



ORIGINAL RESEARCH ARTICLE

Identification of single nucleotide polymorphisms (SNPs) and other sequence changes and estimation of nucleotide diversity in coding and flanking regions of the NMDAR1 receptor gene in schizophrenic patients

SR Rice¹, N Niu¹, DB Berman¹, LL Heston² and JL Sobell¹

¹Division of Molecular Medicine, City of Hope National Medical Center, Duarte, CA; ²Department of Psychiatry, University of Washington, Seattle, WA, USA

Glutamatergic dysregulation has been hypothesized to play a role in schizophrenia. The *N*-methyl-D-aspartate (NMDA) type of glutamate receptor especially is of interest because, in addition to binding sites for glutamate and glycine, a necessary co-agonist, this receptor also contains noncompetitive binding sites for the psychotomimetics phencyclidine (PCP), MK-801, and ketamine. PCP-induced psychosis has been a useful disease model in that both the positive as well as the negative symptomatology seen in schizophrenia are observed. Recently, a mouse deficient in expression of the NR1 subunit gene (NMDAR1) of the heteromeric receptor has been developed and shown to display aberrant behaviors, with reduced social and sexual interactions as well as increased stereotypic motor activity. In an extensive examination of the NMDAR1 gene in our laboratory in approximately 100 chronic schizophrenic patients, 28 unique sequence changes were identified, including eight single nucleotide polymorphisms (SNPs) in the 5' untranslated region (5'UTR), six SNPs in coding regions (cSNPs), eleven intronic SNPs, two intronic deletions of 7 and 30 bp, and an intronic microinsertion/deletion. With the exception of one previously reported cSNP, all of the identified changes were novel. The frequency of polymorphisms differed significantly by ethnicity and several appeared to be in linkage disequilibrium. None of the changes appeared likely to be of functional significance, thus suggesting that changes in the genomic NMDAR1 are unlikely to contribute to the etiology of schizophrenia. Estimates of nucleotide diversity are comparable to those observed in studies of other genes. *Molecular Psychiatry* (2001) 6, 274–284.

Keywords: schizophrenia; glutamate receptor; *N*-methyl-D-aspartate; single nucleotide polymorphism; rare variant; nucleotide diversity; putative promoter

Introduction

Glutamate is the major excitatory neurotransmitter in vertebrate central nervous systems (CNS) and is bound by several receptor types, including the NMDA-type of ligand gated receptor. Features of the NMDA-type receptor include the ability to gate calcium, the requirement for glycine as a co-agonist, and the existence of a voltage-dependent magnesium block. The NMDA receptor appears to play a significant role in the organization of synaptic networks during development, in the mediation of certain forms of synaptic plasticity such as long-term potentiation, and in the regulation of excitotoxicity.^{1,2}

The NMDA-type receptor is of interest in schizophrenia due to its role in synaptic formation and in CNS plasticity.³ Additionally, the receptor includes a

specific and noncompetitive binding site for phencyclidine (PCP), a psychotomimetic. The behavioral changes induced by PCP mimic those observed in schizophrenic patients, including both positive as well as negative symptoms. In addition, schizophrenia-like cognitive and neuropsychological deficits may be induced (eg, PCP inhibits spatial learning in rats and paired associative learning in monkeys).⁴ Certain classes of NMDA-blocking drugs, designed for the treatment of NMDA-mediated neurotoxicity following stroke or prolonged seizures, induce psychosis.⁵ Sub-clinical doses of ketamine, an anesthetic agent, have been demonstrated to produce psychosis in normal subjects and to exacerbate symptoms in psychiatric patients.⁶

The stoichiometry of the native NMDA receptor is not certain, although recent data suggest that the complex is a tetramer with two NR1 subunits and any two of four known NR2 subunits (NR2A, B, C, or D)³ which function to modulate channel activity. A third family of NR subunits, NR3, has been reported.⁷ The NR3 sub-

unit appears to have inhibitory properties and may be expressed during development.⁷ Functional receptor diversity is further enhanced by the existence of eight splice variants for NR1, created by the alternative use of one exon in the N-terminal domain (N1 cassette) and two exons in the C-terminal domain (C1 and C2 cassettes).⁸

Our laboratory is systematically investigating the potential role of the NMDA receptor in schizophrenia through direct examination of each of the subunit genes in a cohort of patients in search of causative mutations. This paper reports findings in our examination of all 22 exons and portions of the 5' and 3'UTRs of the NMDAR1 subunit gene. Recently, mice expressing only 5% of normal levels of NMDAR1 have been generated and observed to display behavioral abnormalities, including decreased social and sexual interactions as well as increased motor activity and stereotypic actions, that are ameliorated by neuroleptic treatment.⁹

Materials and methods

Patient samples

All schizophrenic patients met disease criteria as defined by the Diagnostic and Statistical Manual, Third Edition, Revised (DSM-III-R), as described previously.¹⁰ The majority of patients were ascertained through state mental institutions in Minnesota, Washington, and Oregon. A set of 96 individuals were *screened* for sequence changes in all gene regions of interest, with an additional 18 *screened* in only the putative promoter and 5'UTR. The distribution of the *screened* patients by race and gender is shown in Table 1. Once sequence changes were identified in the screened samples, additional schizophrenic patients were *scored* for the presence of the identified change. The distribution of the *scored* patients by race and gender is shown in Table 2.

The control pool, as described elsewhere^{10,11} consists of individuals aged 45 and above who were ascertained as community medicine patients at the Mayo Clinic and who were representative of the county population in terms of ethnic background. Each individual completed a brief screening questionnaire on personal and family history of psychiatric illness. Medical history

Table 1 Distribution of screened schizophrenic patients by gender and race^a

	<i>Caucasian</i>	<i>African American</i>	<i>Asian</i>	<i>Native American</i>	<i>Total</i>
Male	41 (11)	8 (0)	4 (0)	4 (1)	57 (12)
Female	30 (4)	7 (1)	1 (0)	1 (1)	39 (6)
Total	71 (15)	15 (1)	5 (0)	5 (2)	96 (18)

^aA core group of 96 patients were screened for sequence changes in all regions; an additional 18 patients (in parentheses) were screened for sequence changes in the 5'UTR only.

Table 2 Distribution of scored schizophrenic patients by gender and race^a

	<i>Caucasian</i>	<i>African American</i>	<i>Asian</i>	<i>Native American</i>	<i>Total</i>
Male	69	28	2	6	106
Female	18	4	0	1	23
Total	87	32	2	8	129

^aNot all of the additional patients were scored for every sequence change. The exact numbers scored for each sequence change are listed in Table 4.

data were available from the Medical Index (MI). The MI is a longstanding Mayo Clinic resource that was developed as part of the Rochester Epidemiology Project.¹² Briefly, after each patient visit to the Clinic, the physician is required to enter all diagnoses on a Master Sheet. These diagnoses are then codified by a team of medical nosologists and entered into the MI. Extensive medical records are available for each of the individuals included in the control DNA bank. These extensive records exist because: (a) the Mayo Clinic represents the largest provider of medical care for the surrounding Olmsted county residents; (b) the county population is stable and, consequently, individuals often have lifetime medical records available; and (c) the Mayo Clinic maintains a central medical record on each patient that includes findings from all physician office visits (including psychiatry/psychology); hospitalizations; emergency room visits; prescription drug data; and all pathology, radiology, and other laboratory tests. In reference to psychiatric services, the Mayo Clinic represents the largest source of psychiatric care in the community and maintains the only acute care psychiatric hospital beds in and beyond the county boundaries. Evidence of psychotic illness, including affective psychosis, brief reactive psychosis, drug-induced psychosis, and the like can be readily retrieved from the MI or the actual medical records and individuals with evidence of psychotic disorder, or other major psychiatric disorder (eg, major depressive episode), use of prescription drugs such as antipsychotics, antidepressants, antimania, or other major psychoactive drugs can be excluded.

Gene screening and genotype scoring

Regions of interest included all 22 exons and flanking intronic sequences; 1518 bp of the 5'UTR; and 64 bp of the 3'UTR. The size of the 5' and 3'UTR fragments reflected the number of basepairs available in GenBank at the commencement of the project.

Nomenclature The nomenclature used herein for nucleotide numbering corresponds to sequences contained in GenBank. Accession number Z32772 includes the 5'UTR through the 5' portion of intron 2; Z32773 includes the 3' portion of intron 2 through the 5' portion of intron 5; and Z32774 includes the 3' portion of intron 5 through the 3'UTR. Based on these

sequences, the alternatively spliced cassettes correspond to exons 4 (N1), 20 (C1) and 21 (C2). The GenBank protein accession file NP_000823 was the basis for amino acid numbering; however, as this file does not include the 21 amino acids of the alternatively spliced exon 4, all codon numbers beyond 190 (the last codon of exon 3) have been increased by 21. This numbering system was used so that the full genomic coding sequence would be represented in the protein translation.

DNA amplification DNA was extracted by standard commercial protocol (Analytical Genetic Testing Center, Denver, CO, USA). Most PCR amplifications were carried out in a total volume of 25 λ with 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5–3.5 mM MgCl₂, 200 μ M of each deoxyribonucleoside triphosphate, 0.5–1.0 U Taq DNA polymerase (Boehringer-Mannheim, Indianapolis, IN, USA), 0–10% dimethyl sulfoxide, 0.1–0.25 pmol of each primer and 100 ng of genomic DNA. For specific regions with a high G+C content, 16 mM (NH₄)₂SO₄, 67 mM Tris-HCl (pH 8.8), 0.01% Tween-20 and 1.25 U Biolase Platinum (Intermountain Scientific, Kaysville, UT, USA) were substituted with the other parameters unchanged. Cycling conditions were an initial denaturation at 95°C for 1 min followed by 30–35 cycles of denaturation at 95°C for 20 s, annealing at 50–65°C for 30 s–1 min, and elongation at 72°C for 1–2 min with the Perkin-Elmer GeneAmp PCR system 9600.

Some fragments (1.8–2.5 kb) were amplified using Expand High Fidelity 10 \times Buffer, 4.5% dimethyl sulfoxide, 200 μ M of each deoxyribonucleoside triphosphate, 12.8–13.8 μ l ddH₂O, 100–200 ng genomic DNA and 1.4–1.75 U of Expand High Fidelity enzyme (Boehringer-Mannheim). Cycling conditions in these assays were an initial denaturation at 95°C for 3 min followed by 10 cycles of denaturation at 95°C for 15 s, annealing at 57–60°C for 30 s and elongation at 72°C for 3 min, then 20 cycles of denaturation at 95°C for 20 s, annealing at 57–60°C for 30 s and elongation at 72°C for 3 min. There was an incremental time increase of 20 s/cycle in the elongation step. A final extension step at 72°C for 7 min was performed.

Nested PCR was utilized to further increase the specificity of PCR products. For some regions, the first round PCR generated fragments between 1.8 and 2.5 kb, with the subsequent nested PCR generating final fragment sizes between 300–500 bp. In a few regions, the first round PCR included spurious amplification products, so a nested PCR was performed to generate a clean fragment just slightly smaller than the initial product. The template for all nested amplifications was a 5- to 10-fold diluted product from the first round PCR. Conditions for the nested amplifications were as described above for the assays using Taq polymerase. All oligonucleotide primer sequences for first round and nested amplification are available upon request.

Dideoxy fingerprinting (ddF) and bi-directional dideoxy fingerprinting (bi-ddF) Briefly, ddF is a hybrid

technique combining single strand conformation polymorphism (SSCP) analysis and dideoxy sequencing. In ddF, only one of the four standard dideoxy sequencing reactions is performed, and the products are resolved by electrophoresis on a non-denaturing matrix. Mutations result in extra, diminished/missing, or shifted bands. Herein, samples were sequenced with one dideoxyribonucleotide in the downstream orientation and, in a second reaction, with another dideoxy in the upstream orientation to increase the likelihood of finding sequence polymorphisms (eg, downstream sequencing with ddC and upstream sequencing with ddA; downstream sequencing with ddG and upstream sequencing with ddT).

Bi-ddF is an extension of ddF that allows both the downstream and upstream sequences to be examined simultaneously. In the sequencing component of bi-ddF, both the downstream and upstream sequencing are included in a single reaction tube, with one dideoxynucleotide. The reaction product is run in a single lane during the electrophoresis. As with ddF, mutations result in extra, diminished/missing, or shifted bands that may be detected in either orientation. For example, in a bi-ddF performed with ddA, a G→A polymorphism could be detected as an added band in the downstream orientation while a G→T change could be detected in the upstream orientation as an added band (C→A). By conducting a second sequencing reaction using a different dideoxynucleotide with both downstream and upstream primers and electrophoresing the product in a single lane, all possible sequence changes theoretically can be detected.¹³ Using the above example of a G→A polymorphism, a second sequencing reaction using ddC could be detected by a diminished band in the upstream orientation (C→T). The conditions for the bi-ddF assays were identical to those utilized in unidirectional ddF except that both the downstream and upstream primers were added to the cycling mixture.

ddF was utilized in most gene regions, with the exception of the 5'UTR, exons 10–11, and exons 18–21b, in which the DNA was examined by bi-ddF. Extensive analyses have demonstrated that ddF and bi-ddF detect virtually 100% of mutations with high specificity.^{14–17}

For either ddF or bi-ddF, PCR products were pre-treated with 10 U exonuclease I and 2 U shrimp alkaline phosphatase and incubated at 37°C for 15 min followed by a termination step of 80°C for 15 min. A total of 50–500 ng of the treated product was used in conjunction with the Thermosequenase Radiolabeled Terminator Cycle Sequencing kit (Amersham, Arlington Heights, IL, USA) according to manufacturer's specifications to generate one of the four standard dideoxy sequencing reaction products. Cycling conditions were denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and an extension step of 60°C for 7 min. Deoxyinositol triphosphate (dITP) was substituted for the guanine triphosphate (dGTP) when the G+C content was greater than 50%. Reactions were terminated by the addition of 4 μ l of the proprietary stop solution and

denatured at 95°C for 5 min. Samples were transferred to ice for 5 min prior to loading.

Initially, prior to the availability of the Thermosequenase kit with radiolabeled terminators, end-labeled primers were used in standard cycle sequencing reactions in a total volume of 8 μ l, with 50 mM Tris-HCl, 50 mM KCl (pH 8.8), 10 μ M of each dNTP, one ddNTP (either ddATP (400 μ M), ddTTP (600 μ M), ddGTP (60 μ M), or ddCTP (200 μ M), 100 ng of the amplified DNA, and 0.003–0.006 pmol end-labeled sequencing primer. For regions with >55% G+C content, the concentrations of ddGTP and ddCTP were reduced to 50%, 5–10% dimethyl sulfoxide was added, and 7-deaza-2'-deoxy-GTP was used to minimize compression. Sequencing primers were end-labelled with γ^{32} P dATP or γ^{32} P dATP by T4 polynucleotide kinase at 37°C for 30 min and heated at 65°C for 5 min to stop the reaction. The ddF PCR cycling conditions were denaturation at 95°C for 20 s, annealing at 60°C for 30 s and elongation at 72°C for 1 min for a total of 30 cycles. Reactions were terminated by heating for 3 min with 7 μ l of stop/loading buffer (7 M urea, 50% deionized formamide, 2 mM EDTA, 7.5% xylene cyanol, and 7.5% bromophenol blue), and the samples were chilled on ice for 5 min. For regions of high G+C content (>60%), the Thermosequenase kit provided better results than the standard cycle sequencing protocol. Subsequently, the kit was utilized regardless of template sequence because consistently better products were achieved.

A total of 1.5–2.5 μ l of the cycle sequencing products were loaded on a 0.5 \times MDE (FMC BioProducts, Rockland, ME, USA) gel with 0–2.0 M urea. Electrophoresis was performed at 27 W for 4–7 h at 12°C. The gels were transferred to Whatman blotting filter paper, dried and exposed to X-ray film.

The utilization of either ddF or bi-ddF in the present study was dictated primarily by the G+C content of the target sequence and by technician skill in interpreting the often complex bi-ddF patterns. The techniques are of similar high sensitivity even for fragments as large as 600 basepairs.^{13,14} If a complex pattern resulted from the bi-ddF due to several different sequence changes in the region, lanes with similar patterns were grouped and representative samples were directly sequenced to determine the exact polymorphism(s). Similarly, in the ddFs, samples with unusual fingerprints were manually sequenced to locate the exact change.

Sequencing As noted, samples with variant banding patterns were re-amplified from a new aliquot of genomic DNA by PCR and were analyzed by cycle sequencing in both the sense and antisense orientations using the Amersham Thermosequenase Radiolabeled Terminator Cycle Sequencing kit according to manufacturer's specifications. Once sequence changes were verified, additional patient samples were scored for genotype. The scoring was done by abbreviated cycle sequencing (two dideoxynucleotides only) as described above.

Data analysis

Assessment of likely functional significance Sequence changes were examined in the context of the flanking sequence to determine if functional consequences of the changes were likely (eg, creation of cryptic splice sites, obliteration of normal splice junction, alteration of consensus sequence for transcription factor binding). In the examination of sequence in the 5'UTR, multiple analyses with public domain software were conducted, including MatInspector Version 2.2,¹⁸ GeneExpress¹⁹ and TFSearch/TRANSFAC.²⁰

Estimates of nucleotide diversity Nucleotide diversity (θ) and standard deviations ($S(\theta)$) were calculated for coding and non-coding regions under the assumption of an infinite site neutral model^{21,22} where $\theta = K/aL$ and $S(\theta) = \sqrt{a\theta L + b(\theta L)^2/aL}$; $a = \sum_{i=2}^K [1/(i-1)]$, $b = \sum_{i=2}^K [1/(i-1)^2]$; and K = the number of observed SNPs among L basepairs of genomic sequence in a sample of n alleles.

Results

Approximately 650 kb of genomic sequence was screened for novel sequence changes in a core sample of 96 patients (~6.6 kb/individual) (Table 1). All exons, including at least 50 bp of 5' and 3' flanking intronic sequence, as well as approximately 1.5 kb of the 5' untranslated region and 64 bp of the 3'UTR were screened initially in both the sense and antisense orientations (~1.3 megabases total). In the 5'UTR, an additional 18 schizophrenic patients were screened and the initial 96 patients were rescreened following the report of a mouse model of schizophrenia with markedly reduced NR1 expression (5–10% of normal) for a total of 114 subjects screened in this region (Table 1). Variants in ddF or bi-ddF pattern were sequenced in both directions to identify the change. To confirm the changes, DNA from individuals with the variant were reamplified from stock supplies and resequenced in both orientations to confirm the findings.

Once sequence changes were identified, additional patient samples were scored for the presence of the change to increase the sample size for estimation of allele frequency (Table 2). In several circumstances, the scoring assay resulted in the detection of additional novel sequence changes. All sequence changes, including those first found during scoring, are included in Tables 3 and 4.

In the combined analysis (initial screening as well as scoring assays), a total of 28 different sequence changes were identified (Tables 3 and 4). These included eight single nucleotide polymorphisms (SNPs) in the 5'UTR, five novel and one previously observed SNPs in coding regions (cSNPs), eleven novel intronic SNPs, two intronic deletions of 7 and 30 bp, and a novel intronic microinsertion/deletion (Tables 3 and 4). The frequency of changes across all examined regions was 1/236, with a nucleotide diversity of 7.3×10^{-4}

Table 3 Frequency of sequence changes in coding and non-coding regions of NMDAR1^a

Region	Number (%) of bp screened per individual	Number of sequence changes	Frequency of changes per basepair	Nucleotide diversity $\times 10^{-4}$ (\pm SD)
1. Non-coding ^b	3665 (0.555)	22	1/167	10.3 (3.1)
5'UTR	1500	8	1/188	8.9 (3.7)
3'UTR	65	0	0	0
Intronic	2100	14	1/150	11.4 (4.0)
2. Non-coding ^b	3665 (0.555)	19	1/193	8.9 (2.8)
5' UTR	1500	8	1/188	8.9 (3.7)
3' UTR	65	0	0	0
Intronic	2100	11	1/191	9.0 (3.3)
3. Coding	2940 (0.445)	6	1/490	3.5 (1.6)
Synonymous		6	1/490	3.5 (1.6)
Non-synonymous		0	0	0
Total (all changes)	6605 (1.00)	28	1/236	7.3 (2.1)

^aAll but three of the 28 total changes were single nucleotide polymorphisms. Nine of the 28 changes were rare variants occurring in <0.01 of the racial group in which identified. For most regions, 96 individuals (192 alleles) were screened, except in the 5'UTR where the total was 114 individuals (228 alleles). In total, approximately 1.3 mb of sequence was examined in both the sense and antisense orientations.

^bRow 1 shows $n=22$ sequence changes when all alterations (19 SNPs, two observed intronic deletions, one intronic deletion/insertion) are included. Row 2 shows $n=19$ when only SNPs are considered.

(± 2.1 SD).²¹ Six (21.4%) of the 28 total changes occurred in coding regions, which comprised about half (44.5%) of the examined nucleotides (~2940 bp coding/6605 bp total for each of 96 subjects or 282.2 kb/634.1 kb). The observed frequency of cSNPs per basepair in coding regions was 1/490 and the nucleotide diversity was 3.5×10^{-4} (± 1.6). As expected, the majority (78.6%) of sequence changes occurred among nucleotides in non-coding regions (55.5% of the total bp examined). The frequency of changes in non-coding regions was 1/167 bp if all changes, including deletions and insertions, were considered. The corresponding nucleotide diversity was 10.3×10^{-4} (± 3.1). If only single nucleotide changes are included, the frequency of SNPs per basepair in non-coding regions was 1/193 and the nucleotide diversity was 8.9×10^{-4} (± 2.8).

Coding region changes Six cSNPs were found, including one in exon 6 that had been identified previously⁶ (Table 4). All six changes involved alteration of the third basepair in the codon and resulted only in synonymous changes. Exons 3, 6, 7, and 16 each contained one synonymous change, while exon 17 contained two consecutive synonymous changes. Two of the six cSNPs (exon 6 (P284P); exon 7 (V306V) were common among Caucasians (28.6% and 29.1%, respectively) and also were observed among African Americans (5.9% and 5.6%, respectively). Although the sample sizes were quite limited, neither the exon 6 nor the exon 7 cSNP was observed among Asians, while each had a frequency of 25% among the small number of Native Americans analyzed ($n=6$). Further, in the three racial groups in which the exon 6 and exon 7 cSNPs were seen, the polymorphisms were in linkage disequilibrium, along with an intron 7 SNP (see below).

Exon 17 synonymous cSNPs (V768V; T769T) were observed only in African Americans (5.6% each), as was an exon 3 silent change (A175A) (1.6%). Lastly, a rare synonymous cSNP in exon 16 (D711D) was observed only among Caucasians (0.5%).

Non-coding region changes Non-coding regions examined per each of 96 subjects included approximately 2100 bp of intronic flanking sequence (~201.6 kb); 1500 bp (~171 kb) of 5'UTR sequence ($n=114$ subjects), and 65 bp (6.24 kb) of 3'UTR sequence. Eleven SNPs were identified in intronic regions (Table 4). Of these, seven were observed in at least 1% of the ethnic group in which identified, while the remaining four changes were found in frequencies below 1% and are designated as rare variants. Excluding these rare variants, the most frequent SNP observed was in intron 10 (G6435A), with a minor allele frequency of nearly one-fourth among Caucasians, 6.7% among African Americans, and 20.0% among the small Native American sample ($n=5$). As noted earlier, this change appeared to be in linkage disequilibrium with the exon 6 and exon 7 synonymous cSNPs. As with these cSNPs, the intron 10 minor allele was not observed among the small sample of Asians ($n=5$). The next most commonly observed intronic SNP was in intron 2 (A290G), with frequencies of 7.0% among Caucasians, 6.5% among African Americans, 7.1% among Asians, and 7.7% among Native Americans. With the exception of one change in intron 11 (A6608G) which was observed in Caucasians (7.0%) as well as in African Americans (3.3%), the remainder of the SNPs with frequencies of at least 1% were observed only in one racial group. All of the rare variants (<1% frequency) were observed only in Caucasians where the sample size was sufficient for the detection of low prevalence alleles.

Table 4 Summary of NMDAR1 sequence changes identified among schizophrenic subjects

Gene region and nucleotide change ^a	Type of change ^b	Structural change (codon)	Number of patients and frequency of minor allele by race ^{c,d}						
			Race	Screened Freq	(No.)	Scored Freq	(No.)	Total Freq	(No.)
5'UTR C112T	SNP	None	Cauc	0.006	(86)	—	—	0.006	(86)
			Afr Am	0	(16)	—	—	0	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(7)	—	—	0	(7)
5'UTR C113T	SNP	None	Cauc	0.291	(86)	—	—	0.291	(86)
			Afr Am	0.063	(16)	—	—	0.063	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0.214	(7)	—	—	0.214	(7)
5'UTR C319G	SNP	None	Cauc	0	(86)	—	—	0	(86)
			Afr Am	0.031	(16)	—	—	0.031	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(7)	—	—	0	(7)
5'UTR G716A	SNP	None	Cauc	0	(86)	—	—	0	(86)
			Afr Am	0	(16)	—	—	0	(16)
			Asian	0.100	(5)	—	—	0.100	(5)
			Nat Am	0	(7)	—	—	0	(7)
5'UTR A750G	SNP	None	Cauc	0.012	(86)	—	—	0.012	(86)
			Afr Am	0	(16)	—	—	0	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(7)	—	—	0	(7)
5'UTR G1001C	SNP	None	Cauc	0.122	(86)	—	—	0.122	(86)
			Afr Am	0.063	(16)	—	—	0.063	(16)
			Asian	0.200	(5)	—	—	0.200	(5)
			Nat Am	0.143	(7)	—	—	0.143	(7)
5'UTR T1135G	SNP	None	Cauc	0	(84)	—	—	0	(84)
			Afr Am	0.031	(16)	—	—	0.031	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(7)	—	—	0	(7)
5'UTR A1289G	SNP	None	Cauc	0.006	(84)	—	—	0.006	(84)
			Afr Am	0	(16)	—	—	0	(16)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(7)	—	—	0	(7)
Intron 2 G4553A	SNP	None	Cauc	0.014	(71)	0.014	(37)	0.014	(108)
			Afr Am	0	(15)	0	(2)	0	(17)
			Asian	0	(5)	—	—	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Intron 2 A280G	SNP	None	Cauc	0	(71)	0	(80)	0	(151)
			Afr Am	0.067	(15)	0.063	(16)	0.065	(31)
			Asian	0	(5)	0	(2)	0	(7)
			Nat Am	0	(5)	0	(8)	0	(13)
Intron 2 A290G	SNP	None	Cauc	0.064	(71)	0.076	(79)	0.070	(150)
			Afr Am	0	(15)	0.125	(16)	0.065	(31)
			Asian	0	(5)	0.250	(2)	0.071	(7)
			Nat Am	0	(5)	0.125	(8)	0.077	(13)
Intron 2 G301A	SNP	None	Cauc	0	(71)	0	(80)	0	(151)
			Afr Am	0	(15)	0	(16)	0	(31)
			Asian	0.100	(5)	0	(2)	0.071	(7)
			Nat Am	0	(5)	0	(8)	0	(13)

Continued

Table 4 *Continued*

Gene region and nucleotide change ^a	Type of change ^b	Structural change (codon)	Number of patients and frequency of minor allele by race ^{c,d}						
			Race	Screened Freq	(No.)	Scored Freq	(No.)	Total Freq	(No.)
Intron 2 G308A	SNP	None	Cauc	0	(71)	0.006	(80)	0.003	(151)
			Afr Am	0	(15)	0	(16)	0	(31)
			Asian	0	(5)	0	(2)	0	(7)
			Nat Am	0	(5)	0	(8)	0	(13)
Exon 3 G443A	cSNP	Silent (A175A)	Cauc	0	(71)	0	(80)	0	(151)
			Afr Am	0	(15)	0.031	(16)	0.016	(31)
			Asian	0	(5)	0	(2)	0	(7)
			Nat Am	0	(5)	0	(8)	0	(13)
Exon 6 A1970G	cSNP	Silent (P284P)	Cauc	0.246	(71)	0.359	(39)	0.286	(110)
			Afr Am	0.067	(15)	0	(2)	0.059	(17)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0.200	(5)	0.500	(1)	0.250	(6)
Exon 7 G2108A	cSNP	Silent (V306V)	Cauc	0.246	(71)	0.359	(46)	0.291	(117)
			Afr Am	0.067	(15)	0	(3)	0.056	(18)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0.200	(5)	0.500	(1)	0.250	(6)
Intron 7 del 2278–2307	del	None	Cauc	0.007	(71)	0	(46)	0.004	(117)
			Afr Am	0	(15)	0	(3)	0	(18)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Intron 9 T3915G	SNP	None	Cauc	0.007	(71)	–	–	0.007	(71)
			Afr Am	0	(15)	–	–	0	(15)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	–	–	0	(5)
Intron 10 G6435A	SNP	None	Cauc	0.246	(71)	–	–	0.246	(71)
			Afr Am	0.067	(15)	–	–	0.067	(15)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0.200	(5)	–	–	0.200	(5)
Intron 10 C6456GTAA del C; ins GTAA	del/ins	None	Cauc	0	(71)	–	–	0	(71)
			Afr Am	0.033	(15)	–	–	0.033	(15)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	–	–	0	(5)
Intron 11 A6608G	SNP	None	Cauc	0.070	(71)	–	–	0.070	(71)
			Afr Am	0.033	(15)	–	–	0.033	(15)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	–	–	0	(5)
Intron 11 G7027C	SNP	None	Cauc	0	(71)	0.010	(50)	0.004	(121)
			Afr Am	0	(15)	0	(2)	0	(17)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Intron 11 C7033T	SNP	None	Cauc	0.014	(71)	0	(50)	0.008	(121)
			Afr Am	0	(15)	0	(2)	0	(17)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Intron 12 del 7270–7276	del	None	Cauc	0	(71)	0.010	(50)	0.004	(121)
			Afr Am	0	(15)	0	(2)	0	(17)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)

Continued

Table 4 *Continued*

Gene region and nucleotide change ^a	Type of change ^b	Structural change (codon)	Number of patients and frequency of minor allele by race ^{c,d}						
			Race	Screened		Scored		Total	
				Freq	(No.)	Freq	(No.)	Freq	(No.)
Intron 15 A7946G	SNP	None	Cauc	0	(71)	0	(48)	0	(119)
			Afr Am	0.033	(15)	0	(2)	0.029	(17)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Exon 16 T8070C	cSNP	Silent (D711D)	Cauc	0	(71)	0.013	(38)	0.005	(109)
			Afr Am	0	(15)	0	(3)	0	(18)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	–	–	0	(5)
Exon 17 G8406C	cSNP	Silent (V768V)	Cauc	0	(71)	0	(46)	0	(117)
			Afr Am	0.067	(15)	0	(3)	0.056	(18)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)
Exon 17 G8409C	cSNP	Silent (T769T)	Cauc	0	(71)	0	(46)	0	(117)
			Afr Am	0.067	(15)	0	(3)	0.056	(18)
			Asian	0	(5)	–	–	0	(5)
			Nat Am	0	(5)	0	(1)	0	(6)

^aNucleotide numbering is based on GenBank sequence in which accession number Z32772 includes the 5'UTR through the 5' portion of intron 2; Z32773 includes the 3' portion of intron 2 through the 5' portion of intron 5; and Z32774 includes the 3' portion of intron 5 through the 3' UTR. Codon numbering is based on GenBank accession number NP_000823; however, as this file does not include the 21 amino acids of the alternatively spliced exon 4, all codon numbers beyond 190 (the last codon of exon 3) have been increased by 21.

^bSNP = single nucleotide polymorphism; cSNP = single nucleotide polymorphism in coding region; del = deletion; del/ins = deletion and insertion.

^cCaucasians were of western European descent.

^dA core sample of 71 Caucasians, 15 African American, five Asian, and five Native American ($n = 96$) schizophrenic patients had ddF or bi-ddF *screening* for polymorphism discovery in all gene regions, and an additional 15 Caucasians, one African American, and two Native Americans ($n = 18$) had bi-ddF *screening* in the 5' UTR only (Table 1). Once sequence changes were identified, samples from additional patients were *scored* for the presence of the changes. During the *scoring*, some additional novel sequence changes were identified in the scored sample. Estimates of minor allele frequency are presented for the *screened* sample, the *scored* sample, and the total sample.

In addition to the single nucleotide changes, two rare deletions and one presumed deletion/insertion event occurred in intronic regions (Table 4). The observed deletions were of 30 basepairs in intron 7 (0.4% among Caucasians) and of seven basepairs in intron 12 (0.4% among Caucasians). In intron 10, a cysteine nucleotide was replaced by four nucleotides (GTAA) (3.3% among African Americans).

Among the 166.5 kb of examined 5'UTR sequence, including the putative proximal promoter, eight SNPs were identified (Table 4). Two were rare variants (C112T; A1289G) observed only among Caucasians (0.6% each). Of the six SNPs with frequencies of at least 1%, the most commonly observed was C113T, with minor allele frequencies of 29.1% among Caucasians, 6.3% among African Americans, and 21.4% among Native Americans. The next most common SNP was G1001C, with minor allele frequencies of 12.2% among Caucasians, 6.3% among African Americans, 20.0% among Asians, and 14.3% among Native Americans. Of the remaining SNPs with frequencies of at least 1%, one (G716A) was found only among Asians (10.0%), two (C319G; T1135G) were found only among

African Americans (3.1% each), and one (A750G) was observed only in Caucasians (1.2%).

In the 3'UTR, only 6.24 kb of sequence were examined. No sequence changes were identified.

Discussion

Examination of the entire coding region, flanking intronic sequence, and both 5' and 3' UTR sequence of the NR1 gene among approximately 100 schizophrenic patients failed to reveal sequence changes of likely functional significance. In coding regions, six single nucleotide polymorphisms were identified, but all resulted in synonymous substitutions. Based on nucleotide composition and codon position, none appeared likely to create cryptic splice sites or to disrupt splicing enhancer or repressor motifs.^{23–27} The identified synonymous cSNPs occurred in exons 3, 6, 7, 16, and 17, none of which is alternatively spliced. Further, none of the nucleotide changes appear to create either better splice donor or acceptor consensus sequences, based on a visual examination of linear DNA sequence con-

text. The majority of NR1 splice junctions adhere to the GT-AG rule.

In screening for sequence alterations in flanking intronic regions, oligonucleotide primer sets were designed so that 50–100 bp of both 5' and 3' sequence, including the splice donor and acceptor sites as well as putative branch site consensus sequences, were examined. In intron 10, a presumed microdeletion/insertion was found in an African-American male patient. A cysteine nucleotide appeared to have been deleted and four bases, GTAA, were inserted. Alternatively, a C→G transversion may have occurred as well as a three nucleotide insertion. However, as no other individuals were found to have only a C→G transversion among any of the examined racial groups, an insertion/deletion event may be the more likely explanation. The sequence alteration appeared to create a stronger consensus sequence for selection of the lariat branch site during pre-mRNA processing (described in Hammond *et al*, in preparation). However, the change did not correlate with disease status and thus is unlikely to be a deleterious change related to the schizophrenia phenotype (Hammond *et al*, in preparation).

5'UTR The human NMDAR1 promoter has not yet been functionally characterized. As reviewed by Myers and colleagues, the promoter region appears to have characteristics in common with other glutamate receptors (eg, GluR1, GluR2, KA2, NR2B, NR2C), including lack of TATA and CAAT boxes and neural specific expression.²⁸ NMDAR1 is ubiquitously expressed in all brain regions and throughout early development and adult life. A search for sequence alterations in the 1.5-kb region upstream of the translation initiation site in 114 schizophrenic patients revealed eight single nucleotide changes. All but two were of low frequency. The C113T transition had a minor allele frequency of nearly 30% among Caucasians, 21.4% among Native Americans, and 6.3% among African Americans. The change was not observed among the ten Asian alleles that were examined. It should be noted that a thymine nucleotide at position 113 is reported as wildtype in GenBank Z32772, but, based on our findings, this represents the minor allele. All ethnic groups had the G1001C transversion, with minor allele frequencies of 12.2% among Caucasians, 6.3% among African Americans, 20% among Asians and 14.3% among Native Americans. The G1001C change appeared to alter the +1 G of the transcription factor NF Kappa B50 consensus sequence (GGGG) based on analyses performed with MatInspector Version 2.2.¹⁸ However, further analyses with GeneExpress did not predict a binding site at this sequence.¹⁹ Thus, it was reasoned that this alteration, especially given the relatively high frequencies of the minor allele, was not likely to functionally alter gene expression.

A rare variant, C112T, was identified in one Caucasian patient (0.6%) who was also homozygous for the C113T variant. As assessed through the MatInspector Version 2.2 software, the C112T change falls at position

+4 of the consensus sequence for a putative AP1 binding site (TGAC). Taken together, these changes also potentially could disrupt two nucleotides in the consensus binding sites for CREBP1cJUN (ACCT) (MatInspector Version 2.2) as well as for COUP TF (TGACCT) (GeneExpress).¹⁹ DNA samples and psychiatric diagnostic data were available for nearly 20 family members of this individual, including a full sibling with a diagnosis of schizoaffective disorder. However, based on the finding that this sibling was homozygous for the wildtype C112 allele and that the variant T113 allele has a frequency of nearly 30% of Caucasians, these sequence alterations do not appear to be related to the schizophrenia phenotype.

Nucleotide diversity Estimates of the nucleotide diversity and standard deviations (Table 3) for the NMDAR1 gene were calculated by the same methods as utilized by Halushka and others.²² Halushka examined 75 genes related to blood pressure homeostasis in 80 African American chromosomes and 68 northern European chromosomes derived from a group of individuals with a range of blood pressure phenotypes. High density variant detection arrays were utilized to identify single nucleotide changes. In the coding regions, nucleotide diversity was estimated as 8.0×10^{-4} (± 1.9), but ranged from a high of 15.1×10^{-4} (± 3.6) for synonymous changes to a low of 5.7×10^{-4} (± 1.4) for non-synonymous changes. Non-coding regions had an overall estimated nucleotide diversity of 8.5×10^{-4} (± 2.0). When divided into specific regions, the estimate for intronic regions was the highest (10.5×10^{-4} (± 2.6)), followed by that for the 3'UTR (8.4×10^{-4} (± 2.0)) and then the 5'UTR (6.8×10^{-4} (± 1.7)). Cargill and colleagues examined 106 genes of cardiovascular, endocrine, or neuropsychiatric relevance in a sample of 114 alleles from normal individuals of mixed European, African American, African Pygmy, or Asian descent.²⁹ Sequence changes were identified independently by either variant detector arrays or denaturing high performance liquid chromatography (DHPLC). All initially identified SNPs were confirmed by sequencing. Nucleotide diversities of 5.43×10^{-4} (± 1.36) for coding regions and a similar frequency of 5.30×10^{-4} (± 1.33) for non-coding regions were estimated. In comparing their data, both Cargill and Halushka noted that their results were in overall agreement, especially when estimates of nucleotide diversities among genes derived from northern European descent were compared.^{22,29}

In examination of the NMDAR1 gene, the highly sensitive techniques of ddF and biddF^{13,15} were utilized, followed by direct sequencing of samples with aberrant patterns. All mutation types, including insertions and deletions, were detectable. When all types of sequence changes were included in the calculation of nucleotide diversity, the estimate for non-coding regions was 10.3×10^{-4} (± 3.1) as compared to 8.9×10^{-4} (± 2.8) when only single nucleotide polymorphisms were included (Table 3). Considering only the SNPs and not other types of sequence change (Table 3, row

2), the overall nucleotide diversity estimate of 8.9×10^{-4} (± 2.8) for non-coding regions is comparable to that reported by Halushka (8.5×10^{-4} (± 2.0)) for 75 genes.²² Within the non-coding sequence, the nucleotide diversity in intronic regions was 9.0×10^{-4} (± 3.3) for NMDAR1 as compared to the Halushka estimate of 10.5×10^{-4} (± 2.6). In the 5'UTR, an estimated nucleotide diversity of 8.9×10^{-4} (± 3.7) was calculated with the NMDAR1 SNP data alone, which again is equivalent to that estimated by Halushka (6.8×10^{-4} (± 1.7)). Estimates of nucleotide diversity in the 3'UTR also were calculated by Halushka, but since only a small region of 3'UTR sequence was examined in the present study with no changes identified, reasonable comparisons can not be made.

Nickerson examined approximately 10 kb of the lipoprotein lipase (LPL) gene through direct sequencing in 142 alleles from normal individuals representing Caucasian, African-American, and Finnish ancestry and reported a nucleotide diversity of 5.0×10^{-4} (± 5.0) for coding regions and as high as 21×10^{-4} (± 10) for non-coding regions.³⁰ The LPL gene contains several regions of interspersed repeats (eg, ALU, L1, and simple repeats) and nucleotide diversity was highest among these regions (32×10^{-4} (± 18)). By utilizing direct sequencing as the method of polymorphism discovery, changes of all types could be identified. Of the nine insertions or deletions identified in the LPL gene, eight were in regions of repeat DNA.³⁰

Our estimate of nucleotide diversity in coding regions (3.5×10^{-4} (± 1.6)) is similar to that found by Nickerson (5.0×10^{-4} (± 5.0)) and Cargill (5.43×10^{-4} (± 1.4)), with the estimate of Halushka (8.0×10^{-4} (± 1.9)) being the highest. These estimates are for all races combined; perhaps the data would correlate even more closely if race-specific figures were compared. However, due to the limited number of non-Caucasian alleles in our sample, robust estimates can not be derived. Overall, these estimates of nucleotide diversity in coding regions are not widely divergent.

For non-coding regions and all types of sequence changes, the Nickerson estimate was 21×10^{-4} (± 10.0) compared to our estimate of 10.3×10^{-4} (± 3.1) (Table 3, row 1). The difference in estimates most likely is a reflection of the increased rate of polymorphism in the repeated sequences in the LPL gene and the fact that Nickerson sequenced complete intronic regions while the present work included only flanking regions containing important regulatory sequences (eg, splice junctions and branch site consensus sequences). Additionally, given the large reported standard deviation (± 10.0) for the Nickerson point estimate of nucleotide diversity, the confidence intervals around their estimate and ours do overlap, signifying a lack of statistically significant differences. The lower estimates of nucleotide diversity in non-coding regions by Halushka (8.5×10^{-4} (± 2.0)) and by Cargill (5.3×10^{-4} (± 1.3)) include only single nucleotide polymorphisms. Other variables, such as differences in the racial substructure of the sample set studied, differences in the distribution of SNPs between genes²⁹ and the power to

detect changes given the sample size, may influence observed differences in nucleotide diversity estimates between studies.

Lack of support for disease association The estimates of nucleotide diversity observed for the NMDAR1 gene in our sample of western Europeans, African Americans, Asians, and Native Americans are roughly comparable to those calculated in other, often more extensive, analyses. Based on these data, along with the reported sensitivities of ddF and bi-ddF,^{14–17} a comprehensive screening of all coding regions, flanking intronic regions, 5'UTR, and a small portion of the 3'UTR has been performed among a group of chronically ill schizophrenic subjects. No sequence changes of likely functional significance, such as missense substitutions in highly conserved residues or nonsense, frameshift, or splice junction changes, were observed.

In a screening of approximately 1.5 kb upstream of the start of translation, several sequence changes that may interrupt consensus binding sites for transcription factors were identified, but these data are speculative and the results of analyses with different site recognition software tools (eg, GeneExpress, MatInspector) varied. Furthermore, a two basepair change in a four basepair recognition site for COUP transcription factor, identified in one schizophrenic individual, was not observed in a schizoaffective full sibling.

Polymorphisms identified in the schizophrenic subjects of western European descent (ie, the subjects designated as Caucasian) also were observed in control samples. These controls, as described in Materials and Methods, were unaffected, unrelated individuals of western European descent who were used only for scoring of sequence variations identified in the polymorphism discovery screening of cases.

Both because only polymorphisms not likely to be of functional significance (eg, no missense changes, no nonsense changes, no frameshift mutations, no disruption of likely promoter elements) were identified in the NMDAR1 gene among our schizophrenic patients and because, among the Caucasian subjects, these same changes were observed in unrelated, unaffected controls of similar ethnicity, we conclude that the observed polymorphisms are unlikely to be associated with increased risk of schizophrenia. However, the rare variants identified in the patients were not searched for in the controls, nor were the polymorphisms that were identified in only the minority samples (African American, Native American, and Asian) as a sufficient number of control subjects of similar ethnicity were not available.

Although it seems unlikely that any of the sequence variants identified in our examination of NMDAR1 is associated with a significantly increased risk of schizophrenia, we cannot rule out the possibility that a particular variant is associated with disease in only one family. Additional affected and unaffected family members would be required to address that possibility.

Because our findings do not support an association between genomic changes in the NMDAR1 gene and

schizophrenia, in this context, the observed nucleotide diversities should provide unbiased estimates with regard to the affection status of the study subjects.

Acknowledgements

This work was supported in part by grants from the National Alliance for Research on Schizophrenia and Depression (NARSAD), the National Institute of Mental Health (MH54232), and the Beckman Research Institute at the City of Hope National Medical Center. We graciously thank all research participants and acknowledge the support and cooperation of staff from the Mayo Clinic (MN), Zumbro Valley Mental Health Center (MN), St Peter Regional Treatment Center (MN), Anoka-Metro Regional Treatment Center (MN), Western and Eastern State Hospitals (WA), and Dammasch State Hospital (OR).

References

- 1 Wood MW, VanDongen HMA, VanDongen AMJ. The 5'-untranslated region of the n-methyl-d-aspartate receptor. NR2A subunit controls efficiency of translation. *J Biol Chem* 1996; **271**: 8115–8120.
- 2 Laube B, Kuhse J, Betz H. Evidence for a tetrameric structure of recombinant NMDA receptors. *J Neurosci* 1998; **18**: 2954–2961.
- 3 Tamminga CA. Schizophrenia and glutamatergic transmission. *Crit Rev Neurobiol* 1998; **12**: 21–36.
- 4 Javitt DC, Zukin SR. Recent advances in the phencyclidine model of schizophrenia. *Am J Psychiatry* 1991; **148**: 1301–1308.
- 5 Kandel E. Disorders of thought: schizophrenia. In: Kandel ER, Schwartz JH, Jessell TM (eds). *Principles of Neural Science*, 3rd edn. Elsevier Press, New York, 1991, p 866.
- 6 Malhotra A. Park City Molecular Psychiatry Conference, Feb 2–4, 1997. Meeting report: *Psych Genet* 1997; **7**: 53–56.
- 7 Das S, Sasaki YF, Rothe T, Premkumar LS, Takasu M, Crandall JE *et al*. Increased NMDA current and spine density in mice lacking the NMDA receptor subunit NR3A. *Nature* 1998; **393**: 377–381.
- 8 Hollmann M, Heinemann S. Cloned glutamate receptors. *Annu Rev Neurosci* 1994; **17**: 31–108.
- 9 Mohn AR, Gainetdinov RR, Caron MG, Koller BH. Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. *Cell* 1999; **98**: 427–436.
- 10 Sobell J, Heston L, Sommer S. Novel association approach for determining the genetic predisposition to schizophrenia: case-control resource and testing of a candidate gene. *Am J Med Genet (Neuropsych Genet)* 1993; **48**: 28–35.
- 11 Sobell JL, Lind TJ, Sigurdson DC, Zald DH, Snitz BE, Grove WM *et al*. The D5 dopamine receptor gene in schizophrenia: identification of a nonsense mutation and multiple missense changes but lack of association with disease. *Hum Mol Genet* 1995; **4**: 507–514.
- 12 Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am* 1981; **245**: 54–63.
- 13 Liu Q, Feng J, Sommer S. Bi-directional dideoxy fingerprinting (BiddF): a rapid method for quantitative detection of mutations in genomic regions of 300–600 bp. *Hum Mol Genet* 1996; **5**: 107–114.
- 14 Sarkar G, Yoon H-S, Sommer S. Dideoxy fingerprinting (ddF): a rapid and efficient screen for the presence of mutations. *Genomics* 1992; **13**: 441–443.
- 15 Liu Q, Sommer S. Parameters affecting the sensitivities of dideoxy fingerprinting and SSCP. *PCR Meth Applic* 1994; **4**: 97–108.
- 16 Lancaster JM, Berchuck A, Futreal PA, Wiseman RW. Dideoxy fingerprinting assay for BRCA1 mutation analysis. *Mol Carcinog* 1997; **19**: 176–179.
- 17 Puck JM, Middleton L, Pepper AE. Carrier and prenatal diagnosis of X-linked severe combined immunodeficiency: mutation detection methods and utilization. *Hum Genet* 1997; **99**: 628–633.
- 18 Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 1995; **23**: 4878–4884.
- 19 Kolchanov NA, Ponomarenko MP, Frolov AS, Ananko EA, Kolkpakov FA, Ignatieva EV *et al*. Integrated databases and computer systems for studying eukaryotic gene expression. *Bioinformatics* 1999; **15**: 669–686.
- 20 Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV *et al*. Databases on transcriptional regulation: TRANSFAC, TRRD, and COMPEL. *Nucleic Acids Res* 1998; **26**: 364–370.
- 21 Li W-H. *Molecular Evolution*. Sinauer Associates, Sunderland, MA, 1997, pp 237–238.
- 22 Halushka MK, Fan J-B, Bentley K, Hsie L, Shen N, Weder A *et al*. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 1999; **22**: 239–247.
- 23 D'Souza I, Poorkaj P, Hong M, Nochlin D, Lee VM-Y, Bird TD *et al*. Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc Natl Acad Sci USA* 1999; **96**: 5598–5603.
- 24 Chew SL, Liu HX, Mayeda A, Krainer AR. Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. *Proc Natl Acad Sci USA* 1999; **96**: 10655–10660.
- 25 Kuo BA, Norton PA. Accurate selection of a 5' splice site requires sequences within fibronectin alternative exon B. *Nucl Acid Res* 1999; **27**: 3945–3952.
- 26 Modafferi EF, Black DL. Combinatorial control of a neuron-specific exon. *RNA* 1999; **5**: 687–706.
- 27 Liu W, Qian C, Francke U. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nat Genet* 1997; **16**: 328–329.
- 28 Myers SJ, Dingledine R, Borges K. Genetic regulation of glutamate receptor ion channels. *Annu Rev Pharmacol Toxicol* 1999; **39**: 221–241.
- 29 Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N *et al*. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; **22**: 231–238.
- 30 Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J *et al*. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; **3**: 233–240.