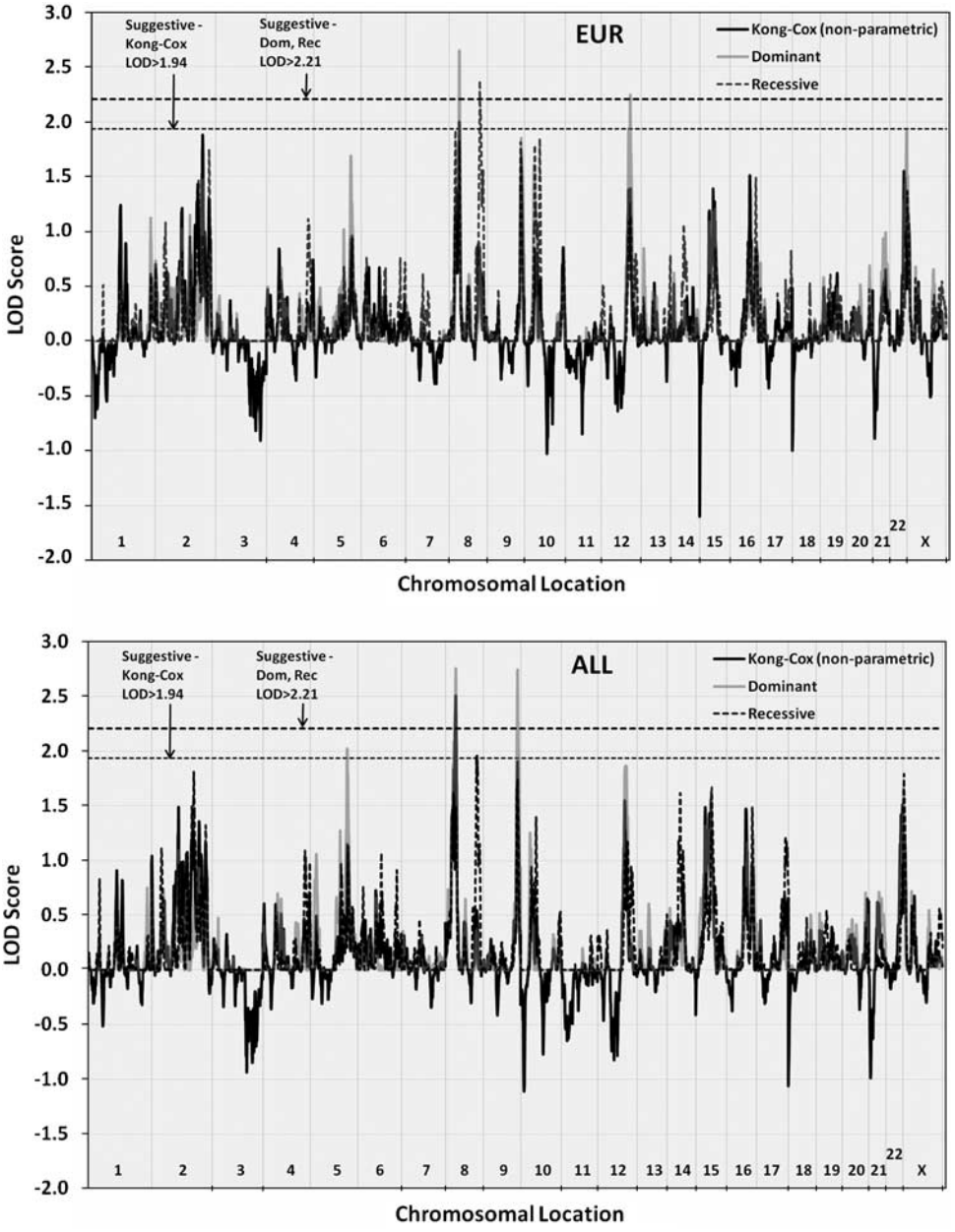

Figure 1 Genomewide linkage results. Shown for 707 European-ancestry families (top) and for all 807 families (bottom) are linkage results across the genome. The *x* axis values are cumulative chromosomal locations in centimorgans (deCODE map), with chromosome boundaries shown as vertical gridlines. The *y* axis values are Kong–Cox LOD scores for nonparametric analyses, or heterogeneity LOD (hLOD) scores for parametric analyses. Black solid lines represent nonparametric LOD scores, gray lines (red in the online html version) hLOD scores under a dominant model and dashed lines (purple online) hLOD scores under a recessive model (see text for details of the models). Dotted lines show the empirical thresholds for genomewide suggestive evidence for linkage (less than 1 peak of this magnitude expected by chance, in the absence of linkage) for nonparametric and parametric analyses. The parametric threshold takes into account that two such tests (dominant and recessive) were performed.

genomewide evidence for linkage with heterogeneity was observed on chromosome 10p12 (see Table 3 legend for additional details). Chromosome 8p21 again produced suggestive evidence for linkage both with and without heterogeneity in this analysis. Results of tests for intersite heterogeneity (that is, the difference between LODs with and without allowing for heterogeneity) are shown in online Supplementary Table 1. Significant heterogeneity was observed on chromosomes 10p (45.6 cM) and 22q11.1 (0 cM).

Supplementary Online Files provide details of parametric and nonparametric linkage scores, genetic location and information content for each analyzed point for each analysis in the entire sample for EUR and ALL families, as well as the full and No-LD

Table 3 Logistic regression analysis of linkage allowing for intersite heterogeneity

Chr	*Linkage (homogeneity)*			*Linkage allowing for heterogeneity*		
	Max LOD	*cM*	*No. of GW peaks*	*Max LOD*	*cM*	*No. of GW peaks*
1	1.14	122.4	6.96	2.81	47.1	14.76
2	1.51	116.5	3.36	4.47	152.1	1.66
3	0.63	10.3	19.86	3.71	15.1	4.76
4	0.81	177.6	13.84	2.69	179.3	17.01
5	0.84	160.9	13.06	3.46	130.6	6.46
6	0.9	33	11.55	3.12	183.5	10.08
7	0.53	10	24.53	2.06	86.1	34.61
8	**2.95**	**46.3**	**0.15**	**5.11**	**46.5**	**0.74**
9	1.41	146.5	4.02	4.1	2.5	2.74
10	1.46	45.6	3.69	***8.32***	***45.6***	***0.001***
11	0.15	63.3	57.57	2.41	63.3	23.62
12	1.71	119.1	2.2	3.08	119.1	10.55
13	0.41	12.4	31.63	2.94	128.1	12.56
14	0.6	14.2	21.27	3.16	47.8	9.58
15	1.13	41.4	7.08	4.06	42.2	2.91
16	1.26	83.6	5.48	3.48	115.4	6.34
17	0.63	136.9	19.86	1.97	0.4	38.06
18	0.36	78.5	35.48	2.2	78.5	29.65
19	0.82	73	13.56	3.13	73	9.97
20	0.44	60.9	29.69	3.9	60.9	3.64
21	1.08	55	7.86	2.88	49.7	13.52
22	2.08	73.7	1.06	4.81	72.6	1.09

Shown are maximum LOD scores on each autosome, by logistic regression analysis of IBD allele sharing of affected relative pairs, assuming either homogeneity across the 8 sites (no covariates) or allowing for heterogeneity (covariates for sites, 7 d.f.). 'No. of GW peaks' refers to the number of peaks (at least 30 cM apart) of this size observed genomewide (autosomes) in 20 000 simulated replicates of chromosome 22 followed by adjustment for autosomal genome length. Results allowing for heterogeneity on chromosome 10p12 (bold italics; at 21.275 Mb, near rs 893882) would be expected by chance once per thousand genome scans, indicating genomewide significance. Results on chromosome 8 would be expected by chance less than once per genome scan (bold text suggestive evidence for linkage). Online Supplementary Table 2 lists IBD proportions and LOD scores by site for the peaks on chromosomes 10p and 8p.

marker maps. Online files for the companion meta-analysis paper[1] provide ranked results for each of our eight samples separately and for EUR and ALL families separately.

Other analyses

No significant chromosomewide effects were observed for sex of the affected pair or parent of origin (online Supplementary Tables 3 and 4). Online Supplementary Table 5 shows results of interaction analyses for all pairs of 18 regions with Kong–Cox LOD scores greater than 1. No genomewide significant interactions were observed. The online table legend includes a list of the most significant empirical interaction *P*-values.

Discussion

Suggestive evidence for linkage was detected on chromosome 8p21 in multiple analyses: in nonparametric, dominant and recessive analyses of 707 European-ancestry families, and in nonparametric and dominant analyses of all 807 families.

This same region produced suggestive evidence for linkage (and the largest peak), in the independent Molecular Genetics of Schizophrenia (MGS) sample[35] of 409 European-ancestry and African-American families. Our peak results were between 45.9 and 46.8 cM (between rs1561817 and rs9797, 26.59–27.65 Mb; deCODE linkage map and genome build 36.3 physical locations). The MGS peak was at 43.3 cM for all families (near rs196886 at 24.79 Mb), whereas in European-ancestry families it was at 15.3 cM (8p23, near rs7834209 at 6.9 Mb), with a slightly smaller peak at 34.6 cM (8p21, near rs34393111 at 20.28 Mb), and suggestive evidence for linkage extended beyond our peak scores. Pulver *et al*.[36] were the first to report preliminary and then strongly suggestive evidence[10] for linkage of SCZ to chromosome 8p markers in much of the JHU sample that is included here. We previously reported support for 8p linkage in a study of microsatellite markers in a majority of the families in the present analysis,[13] consistent with results in this enlarged sample.

The most widely studied 8p candidate gene is *NRG1* (neuregulin 1), found to be associated with SCZ by Stefansson *et al*.[37] in an LD mapping study of a

suggestive linkage peak observed in Icelandic families (there were two 8p peaks in that analysis, with the second one closer to ours), with supportive evidence in some data sets.[38] There are several indications that if there is linkage on chromosome 8p it is not entirely explained by *NRG1*. Here, LOD scores within one unit of the maximum (1-LOD interval) were observed between 21.37 and 29.36 Mb, whereas *NRG1* is between 32.53 and 32.74 Mb. (The 1-LOD interval is a reliable confidence interval in studies of Mendelian disorders, but not for complex disorders.) In the companion meta-analysis paper,[1] the second 'bin' on chromosome 8 (8.2, 28.1–56.2 cM, ~15.7–33 Mb) produced the strongest (suggestive) evidence for linkage in 22 European-ancestry data sets, and was ranked eighth in the analysis of all 32 data sets. *NRG1* is at the centromeric edge of that bin (~55.7 cM), so one would expect that if it explained the linkage, the signal would extend equally in the centromeric and telomeric directions, but support for linkage was not observed in more centromeric bins (bin 8.3 in the primary analysis, from 56.2 to 84.3 cM; bin 8.4 in the '20 cM' analysis from 56.2 to 75 cM; or bin 8.3 in the '30 cM' shifted analysis from 42.15 to 70.25 cM).[1] We hypothesize that there is weak linkage to SCZ on chromosome 8p, due to one or more loci in which there are multiple rare risk-associated SNPs and/or structural variants and/or multiple associated common SNPs. There are other candidate genes on 8p (see discussion in the meta-analysis paper[1]), but it is not yet clear what accounts for the evidence for linkage in this region.

Suggestive evidence for linkage was observed on chromosome 9q in the dominant analysis of all families. Support for this region in other analyses was modest, but not substantially different than the evidence for 8p. This region is not supported by previous linkage findings or the meta-analysis.[1]

Genomewide significant evidence for linkage allowing for intersite heterogeneity was observed on chromosome 10p12 at 45.6 cM (21.28 Mb). We previously reported modest evidence for heterogeneity in this region,[14] and in that report we also reviewed the evidence for 10p linkage reported previously in the NIMH-SGI, VCU/Ireland and part of the Bonn/Perth samples studied here. A significant signal is now seen in the present expanded sample, with a denser marker map, due to allele sharing in the Paris/CNRS, NIMH-SGI and (to a lesser degree) the VCU samples (online Supplementary Table 2). There is no indication of a high-penetrance signal from a small subset of families: the NIMH sample includes small nuclear families from the general US population; and although there are some large, extended pedigrees in the Paris/CNRS sample from La Réunion Island, most of the families with positive LOD scores were small families from the general French population, and no single family had a LOD score (Kong–Cox, dominant or recessive) greater than 1.4. Because we combined families from eight previously collected data sets, we do not have a consistent set of clinical ratings across samples to search for a possible clinical basis for linkage heterogeneity. The 10p peak is not supported by meta-analysis,[1] and is far from the chromosome 10q peaks observed between 100 and 110 cM in two independent studies.[39,40]

Significant heterogeneity (but not linkage with heterogeneity) was seen on chromosome 22q at 15 Mb, adjacent to the typical region (17–21 Mb) of the 22q11 deletion syndromes whose manifestations include SCZ in approximately 20% of cases.[41] This deletion was detected in less than 0.5% of SCZ cases in two recent large studies.[42,43] No consistent association signals have been observed to date between SCZ and common SNPs in candidate genes within the deletion region.

Two other regions, on chromosomes 8q24.1 and 12q24.1, produced suggestive evidence for linkage in at least one analysis, both reportedly linked to mood disorders rather than SCZ. On 8q, a combined analysis of genotypes from 11 linkage scans (1067 families) produced a nonparametric LOD score of 3.40 at 134.5 Mb, just telomeric to our 1-LOD interval, in an analysis of bipolar I and bipolar II cases, but the signal was much smaller in an analysis of only bipolar I.[44] Given that by definition only bipolar I can include psychosis (usually in around half of cases), one would not predict that the same locus in this region would account for linkage signals to bipolar disorder and SCZ. On chromosome 12q, there have been reports of linkage to major depressive[45,46] and bipolar disorders (see review by Barden *et al*.[47]) with peak locations ranging from 97.4–126.5 Mb to 116–126 Mb in bipolar studies, close to our peak at 111 Mb. Neither region was supported by the SCZ linkage meta-analysis.[1]

In the linkage meta-analysis,[1] genomewide significant evidence for linkage was detected on chromosome 2q (132–162 cM, 121–152 Mb), with some support for linkage across a broad region (118–176 and 206–235 cM). In the present study, we see a jagged line across chromosome 2q (Figure 1), reflecting diverse peaks in different samples, although without statistically significant evidence for heterogeneity. Our largest peak was in the nonparametric EUR analysis at 206.6 cM (210.87 Mb). Thus, in our data and in the meta-analysis of 32 data sets, linkage evidence on 2q is intriguing but poorly localized. It was recently reported that an SNP in ZNF804A, at 185 Mb on 2q, produced genomewide significant evidence for linkage when a large collaborative SCZ association sample was combined with bipolar disorders cases from the Wellcome Trust Case Control Consortium project.[48]

What is the relevance of linkage studies as the field moves on to GWAS and large-scale resequencing methods? Meta-analysis provides some support for quite modest linkage signals.[1] Thus, no gene is likely to have a large effect on overall population risk. In this situation, GWAS methods have better power,[2] but (currently) only for common SNPs. GWAS technologies can also detect some but not all copy number variants (CNVs). Recent studies suggest that rare

deletions on chromosomes 1q and 15q (as well as 22q11) predispose to SCZ;[42,43,49,50] and that SCZ cases also have a small but significant excess of very rare CNVs, some of which might therefore also be pathogenic. These findings support the more general hypothesis of multiple rare genomic events (SNPs, CNVs, other structural changes) influencing risk for a common disease.[51–53]

High-penetrance CNVs like those on 1q and 15q have effects such as mental retardation and/or autism, consistent with the observation that they reduce fertility and thus are usually *de novo* mutations rather than transmitted in families. But most SCZ risk variants probably have smaller effects: the risk to probands' siblings is around 5%,[20] and if one allows for a small proportion of cases to be due to high-penetrance CNVs, the remaining risk should be due to lower-penetrance variants that would thus be transmitted in families. It is possible that weak SCZ linkage signals are in regions where there are multiple rare as well as common risk variants, whose aggregate frequency and effects are sufficient to produce a linkage signal, and whose effects on fertility are not too severe. We refer here to both deleterious transmitted and/or recurring sequence and structural polymorphisms with low population frequencies, and to very rare and thus very deleterious variants that segregate in different families, that is, extreme allelic heterogeneity.

One approach to finding these variants would be high-throughput resequencing studies of linkage regions. For example, significant differences have been found in the proportions of high- and low-risk individuals carrying very rare nonsynonymous coding SNPs for some diseases.[54–55] This approach has not yet been attempted for SCZ in a large sample, thus we lack information to predict the power or optimal design of such studies. If a region in fact contained a sufficient number of rare high-risk variants to produce a linkage signal, then it might be possible to detect them by resequencing, although success would depend on the proportion of subjects of families carrying such variants, and by the extent of locus heterogeneity, that is, if a small proportion of cases carried rare risk variants at a large number of loci in a linkage region, studies of a feasible sample size might not detect them. It is not known whether it will prove most productive to resequence exons, entire genes with their nearby regulatory regions or entire linkage regions (given that there are likely to be relevant unannotated intergenic regulatory sequences). Family-based samples might be particularly useful for resequencing studies of linkage peaks, if rare variants were contributing to the signal. But it is also possible that these variants are rare precisely because they reduce fertility, they could be more easily found in case–control samples, which are also larger. In our view, multiple strategies should be attempted.

It has also been suggested that the power of GWAS can be increased by upweighting evidence for association based on linkage scores (resulting in a small downweighting of other regions).[56] Whether or not this formal approach is used, it would be reasonable to consider linkage findings when selecting genes and regions for dense LD mapping and large-scale resequencing studies.

Acknowledgments

We thank the many family members who participated in the studies that recruited these samples. This work was supported by National Institute of Mental Health grants 7R01MH062276 (to DFL, CL, MO and DW), 5R01MH068922 (to PG), 5R01MH068921 (to AEP) and 5R01MH068881 (to BR). For the NIMH sample, data and biomaterials were collected in three projects that participated in the National Institute of Mental Health (NIMH) Schizophrenia Genetics Initiative. From 1991 to 97, the principal investigators and co-investigators were: Harvard University, Boston, MA, U01 MH46318, Ming T Tsuang, Stephen Faraone and John Pepple; Washington University, St Louis, MO, U01 MH46276, C Robert Cloninger, Theodore Reich and Dragan Svrakic; Columbia University, New York, NY U01 MH46289, Charles Kaufmann, Dolores Malaspina and Jill Harkavy Friedman. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number N01-HG-65403. The NIMH Cell Repository at Rutgers University (Dr Douglas Fugman and Dr Jay Tischfield) and the NIMH Center for Collaborative Genetic Studies on Mental Disorders (Dr John Rice) made important contributions to this project.

References

1 Ng MYM, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T *et al.* Meta-analysis of 32 genomewide linkage studies of schizophrenia. *Mol Psychiatry* 2008; advance online publication 30 December 2008; doi:10.1038/mp.2008.135.
2 Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008; **118**: 1590–1605.
3 Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A *et al.* Genome scan of schizophrenia. *Am J Psychiatry* 1998; **155**: 741–750.
4 Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW *et al.* Identification and analysis of error types in high-throughput genotyping. *Am J Hum Genet* 2000; **67**: 727–736.
5 Schwab SG, Hallmayer J, Albus M, Lerer B, Eckstein GN, Borrmann M *et al.* A genome-wide autosomal screen for schizophrenia susceptibility loci in 71 families with affected siblings: support for loci on chromosome 10p and 6. *Mol Psychiatry* 2000; **5**: 638–649.
6 Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA *et al.* A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. *Hum Mol Genet* 1999; **8**: 1729–1739.
7 Cao Q, Martinez M, Zhang J, Sanders AR, Badner JA, Cravchik A *et al.* Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. *Genomics* 1997; **43**: 1–8.

8 Campion D, d'Amato T, Bastard C, Laurent C, Guedj F, Jay M *et al.* Genetic study of dopamine D1, D2, and D4 receptors in schizophrenia. *Psychiatry Res* 1994; **51**: 215–230.
9 Bonnet-Brilhault F, Laurent C, Campion D, Thibaut F, Lafargue C, Charbonnier F *et al.* No evidence for involvement of KCNN3 (hSKCa3) potassium channel gene in familial and isolated cases of schizophrenia. *Eur J Hum Genet* 1999; **7**: 247–250.
10 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al.* Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 1998; **20**: 70–73.
11 Kendler KS, FA ON, Burke J, Murphy B, Duke F, Straub RE *et al.* Irish study on high-density schizophrenia families: field methods and power to detect linkage. *Am J Med Genet* 1996; **67**: 179–190.
12 Cloninger CR, Kaufmann CA, Faraone SV, Malaspina D, Svrakic DM, Harkavy-Friedman J *et al.* Genome-wide search for schizophrenia susceptibility loci: the NIMH Genetics Initiative and Millennium Consortium. *Am J Med Genet* 1998; **81**: 275–281.
13 Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8. Additional support for schizophrenia linkage on chromosomes six and eight: a multicenter study. *Am J Med Genet* 1996; **67**: 580–594.
14 Levinson DF, Holmans P, Straub RE, Owen MJ, Wildenauer DB, Gejman PV *et al.* Multicenter linkage study of schizophrenia candidate regions on chromosomes 5q, 6q, 10p, and 13q: schizophrenia linkage collaborative group III. *Am J Hum Genet* 2000; **67**: 652–663.
15 Levinson DF, Holmans PA, Laurent C, Riley B, Pulver AE, Gejman PV *et al.* No major schizophrenia locus detected on chromosome 1q in a large multicenter sample. *Science* 2002; **296**: 739–741.
16 Mowry BJ, Holmans PA, Pulver AE, Gejman PV, Riley B, Williams NM *et al.* Multicenter linkage study of schizophrenia loci on chromosome 22q. *Mol Psychiatry* 2004; **9**: 784–795.
17 Williams NM, Norton N, Williams H, Ekholm B, Hamshere ML, Lindblom Y *et al.* A systematic genomewide linkage study in 353 sib pairs with schizophrenia. *Am J Hum Genet* 2003; **73**: 1355–1367.
18 Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B *et al.* Genome scan of European–American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. *Am J Med Genet* 1998; **81**: 290–295.
19 Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD *et al.* The NIMH Genetics Initiative Millennium Schizophrenia Consortium: linkage analysis of African–American pedigrees. *Am J Med Genet* 1998; **81**: 282–289.
20 Maier W, Lichtermann D, Minges J, Hallmayer J, Heun R, Benkert O *et al.* Continuity and discontinuity of affective disorders and schizophrenia: results of a controlled family study. *Arch Gen Psychiatry* 1993; **50**: 871–883.
21 Faraone SV, Blehar M, Pepple J, Moldin SO, Norton J, Nurnberger JI *et al.* Diagnostic accuracy and confusability analyses: an application to the Diagnostic Interview for Genetic Studies. *Psychol Med* 1996; **26**: 401–410.
22 Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002; **32S**: 60–61.
23 Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B *et al.* A high-resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
24 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
25 Sun L, Wilder K, McPeek MS. Enhanced pedigree error detection. *Hum Hered* 2002; **54**: 99–110.
26 Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.
27 Freimer NB, Sandkuijl LA, Blower SM. Incorrect specification of marker allele frequencies: effects on linkage analysis. *Am J Hum Genet* 1993; **52**: 1102–1110.
28 Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
29 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
30 Levinson DF, Holmans P. The effect of linkage disequilibrium on linkage analysis of incomplete pedigrees. *BMC Genet* 2005; **30**(Suppl 1): S6.
31 Abecasis GR, Wigginton JE. Handling marker–marker linkage disequilibrium: pedigree analysis with clustered markers. *Am J Hum Genet* 2005; **77**: 754–767.
32 Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; **25**: 12–13.
33 Holmans P. Detecting gene–gene interactions using affected sib pair analysis with covariates. *Hum Hered* 2002; **53**: 92–102.
34 Holmans P, Weissman MM, Zubenko GS, Scheftner WA, Crowe RR, Depaulo Jr JR *et al.* Genetics of recurrent early-onset major depression (GenRED): final genome scan report. *Am J Psychiatry* 2007; **164**: 248–258.
35 Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C *et al.* Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am J Hum Genet* 2006; **78**: 315–333.
36 Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL *et al.* Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am J Med Genet* 1995; **60**: 252–260.
37 Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002; **71**: 877–892.
38 Munafo MR, Thiselton DL, Clark TG, Flint J. Association of the *NRG1* gene and schizophrenia: a meta-analysis. *Mol Psychiatry* 2006; **11**: 539–546.
39 Fallin MD, Lasseter VK, Wolyniec PS, McGrath JA, Nestadt G, Valle D *et al.* Genomewide linkage scan for schizophrenia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 10q22. *Am J Hum Genet* 2003; **73**: 601–611.
40 Faraone SV, Hwu HG, Liu CM, Chen WJ, Tsuang MM, Liu SK *et al.* Genome scan of Han Chinese schizophrenia families from Taiwan: confirmation of linkage to 10q22.3. *Am J Psychiatry* 2006; **163**: 1760–1766.
41 Williams NM, Owen MJ. Genetic abnormalities of chromosome 22 and the development of psychosis. *Curr Psychiatry Rep* 2004; **6**: 176–182.
42 Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**: 232–236.
43 International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
44 McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW *et al.* Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet* 2005; **77**: 582–595.
45 Abkevich V, Camp NJ, Hensel CH, Neff CD, Russell DL, Hughes DC *et al.* Predisposition locus for major depression at chromosome 12q22–12q23.2. *Am J Hum Genet* 2003; **73**: 1271–1281.
46 McGuffin P, Knight J, Breen G, Brewster S, Boyd PR, Craddock N *et al.* Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Hum Mol Genet* 2005; **14**: 3337–3345.
47 Barden N, Harvey M, Gagne B, Shink E, Tremblay M, Raymond C *et al.* Analysis of single nucleotide polymorphisms in genes in the chromosome 12q24.31 region points to *P2RX7* as a susceptibility gene to bipolar affective disorder. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141**: 374–382.
48 O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40**: 1053–1055.

49 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al*. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
50 Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; **40**: 880–885.
51 Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant… or not? *Hum Mol Genet* 2002; **11**: 2417–2423.
52 Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 2007; **80**: 727–739.
53 Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003; **33**(Suppl): 228–237.
54 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; **305**: 869–872.
55 Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB *et al*. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008; **40**: 495##.
56 Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. *Am J Hum Genet* 2006; **78**: 243–252.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)