Neuroscience Letters 475 (2010) 80–84

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Rare *NRXN1* promoter variants in patients with schizophrenia

Abhishek K. Shah[a], Nina M. Tioleco[a], Karen Nolan[b,c], Joseph Locker[d], Katherine Groh[a], Catalina Villa[a], Pavla Stopkova[e], Erika Pedrosa[a], Herbert M. Lachman[a,*]

[a] Department of Psychiatry and Behavioral Sciences, Division of Basic Research, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, New York 10461, United States
[b] Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York, United States
[c] Department of Psychiatry, New York University School of Medicine, New York, United States
[d] Department of Pathology, Albert Einstein College of Medicine, United States
[e] Prague Psychiatric Center, Prague, Czech Republic

ARTICLE INFO

Article history:
Received 23 December 2009
Received in revised form 10 March 2010
Accepted 18 March 2010

Keywords:
Neurexin
Schizophrenia
Bipolar disorder
Autism
CNV
Copy variant

ABSTRACT

Copy number variants (CNVs) affecting the neurexin 1 (*NRXN1*) gene have been found in a subgroup of patients with schizophrenia (SZ). *NRXN1* expression is complex, with multiple alternative splice forms generated from two major transcripts; *NRXN1α* and *NRXN1β*. The majority of CNVs in SZ are deletions affecting the proximal *NRXN1α* exons and promoter region. Rare chromosomal events are useful in understanding the genetic basis of complex psychiatric disorders since affected genes become feasible targets to analyze for more subtle genetic alterations. As a first step towards this goal, we resequenced the *NRXN1α* promoter region in 170 patients with SZ and a similar number of controls. Two rare mutations were identified in the patient population. One previously unknown single nucleotide polymorphism (SNP) was found in controls. Bioinformatics analysis suggests that binding to several transcription factors may be affected by the minor alleles. The findings suggest that in addition to chromosomal alterations disrupting the *NRXN1α* promoter, rare point mutations in the region may also be involved in SZ pathogenesis.

The three members of the neurexin gene family, *NRXN1*, *NRXN2* and *NRXN3*, code for proteins that together with its binding partners, the neuroligins, are involved in presynaptic and postsynaptic differentiation at GABAergic and glutamatergic synapses [6,10,16,22]. There is mounting evidence implicating both neurexin and neuroligin encoding genes as primary factors in neuropsychiatric disorders. In addiction, for example, *NRXN1* and *NRXN3* have been implicated in alcoholism, and addiction to nicotine and illicit drugs [17,14,2,24]. Recently, several groups have identified patients with SZ who have deletions disrupting the *NRXN1α* promoter region and proximal (N-terminal encoding) exons [7,8,12,23,25,27,28]. In addition, several groups have reported deletions and translocations affecting *NRXN1* in autism spectrum disorders (ASD) [4,11,20]. The identification of structural variants affecting the *NRXN1α* promoter region suggests that altered expression of the long alpha isoform contributes to SZ susceptibility. These observations also suggest that more subtle disruptions of gene expression, such as promoter mutations, might be involved in disease susceptibility in a subgroup of SZ patients. To test this hypothesis, we resequenced the *NRXN1α* promoter region in patients and controls.

Patients with SZ ($n = 170$) were recruited from the Nathan Kline Institute/Rockland Psychiatric Center (NKI/RPC). Clinical diagnosis was established according to DSM-IV criteria using the SCID (research version, $N = 122$) or chart review of chronic inpatients with an admitting diagnosis of SZ ($N = 48$). U.S. controls were Caucasian blood-bank donors. In the SZ sample, 31% were female with a mean age of 42 ± 10, while 45% of controls were female with a mean age of 48 ± 13. Patients with bipolar disorder (BD) from the Czech Republic were recruited from the Prague Psychiatric Center. Patients were diagnosed on the basis of either a Schedule for Affective Disorders and Schizophrenia-Lifetime (SADS-L) interview or by clinical interview modified from SADS-L using Research Diagnostic Criteria (RDC). Controls were blood-bank donors and patients hospitalized for medical reasons. Fifty-four percent of bipolar patients were female, with a mean age of 49 ± 17 years, compared to 40% females in the control sample with a mean age of 47 ± 16. The cohort consisted of 78% bipolar I and 22% bipolar II.

All patients signed an informed consent approved by the Ethical Committee on Clinical Investigation (Czech samples) and the AECOM and NKI/RPC IRBs (U.S. samples), which comply with guidelines established for human experimentation by the Declaration of Helsinki.

An 814 base pair (bp) PCR fragment was generated using primers P1 (cagctttccatgggtctagcaggggcct) and P2 (aggcttcatgcaaaacaacc).

* Corresponding author. Tel.: +1 718 430 2428.
E-mail address: Herb.Lachman@einstein.yu.edu (H.M. Lachman).

0304-3940/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2010.03.047

DNA samples were amplified using "hot-start" polymerase (manufacturer's instructions; Qiagen, HotStarTaq). Cycling parameters were 95 °C for 15 min, then 30 cycles of 94 °C (30 s), 49 °C (30 s) and 72 °C (2 min). Samples were purified using QIAquick® (Qiagen Sciences). Sequencing was carried out using primer P2.

Genotyping for rs2287235 was carried out using TaqMan® Allelic Discrimination. Samples were amplified by PCR using a 7900HT cycler and analyzed using SDS 2.1 software (Applied Biosystems).

StatXACT-5 (Cytel Software Corporation, Cambridge, MA) was used to compute χ^2 statistics. The level of significance was set at $p < 0.05$. Hardy–Weinberg equilibrium (HWE) was computed using goodness-of-fit χ^2.

Promoter variants were analyzed computationally using TRANSFAC suite from Biobase Biological databases (https://portal.biobase-international.com) to identify potential cis-regulatory factors that might be altered by allelic differences.

NRXN1α gene expression initiates from non-coding exon 1. This exon is in very close juxtaposition (<100 bp) to the transcription start site of a long non-coding RNA, AK127244 (Fig. 1). Previous work in our lab showed that exon 1 contains a dual, bidirectional promoter with activity in the direction of both *NRXN1* and AK127244 [26]. Most of the *NRXN1α* CNVs that have been found in patients with SZ disrupt exon 1, which was the focus of this resequencing effort. An 814 bp fragment was generated and sequenced using primer P2. The amplified fragment contains a polymorphic CT tract (AG on plus strand), which is at the most distal *NRXN1* transcription start site. Using the P2 primer, we obtained excellent sequence data for a portion of intron 1, the intron 1/exon 1 junction, and the entire length of exon 1, up until the polymorphic CT/AG repeat. The amplified region contains two other known polymorphisms, the SNPs rs2287235 and rs72828367, the latter of which is a T/C transition within the repeat element. The DNA sequencing data we obtained allowed us to accurately genotype rs2287235, which is 3737 bases away from the translation start site in exon 2 (−3737T/C; note that all polymorphic bases and rare mutations will be described from the perspective of the minus strand, since *NRXN1* is transcribed from that strand). An accurate assessment of rs72828367 and the polymorphic CT/AG repeat was not possible, however, since these are at the very end of the sequencing run, resulting in low amplitude fluorescent peaks and difficulty genotyping heterozygotes.

We identified three new variants after sequencing 170 SZ patients and 160 controls (Figs. 1 and 2). One was −3797G/A, which was found in three controls and no patients, suggesting that it is a previously unidentified SNP. The other two were −3781T/C, which was found in one patient, and −3681G/T, which was found in another. The variants −3797G/A and −3781T/C map to exon 1 and −3681G/T maps to intron 1, 27 bases away from NRXN1 exon 1/intron1 junction.

The SNP rs2287235 (−3737 T/C) was genotyped in cases and controls using the sequence data. As seen in Fig. 2B, there was no significant difference in the genotype or allele distribution. Considering the genetic overlap between SZ and BD featured at many genetic loci and candidate genes, we also genotyped a cohort of patients with BD and controls from the Czech Republic. For this analysis, a Taqman allele discrimination system was employed since the bipolar samples were not sequenced. There was no difference in the allele or genotype distribution in bipolar patients and controls. In addition, there was no deviation in the genotype distribution from that expected from a Hardy–Weinberg equilibrium in any of the patient and control cohorts. The allele frequencies were similar to that found in the HapMap CEU families (CEPH – Utah residents with ancestry from northern and western Europe) ("A" allele frequency = 0.853; "G" allele frequency = 0.147) (http://hapmap.ncbi.nlm.nih.gov/).

These findings suggest that rs2287235 does not play a role in either SZ or BD pathogenesis, although the samples we analyzed are underpowered to detect a genetic contribution for this SNP if the effect size is small. The finding is also consistent with the data reported by Rujescu et al. who did not detect any significant association to 2p16.3 polymorphic markers in their large SZ and control cohort – the same data set in which a significant increase in CNVs disrupting *NRXN1* in patients was found [27].

These three variants, as well as rs2287235, were subjected to TRANSFAC analysis to assess the potential for allele-specific differences in transcription factor binding sites, the results of which are summarized in Fig. 3. Analysis used the MATCH program with the set of high quality matrices for vertebrate transcription factors. Sequences were first scanned using settings that minimized false negative matches. They were then rescanned with settings that minimized false negative matches. Separate analysis of each variant pair was compared, to find significant matches that discriminated the two alleles. The most significant finding was for −3781, in which a high motif score was given for binding to the transcription factor Runx2 when the mutant allele was present, but no score for the major allele. Other allele-specific differences were found for YY1 and Cux1, ubiquitously expressed transcription factors.

Interesting differences were also found for the other patient-specific variant, −3681C/A. Four transcription factors – SREBP1, WT1, MAZ, and ZBTB7 – showed perfect matrix scores of 1.00 for the core and/or total binding sites. All showed significant allele-specific differences with the minor allele showing lower scores, except for SREBP1, which showed a lower score for the major allele.

The −3681 variant could potentially have an effect at the RNA level too. The mutation is within a polypurine tract of 32/37 A's and G's; intronic G-rich sequences have been shown to have a role in promoting splicing and runs of triple "G" nucleotides can act as intron splicing enhancers and suppressors [13,15,21].

In addition to the allele-specific findings for the SZ-specific variants, analysis using TRANSFAC identified potential differences in the −3797 SNP found in three controls, and rs2287235. The most substantial allele-specific difference was for Myb, the prototypical member of the Myb family of transcription factors, in which moderately high scores were obtained for the minor allele, and no score was detected for the major allele. Rfx1 and Dbp show marginal allele-specific differences. Allele-specific differences in binding to WT1, ETS1, and CP2 were predicted for rs2287235.

Evidence for involvement of *NRXN1* in SZ and ASD is mounting, stemming from the finding that copy loss in the form of partial deletions clustered in the proximal exons, as well as some gene duplications, have been found in a statistically significant proportion of patients (0.47% vs 0.15% controls) [27]. The finding of CNVs in control populations shows that disruptions in the *NRXN1α* locus are not completely penetrant. One important objective in identifying patient-specific CNVs as disease-causing variants is that the affected genes become plausible targets to analyze by resequencing to determine whether disease risk can be attributed to more subtle types of genetic variation, such as single base substitutions or small in/dels. This type of analysis has recently been carried out in 57 patients with ASD; two missense mutations in *NRXN1α* were found in the coding exons [11]. In a more extensive study by Yan et al., also focusing on coding exons, nearly 5% of patients were found to harbor ultra-rare point mutations [30].

We focused our resequencing effort on exon 1, which is non-coding and shares a dual, bidirectional promoter with a long non-coding RNA gene, AK127244 [26]. Although many different CNVs of varying length have been found in patients with SZ and autism, most cluster around exons 1 and 2 and the immediate upstream region, which includes the 5′ region of AK127244 [27]. CNVs in this region are expected to reduce expression of *NRXN1α*. Analysis of more than 300 subjects showed that genetic variation