

ORIGINAL ARTICLE

Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia

T Lencz^{1,2,3}, TV Morgan⁴, M Athanasiou⁵, B Dain⁵, CR Reed⁵, JM Kane^{1,2,3}, R Kucherlapati^{4,6} and AK Malhotra^{1,2,3}

¹Department of Psychiatry Research, The Zucker Hillside Hospital, Division of the North Shore–Long Island Jewish Health System, Glen Oaks, NY, USA; ²The Feinstein Institute for Medical Research, Manhasset, NY, USA; ³Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY, USA; ⁴Harvard Partners Center for Genetics and Genomics, Cambridge, MA, USA; ⁵PGxHealth, a Division of Clinical Data, Inc., New Haven, CT, USA and ⁶Department of Genetics, Harvard Medical School, Boston, MA, USA

Schizophrenia is a strongly heritable disorder, and identification of potential candidate genes has accelerated in recent years. Genomewide scans have identified multiple large linkage regions across the genome, with fine-mapping studies and other investigations of biologically plausible targets demonstrating several promising candidate genes of modest effect. The recent introduction of technological platforms for whole-genome association (WGA) studies can provide an opportunity to rapidly identify novel targets, although no WGA studies have been reported in the psychiatric literature to date. We report results of a case–control WGA study in schizophrenia, examining ~500 000 markers, which revealed a strong effect ($P=3.7 \times 10^{-7}$) of a novel locus (rs4129148) near the *CSF2RA* (colony stimulating factor, receptor 2 alpha) gene in the pseudoautosomal region. Sequencing of *CSF2RA* and its neighbor, *IL3RA* (interleukin 3 receptor alpha) in an independent case–control cohort revealed both common intronic haplotypes and several novel, rare missense variants associated with schizophrenia. The presence of cytokine receptor abnormalities in schizophrenia may help explain prior epidemiologic data relating the risk for this illness to altered rates of autoimmune disorders, prenatal infection and familial leukemia.

Molecular Psychiatry (2007) 12, 572–580; doi:10.1038/sj.mp.4001983; published online 20 March 2007

Keywords: whole-genome association; pseudoautosomal region; colony-stimulating factor; interleukin 3; schizophrenia; cytokine receptor

Introduction

Schizophrenia is a disease with estimated lifetime morbid risk approaching 1% worldwide.¹ Its course is commonly chronic and severely disabling, accounting for nearly 3% of all years lived with disability; amongst individuals aged 15–44, SCZ is the third-leading cause of disability.² Schizophrenia (SCZ) is associated with an increase of at least 50% in mortality rates compared with the general population,³ including a suicide rate of approximately 5%,⁴ resulting in a 10-year average lifespan reduction.² Despite recent advances, pathophysiology remains poorly understood, and current treatments remain unsatisfactory.⁵

Genetic epidemiologic studies have revealed high heritability estimates (~80%) for schizophrenia and elevated relative risk ($\lambda \sim 10$) in first-degree relatives.⁶

The discovery of susceptibility genes for SCZ has proven challenging; however, research to date has identified genes with, at best, modest individual effects. As with other complex diseases, linkage studies have revealed multiple candidate regions with marginal LOD scores spanning large regions of the genome,⁷ whereas studies of individual candidate genes are inherently limited in their scope and may miss unexpected loci of strong effect. By contrast, the development of whole-genome association (WGA) technology provides an opportunity to rapidly identify novel susceptibility genes for complex phenotypes, even with relatively small samples, as recently demonstrated in macular degeneration,^{8,9} and QT interval prolongation,¹⁰ for example. These WGA studies, which examined approximately 100 000 single-nucleotide polymorphisms (SNPs), also demonstrated the importance of subsequent fine mapping and examination of independent cohorts in clarifying and validating initial association signals obtained with WGA. Because WGA studies are unlikely to directly reveal functional variants, our goal in WGA was to identify SNPs in or near genes that could then be independently examined through

Correspondence: Dr T Lencz, Department of Psychiatry Research, The Zucker Hillside Hospital, 75-59 263rd Street, Glen Oaks, NY 11004, USA.

E-mail: lencz@lij.edu

Received 26 January 2007; revised 2 February 2007; accepted 4 February 2007; published online 20 March 2007

deep sequencing. Thus, we report the results of a WGA study in a schizophrenia case–control cohort using a ~500 000 SNP array and data from exonic sequencing in an independent case–control cohort.

Materials and methods

The WGA sample

For the WGA study, patients with schizophrenia-spectrum disorders ($n = 178$, including schizophrenia, $n = 158$; schizoaffective disorder, $n = 13$; or schizophreniform disorder, $n = 7$) were recruited from the inpatient and outpatient clinical services of The Zucker Hillside Hospital, a division of the North Shore–Long Island Jewish Health System. After providing written informed consent, the Structured Clinical Interview for DSM-IV Axis I disorders (SCID, version 2.0) was administered by trained raters. Information obtained from the SCID was supplemented by a review of medical records and interviews with family informants when possible; all diagnostic information was compiled into a narrative case summary and presented to a consensus diagnostic committee, consisting of a minimum of three senior faculty. Healthy controls ($n = 144$) were recruited by use of local newspaper advertisements, flyers and community Internet resources, and underwent initial telephone screening to assess eligibility criteria. The nonpatient SCID (SCID-NP) was administered to subjects who met eligibility criteria, to rule out the presence of an Axis I psychiatric disorder; a urine toxicology screen for drug use and an assessment of the subject's family history of psychiatric disorders were also performed. Exclusion criteria included (current or past) Axis I psychiatric disorder, psychotropic drug treatment, substance abuse, a first-degree family member with an Axis I psychiatric disorder, or the inability to provide written informed consent.

Patients (65 female/113 male) and controls (63 F/81 M) did not significantly differ in sex distribution ($P > 0.05$). All self-identified as Caucasian, non-Hispanic. Population structure was tested by examination of 210 ancestry informative markers (AIMs). AIMs included all SNPs on the array that passed initial quality control procedures (described below) and demonstrated a frequency difference of ≥ 0.5 in comparisons between Caucasian individuals and Asians or African-Americans in publicly available data from Shriver and colleagues.¹¹ Two tests of structure were performed, both of which indicated no significant stratification. First, analysis with the STRUCTURE program¹² confirmed that all subjects were drawn from a single population; second, comparison of cases and controls on allelic frequency across the 210 AIMs revealed no differences beyond those expected by chance.

The sequencing sample

The sample for the sequencing study was drawn from a larger sample of 151 SCZ patients and healthy subjects of various self-reported ethnicities who

provided written, informed consent. All 151 subjects were genotyped at 67 AIMs and analyzed using the STRUCTURE program.¹² The 71 cases (28 F/43 M) and 31 controls (18 F/13 M) who were at least 90% Caucasian were used in the primary case–control analysis; groups did not significantly differ on sex distribution ($P > 0.05$). All cases were diagnosed with schizophrenia and were drawn from a nationwide population of clozapine-treated patients in the United States as part of a study on clozapine-induced agranulocytosis. As a secondary analysis, to provide comparable sample sizes between groups, 41 additional healthy individuals of Asian, Hispanic or Native American ancestry were also examined.

Whole-genome genotyping

For the WGA study, genomic DNA extracted from whole blood was hybridized to two chips containing ~262 000 and ~238 000 SNPs as per manufacturer's (Affymetrix, Santa Clara, CA, USA) specifications. For all preliminary steps (e.g. polymerase chain reactions), patient and control samples were proportionally represented on each 96-well plate, so as to avoid the potential confound of plate artifacts. Genotype calls were obtained using the Bayesian Robust Linear Model with Mahalanobis distance classifier algorithm threshold at 0.5 applied to batches of 100 samples. Quality control procedures followed several steps. First, samples that obtained mean call rates $< 90\%$ across both chips (or $< 85\%$ for a single chip) were rejected. Mean call rate of remaining samples (total $n = 322$) was 97%. Twenty-two of these cases were successfully repeated, and concordance of the two calls (reliability) for each SNP was evaluated. SNPs with > 1 discrepancy were excluded from further analyses. Concordance across the remaining 454 699 SNPs exceeded 99%. An additional 1526 SNPs with call rates < 0.85 across all valid cases were excluded, as were 13 662 SNPs not in Hardy–Weinberg equilibrium ($P < 0.001$) in the control sample, yielding 439 511 SNPs available for analysis.

Whole-genome association statistical analysis

The primary analytic modality involved computation of likelihood ratios ($df = 1$) for the best possible genotypic split (e.g. recessive or dominant models) for each SNP, with the constraint that a minimum of 10 subjects populate each split group (thereby excluding monomorphic and very rare SNPs), yielding 362 188 splits plotted in Figure 1. Allelic χ^2 -squares were also examined for each SNP that passed quality control procedures described above. All WGA statistical analyses were conducted using HelixTree 5.0.2 software (Golden Helix, Bozeman, MT, USA).

Because strict Bonferroni correction is considered overly conservative, a Bayesian formula was applied to obtain ≥ 0.95 posterior probability of a correct inference of association to a particular gene; this approach was advocated by Freimer and Sabatti,¹³ and modified to take into account recent estimates of

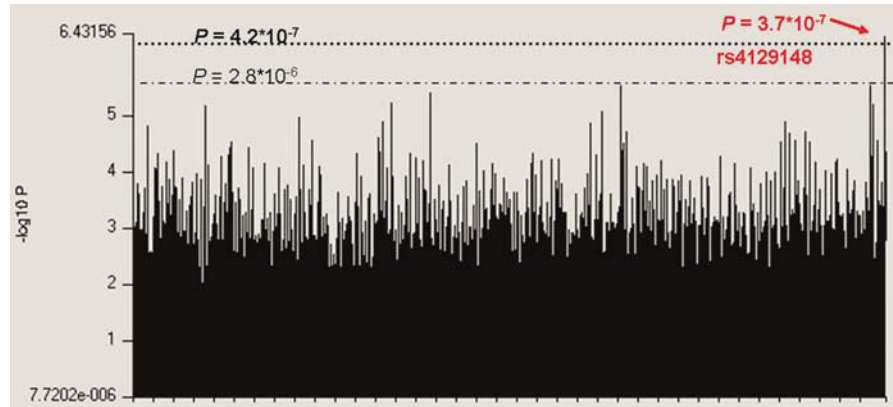


Figure 1 Plot of statistical significance ($-\log_{10} P$) values for successfully genotyped SNPs in the WGA study, arranged in order of chromosomal position (1–22, X). Heavy reference line indicates threshold for genomewide significance ($P = 4.2 \times 10^{-7}$). One SNP (rs4129148) indicated with red arrow crosses this threshold ($P = 3.7 \times 10^{-7}$). Thin reference line indicates P -value of next highest SNP ($P = 2.8 \times 10^{-6}$), separated by approximately one order of magnitude.

the total number of genes in the human genome to approximately 20 000, resulting in a P -value threshold of $\sim 4.2 \times 10^{-7}$. This gene-wise approach is particularly applicable to the current study design, in which the WGA analysis is conducted with the assumption that significant marker(s) are not themselves functional variants but rather indicate neighboring genetic loci to be examined in greater detail. It could be argued that this threshold is also overly conservative, in that it assumes a single genetic locus and does not take into account prior information (e.g. cytogenetic and linkage data) that may provide greater likelihood in particular regions of the genome. However, such information is difficult to quantify¹³ and therefore was not included in deriving the threshold. Finally, it should be noted that empirical testing of all 362 188 P -values from the primary analysis, using the Q -value program,¹⁴ yielded a false discovery rate $q < 0.05$ for the threshold applied in the present study.

Sequencing methods

As described below, WGA data yielded an SNP near the colony stimulating factor 2 receptor alpha (*CSF2RA*) gene which was examined in the independent sequencing sample ($n = 102$). Notably, no HapMap data are available for *CSF2RA* to guide SNP selection; all SNPs analyzed in the sequencing sample are derived from sequencing of exons and their immediate flanking regions. Because *CSF2RA* forms a cytokine receptor gene cluster with its neighboring gene (interleukin 3 receptor alpha; *IL3RA*), we also examined sequencing data for *IL3RA* in the same 102 individuals. The following regions were sequenced: *CSF2RA*, exons 3, 7, 8, 9, 12 and an upstream region; *IL3RA*, exons 1–3, 5–8, 10 and 12 (reference GenBank mRNA accession numbers NM_006140.3 and NM_002183.2, respectively). Fragments (500bp) were amplified using primers displayed in Table 1; note that intronic regions flanking each exon were thus included in the sequencing analysis. Primers were

tailed with M13F for forward primers and with M13R for reverse primers. The sequencing reactions were carried out in both directions using BigDye Terminator Cycle Sequencing, and electrophoresis was run on the ABI Prism 3700 DNA Analyzer, according to standard procedures. Fragments were blasted against the target sequence and polymorphisms were scored for fragments with an average Phred score > 30 . As a control, Mendelian inheritance of each common polymorphism was assessed in one Caucasian family (four grandparents, two parents, four children) and one African American family (two parents, five children) and population specific Hardy–Weinberg equilibrium was determined for 79 unrelated individuals.

Results

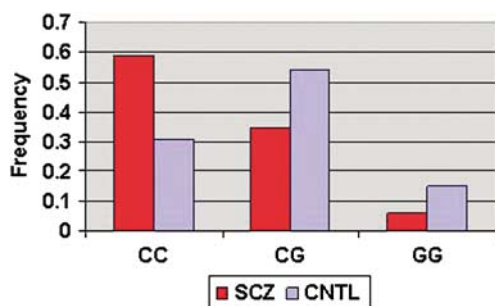
WGA results

As shown in Figure 1, the WGA study yielded one SNP demonstrating an association beyond the genomewide significance threshold: rs4129148, neighboring *CSF2RA*. *CSF2RA* is located in the pseudoautosomal region (PAR1) of the X and Y chromosomes (Xp22.32/Yp11.3). Under a recessive genetic model, homozygosity for the C allele (–strand) was significantly associated with schizophrenia. As shown in Figure 2, 59% of cases (105/178), but only 31% of controls (44/143, one subject not called) were CC homozygotes (odds ratio = 3.23; 95% confidence interval = 2.04–5.15; population attributable risk = 23.5%).

Additional tests were performed to rule out several potential confounds that might artifactually inflate the P -value of rs4129148. First, as noted above, there was no evidence of population stratification; the maximal difference between cases and controls on 210 AIMs was more than four orders of magnitude smaller than that observed for rs4129148. This distinction is important insofar as the HapMap

Table 1 Primers used in the sequencing study

Gene	Region	Direction	Primer sequence
CSF2RA	upstream	Forward	GTGTGGGGATGTGTGTGTGC
		Reverse	CTGAGTAGCTCCCTTCAGAGTTCC
	exon 3	Forward	GCCATGACCCCTCATTTGACC
		Reverse	CTCTGAGCTTCTGGTGGGTTGG
	exon 7	Forward	TTGTTTTCTGATTGCTCTC
		Reverse	GTTTCCCATGTTGGTCAGG
	exon 8	Forward	CCCAGTGTAGTCAAAGAGGTGTCC
		Reverse	CTTCACCTCTTCCCTCCGCTACC
	exon 9	Forward	TCTTGTAGATTCGGGGTTTGTGG
		Reverse	AGTAGGGGACCGTAGGGTTACGC
	exon 12	Forward	CGGCCTAGTTCCTTAAGATTTTCG
		Reverse	GGATACAGATTCCAGCCACACC
IL3RA	exon 1	Forward	TTAATAGAGACGGGGTTTCACC
		Reverse	CAGCTTCGTAACCATGAACC
	exon 2	Forward	AAGCAGCTCCTTCTTCTGTGAGC
		Reverse	CCGACCACTGCATGAGACTCTGCC
	exon 3	Forward	AAAAGAACTCCACCTCCCAAAGG
		Reverse	AGGTGGAGGTTGCAGAGAGC
	exon 5	Forward	GAGATTGGGTTTCTCCATGTTGG
		Reverse	AGTCGAGAGATGTCATCGAAACG
	exon 6	Forward	CTGACCTGCTGGATTTCATGACC
		Reverse	CTCTCCAAGCCAGGAAGAGACC
	exon 7	Forward	GTTTCTTTGTTCCCTTTCCTTCC
		Reverse	AGATTGTTGCCCGTTTCAAGG
exon 8	Forward	CTCATTATTTTGGCCCCACTGC	
	Reverse	CTTTTCCCATCCCTCAATTTGG	
exon 10	Forward	GTGGCACTACTGGGGTGTC	
	Reverse	AGATCGTGCCACTGCACTCC	
exon 12	Forward	CCAAAGTGCTGGGATGACAGG	
	Reverse	AGCTTCGGACGAAATTACACAGG	

**Figure 2** Genotype frequencies for rs4129148 for patients (SCZ) and controls from the WGA study. Note that alleles are reported from the (-) strand.

reveals significant allelic frequency differences between populations for rs4129148. Moreover, the genotype and allelic frequencies in the WGA control sample are comparable to those reported in the HapMap CEU sample, and neither cases nor controls in the WGA sample exhibit genotypic or allelic frequencies comparable to the other HapMap ethnic groups.

Second, because rs4129148 is located on the X/Y chromosomes, case-control comparisons were performed separately for males and females. For both sexes, the recessive model described above significantly differentiated cases and controls ($P \leq 0.001$), with comparable genotypic and allelic frequencies across sexes. Moreover, Hardy-Weinberg equilibrium was observed in both sexes, and only one subject in the entire sample was not called. Thus, it was unlikely that any sex difference in calling of alleles or observation of heterozygosity could account for the significant results for this pseudoautosomal (not X-linked) SNP.

Third, by selecting amongst more than one genotypic model (i.e. dominant and recessive), it is possible that P -values were artificially inflated. Therefore, we separately performed genomewide allelic χ^2 -tests, which again demonstrated that rs4129148 achieved the strongest P -value, with virtually the same strength ($P = 4.4 \times 10^{-7}$) as the genotypic P -value for the recessive model. Moreover, examination of the distribution of P -values for all allelic tests across the genome demonstrated no excess of significant results. Examining all 439 511

SNPs that passed quality control procedures, 44 crossed a significance threshold of $P < 0.0001$, 429 were significant at $P < 0.001$, and 4346 were significant at $P < 0.01$; these results demonstrate no inflation of the family-wise error rate.

Sequencing results

Next, we sought to extend the WGA finding by conducting two sets of analyses in the sequencing sample: (1) haplotype analysis of common SNPs (minor allele frequency ≥ 0.10); and (2) frequency analysis of all rare missense variants. Common SNPs were examined using Haploview 3.32¹⁵ for patterns of linkage disequilibrium (LD) and identification of haplotype blocks. Figure 3 demonstrates the patterns of LD for common variants in *CSF2RA* and *IL3RA*. Three haplotype blocks were identified with solid spines of $D' > 0.70$.

Our case-control analysis followed a hierarchical approach, to limit multiple testing concerns. First, we performed global χ^2 -tests for each of the three haplotype blocks, controlling overall alpha at 0.05 using the false discovery rate method of Benjamini and Hochberg.¹⁶ For each significant haplotype block, we then conducted planned *post hoc* analyses to determine which haplotypes (and SNPs) within the block were accounting for the signal. Haplotype block 1, consisting of three common SNPs in *CSF2RA* intron 8, was significantly associated with schizophrenia (global $P = 0.016$, Table 2). As shown in Table 2, the CGC haplotype within this block was significantly under-represented in patients ($P = 0.0017$), and two SNPs in the block were individually associated with SCZ (rs28694718, $P = 0.0175$; rs28414810, $P = 0.0084$). Haplotype block 2, containing three SNPs spanning introns 4, 5 and 6 of *IL3RA*, was also

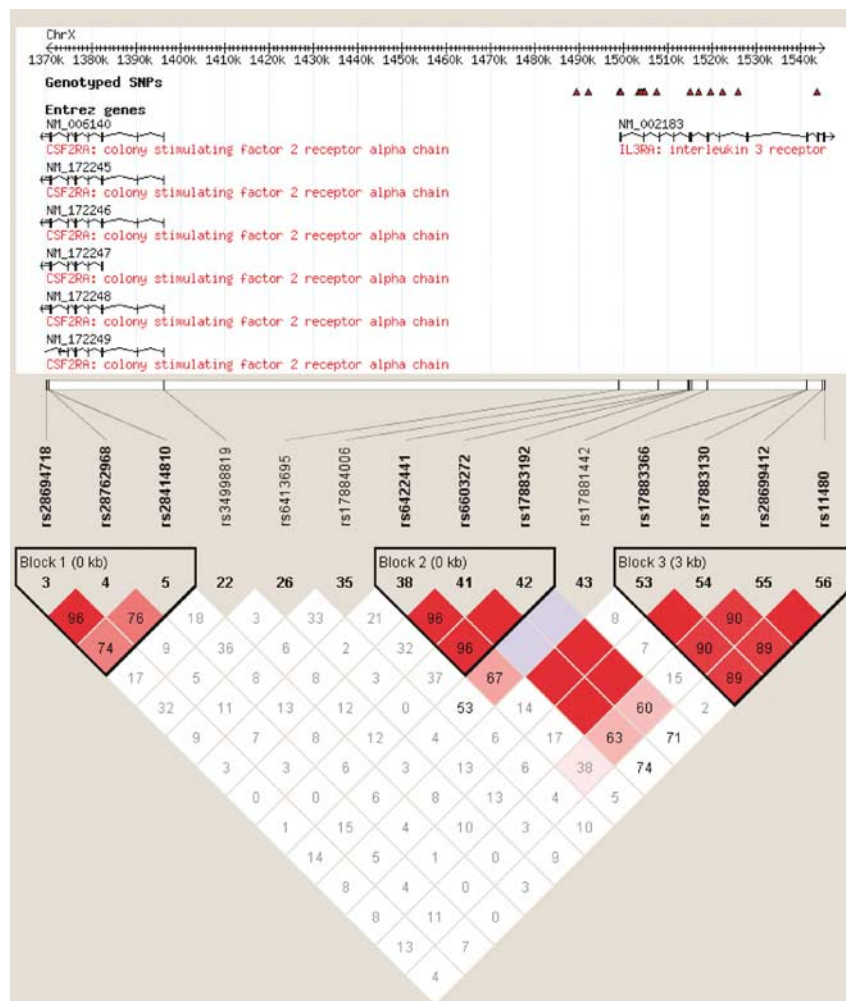


Figure 3 LD plot of common SNPs in *CSF2RA* and *IL3RA* in the sequencing sample. Red shading indicates relatively strong LD as measured by D' , also indicated by the number inside each compartment. Blue compartments represent pairs with high D' but low r^2 -values. Three haplotype blocks are also shown. Note that the upper portion of the figure (white background) is downloaded from the HapMap (Release 21a); multiple splice variants for *CSF2RA* are depicted, and the 'genotyped SNPs' represented by red triangles below the number line represent publicly available SNPs in the HapMap dataset. SNPs genotyped in the sequencing study are listed in the lower (gray background) portion of the figure.

significantly associated with illness (global $P=0.027$). A haplotype containing the minor alleles of these SNPs (CGC for rs6422441, rs6603272 and rs17883192) was over-represented in patients ($P=0.0086$), and each of these SNPs was significantly associated with SCZ ($P's < 0.05$). Of note, when patients who had developed agranulocytosis while undergoing treatment with clozapine ($n=26$) were removed from the analysis, haplotype comparisons of SCZ patients vs controls remained significant (CGC in *CSF2RA*, $P=1 \times 10^{-4}$; CGC in *IL3RA*, $P=0.02$).

We also identified seven novel exonic mutations within these two genes that resulted in amino acid substitutions (Table 3). A total of 15 substitutions were detected in cases, with only one detected in

controls (Fisher's exact $P=0.031$). Moreover, all but one of the rare missense variants shown in Table 3 were observed in patients without agranulocytosis, again indicating that these results are not a reflection of susceptibility to agranulocytosis. Finally, because it could be argued that the control group sample size was too small for adequate comparison of rare events, we obtained sequence data from an additional 41 healthy individuals of Asian, Hispanic or Native American ancestry, resulting in total sample size ($n=72$ controls) directly comparable to the patient group. Notably, six of the seven rare missense alleles listed in Table 3 were not detected in any of these healthy individuals. Moreover, the rare allele of the other mutation locus (Ala17Gly) was detected in only

Table 2 Haplotype blocks, common haplotypes, and association to illness in the sequencing sample

Haplotype/SNP allele(s)	Global χ^2	Global P-value	Case, control frequencies	χ^2 (df=1)	P-value	Odds ratio
Block 1	10.37 (df=3)	0.016				
TAG			0.36, 0.24	2.77	0.096	
CGG			0.30, 0.29	0.02	0.877	
CGC			0.16, 0.35	9.90	0.0017	0.33
TGG			0.14, 0.10	0.78	0.378	
rs28694718 T			0.54, 0.35	5.65	0.0175	2.11
rs28414810 G			0.80, 0.63	6.95	0.0084	2.40
Block 2	7.2 (df=2)	0.027				
TTG			0.58, 0.74	4.76	0.029	0.48
CGC			0.28, 0.11	6.90	0.0086	3.08
CTG			0.13, 0.13	0	0.989	
rs6422441 C			0.41, 0.26	4.518	0.0335	2.03
rs6603272 G			0.29, 0.13	6.031	0.0141	2.74
rs17883192 C			0.29, 0.12	6.893	0.0087	3.07
Block 3	2.57 (df=2)	0.277				

Significant individual SNPs within each haplotype block are also displayed.

Table 3 Novel missense variants in *CSF2RA* and *IL3RA* associated with schizophrenia in the sequencing sample

Gene	Exon	ATG offset (bases)	Alleles	Amino acid	Minor allele carriers (n carriers/total n)	
					Cases	Controls
<i>CSF2RA</i>	3	50	C/G	Ala 17 Gly	8 ^a /70	1/31
<i>CSF2RA</i>	7	7651	G/A	Arg 164 Gln	1/71	0/31
<i>CSF2RA</i>	7	7725	C/T	Arg 189 Cys	1/71	0/31
<i>CSF2RA</i>	8	11 670	C/T	Thr 231 Met	1/70	0/31
<i>IL3RA</i>	5	10 429	G/A	Ala 132 Thr	1/71	0/31
<i>IL3RA</i>	7	14 515	G/A	Trp 226 Ter	1/70	0/31
<i>IL3RA</i>	7	14 538	C/G	Arg 234 Gly	1/69	0/31
Total carriers/total subjects					11/71	1/31
Total rare alleles/total called alleles					15/984	1/434

^aIncludes 1 G/G homozygote.

three of the additional control subjects. No additional mutations were detected in the additional subjects. Thus, the rate of mutation was four times higher in patients than in the combined control sample, and the results remain statistically significant ($P=0.01$).

Using similar methods, we also sequenced a small region surrounding rs4129148 in the original 102 Caucasian individuals. There were no significant differences in rs4129148 genotype ($P=0.28$) or allele ($P=0.14$) frequency between patients and controls in this sample. Evidence for long-range LD between rs4129148 and the genic SNPs in *CSF2RA* and *IL3RA* in the sequenced cohort was present, though limited. In the control sample, there was modest evidence of LD between rs4129148 and SNPs in haplotype block 2 ($D'=0.5$, $r^2=0.12$), which was not observed in the patient group. In both samples, there was some evidence of strong LD (D' as high as 1, but with modest r^2) between rs4129148 and several SNPs with low minor allele frequency ($0.05 < \text{MAF} < 0.10$) within the two genes. It should be noted that these relationships could not be tested in the WGA sample, because the Affymetrix 500K array has very limited coverage of the PAR1; there are no SNPs on the array within *CSF2RA*, and only two SNPs within *IL3RA*, neither of which passed quality control screens.

Discussion

This study identified three converging lines of evidence, from two independent samples, suggesting the presence of a novel schizophrenia susceptibility locus in the PAR1. First, an SNP neighboring *CSF2RA* survived a genomewide test of significance in the WGA sample, with an odds ratio greater than 3 and a population-attributable risk approaching 25%. Second, intronic haplotype blocks within *CSF2RA* and its immediate neighbor *IL3RA* were significantly associated with SCZ in the independent sequencing sample. Third, seven novel missense mutations in these two genes were also significantly over-represented in patients with schizophrenia in the sequencing sample.

Whereas there is some prior cytogenetic support for a PAR1 locus,¹⁷ linkage studies have reported mixed results^{18–20} and have often not included coverage of the PAR1 region. To our knowledge, neither *CSF2RA* nor *IL3RA* have been previously examined in candidate gene association studies of schizophrenia, and studies of other cytokine genes in schizophrenia have been equivocal. Intriguingly, a very recent study has identified several schizophrenia-related SNPs and haplotypes both in and near the gene for interleukin 3 (the ligand for *IL3RA*), which is located within a commonly identified linkage peak on chromosome 5q.²¹

Results of the present study are also notable in that associations to illness were found for common variants in non-coding regions as well as rare missense mutations. Whereas hereditary patterns and linkage data in schizophrenia are consistent

with a polygenic, common disease/common variant model,²² it has also been suggested that heterogeneous rare mutations may play an important role.²³ Results of the present study suggest that these mechanisms are not mutually exclusive in the genetics of schizophrenia.

A substantial role for cytokine-related genes such as *CSF2RA* and *IL3RA* in mediating schizophrenia risk might help connect several disparate lines of epidemiologic and biological data in SCZ. First, it has long been observed that monozygotic twin concordance is approximately 50%, indicating both strong genetic and environmental components of risk.⁶ Well-replicated findings of increased SCZ risk for winter-born and urban-born/dwelling individuals, combined with large-scale studies demonstrating increased risk following maternal exposure to a variety of infectious pathogens, have suggested that prenatal infection may represent a significant environmental risk factor.²⁴ The lack of specificity to particular teratogens, including perinatal hypoxic insult or trauma as well as maternal stress, points to a common pathway by which these varying exposures may equally influence schizophrenia liability; one such pathway is the cytokine-mediated inflammatory response.²⁵ Moreover, increased levels of pro-inflammatory cytokines reported in the peripheral blood and cerebrospinal fluid of patients with schizophrenia suggest that activation of this pathway in adulthood may form a component of SCZ pathophysiology.²⁶ Furthermore, family-based epidemiologic studies suggest that relatives of patients with schizophrenia are psychiatrically²⁷ and neuroanatomically²⁸ more sensitive to negative sequelae of teratogenic exposures. The present findings point toward an explanation for such gene–environment interactions: genetic variation in the receptor structure or expression of pro-inflammatory cytokines, such as *CSF2RA* and *IL3RA* and their respective ligands, granulocyte-macrophage colony stimulating factor (GM-CSF) and interleukin-3 (IL-3), may serve to heighten sensitivity to the teratogenic effects of prenatal infections and perinatal insults,²⁹ and may contribute to the variable course and episodic nature of the adult disorder.³⁰

Abnormalities of GM-CSF, IL-3 and their receptors have been most commonly associated with lymphoma and leukemia.^{31,32} A previously unexplained (though replicated) epidemiologic finding in schizophrenia has been the reduced incidence of these two cancers in relatives of SCZ patients.^{33,34} Similarly, an elevated level of GM-CSF in synovial fluid has been implicated in the inflammatory pathophysiology of rheumatoid arthritis (RA),³⁵ and SCZ patients have consistently demonstrated a reduced prevalence of RA.³⁶ Moreover, SCZ patients show increased incidence of other inflammatory autoimmune disorders such as celiac disease and Sjögren's syndrome before SCZ onset, and prevalence of many autoimmune disorders is increased in parents of SCZ patients.³⁶ These patterns of comorbidity, both in patients and first-degree relatives, might also be explained by

gene–environment interactions of abnormal cytokine receptor genes and varying levels of environmental exposures.

Finally, whereas GM-CSF is most commonly associated with proliferation and differentiation of granulocytes and macrophages, there is accumulating evidence of its critical role in the central nervous system repair via expression of brain-derived neurotrophic factor,³⁷ a trophic factor that has been associated with psychiatric illness. This is consistent with broader evidence of a role for GM-CSF and IL-3 in the central nervous system, including neuroprotection,³⁸ communication across the blood–brain barrier,³⁹ and neurotransmitter modulation (particularly of acetylcholine and GABA levels).⁴⁰ Further studies are needed to determine whether the association of this cytokine receptor genetic locus with schizophrenia is mediated by immune response to infectious agents, autoimmune or inflammatory processes, trophic factors or a combination of these mechanisms.

As with any genetic association study, additional independent replications from other laboratories will be critical to confirm these novel findings. The lack of significant association to rs4129148 in the small sequencing cohort provides a note of caution in interpreting results of the WGA study. Nevertheless, the genomewide-significant result of the WGA study appeared robust to several potential forms of artifact, including population stratification and gender bias. Notably, rs4129148 showed zero errors in the reproducibility analysis, and had a high minor allele frequency resulting in apparently robust calling of both heterozygotes and homozygotes. Still, it is important to note that even multiply replicated findings can be subject to subsequent failures to replicate,^{41–44} presumably owing to the multiple sources of heterogeneity underlying the pathophysiology of complex disease.

Understanding the role of rs4129148 will require further explication of its relationship to the neighboring genes *CSF2RA* and *IL3RA*. Whereas the sequencing sample demonstrated only modest evidence of long-range LD, difficulties in sequencing and genotyping this region (as demonstrated by gaps in coverage in the reference sequence,⁴⁵ the HapMap,⁴⁶ and commercially available arrays such as the Affymetrix 500K) limit our current understanding of the LD structure in this region. Other roles of rs4129148 distinct from any potential relationship to *CSF2RA* and *IL3RA*, such as regulatory functions or coding of small RNAs, cannot be ruled out at this time. Additionally, it should be noted that an additional cytokine receptor gene, cytokine receptor-like factor 2 (*CRLF2*), may lie between rs4129148 and *CSF2RA*, although its RefSeq status is currently listed as provisional.

Despite these limitations, the design of the present study did permit the WGA result to direct attention to PAR1. The sequencing study identified two common haplotypes in *CSF2RA* and *IL3RA* that were significantly associated with schizophrenia; we also

identified seven previously unreported missense mutations, which were collectively associated with illness. Future biological studies will be needed to explicate potential consequences of these haplotypes and amino acid substitutions on receptor transcription, structure and function.

Acknowledgments

This work was supported by the Donald and Barbara Zucker Foundation, internal funding from the North Shore – Long Island Jewish Health System, Clinical Data, Inc., a Keyspan Fellowship Award (TL) and grants from the Stanley Foundation (AKM), the National Alliance for Research on Schizophrenia and Depression (AKM), and the NIH (MH065580 to TL; MH074543 to JMK; MH001760 to AKM). We thank Christophe Lambert and Josh Forsythe of Golden Helix, Inc., for their computing expertise and technical support.

References

- 1 Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2005; **2**: 413–433.
- 2 World Health Organization. *International Classification of Functioning, Disability and Health – ICF*. World Health Organization: Geneva, 2001.
- 3 Goff DC, Cather C, Evins AE, Henderson DC, Freudenreich O, Copeland PM *et al*. Medical morbidity and mortality in schizophrenia: guidelines for psychiatrists. *J Clin Psychiatry* 2005; **66**: 183–194.
- 4 Palmer BA, Pankratz VS, Bostwick JM. The lifetime risk of suicide in schizophrenia: a reexamination. *Arch Gen Psychiatry* 2005; **62**: 247–253.
- 5 Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO *et al*. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005; **353**: 1209–1223.
- 6 Gottesman II. *Schizophrenia Genesis*. Freeman: New York, 1991.
- 7 Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I *et al*. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. *Am J Hum Genet* 2003; **73**: 34–48.
- 8 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C *et al*. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**: 385–389.
- 9 DeWan A, Liu M, Hartman S, Zhang SS, Liu DTL, Zhao C *et al*. *HTRA1* promoter polymorphism in wet age-related macular degeneration. *Science* 2006; **314**: 989–992.
- 10 Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M *et al*. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet* 2006; **38**: 644–651.
- 11 Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C *et al*. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 2003; **112**: 387–399.
- 12 Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.
- 13 Freimer N, Sabatti C. The use of pedigree, sib-pair, and association studies of common diseases for genetic mapping and epidemiology. *Nat Genet* 2004; **36**: 1045–1051.
- 14 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.
- 15 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 16 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995; **57**: 289–300.

- 17 Demirhan O, Tastemir D. Chromosome aberrations in a schizophrenia population. *Schizophr Res* 2003; **65**: 1–7.
- 18 Collinge J, DeLisi LE, Boccio A, Johnstone EC, Lane A, Larkin C *et al*. Evidence for a pseudo-autosomal locus for schizophrenia using the method of affected sibling pairs. *Br J Psychiatry* 1991; **158**: 624–629.
- 19 Crow TJ, DeLisi LE, Lofthouse R, Poulter M, Lehner T, Bass N *et al*. An examination of linkage of schizophrenia and schizoaffective disorder to the pseudoautosomal region (Xp22.3). *Br J Psychiatry* 1994; **164**: 159–164.
- 20 Barr CL, Kennedy JL, Pakstis AJ, Castiglione CM, Kidd JR, Wetterberg L *et al*. Linkage study of a susceptibility locus for schizophrenia in the pseudoautosomal region. *Schizophr Bull* 1994; **20**: 277–286.
- 21 Chen X, Wang X, Hossain S, O'Neill FA, Walsh D, van den Oord E *et al*. Interleukin 3 and schizophrenia: the impact of sex and family history. *Mol Psychiatry* 2007; **12**: 273–282.
- 22 Hirschhorn JN, Daly MJ. Genome-wide association studies for common disease and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
- 23 McClellan JM, Susser E, King MC. Maternal famine, *de novo* mutations, and schizophrenia. *JAMA* 2006; **296**: 582–584.
- 24 Brown AS. Prenatal infection as a risk factor for schizophrenia. *Schizophr Bull* 2006; **32**: 200–202.
- 25 Hanson DR, Gottesman II. Theories of schizophrenia: a genetic-inflammatory-vascular synthesis. *BMC Med Genet* 2006; **6**: 7.
- 26 Sperner-Unterwieser B. Immunological aetiology of major psychiatric disorders. *Drugs* 2005; **65**: 1493–1520.
- 27 van Os J, Pedersen CB, Mortensen PB. Confirmation of synergy between urbanicity and familial liability in the causation of psychosis. *Am J Psychiatry* 2004; **161**: 2312–2314.
- 28 Cannon TD, van Erp TG, Rosso IM, Huttunen M, Lonnqvist J, Pirkola T *et al*. Fetal hypoxia and structural brain abnormalities in schizophrenia patients, their siblings, and controls. *Arch Gen Psychiatry* 2002; **59**: 35–41.
- 29 Martinez-Moczygemba M, Huston DP. Biology of common β receptor-signaling cytokines: IL-3, IL-5, and GM-CSF. *J Allergy Clin Immunol* 2003; **112**: 653–655.
- 30 Boulanger LM, Schatz CJ. Immune signalling in neural development, synaptic plasticity and disease. *Nat Rev Neurosci* 2004; **5**: 521–531.
- 31 Trus MR, Bordeleau L, Pihl C, McGeer A, Prevost J, Minden MD *et al*. Clinical manifestations associated with the aberrant expression of the soluble granulocyte-macrophage colony stimulating factor receptor in patients presenting with haematological malignancies. *Br J Haematol* 2003; **121**: 86–93.
- 32 Testa U, Riccioni R, Diverio D, Rossini A, Lo Coco F, Peschle C. Interleukin-3 receptor in acute leukemia. *Leukemia* 2004; **18**: 219–226.
- 33 Dalton SO, Laursen TM, Mellekjær L, Johansen C, Mortensen PB. Risk for cancer in parents of patients with schizophrenia. *Am J Psychiatry* 2004; **161**: 903–908.
- 34 Lichtermann D, Ekelund J, Pukkala E, Tanskanen A, Lönnqvist J. Incidence of cancer among persons with schizophrenia and their relatives. *Arch Gen Psychiatry* 2001; **58**: 573–578.
- 35 Barreda DR, Hanington PC, Belosevic M. Regulation of myeloid development and function by colony stimulating factors. *Dev Comp Immunol* 2004; **28**: 509–554.
- 36 Eaton WW, Byrne M, Ewald H, Mors O, Chen CY, Agerbo E *et al*. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry* 2006; **163**: 521–528.
- 37 Bouhy D, Malgrange B, Multon S, Poirrier AL, Scholtes F, Schoenen J *et al*. Delayed GM-CSF treatment stimulate axonal regeneration and functional recovery in paraplegic rats via an increased BDNF expression by endogenous macrophages. *FASEB J* 2006; **20**: 1239–1241.
- 38 Wen TC, Tanaka J, Peng H, Desaki J, Matsuda S, Maeda N *et al*. Interleukin-3 prevents delayed neuronal death in the hippocampal CA1 field. *J Exp Med* 1988; **188**: 625–649.
- 39 Verna S, Nakaoke R, Dohgu S, Banks WA. Release of cytokines by brain endothelial cells: a polarized response to lipopolysaccharide. *Brain Behav Immun* 2006; **20**: 449–455.
- 40 Tabira T, Konishi Y, Gallyas F. Neurotrophic effect of hematopoietic cytokines on cholinergic and other neurons *in vitro*. *Int J Dev Neurosci* 1995; **13**: 241–252.
- 41 Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T *et al*. A common genetic variant is associated with adult and childhood obesity. *Science* 2006; **312**: 279–283.
- 42 Rosskopf D, Bornhorst A, Rimmbach C, Schwahn C, Kayser A, Kruger A *et al*. Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187.
- 43 Loos RJ, Barroso I, O'rahilly S, Wareham NJ. Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187.
- 44 Dina C, Meyre D, Samson C, Tichet J, Marre M, Jouret B. Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187.
- 45 Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D *et al*. The DNA sequence of the human X chromosome. *Nature* 2005; **434**: 325–337.
- 46 The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.