Genetic variation at the 22q11 *PRODH2/DGCR6* locus presents an unusual pattern and increases susceptibility to schizophrenia

Hui Liu[†], Simon C. Heath[‡], Christina Sobin[†], J. Louw Roos[§], Brandi L. Galke[†], Maude L. Blundell[†], Marge Lenane[¶], Brian Robertson[‖], Ellen M. Wijsman[**], Judith L. Rapoport[¶], Joseph A. Gogos[††], and Maria Karayiorgou[†‡‡]

[†]Human Neurogenetics Laboratory, The Rockefeller University, New York, NY 10021; [‡]Memorial Sloan Kettering Cancer Center, New York, NY 10021; [§]Department of Psychiatry and Weskoppies Hospital, University of Pretoria, Pretoria 0083, Republic of South Africa; [¶]Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD 20892; [‖]Department of Psychiatry and Valkenberg Hospital, University of Capetown, Capetown 7925, Republic of South Africa; [**]Division of Medical Genetics and Department of Biostatistics, University of Washington, Seattle, WA 98195; and [††]College of Physicians and Surgeons, Department of Physiology and Cellular Biophysics, Center for Neurobiology and Behavior, Columbia University, New York, NY 10032

Communicated by David E. Housman, Massachusetts Institute of Technology, Cambridge, MA, December 26, 2001 (received for review September 25, 2001)

The location of a schizophrenia susceptibility locus at chromosome 22q11 has been suggested by genome-wide linkage studies. Additional support was provided by the observation of a higher-than-expected frequency of 22q11 microdeletions in patients with schizophrenia and the demonstration that ≈20–30% of individuals with 22q11 microdeletions develop schizophrenia or schizoaffective disorder in adolescence and adulthood. Analysis of the extent of these microdeletions by using polymorphic markers afforded further refinement of this locus to a region of ≈1.5 Mb. Recently, a high rate of 22q11 microdeletions was also reported for a cohort of 47 patients with Childhood Onset Schizophrenia, a rare and severe form of schizophrenia with onset by age 13. It is therefore likely that this 1.5-Mb region contains one or more genes that predispose to schizophrenia. In three independent samples, we provide evidence for a contribution of the *PRODH2/DGCR6* locus in 22q11-associated schizophrenia. We also uncover an unusual pattern of *PRODH2* gene variation that mimics the sequence of a linked pseudogene. Several of the pseudogene-like variants we identified result in missense changes at conserved residues and may prevent synthesis of a fully functional enzyme. Our results have implications for understanding the genetic basis of the 22q11-associated psychiatric phenotypes and provide further insights into the genomic instability of this region.

Schizophrenia is a common severe mental illness that affects 1% of the population and is characterized by disordered thinking as well as deficits in emotional and social behavior. Microdeletions of chromosome 22q11 are associated with variable phenotypic expression that often includes learning disabilities, palatal abnormalities, congenital heart defects, and mildly dysmorphic facial features (1). One of the first clues that psychiatric symptomatology may also be associated with this microdeletion was a report of expressionless face, monotonous speech, and flattened affect among children with the microdeletion (2). In light of the evidence for suggestive linkage for schizophrenia on chromosome 22 (reviewed in ref. 3), patients with the 22q11 microdeletion were evaluated for psychiatric symptoms or disorders, and a relatively high frequency of severe mental illness has since been reported. Specifically, two independent studies have reported that 25–31% of patients with the 22q11 microdeletion met diagnostic criteria for schizophrenia or schizoaffective disorder (4, 5). Although the microdeletion occurs in the population at a rate of 0.025%, it has been found in 2% of adult schizophrenic patients (6) and in 6% of cases with childhood onset schizophrenia (COS) (7). In addition, several studies have described 22q11 microdeletions among schizophrenic patients of various ethnic origins (8). These studies collectively suggest that the morbid risk of schizophrenia for a patient with a 22q11 microdeletion may be approximately 20–30 times the general population risk of 1%, and that the rate of 22q11 microdeletions in schizophrenia, although relatively low, may be approximately 80 times the estimated general population rate. It therefore seems likely that the 22q11 region harbors genes that alone, or in combination, are causally implicated in schizophrenia in a certain proportion of patients. The overwhelming majority of the 22q11 deletions are 3 Mb in size, whereas a smaller percentage (≈8%) involve the same proximal breakpoint but a different distal breakpoint resulting in a smaller 1.5-Mb deletion (9, 10). At least one schizophrenic patient has been described as carrying the smaller 22q11 microdeletion, and therefore the "schizophrenia critical region" has been defined to 1.5 Mb (ref. 6). The majority of the genes in the region are known (11). To address the role of individual genes from this chromosomal region, we are undertaking systematic mutational screening of these genes in combination with linkage disequilibrium (LD) studies in family samples (triads) that test for preferential transmission of single-nucleotide polymorphisms (SNPs) and multi-SNP haplotypes from parents to affected individuals.

GENETICS

Materials and Methods

Patient Samples. Detailed information about the adult schizophrenic (AS) sample is provided in ref. 12. The South African sample is part of our ongoing collection of schizophrenia patients of Afrikaner origin and will be described in detail elsewhere (M.K., M. Torrington, C.S., B.R., S.C.H., M.L.B., H. Pretorius, S. Lay, J.A.G., and J.L.R., unpublished work). Probands in both samples met lifetime criteria for Diagnostic and Statistical Manual of Mental Disorders, 4th Ed. (DSM-IV) (13) schizophrenia or schizoaffective disorder. Participants were interviewed by specially trained clinicians by using the Diagnostic Interview for Genetic Studies (DIGS) (14). Detailed information about the COS sample is provided in refs. 7 and 15. All COS probands met unmodified criteria for schizophrenia with onset of psychotic symptoms before their 13th birthday and mean age of onset of psychosis at 10.1 (±1.8 yr). The protocol and the consent forms were approved by the Institutional Review Boards (IRBs) at all participating sites. The National Institute of Mental Health (NIMH) samples have been obtained from the NIMH Human Genetics Inititative dataset (http://zork.wustl.edu/nimh). [Acknowledgment for this sample can be found in the *Appendix*, which is published as supporting information on the PNAS web site (www.pnas.org)].

Abbreviations: LD, linkage disequilibrium; COS, childhood onset schizophrenia; SNP, single-nucleotide polymorphism; AS, adult schizophrenic; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, fourth edition; NIMH, National Institute of Mental Health; LCR, low copy repeat sequence; RFLP, restriction fragment length polymorphism; HHRR, haplotype-based haplotype relative risk; TDT, Transmission Disequilibrium Test; T/nT, Transmitted/non-Transmitted; CI, confidence interval.

[‡‡]To whom reprint requests should be addressed. E-mail: karayim@mail.rockefeller.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

SNP Identification and Genotyping. Methods (16, 17), primers, and PCR conditions can be found in the *Appendix*, which is published as supporting information on the PNAS web site, www.pnas.org.

Somatic Cell Hybrids. Somatic cell hybrid methods are as described in ref. 6.

Statistical Analysis. Haplotype probabilities for each individual were calculated assuming no recombination between SNPs, summing over ambiguous haplotypes caused by missing phase information or genotype data. Haplotype frequences were estimated by using the Expectation Maximization algorithm (18) simultaneously with the calculation of haplotype probabilities. For the analyses on multilocus SNP haplotypes, trios with ambiguous haplotype reconstructions were used only if there was one reconstruction with a probability of >0.95. Transmissions of single SNPs were analyzed by using both the haplotype-based Haplotype Relative Risk (HHRR) statistic (19) and the Transmission Disequilibrium Test (TDT) (20) (where the TDT considered only transmissions from heterozygous parents). The significance of the TDT analyses was assessed by using a standard 1 degree of freedom χ^2 test, and the significance of the HHRR tests was assessed by using Fisher's exact test. Multilocus SNP haplotype transmissions were analyzed by using the HHRR, with the significance being estimated by using Fisher's exact test. Because of the relatively small size of the COS sample and the stratified adult samples, exact *P* values are given only for $P > 0.001$.

Results

Association Between *PRODH2*/*HsPOX2* Locus Variation and Schizophrenia. A first-stage scan of the 22q11 region involved a set of 18 SNPs from nine gene loci distributed across the entire 1.5 Mb (Fig. 1*b*, green shading). The inheritance of these variants by probands affected with schizophrenia from parents was examined in 107 independent triads from the U.S. (AS sample). Analysis of the transmission by using the HHRR statistic (19) revealed a modest nominally significant association between schizophrenia and *PRODH2**1945 allele 2 ($P = 0.04$, uncorrected for multiple testing). The TDT test (20), a test for association in the presence of linkage, provided similar results [($P = 0.035$; Transmitted/non-Transmitted (T/nT), 41/24]. No significant evidence of transmission disequilibrium was observed for any of the other 17 SNPs (Fig. 1*b*).

The *PRODH2* gene encodes for proline dehydrogenase, a mitochondrial enzyme that converts proline to Δ^1-pyrroline-5-carboxylate and is involved in transfer of redox potential across the mitochondrial membrane (21). The gene is widely expressed in the brain as well as in other tissues (21), consists of 14 coding exons, and is located in the most centromeric part of the 22q11 critical region, where, along with the *DGCR6* gene, it resides within a 75-kb module (previously referred to as sc11.1), which is part of a ≈350-kb low copy repeat (LCR) sequence (LCR-A) (10). This module is duplicated ≈1.5 Mb distally as part of another ≈135-kb LCR (LCR-B) (10). The duplicated copy of the gene, *PRODH2-P* (AC007663.29), has accumulated point mutations and small deletions and insertions, as well as a large deletion that removes exons 2, 3, 4, 5, and part of exon 6 (ref. 11). Reverse transcription–PCR analysis from total cellular RNA extracted from lymphocytes indicates that the duplicated copy is not transcribed and therefore has become a nonexpressed pseudogene (unpublished observations). *PRODH2* and *PRODH2-P* share over 95% sequence similarity.

We sought to follow up the initial finding of association between schizophrenia and the *PRODH2**1945 SNP by, (*i*) identifying additional SNPs from the *PRODH2* locus, (*ii*) better defining the nature of the potential association, and (*iii*) attempting to replicate it in two additional samples. Exons 2–14 and surrounding intronic sequences were examined at this second stage in 48 AS patients, resulting in the identification of five additional SNPs (Fig. 1 *a* and *b*, blue shading). Significant evidence for pairwise LD ($P < 0.01$)

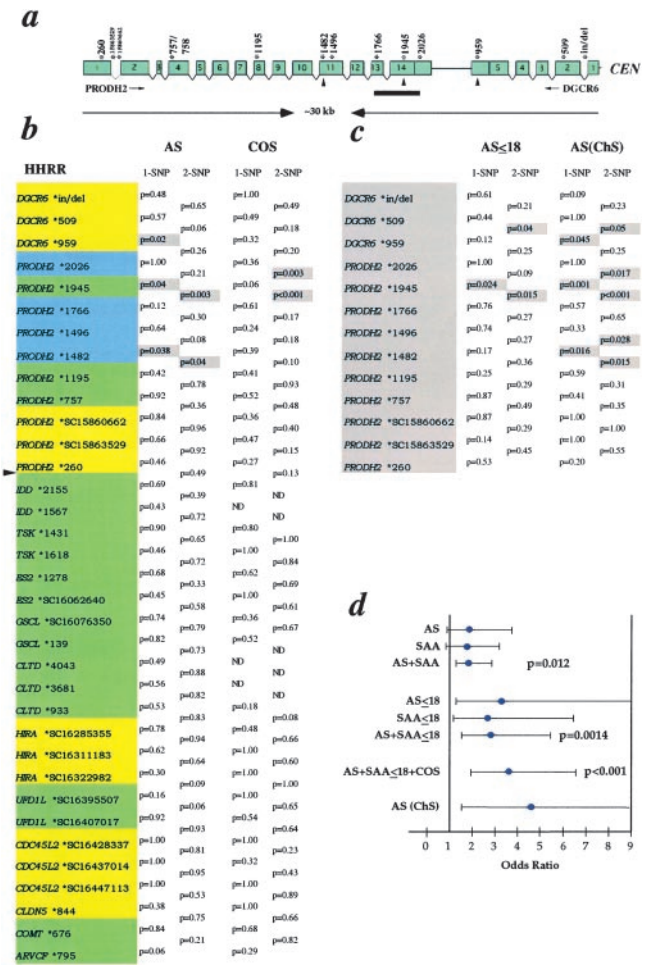

Fig. 1. (*a*) Genomic structure of the *PRODH2*/*DGCR6* locus. The approximate distribution of the identified polymorphisms in both genes is indicated (*Top*). Polymorphisms demonstrating significant overtransmission in the AS sample are indicated by arrowheads. Arrows indicate direction of transcription. The black horizontal bar signifies the *PRODH2**1766/1945/2026 SNP combination that defines the risk haplotype presented in *d*. The approximate size of the locus is also indicated. CEN, centromere. (*b* and *c*) Linkage disequilibrium results in the AS and COS (*b*), as well as stratified AS families (*c*). In a total of 133 families, transmission to the affected probands of alleles and haplotypes defined by the *PRODH2**1482, *1766, *1945, *2026, and *DGCR6**959 SNPs deviates significantly from expectation under the null hypothesis and provides substantial support for a contribution of the *PRODH2*/*DGCR6* locus variation to schizophrenia susceptibility. By contrast, no such association was detected in a sample of 80 North American families not afflicted with schizophrenia (not shown), arguing against a generalized transmission distortion at this locus. Color coding reflects the three stages of our analysis (see text). Markers are listed in order from centromere (*Top*) to telomere (*Bottom*). SNPs located in expressed sequences are referenced according to their position in corresponding cDNAs (accession nos. are available in the *Appendix*). Intronic SNPs are referenced according to their position in the University of California, Santa Cruz public database (October 7, 2000, freeze; http://genome.ucsc.edu). Polymorphism *PRODH2**758 is not included in the analysis, because it was in complete LD with *PRODH2**757 (see Fig. 3, which is published as supporting information on the PNAS web site). ND, the polymorphism was not typed in this group. Significant associations are shaded. Arrowhead indicates the relative position of microsatellite marker D22S1638. (*d*) Estimated risk (with 95% CI) for *PRODH2**1766/1945/2026 2–2-1 haplotype relative to the other combinations. For each study, the circle represents the estimated risk, and the line indicates the extent of the 95% CI around this estimate. Data shown for odds ratios are based on haplotype counts (for the case-control study) and the transmission (HHRR) ratios in trios. AS(ChS), subset of AS patients with early childhood deviant behaviors (12); AS ≤ 18, subset of AS patients with DSM-IV disease onset before or at age 18; SAA, South African Afrikaner case/control sample.

between the majority of all eight *PRODH2* SNP loci was found over a distance of ≈12 kb by analysis of the 214 unrelated parents of the AS sample. A total of 17 haplotypes were detected, but only 4 haplotypes accounted for ≈75% of all those seen. The most common haplotype (30%) was comprised of the most common alleles of all eight markers (see Fig. 3). Two of the additional variants demonstrated evidence for transmission disequilibrium: *PRODH2**1482 (Asp-426→Asp) in exon 11 showed a significant 2-fold overtransmission of the less common *PRODH2**1482 allele 2 (HHRR, $P = 0.038$; TDT, $P = 0.025$; T/nT, 30/15), whereas *PRODH2**1766 (Gln-521→Arg) in exon 13 showed a trend for overtransmission of the most common *PRODH2**1766 allele 2 (HHRR, $P = 0.12$; TDT, $P = 0.07$; T/nT, 6/14). To incorporate information from adjacent SNPs, we analyzed transmission of haplotypes by using a two-locus moving-window strategy (Fig. 1*b*). Of the seven haplotype combinations distributed over the entire gene from exon 2 to the 3′-UTR, only one combination (*PRODH2**1766/1945) showed evidence of transmission disequilibrium more significant than each SNP by itself ($P = 0.003$), primarily because of preferential transmission of the *PRODH2**1766/1945 2–2 haplotype (T/nT, 36/17).

***PRODH2* Locus Variation and COS.** Given the apparent enrichment of 22q11 microdeletions in patients with COS (7), we sought to replicate the identified association in an independent sample afflicted with COS. Twenty-nine triads (probands with both biological parents) were used in this analysis. In three of these families, COS probands were previously shown to carry a 22q11 microdeletion by fluorescence *in situ* hybridization (7). In addition, polymorphic marker analysis showed that, in all three probands, the *PRODH2* gene is hemizygously deleted (data not shown). Analysis of the inheritance of the *PRODH2* alleles and two-locus haplotypes in the remaining 26 families revealed a pattern of transmission distortion remarkably similar to that observed in the AS sample (Fig. 1*b*). Specifically, and despite the small size of the COS sample, a preferential transmission of *PRODH2**1945 allele 2 to COS probands was observed and provided substantiating evidence for an allelic association (HHRR, $P = 0.06$; TDT, $P = 0.03$; T/nT, 11/3). The *PRODH2**1766/1945 combination showed again significant evidence for transmission disequilibrium ($P < 0.001$) because of preferential transmission of the 2–2 haplotype (T/nT, 10/0). Transmission of the adjacent combination also deviated significantly from neutral ($P = 0.003$) because of overtransmission of the *PRODH2**1945/2026 2–1 haplotype (T/nT, 11/1). No other two-SNP haplotypes provided significant association results.

***PRODH2* Locus Variation and Schizophrenia with Early Onset Features.** Early behavioral and cognitive abnormalities are a fundamental component of the 22q11 microdeletion syndrome and are present more consistently than the physical features of the syndrome (such as palatal abnormalities and heart defects) among schizophrenic patients carrying the 22q11 microdeletions (22). On the basis of these observations and the apparent enrichment of 22q11 microdeletions in patients with COS (7), we reasoned that the contribution of the *PRODH2* locus to schizophrenia susceptibility may be more significant among the subgroup of our AS patients with early disease onset or early deviant behaviors. To address this possibility, we stratified the AS families according to two criteria: First, by age at which full diagnostic criteria for schizophrenia according to the DSM-IV (13) are met. We set the cutoff point before or equal to age 18. Because age at clinical syndrome onset might not be an accurate indicator of the biological onset of the disease, we also stratified the families by history of early deviant behaviors occurring several years before the onset of the syndrome (before age 10) (ref. 12). The overall significant association in the entire AS sample justified further HHRR analysis including only AS probands stratified according to early onset profiles. Despite the decrease in sample size, this stratified analysis provided significant evidence for association (Fig. 1*c*). In the subset of AS probands with early childhood deviant behavioral profiles, transmission of the three-marker haplotype *PRODH2**1766/1945/2026 2–2-1 was associated with an odds ratio (approximate relative risk) of 4.6 [95% confidence interval (CI), 1.4–14.2], a 2.4-fold increase compared with the entire AS sample. When families were stratified according to age of DSM-IV schizophrenia onset (AS ≤ 18), transmission of the 2–2-1 risk haplotype was also associated with an increase in the relative risk to 3.3 (95% CI, 1.1–9.8) (Fig. 1*d*).

Delineation of the Associated Region and Analysis of the Contribution of *DGCR6*. Evidence for association is weak or absent at the 5′ half of the *PRODH2* gene. By screening a sample of 18 patient carriers of the risk haplotype (Fig. 3), we identified three additional SNPs from the 5′ end of the gene [*PRODH2**260 (Pro-19→Gln) from coding exon 1, *PRODH2**SC15863529 and *SC15860662 from the adjacent intron; Fig. 1 *a* and *b*, yellow shading] and typed them in the AS and COS families. Although one of these variants (*SC15860662) showed strong disequilibrium with *PRODH2**1945 ($P < 10^{-4}$), none yielded any evidence for association. We also typed both family samples for the D22S1638 microsatellite marker, which maps ≈50 kb distal to *PRODH2*. This marker is not in LD with *PRODH2**1945 ($P = 0.59$), and it did not yield any evidence for association (data not shown). DNA sequence analysis (not shown), as well as a search of public databases (http://genome.ucsc.edu), indicates that the ≈100-kb region immediately distal to the *PRODH2* gene is devoid of functional genes (a HERV-K101 provirus and several expressed sequence tags (ESTs) representing noncoding mRNAs, likely originating from expressed pseudogenes, have been mapped in the region between *PRODH2* and *IDD*; ref. 11 and data not shown). By contrast, the adjacent region, from *IDD* to *CLDN5* (spanning 488 kb) is highly gene-rich (11). In addition to the 13 SNPs from this interval tested in the first stage of our screening, we identified and typed seven additional SNPs (Fig. 1*b*, yellow shading). Analysis of transmission of all 20 individual SNPs and two-SNP combinations did not provide any significant results in both AS and COS samples. Therefore, this analysis does not support the region distal to the *PRODH2* as the source of the observed association.

Evidence for transmission disequilibrium appears to be stronger at the 3′ end of the *PRODH2* gene. This region is particularly interesting, because it contains the *DGCR6* gene, which encodes for a product that shares homology with the *Drosophila melanogaster* gonadal protein and has weak characteristics of a cell adhesion molecule (23). The gene consists of at least five exons distributed over a region of ≈5 kb in LCR-A (10), well within the LD range for the identified *PRODH2* risk haplotype (the two genes share an ≈1.5 kb 3′ intergenic region; Fig. 1*a*). A duplicated copy of the *DGCR6* gene [designated *DGCR6*-like (L) (AC008103.22)] exists in LCR-B adjacent to *PRODH2-P*. *DGCR6-L* shares over 95% sequence similarity with *DGCR6* and has accumulated several point mutations. At least five of them result in missense changes (not shown), but none introduces a premature stop codon. Analysis of EST databases indicates that *DGCR6-L* is transcribed in a variety of tissues, and it is unclear at present whether it represents a nonfunctional pseudogene. We scanned the ≈1.5-kb intergenic region, as well as all exons and flanking intronic sequences of the *DGCR6* gene in 18 patient carriers of the risk haplotype, for common sequence variations. This analysis identified two SNPs (*DGCR6**509/G→A, *DGCR6**959/C→T) and one intronic 9-bp insertion/deletion (in/del) polymorphism (*DGCR6**in/del) 9 bp from the splice donor site of exon 1 (Fig. 1*b*, yellow shading). Allelic frequencies ranged from 3.9 to 12.3%. Not surprisingly, both common variants (*DGCR6**959, *DGCR6**in/del) were in LD with each other (not shown) and also in strong LD with *PRODH2**1945 ($P < 10^{-6}$). Only the rare allele (T, 12.3%) and associated haplotypes of one of these variants (*DGCR6**959, located at the 3′ UTR of the *DGCR6* gene, 1,267 bp proximal to *PRODH2**1945; Fig. 1*a*)

demonstrated evidence for association (Fig. 1*b*). Overall, this stage of our analysis extended the associated region to include the 3′ end of the *DGCR6* gene and delineated the range of significant disequilibrium within a segment in which *PRODH2* and *DGCR6* are the only known genes.

Genetic Variation at the *PRODH2* Locus Presents an Unusual Pattern. In the course of this work, we noticed that the sequence of the infrequent allele of the exon 11 SNP *PRODH2**1496 (a1 = A, 11%, Fig. 3) is identical to the published pseudogene sequence. To test the possibility that the *PRODH2-P* pseudogene can serve as a reservoir of mutations (24), we aligned the gene/pseudogene exonic sequences (exons 7–14) and initially assayed for a subset of differences, namely those detected by simple PCR-RFLP assays predicted to result in missense changes (see Table 3, which is published as supporting information on the PNAS web site).

We analyzed genomic DNA from AS (n = 84–106) and COS (n = 41) samples, and several affected individuals were discovered to be heterozygous carriers of pseudogene-like single base-pair substitutions or clusters of them (Fig. 2 *a* and *b* and Table 1). Interestingly, the majority of the identified pseudogene-like alleles were located in exon 11 of the *PRODH2* gene where the originally identified *PRODH2**1496 polymorphism resides. In all cases where parents were available for genotyping, we confirmed inheritance of pseudogene-like alleles from parental chromosomes, and no *de novo* changes were identified. In addition, several control experiments confirmed that the observed changes are not the result of artifactual hybrid amplicons. First, control PCR amplification and sequencing of the equivalent of exon 11 from all affected carriers by using *PRODH2-P* (pseudogene)-specific primers produced a sequence reflecting the entire set of multiple substitutions expected by the published pseudogene sequence. Second, we used reverse transcription–PCR on total RNA isolated from available transformed lymphocytes to amplify exon 11 sequences by using primers located in flanking exons (data not shown). In all cases tested, the presence of single or multiple adjacent pseudogene-like substitutions was confirmed in the cDNA where *PRODH2-P* is not represented. Third, in one case, we confirmed the presence of a pseudogene-like allele in the paternal chromosome of a heterozygous carrier (haplotype D; see below) by qualitative Southern blot hybridization on DNA isolated from hamster–human somatic cell hybrids retaining only one of the two chromosome 22 copies (Fig. 2*c*). In a similar manner, we also aligned the coding sequences from both copies of the *DGCR6* gene and checked *DGCR6*, by PCR-RFLP assays, for four of the five differences present in the *DGCR6-L* predicted to result in missense changes (see Table 3). We analyzed the same AS and COS samples, but no affected individuals were discovered to carry the predicted mutations.

We determined *PRODH2* exon 11 pseudogene-like haplotypes by sequencing the entire exon from affected carriers and their parents (when available). Seven distinct haplotypes were identified carrying single or clusters of multiple adjacent substitutions (Fig. 2*b*). Conversion of proline to Δ^1-pyrroline-5-carboxylate is a rate-limiting step, and therefore the *PRODH2* gene can be haploinsufficient, in the sense that heterozygous carriers of debilitating mutations may demonstrate plasma proline accumulation (25). For practical reasons, it was not feasible to test the plasma proline levels of all affected carriers. Nevertheless, in the small number of heterozygous carriers tested, plasma proline levels were at the higher end of the normal distribution. The plasma proline level of three adult female carriers of exon 11 haplotypes A, B, and E, respectively, was ≈1.5- to 2-fold higher when compared with the mean value of sex- and age-matched noncarrier controls (Fig. 2*d*). However, it should be emphasized that it is unclear at present whether the plasma proline levels accurately reflect the effect of the identified mutations on the function of the PRODH2 enzyme and the local concentration of proline in brain subregions, given the high-level expression of a homologous proline dehydrogenase

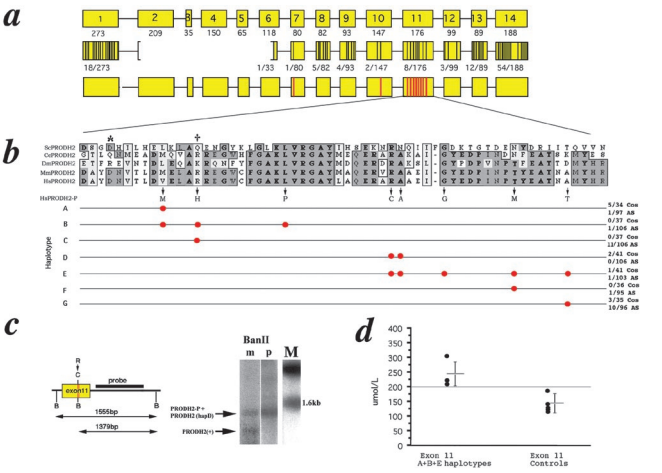

Fig. 2. (*a*) Genomic structure of the *PRODH2* (*Top* and *Bottom*) and *PRODH2-P* locus (*Middle*); colored boxes indicate predicted coding exons; brackets indicate a large deletion that removes predicted exons 2, 3, 4, 5, and part of 6 from *PRODH2-P*. Introns are not to scale. The size of the *PRODH2* exons (in nucleotides), as well as the number of exonic differences between *PRODH2-P* and *PRODH2*, is indicated. Black vertical lines indicate the approximate distribution of these differences. *PRODH2* exons 7, 9, 10, 11, 13, and 14 were assayed for presence of a subset of these differences (n = 12; see Table 3) in the AS and COS patient samples. Red vertical lines (*Bottom*) indicate pseudogene-like substitutions detected in both samples. (*b Top*) Cross-species comparison of the *Homo sapiens PRODH2* exon 11 predicted protein sequence by using the program CLUSTAL-W (as implemented in MacVector, Accelrys, San Diego). Boxed and shaded areas highlight identities or similarities. Bottom row indicates the predicted effect of pseudogene-like base changes on the exon 11 encoded protein sequence. * and † indicate the codons where the synonymous SNP *PRODH2**1482 and the nonsynonymous SNP *PRODH2**1496 reside, respectively. (*Bottom*) Exon 11 haplotypes determined by sequencing the exon from affected carriers and their parents. Seven distinct haplotypes were identified carrying single or clusters of multiple adjacent pseudogene-like substitutions. Filled circles indicate codons carrying pseudogene-like, synonymous (n = 2), and nonsynonymous (n = 6) base-pair substitutions. Solid lines indicate regions of identity to the *PRODH2* reference sequence, shared by all haplotypes. COS patients from all ethnic backgrounds (21 European Americans, 12 African Americans, 8 of other ethnic background) were included in this analysis, whereas only European and African Americans were included in the analysis described in Table 1. (*c*) Southern blot hybridization on DNA isolated from hamster–human somatic cell hybrids retaining only one of the two chromosome 22 copies from a heterozygous carrier of haplotype D (m, maternal copy; p, paternal copy). The Arg-453→Cys substitution, present in the *PRODH2-P* and the paternal copy of *PRODH2* in this patient, is detected by a *Ban*II RFLP (*Ban*II site absent). Genomic DNA was digested with *Ban*II restriction endonuclease, and a 1-kb intronic fragment, indicated by a solid line, was used as a probe. (M, 1-kb ladder DNA marker; B, *Ban*II). (*d*) Distribution of plasma proline level of three adult female carriers of exon 11 haplotypes A, B, and E and five sex- and age-matched noncarriers controls. Blood was drawn from the subjects in the morning, after overnight fasting. Proline levels were determined by the same qualified diagnostic laboratory (Mayo Clinic, Rochester, MN). Horizontal lines denote mean values and vertical lines, standard errors.

(*PRODH1* or *HsPOX1*) (U80018.1) in peripheral tissues but not in the brain.

Transmission and Distribution of Pseudogene-Like Alleles in Patients. In addition to *PRODH2**1496, only one (Ala-472→Thr) of the eight missense pseudogene-like variants identified was sufficiently prevalent in the combined AS and COS sample to be used in family-based association studies. Analysis of the transmission of the Ala-472→Thr variant by using the TDT statistic revealed a modest but significant association between schizophrenia and the rare pseudogene-like allele (P = 0.03; T/nT, 18/7:13/6 in AS, 5/1 in COS). In addition to the preferential inheritance, an enrichment of the Thr-472 substitution is observed in the European American AS (5.8%) and COS (7.1%) probands compared with a European

Table 1. Distribution and Inheritance of *PRODH2* pseudogene-like substitutions

		European American			African American			AS + COS triads
Exon no.	Amino acid change	AS Sample	COS Sample	Controls	NIMH/AS Sample	COS Sample	Controls	T/nT
7	Leu-289 → Met	1/103	0/21	3/93	ND	0/12	ND	ND
10	Pro-406 → Leu	1/84	0/21	0/108	ND	1/12	0/37	ND
11	Val-427 → Met	1/97	1/21	4/108 (1.8%)	7/36 (11%)*	4/12 (16.6%)†	1/37 (1.3%)	ND
11	Arg-431 → His	ND (see Fig. 1)						
11	Leu-441 → Pro	0/105	0/21	ND	2/36	0/12	0/37	ND
11	Arg-453 → Cys	1/102 (0.5%)	3/21 (7.1%)‡	1/115 (0.4%)	ND	0/12	ND	ND
11	Thr-466 → Met	2/95	1/21	0/78	ND	0/12	ND	ND
11	Ala-472 → Thr	11/95 (5.8%)§	3/21 (7.1%)¶	4/93 (2.1%)	ND	0/12	ND	18/7‖

Number of carriers over the total number of patients tested for each substitution is shown. The NIMH/AS sample includes 35 schizophrenia probands of African American descent from the NIMH Human Genetics Initiative dataset and the one African American proband from the AS sample. Allelic frequencies are given in parentheses. ND, not done. *, $P = 0.1$; †, $P = 0.01$; ‡, $P = 0.015$; §, $P = 0.048$; ¶, $P = 0.1$; ‖, $P = 0.03$.

American control sample (2.1%; $P = 0.048$ and 0.1, respectively) (Table 1). The Ala-472→Thr variant is not in disequilibrium with *PRODH2**1945 ($P = 0.124$), indicating an independent association with schizophrenia susceptibility. The distribution of the other rare variants was compared between cases and two racially matched U.S. control samples. The Arg-453→Cys change that affects a highly conserved amino acid seems to be enriched in the European American COS sample (7.1%) compared with both the AS sample (0.5%; $P = 0.02$) and the European American control sample (0.4%; $P = 0.015$). The Thr-466→Met change is present in two AS patients and one COS patient from the European American sample and absent from all of the control chromosomes tested.

The allelic frequency of the Val-427→Met change is significantly higher in the African American subset of the COS sample (16.6%) compared with both an African American control sample (1.3%; $P = 0.01$) and the European American control sample (1.8%; $P = 0.008$). There is no significant difference in the distribution of the Val-427→Met change between the two control samples ($P > 0.7$). Interestingly, the Val-427→Met is present in the only African American patient of the AS sample. Further sequencing revealed that this patient also carries a Leu-441→Pro substitution (Fig. 2*b*, haplotype B) not present in any chromosome in the control African American sample. Leu-441 is a highly conserved amino acid, and the Leu-441→Pro mutation is predicted to result in a protein with drastically reduced enzymatic activity. The apparent enrichment of these two variants among African American patients was investigated further by genotyping 35 independent African American probands extracted from the NIMH Human Genetics Initiative collection of families afflicted with schizophrenia. We identified six additional patient carriers of the Val-427→Met variant (a total of 7/36 as opposed to 1/37 controls; $P = 0.1$) and one additional patient carrier of the Leu-441→Pro variant.

Finally, one additional rare pseudogene-like Pro-406→Leu substitution, which affects a relatively conserved Pro residue at exon 10, was identified in one European American adult and one African American COS patient and was not detected in any control subject.

Replication of the Association Between *PRODH2* Locus Variation and Schizophrenia in a Third Sample. We sought further confirmation of all these results in a third independent sample of 109 unrelated individuals affected with schizophrenia and 75 healthy controls (with no family history of a major psychiatric disorder) of South African Afrikaner origin. We first genotyped *PRODH2**1766, *1945, *2026 SNPs, estimated haplotype frequencies for the three-SNP combinations for cases and controls separately, and compared the distribution of the 2–2-1 risk haplotype in chromosomes of affecteds and controls (Table 2). This follow-up sample demonstrated a strong nonsignificant trend in the same direction and similar odds ratios with the U.S. AS sample (Table 2 and Fig. 1*d*). The frequency of the 2–2-1 haplotype was 9% among control chromosomes, very similar to the frequency observed in the untransmitted chromosomes of the U.S. family sample (8.8%). This frequency increased to 15% among the chromosomes of affecteds, corresponding to a relative risk of 1.7 [95% CI, 0.9–3.3] ($P = 0.055$, one-tailed). Most importantly, and consistent with the pattern observed in the U.S. sample, when only the subset of affecteds with age at onset ≤18 years was considered ($n = 21$), the frequency of the associated haplotype increased to 20%, corresponding to a risk of 2.6 (95% CI, 1.05–6.6) ($P = 0.01$, one-tailed). Only three haplotypes were observed (2–1-1, 2–2-1, and 1–2-2) in this sample, and no other haplotype showed evidence of association. Finally, we compared the distribution in case and control chromosomes of the three exon 11 pseudogene-like variants that were apparently enriched among the European American subset of our U.S. samples. One of these variants (Thr-466→Met) was not present in either case or control chromosomes. However, the other two variants (Arg-453→Cys and Ala-472→Thr) showed a striking enrichment among the disease chromosomes (Table 2). Interestingly, eight of

Table 2. Replication of association in the SAA sample

	*PRODH2**1766		*PRODH2**1945		*PRODH2**2026		*PRODH2**1766/1945/2026		
Allele	1	2	1	2	1	2	221	Other	
SAA	9	209	177	41	209	9	32 (15%)	184	$P = 0.055$*
SAA ≤ 18	2	40	31	11	40	2	9 (20%)	33	$P = 0.01$†
SA controls	5	145	131	19	145	5	14 (9%)	136	
	Arg-453 → Cys				Thr-466 → Met		Ala-472 → Thr		
SAA	5/75 (3.3%)	$P = 0.04$			0/75 (0%)		12/75 (8%)		$P = 0.008$
SA controls	0/50 (0%)				0/50 (0%)		1/50 (1%)		

Number of carriers over the total number of patients tested is shown for pseudogene-like substitutions; one patient carried both variants. Eight of sixteen carriers had age at disease onset ≤18 yr. Allelic frequencies are given in parentheses. *P* values are one-tailed. SAA, South African Afrikaner.

*OR = 1.7 (95% CI, 0.9–3.3); †OR = 2.6 (95% CI, 1.05–6.6).

the sixteen carriers of either variant (50%) had an early age at disease onset (≤18), a 2.5-fold increase over the observed rate of early onset cases in this sample ($P = 0.05$). Moreover, none of the two variants was found in the context of the 2–2-1 risk haplotype, indicating an independent association with schizophrenia susceptibility in this population as well.

Discussion

Using LD analysis in three independent nondeleted patient samples, we provide here strong evidence for an association between variation at the 22q11 locus and schizophrenia with early onset features.

Through fine-scale mapping in heterogeneous U.S. family samples, we delineated the region of significant LD within a segment in which *PRODH2* and *DGCR6* are the only known genes, and we identified a relatively common haplotype where the schizophrenia susceptibility variant(s) likely reside. Until the functional consequences of the associated variation are established in brain samples, our LD studies alone cannot unequivocally distinguish between a contribution from the *PRODH2* gene itself and/or from the neighboring *DGCR6* gene.

Further evidence implicating the *PRODH2* gene itself is provided by the identification of rare pseudogene-like structural variants of the *PRODH2* gene, especially in cases with early onset of the disease. No such variants were identified in the *DGCR6* gene. Several of the variants identified result in missense changes at residues conserved to a different extent among species. Observation of clustering of point mutations suggests that at least some of the pseudogene-like mutations arose as the result of gene conversion events between the *PRODH2* gene and the homologous nonfunctional linked copy (24). Exon 11 of the *PRODH2* gene seems to be a ''hot spot'' for microconversion events, presumably because of the proximity of a strand exchange- promoting sequence (24). Analysis of pseudogene-like haplotypes at this exon indicates that sequence exchanges between the two copies often involve conversion tracts with a minimum size of 18–57 bp. Identification of pseudogene-like alleles may have additional implications. First, it is of interest that recombination between LCR-A (the *PRODH2* locus) and LCR-B (the *PRODH2-P* locus) results in a 1.5-Mb deletion, which accounts for ≈8% of the total 22q11 deletions (9, 10). This rearrangement may involve intrachromosomal pairing between the duplicated modules and formation of a ''stem–loop'' intermediate (9, 10). Therefore, a possible relation between the apparent gene conversion observed in this study and the LCR-A/B-mediated recombination warrants further study. Second, in relation to the origin of mutations in schizophrenia (and other common complex psychiatric disorders), our results implicate the dynamic instability of the genome and the resulting processes of gene duplication and generation of nonfunctional pseudogenes. Pseudogenes may, in turn, serve as a constant and renewable reservoir of variation that can be effectively transferred to linked susceptibility genes via gene conversion, thus partly counteracting the negative effect on reproduction (26).

How does genetic variation in the *PRODH2/DGCR6* locus affect susceptibility to schizophrenia? Some clues are provided by our recent studies of a PRODH-deficient animal model. This analysis was initiated soon after our first positive genetic findings with *PRODH2* to determine the relevance of this gene to the psychiatric phenotype under study. Specifically, comparison of 42 mice homozygous for a truncated form of PRODH with ≈7-fold decreased activity and 26 wild-type littermates revealed a deficit in prepulse inhibition, a measure of sensorimotor gating, a central inhibitory/filtering mechanism affected in patients with psychiatric disorders including schizophrenia (21). Proline itself may serve as a direct modulator of glutamatergic transmission in the brain, a role suggested primarily by the selective expression of a brain-specific high-affinity proline transporter (SLC6A7) in a subset of glutamatergic synapses (27) and supported by neurochemical and electrophysiological studies (21). An alternative model is suggested by recent observations that implicate *PRODH2* in apoptosis (28).

In the present study, we typed one to three SNPs from 13 of the 23 known genes mapping in the entire 22q11 critical region (Fig. 1*b*). The genes examined represent 85% of the known genes located at the proximal half of the critical region, within 780 kb of the *PRODH2/DGCR6* locus. Analysis of the remaining genes is ongoing, with the ultimate goal of examining individual contributions, as well as possible interactions, of all known genes that map in this region.

We are deeply grateful to all of the families who have participated in this research. We thank Alyson Knowles, Lubomira Malinova, and Shoshanna Friedman for technical assistance, Diana Hall for help with the LD calculations, and Markus Stoffel for DNA samples from an African American control group. Support for this work was provided in part by grants from the New York Academy of Medicine, the EJLB Foundation, and the Essel Foundation (M.K.), as well as the Norman and Rosita Winston Foundation (postdoctoral fellowship to H.L.). Clinical work at the Rockefeller University Hospital General Clinical Research Center was supported in part by Grant M01 RR00102.

1. Scambler, P. J. (2000) *Hum. Mol. Genet.* **9,** 2421–2426.
2. Golding-Kushner, K. J., Weller, G. & Shprintzen, R. J. (1985) *J. Cranio-facial Genet. Dev. Biol.* **5,** 259–266.
3. Baron, M. (2001) *Am. J. Hum. Genet.* **68,** 299–312.
4. Pulver, A. E., Nestadt, G., Goldberg, R., Shprintzen, R. J., Lamacz, M., Wolyniec, P. S., Morrow, B., Karayiorgou, M., Antonarakis, S. E., Housman, D. & Kucherlapati, R. (1994) *J. Nerv. Mental Dis.* **182,** 476–478.
5. Murphy, K. C., Jones, L. A. & Owen, M. J. (1999) *Arch. Gen. Psychiatry* **56,** 940–945.
6. Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P. S., Lasseter, V. K., *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7612–7616.
7. Usiskin, S. I., Nicolson, R., Krasnewich, D. M., Yan, W., Lenane, M., Wudarsky, M., Hamburger, S. D. & Rapoport, J. L. (1999) *J. Am. Acad. Child Adolesc. Psychiatry* **38,** 1536–1543.
8. Sugama, S., Namihira, T., Matsuoka, R., Taira, N., Eto, Y. & Maekawa, K. (1999) *J. Neurol. Neurosurg. Psychiatry* **67,** 803–806.
9. Edelmann, L., Pandita, R. K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R. S., Magenis, E., Shprintzen, R. J. & Morrow, B. E. (1999) *Hum. Mol. Genet.* **8,** 1157–1167.
10. Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L., *et al.* (2000) *Hum. Mol. Genet.* **9,** 489–501.
11. Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., *et al.* (1999) *Nature (London)* **402,** 489–495.
12. Sobin, C., Blundell, M. L., Conry, A., Weiller, F., Gavigan, C., Haiman, C. & Karayiorgou, M. (2001) *Psychiatry Res.* **101,** 101–113.
13. American Psychiatric Association (1994) *Diagnostic and Statistical Manual* (Am. Psychiatric Assoc., Washington, DC), 4th Ed.
14. Nurnberger J. I., Jr., Blehar, M. C., Kaufmann, C. A., York-Cooler, C., Simpson, S. G., Harkavy-Friedman, J., Severe, J. B., Malaspina, D. & Reich, T. (1994) *Arch. Gen. Psychiatry* **51,** 849–859.
15. Nicolson, R., Lenane, M., Singaracharlu, S., Malaspina, D., Giedd, J. N., Hamburger, S. D., Gochman, P., Bedwell, J., Thaker, G. K., Fernandez, T., *et al.* (2000) *Am. J. Psychiatry* **157,** 794–800.
16. Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) *Genome Res.* **7,** 996–1005.
17. Chen, X., Levine, L. & Kwok, P. Y. (1999) *Genome Res.* **9,** 492–498.
18. Dempster, A. P., Laird, N. M. & Rubin, D. M. (1997) *J. R. Stat. Soc.* **39B,** 1–38.
19. Terwilliger, J. & Ott, J. (1992) *Hum. Hered.* **42,** 337–346.
20. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993) *Am. J. Hum. Genet.* **52,** 506–516.
21. Gogos, J. A., Santha, M., Takacs, Z., Beck, K. D., Luine, V., Lucas, L. R., Nadler, J. V. & Karayiorgou, M. (1999) *Nat. Genet.* **21,** 434–439.
22. Bassett, A. S., Hodgkinson, K., Chow, E. W., Correia, S., Scutt, L. E. & Weksberg, R. (1998) *Am. J. Med. Genet.* **81,** 328–337.
23. Demczuk, S., Thomas, G. & Aurias, A. (1996) *Hum. Mol. Genet.* **5,** 633–638.
24. Collier, S., Tassabehji, M., Sinnott, P. & Strachan, T. (1993) *Nat. Genet.* **3,** 260–265.
25. Phang, J. M., Yeh, G. C. & Scriver, C. R. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw–Hill, New York), Vol. 1, pp. 1125–1146.
26. Erlenmeyer-Kimling, L. (1978) in *Annual Review of the Schizophrenic Syndrome*, ed. Cancro, R. (Brunner/Mazel, New York), Vol. 5, pp. 298–333.
27. Renick, S. E., Kleven, D. T., Chan, J., Stenius, K., Milner, T. A., Pickel, V. M. & Fremeau, R. T., Jr. (1999) *J. Neurosci.* **19,** 21–33.
28. Maxwell, S. A. & Davis, G. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13009–13014.