Unusually rapid evolution of Neuroligin-4 in mice

Marc F. Bolliger[†], Jimin Pei[‡], Stephan Maxeiner[†], Antony A. Boucard[†], Nick V. Grishin[‡§], and Thomas C. Südhof[†‡¶‖]

Departments of [†]Neuroscience, [§]Biochemistry, and [¶]Molecular Genetics, and [‡]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390

Contributed by Thomas C. Südhof, February 13, 2008 (sent for review January 24, 2008)

Neuroligins (NLs) are postsynaptic cell-adhesion molecules that are implicated in humans in autism spectrum disorders because the genes encoding NL3 and NL4 are mutated in rare cases of familial autism. NLs are highly conserved evolutionarily, except that no *NL4* was detected in the currently available mouse genome sequence assemblies. We now demonstrate that mice express a distant NL4 variant that rapidly evolved from other mammalian *NL4* genes and that exhibits sequence variations even between different mouse strains. Despite its divergence, mouse NL4 binds neurexins and is transported into dendritic spines, suggesting that the core properties of NLs are retained in this divergent NL isoform. The selectively rapid evolution of NL4 in mice suggests that its function in the brain is under less stringent control than that of other NLs, shedding light on why its mutation in autism spectrum disorder patients is not lethal, but instead leads to a discrete developmental brain disorder.

autism | cell-adhesion molecule | neurexin | synapse

Neuroligins (NLs) are postsynaptic cell-adhesion molecules that were discovered as ligands (or receptors, depending on the perspective) for neurexins (1, 2). Neurexins, in turn, are a family of neuronal cell-adhesion molecules that were identified as receptors for the presynaptic neurotoxin α-latrotoxin (3). Neurexins and NLs form a high-affinity complex with each other that is regulated by alternative splicing of both proteins and that mediates intercellular adhesion (4). When nonneuronal cells expressing NLs or neurexins are cocultured with neurons, the NLs and neurexins on the nonneuronal cells potently induce formation of pre- or postsynaptic specializations, respectively, in the cocultured neurons (5–7). These results indicate that NLs and neurexins form a transsynaptic junction with each other and that the formation of this junction is sufficient to signal to the neuron what kind of synaptic specialization to elaborate. Thus, neurexins and NLs probably function either in the initial establishment of synaptic junctions or in the validation of transient junctions formed by other mechanisms. Indeed, knockout experiments revealed that NLs and neurexins are not essential for the establishment of synapses but are required for the functionality of synapses, confirming the second hypothesis (8, 9).

Humans express five NLs from two autosomal genes (*NL1* and *NL2*), two X-chromosomal genes (*NL3* and *NL4*), and one Y-chromosomal gene (*NL5*; 10). NL1, NL2, and NL3 are highly conserved in rodents, but no homologs of NL4 and NL5 were identified. Different NLs appear to be specialized for distinct functions, as evidenced by the distinct properties of NL1 and NL2. NL1 is preferentially localized to excitatory synapses (11), selectively increases excitatory synaptic strength when overexpressed (12), and causes an excitatory synapse deficit when deleted (12). NL2, in contrast, is preferentially localized to inhibitory synapses (6, 13), selectively increases inhibitory synaptic strength when overexpressed (12), and causes an inhibitory synapse deficit when deleted (12).

NLs have aroused considerable attention because mutations in NL genes are found in patients with familial autistic spectrum syndrome (14–17). These mutations include one missense mutation in the *NL3* gene and four missense and two nonsense mutations in the *NL4* gene. In view of the fact that *NL4* is much more frequently affected in human autism spectrum disorder than *NL3*, it is striking that no *NL4* and/or *NL5* equivalent is present in the current draft of the mouse genome sequence. In the present study, we show that mice do contain a NL4 homolog, referred to as NL4*, but that its sequence diverges dramatically from those of NL4 in other species. We demonstrate that the *NL4** gene is located on an unknown autosome and exhibits significant sequence variations among mouse strains. Genomic cloning revealed a high density of repetitive sequences in the *NL4** gene, accounting for its absence from the draft mouse genome sequence. Despite the sequence differences between mouse NL4* and other NLs, mouse NL4* binds to neurexins and is localized to dendritic spines when overexpressed. Our data suggest that NL4 is subject to rapid evolutionary changes in mice.

Results

Cloning of Mouse NL4* cDNA. Using the partial sequences of a previously uncharacterized mouse NL isoform identified in GenBank and Google database searches, we used PCR cloning to obtain the entire cDNA sequence of this isoform. Sequence comparisons revealed that this isoform exhibited similar sequence homology to all five human NL isoforms (55–60% identity; Table 1), without a higher similarity to NL4 or NL5 than to other NLs. However, phylogenetic analyses demonstrated that this isoform is most closely related to NL4, prompting us to refer to it as NL4* (see below). Searches of DNA databases showed that all current assemblies of the mouse genome lack the *NL4** gene, and synteny analyses of genome regions surrounding the *NL4* and *NL5* genes in humans did not identify orthologous syntenies in the current mouse genome sequence assemblies (data not shown).

Mouse NL4*, like other NLs, contains an N-terminal signal peptide, followed by a large extracellular domain homologous to esterases, an *O*-linked glycosylation cassette, a single transmembrane region, and a short cytoplasmic tail. Alignment of the NL4* protein sequence with the sequences of other mouse NLs [supporting information (SI) Fig. S1] or human NL4 (Fig. S2) demonstrated that seven sequence insertions, denoted as i1 to i7, constitute the major difference between NL4* and other NLs. Five of these insertions (i1 to i5; 8–29 residues in length) occur in the extracellular domain and two (i6 and i7) in the cytoplasmic tail. All insertions are strikingly rich in small side-chain residues. However, these insertions are not solely responsible for the relatively low degree of sequence identity among mouse NL4* and other NLs, because removal of these insertions in the protein sequence increases the sequence identity between mouse NL4* and human NL4 only moderately (to 67.5%), well below the sequence identity of dog, opossum, and chicken NL4 with human NL4 (98.8%, 98.2%, and 96.1%, respectively).

Author contributions: T.C.S. designed research; M.F.B., J.P., S.M., and A.A.B. performed research; M.F.B., J.P., S.M., A.A.B., N.V.G., and T.C.S. analyzed data; and M.F.B., N.V.G., and T.C.S. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF692521 and EU350930).

‖To whom correspondence should be addressed at: Department of Neuroscience, University of Texas Southwestern Medical Center, 6000 Harry Hines Boulevard, Dallas, TX 75390-9111. E-mail: thomas.sudhof@utsouthwestern.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801383105/DCSupplemental.

Table 1. Percent identity of mouse and human NL proteins

	mNL1	mNL2	mNL3	mNL4*	hNL1	hNL2	hNL3	hNL4	hNL5
mNL1	100	66.5	71.9	56.4	98.0	66.7	72.8	75.8	74.8
mNL2		100	66.6	55.7	66.4	98.2	66.2	65.0	64.2
mNL3			100	57.0	71.2	66.1	98.6	75.9	72.7
mNL4*				100	55.7	55.7	57.0	59.7	58.7
hNL1					100	66.5	72.5	75.8	74.8
hNL2						100	66.0	65.2	64.3
hNL3							100	75.1	74.2
hNL4								100	97.7
hNL5									100

Sequence identities were calculated from pairwise alignments, with exclusion of the inserts in the alternative splice sites A and B [GenBank accession numbers: NP_619607 (mNL1), NP_942562 (mNL2), CAM24451 (mNL3), ABS19580 (mNL4*), NP_055747 (hNL1), NP_065846 (hNL2), NP_061850 (hNL3), NP_065793 (hNL4), and AAM46113 (hNL5)].

PCR cloning of NL4* from commercial brain cDNA isolated from a pool of 1,000 BALB/c mice identified NL4* variants that differed at 22 sites, with either point mutations or insertions/deletions (Fig. 1*A* and Table 2). Because these variations were found in multiple clones in different combinations, they probably do not represent PCR artifacts. Remarkably, two of these variations change insertion sequences: One alters the number of tandem AGGV motifs in insertion i4, and the other alters the number of tandem GVA motifs in i7. To confirm the unexpected variation we observed in the NL4* cDNAs isolated from a commercial source, we tested genomic DNA purified from three mouse strains. We used PCR primers specific for the three largest exons of the mouse *NL4** gene, covering ≈90% of the *NL4** coding region (Fig. 1*B* and Table 3). The fragments amplified on exon 1 were identical in all mouse strains tested. By contrast, the sequences obtained from exons 6 and 7 differed between BALB/c and 129/Sv mice (C57BL/6 and BALB/c were identical), with eight nucleotide substitutions in exon 6, one of which causes an amino acid substitution (N434D). As a control, we analyzed the corresponding exon of the *NL3* gene but failed to identify any variations in *NL3* between the three mouse strains.

We used the polymorphic nucleotides in the mouse *NL4** gene to test whether *NL4** is X- or Y-chromosomal or autosomal. A male mouse bearing one allele of the "129/Sv-type" and one allele of the "C57BL/6-type" of the *NL4** gene was crossed to a pure 129/Sv female, and the offspring was analyzed (Table S1). Because littermates of both sexes had the C57BL/6 contribution, localization of the *NL4** gene must be autosomal. This was unexpected because most mammalian *NL4* genes are positioned on the X-chromosome (Table 4). Taken together, these findings indicate that *NL4** undergoes rapid evolution in the mouse.

Bioinformatics Analysis of NL Proteins. Based on cDNA sequences encoding mouse and human NLs, we performed BLAST searches to identify close homologs of the NL family in various species (Table S2). In addition to mammalian NLs (Table 4) we identified NLs from chicken (types 1, 3, and 4), frog (type 3), and fishes (all four types). In the phylogenetic tree (Fig. 2), vertebrate NLs form four clearly separable groups that correspond to NL1–NL4. Putative orthologs of NL members were additionally identified in several insect species, sea urchin, and *Caenorhabditis elegans*. In the tree, they serve as outgroup sequences to the four groups of vertebrate NLs and determine the root of these four groups. The root is located between NL2 and the other three groups, indicating that NL2 split early in evolution of the family. The tree shows that NL3 and NL4 are closest to each other, suggesting that their separation is a more recent evolutionary event. The previously unidentified mouse NL4* clearly belongs to the group 4/5, but is placed near the root of this group and not with other mammalian NL4/5 isoforms. We speculate that such a positioning of mouse NL4* is caused by the long-branch attraction artifact of tree reconstruction because of the elevated mutation rate of mouse *NL4** rather than by the ancient origin of *NL4**.

Analysis of NL Genes. Because the mouse *NL4** gene is missing from all current genome databases, we isolated a λ-phage from a genomic 129/Sv mouse library that encodes the 5′ end of the *NL4** gene, determined its sequence, and compared the partial *NL4** gene structure with the structures of other NL genes (Fig. 3). All NL genes have similar exon/intron structures (Fig. 3, Fig. S1, and Table S3), with one notable exception: *NL2*—which is phylogenetically most distant from all other NL isoforms (Fig. 2)—includes an additional intron in exon 6. The conservation of the NL gene structures suggests that the various NL isoforms evolutionarily arose from a single ancestor gene.

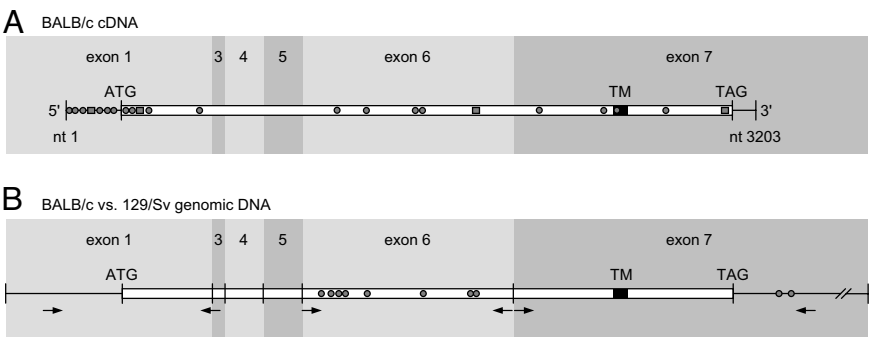

Fig. 1. Sequence variations in *NL4** from different mouse strains. (*A*) The complete coding sequence of *NL4** was amplified by PCR on brain cDNA prepared from a pool of 1,000 BALB/c mice and was cloned and sequenced. A total of 26 clones were analyzed and found to comprise 10 cDNA variants that feature 18 nt substitutions (circles) and four insertions/deletions (rectangles) as listed. The bar shows schematically the mRNA of NL4* with the location of exons in the *NL4** gene (TM, transmembrane region; ATG, initiator codon; TAG, stop codon). See Table 2 for a list of the changes observed. (*B*) Exons 1, 6, and 7 were PCR-amplified on genomic DNA isolated from three mouse strains (arrows indicate the location of PCR primers). The sequences obtained from BALB/c and C57BL/6 mice were identical, whereas the strain 129/Sv sequence exhibited eight changes in exon 6 and two changes in exon 7, with one of the changes leading to an amino acid substitution. See also Table 3.

Table 2. Sequence variations in NL4* cDNAs from BALB/c mice

Position in cDNA	Nucleotide exchange	Amino acid exchange
8	C/T	5′ UTR
12	C/T	5′ UTR
60	A/G	5′ UTR
95–133	ins/del	5′ UTR
147	A/G	5′ UTR
182	C/T	5′ UTR
195	A/G	5′ UTR
273	C/T	P6L
275	C/G	P7A
276–281	ins/del	ALL/V
304	A/G	–
604	C/T	–
1240	A/G	–
1378	C/G	–
1606	C/T	–
1616	A/G	N434D
1886–1898	ins/del	AGGV/Δ
2179	A/G	–
2482	A/G	–
2551	A/G	–
2770	A/G	–
3040	ins/del	GVA/Δ

Nucleotide positions refer to GenBank accession number EF692521, amino acid positions to Fig. S2.

NLs are alternatively spliced at two conserved sites (referred to as splice sites A and B) in the extracellular domain (Fig. 3 and Fig. S1). In NL1 and NL3, two conserved, variably used insert sequences are observed in splice site A (referred to as A1 and A2), whereas in NL2 and NL4/5 (including NL4*), only a single variably used insert sequence corresponding to A2 is detectable. In all NL genes, the alternatively spliced A1 and A2 sequences are encoded by separate exons (exons 2 and 3). In contrast to splice site A, splice site B is present only in *NL1* and is produced by use of alternative splice donor sites in exon 5.

Although the partial *NL4** gene structure indicates that it is identical to the *NL4* gene structure in other species, its intronic sequences feature three unusual repetitive elements (referred to as R1, S1, and T), two of which are also repeated in inverse orientation (R2 and S2; Fig. 4). The three repetitive elements are composed of 21- to 34-bp sequences repeated up to 60 times. These repetitive elements are difficult to sequence and may account for the fact that the *NL4** gene is missing in the current mouse genome assemblies. Databank searches suggest that they are unique in the mouse genome because no similar repetitive elements were detected.

Table 3. Sequence variations in *NL4 genes from BALB/c and 129/Sv mice**

Position in cDNA	Nucleotide exchange	Amino acid exchange
1150	G/T	–
1210	C/T	–
1243	T/G	–
1255	G/A	–
1378	C/G	–
1616	A/G	N434D
1858	G/C	–
1864	C/G	–
3292	G/T	3′ UTR
3387	A/G	3′ UTR

Nucleotide positions refer to GenBank accession number EF692521, amino acid positions to Fig. S2.

Table 4. Chromosomal assignments of NL genes

Genus and species	*NL1*	*NL2*	*NL3*	*NL4*	*NL5*
Homo sapiens	3	17	X	X	Y
Pan troglodytes	3	17	X	X	Y
Macaca mulatta	2	16	X	X	n.d.
Bos taurus	1	19	X	X	n.d.
Canis lupus familiaris	34	5	X	X	n.d.
Rattus norvegicus	2	10	X	n.d.	n.d.
Mus musculus	3	11	X	a	n.d.
Monodelphis domestica	7	2	X	7	n.d.

Listed are the chromosome numbers for the respective NL genes. a, autosomal; X, Y, X and Y chromosomes. Assignments were made based on genomic sequence databases; n.d., gene was not detected.

Biochemical Properties of NL4*. In transfected COS cells, NL4* exhibits an apparent mass of ≈125–130 kDa (Fig. 5*A*). Immunofluorescence analysis of transiently transfected HEK-293T cells confirmed that NL4* is transported to the cell surface (Fig. 5*B*). Binding of soluble Ig-neurexin proteins to NL4* expressed in HEK cells revealed that both neurexin-1α and -1β specifically bound to NL4* (Fig. 5*B*), independent of the presence or absence of an insert in splice site A2 (data not shown). No binding of Ig-control protein to NL4*-expressing HEK cells was detected (Fig. 5*B*), and no binding of any Ig-neurexin fusion protein to cells that do not express NL4* was observed (data not shown).

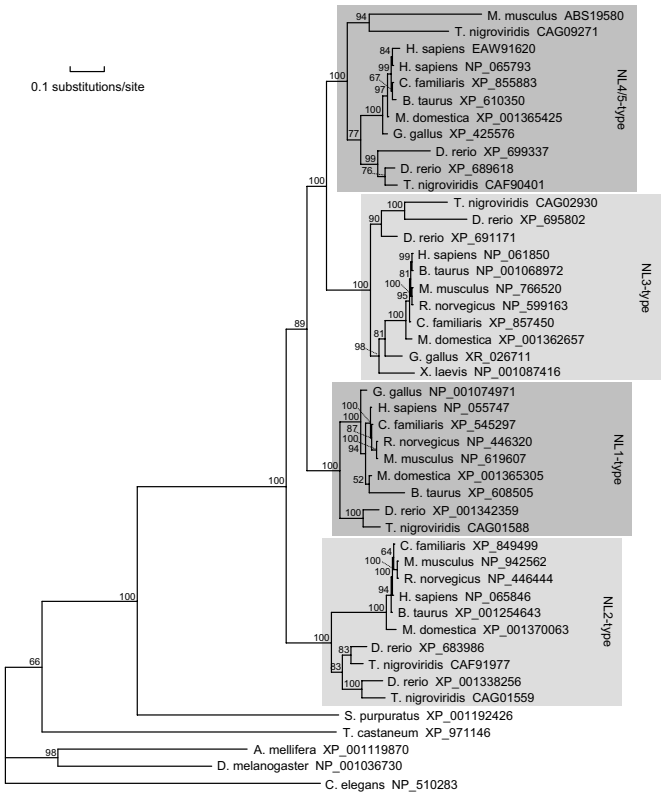

Fig. 2. Phylogenetic analysis of NL sequences. Vertebrate NLs form four groups according to the four major types of NLs. Possible orthologs identified in invertebrates (insects, sea urchin, and *C. elegans*) serve as outgroup sequences and determine the root of the four groups of vertebrate NLs. The tree was built by using the MOLPHY software, and sequences are labeled with species names and GenBank accession numbers. Local bootstrap values are indicated (as percentages) at internal tree nodes.

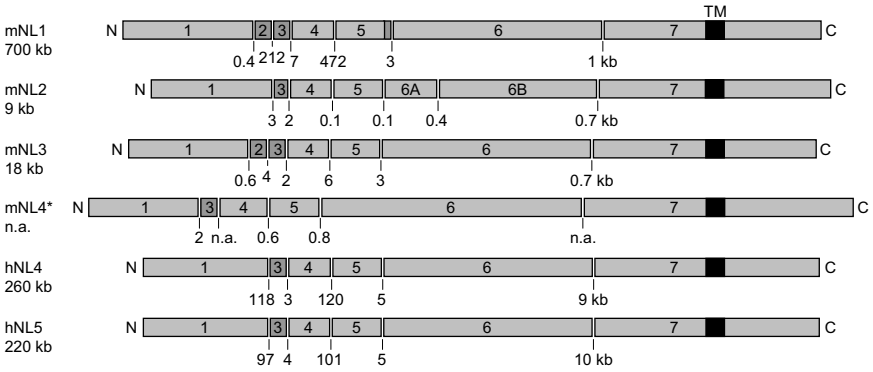

Fig. 3. Organization and size of NL genes. The contribution of the seven coding exons to the four mouse NLs and human NL4 and NL5 proteins is shown (TM, transmembrane region). For simplification, the exon containing the translational start codon is denoted as exon 1, and the exon containing the stop codon is denoted as exon 7 in all NL genes. Most, if not all, NL genes contain at least one 5′ noncoding exon not shown. Exon 2 encoding splice insert A1 is present only in *NL1* and *NL3*, whereas exon 3 (splice insert A2) can be found in all NL genes. Only the *NL1* gene is alternatively spliced at the end of exon 5 to create splice site B. Numbers inserted between exons give the size of corresponding introns in kilobases. Numbers on the left denote gene sizes (as measured from the translational start codon in exon 1 to the stop codon in exon 7). n.a., not available.

Localization of Mouse NL4* in Neurons. RT-PCR revealed that mRNAs encoding NL4*, similar to NL1–NL3, are most highly expressed in brain (Fig. 6*A*). However, we detected broad expression of almost all NL mRNAs in all tissues by RT-PCR, a surprising finding considering the brain-specific expression of NLs as determined by RNA blotting (1, 2). Because RT-PCR can detect minute mRNA levels that do not lead to protein production, we tested by immunoblotting analysis whether NL1, NL2, and NL3 are actually present in the tissues in which RT-PCR detected their mRNA but observed NL1, NL2, or NL3 proteins only in brain samples (Fig. 6*B*). Although this result could not be confirmed for NL4* because no specific antibody was available, these data overall suggest that NL4* probably is also brain-specific.

In a final set of experiments, we examined whether NL4*, similar to NL1-NL3, is transported into spines in transfected cultured hippocampal neurons. Imaging of flag- or myc-tagged NLs expressed in the transfected neurons revealed that all NLs, including mouse NL4*, were transported into dendritic spines (Fig. 6*C*). These spines contained opposing presynaptic terminals, as revealed by synapsin staining, suggesting that they carry synapses and consistent with the conclusion that NL4*, like other NLs, is a postsynaptic protein.

Discussion

The present study addresses the puzzling observation that NL4 is highly conserved in most vertebrates, including fish, but could not be found in rodents. In humans, *NL4* is of prime interest because different mutations in this gene are associated with autism spectrum disorder in multiple families. The human mutations identified in the *NL4* gene include two nonsense (14, 15) and four missense (16) mutations, three of them in the esterase-homologous domain (G99S, K378R, and V403M) and one in the cytoplasmic tail (R704C).

We show that mice contain a *NL4* gene that is highly divergent from *NL4* in other species and is thus referred to as *NL4** (Figs. S1 and S2). NL4* exhibits all of the properties of a standard NL, including the typical NL domain structure, neurexin binding, and transport to dendritic spines in neurons (Figs. 5 and 6). Its gene organization resembles that of other NLs (Fig. 3). G99 and R704 are conserved in the mouse NL4* sequence, and V403 is conservatively replaced by an isoleucine (Fig. S2). Nevertheless, the NL4* sequence is no more similar to those of NL4 in other species than to those of NL1, NL2, or NL3, necessitating a phylogenetic analysis to reveal its relation to NL4 (Fig. 2). Moreover, *NL4** in mice is not located on the X chromosome, whereas *NL4* in most other species is (Table 4). Thus, *NL4* is subject to rapid evolution in mice, suggesting that one reason for the preferential mutation of *NL4* in human patients is that NL4 may be relatively dispensable and that mutations in other NLs have not been observed at all (NL1 and NL2) or not as frequently (NL3) because these proteins perform more central functions in brain.

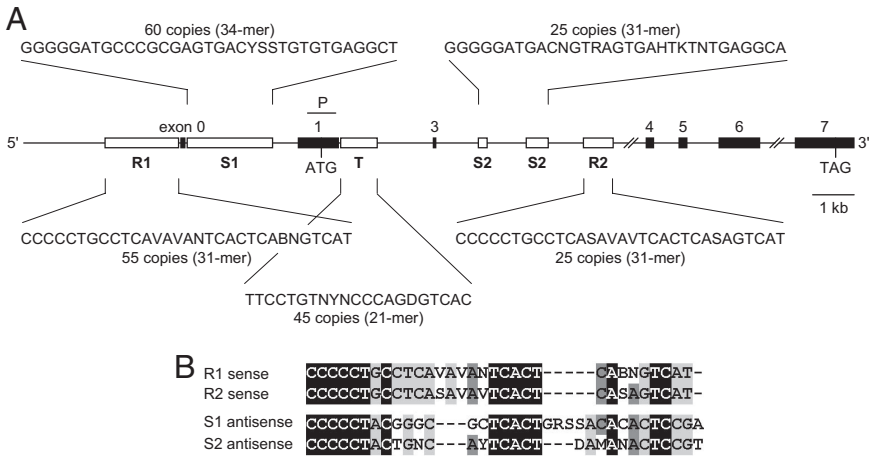

Fig. 4. Organization of the mouse *NL4** gene. (*A*) Schematic diagram of the mouse *NL4** gene structure as determined by sequencing of genomic DNA (GenBank accession number EU350930). The gene contains six coding exons and at least one noncoding (exon 0). Exons are indicated by filled boxes and are numbered. Exon 3 encodes the alternative spliced A2 insert, whereas no exon corresponding to exon 2 in *NL1* and *NL3* (which encodes the alternatively spliced A1 insert) is present. Sequence analyses identified five repetitive DNA regions (R1, R2, S1, S2, and T; open boxes), the first four of which are closely related to each other as shown in the sequence alignment in *B*. The copy numbers and consensus sequences of the repetitive regions (determined with a cut-off of 80%) are indicated (P, location of 681-bp hybridization probe used to isolate the genomic DNA clone). (*B*) Sequence alignment of the repetitive regions R1, R2, S1, and S2 with identical bases boxed on a black background; note that the four repetitive regions form two pairs that are present in opposite orientations.

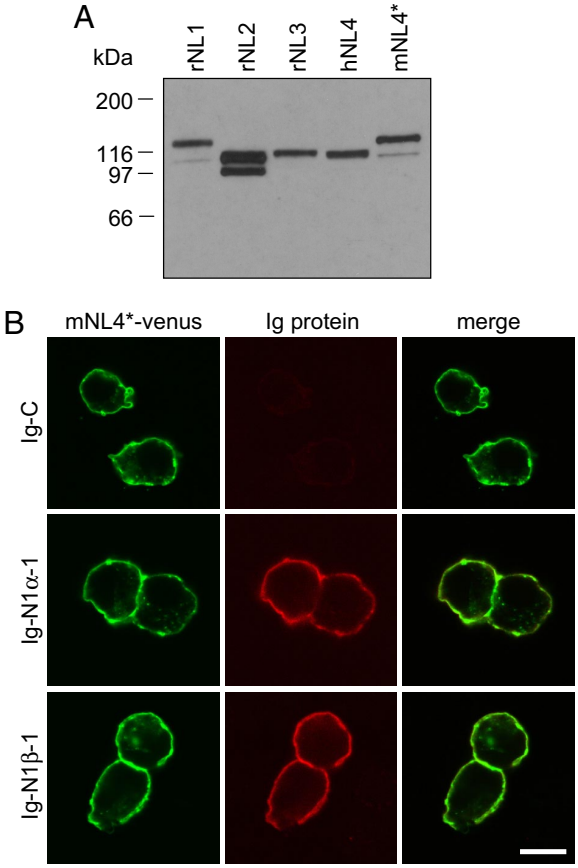

Fig. 5. Expression of NL4* in nonneuronal cells and binding to neurexins. (*A*) COS cells were transiently transfected with expression vectors encoding rat NL1-NL3, human NL4, or mouse NL4*. Protein extracts were analyzed by SDS/PAGE and immunoblotting using the pan-NL antibody 19C. Molecular-mass markers (kDa) are shown on the left. (*B*) HEK-293T cells expressing venus-tagged mouse NL4* (green) were incubated with soluble Ig-neurexin or Ig-control proteins. Cell surface-bound Ig-protein was visualized with fluorescent secondary antibodies (red). This binding assay revealed that NL4* binds both α-neurexin (Ig-N1α-1) and β-neurexin (Ig-N1β-1) but not the Ig-control protein (Ig-C). The images on the right show the merged pictures of neurexin and NL stainings with coincident labeling shown in yellow (Scale bar, 10 μm.)

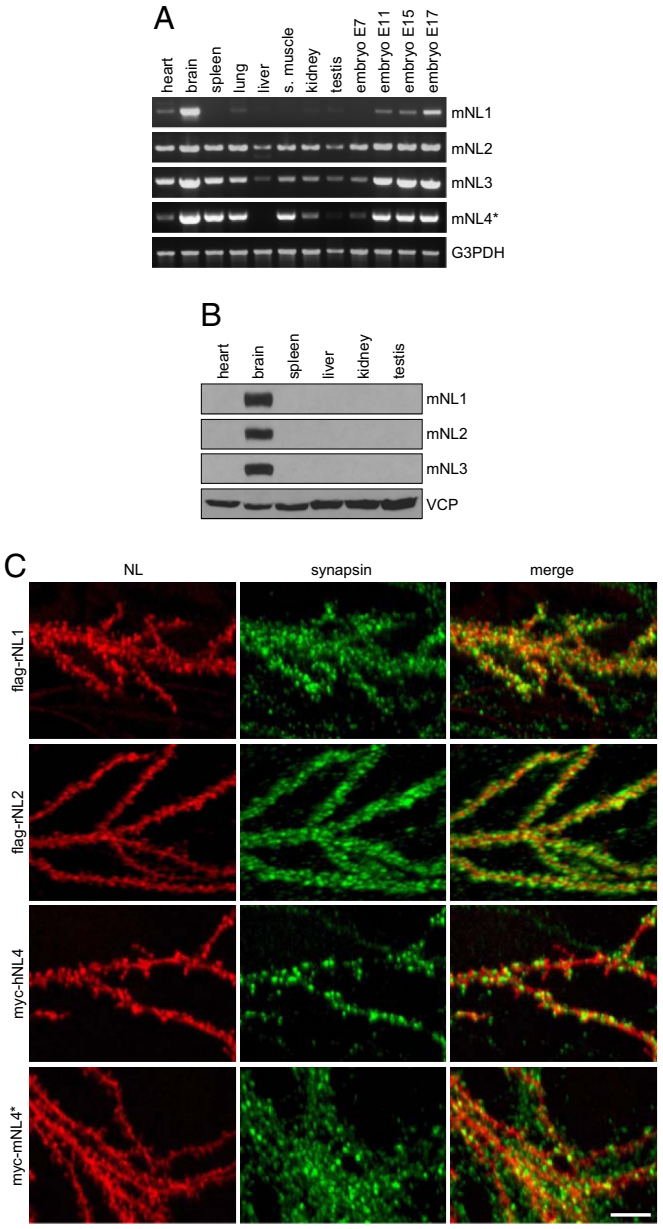

Fig. 6. Tissue expression of NL isoforms and localization in transfected neurons. (*A*) RT-PCR analysis of NLs with primers specific for NL1, NL2, NL3, NL4*, or glyceraldehyde-3-phosphate dehydrogenase (G3PDH, control) was performed on cDNA preparations made from mouse tissue RNA as listed. Products were resolved on 1% agarose gels. (*B*) Immunoblot analysis of NLs in mouse tissues by using isoform-specific antibodies 4C12 (NL1), 169C (NL2), and 639B (NL3). NLs could be detected exclusively in the brain homogenate. Valosin-containing protein (VCP, antibody 443B) served as a loading control. (*C*) Dissociated hippocampal neurons isolated from newborn rat pups were transfected with vectors encoding flag-tagged rat NL1 or NL2, myc-tagged human NL4, or myc-tagged mouse NL4* on DIV 10. Six days later, cultures were fixed and stained with immunofluorescent antibodies to the flag or myc epitopes (red) and synapsin (green; antibody E028). All four NL isoforms localized to dendritic spines, whereas the synapsin immunoreactivity was observed presynaptically (Scale bar, 10 μm.)

What is the nature of the evolutionary change that makes mouse NL4* so different from NL4 in other species? Comparison of the mouse NL4* sequence with those of other mouse NLs and of NL4 in other species identifies seven sequence insertions composed primarily of small side-chain amino acids (glycine, alanine, serine, and proline; Figs. S1 and S2). One mechanism by which these sequences could have been inserted during evolution is by shifts in the exon/intron boundaries. However, analysis of the structure of NL genes revealed that all insertions in mouse *NL4** relative to other NLs are placed within exons and are not located at the exon/intron boundaries. Functionally, the NL4*-specific sequence insertions do not appear to alter the properties of NL4*, as far as tested, and especially do not appear to alter neurexin binding, which is different from the alternatively spliced 9-aa sequence in splice site B of NL1 (18). However, mapping the insertions into the recently published crystal structures of rodent NL1 and human NL4 (19–21) suggests that the insertions are on surface-exposed loops and could possibly alter the binding of NL4* to as-yet-unidentified nonneurexin ligands. Screening for binding partners of NLs and testing the functions of mouse NL4* and of NL4 in other species will clarify these issues.

Materials and Methods

Cloning of the NL4* cDNA, Gene Analyses, and Expression Pattern Studies. PCR was performed with *PfuTurbo* DNA polymerase (Stratagene) on first-strand cDNA produced from RNA isolated from 1,000 pooled male and female BALB/c mouse brains (Clontech) by using primers MB0637 (5′-GGAATTCCGTGACGAAACAGGAAGTGACC-3′) and MB0638 (5′-GGAATTCGTAGCCAAGGCCCCTGCATGTC-3′) specific for the 5′ and 3′ UTRs of NL4*. Twenty-six independent clones containing the PCR product were sequenced, identifying multiple polymorphisms (Fig. 1*A* and Table 2). The 5′ end of the mouse *NL4** gene (fragment of 15

kb with translational start codon in center) was cloned from a genomic λ-library (Stratagene) generated from 129/Sv mice by using a uniformly ^{32}P-labeled single-stranded probe of NL4* (Fig. 4*A*). The *NL4** exons and introns were analyzed by PCR in cloned and in genomic DNA [primers used: exon 1, MB0637 and MB0643 (5′-GGAATTCTCGCTCTGGTCCTGGACGTAG-3′); intron 4, MX0759 (5′-CATCGTCGTCACCGTCAACTACCGGCTCGGCG-3′) and MX0760 (5′-CGTAGTTGC-CCTTGGCGGCCTGGTCGCCCG-3′); intron 5, MX0761 (5′-CGCCTCCTGCGTCAGC-CTCCTCACGCTGTC-3′) and MX0762 (5′-CTCGACAGCGCCGTCCCGCTCTG-GATGATGG-3′); exon 6, MB0784 (5′-GGAATTCCAGAAGGCCATCATCCAGAG-3′) and MB0616 (5′-GGAATTCCGTCTTGGCGAAGTTGGTCC-3′); exon 7, MB0790 (5′-GGAATTCAGGTGGCCTGGGCCAAGTACGAC-3′) and MB0620 (5′-GGAAT-TCAATCCAGTCCAACCCGTGCTGAC-3′)], by using *PfuTurbo* or PrimeSTAR HS DNA polymerase (TaKaRa). Genomic DNA was isolated from tail biopsies of BALB/c, C57BL/6, and 129/Sv mice (two males and two females each). Finally, exon 6 of the mouse *NL3* gene was amplified for sequence analyses by using primers MB0786 (5′-GGAATTCCAGAGGGCCATCATCCAAAG-3′) and MB0787 (5′-GGAAT-TCGGTCTTGGCAAAGTTGGTCC-3′). All PCR fragments were purified and cloned into pBluescript before sequencing. For analysis of mRNA expression by RT-PCR, first-strand cDNA preparations from RNA isolated from various mouse tissues (Clontech) were used as templates to run PCRs with *PfuTurbo* [primers used: NL1, MB0623 (5′-GGAATTCCACATGAGGTGGTTCTTCGGAC-3′) and MB0624 (5′-GGAATTCACATTCCATTGGTGTGTCCCTTGTG-3′); NL2, MB0625 (5′-GGAAT-TCGGAGGAGCTAGTATCGCTGCAGCTG-3′) and MB0626 (5′-GGAATTCGGATGT-GCACACATCACTTCCAG-3′); NL3, MB0627 (5′-GGAATTCGGGTCCCACTCACCAT-GAATGTGAG-3′) and MB0628 (5′-GGAATTCCCTTCCAGAGCTGCTTAGCACTC-3′); NL4*, MB0629 (5′-GGAATTCGGGAGTGCGGTGCAGTGGTCAC-3′) and MB0620; glyceraldehyde-3-phosphate dehydrogenase (control), G3PDH-5′ (5′-TGAAG-GTCGGTGTGAACGGATTTGGC-3′) and G3PDH-3′ (5′-CATGTAGGCCATGAGGTC-CACCAC-3′)]. Products were resolved on 1% agarose gels.

Bioinformatics Analysis. BLAST and PSI-BLAST (22) were used to search GenBank with mammalian NLs as queries. Percent identity numbers were calculated from pairwise alignments of protein sequences by using the LALIGN program (23). Multiple alignments were generated by using the MegAlign sequence analysis software (DNAstar) by the Clustal W method. For phylogenetic analysis, a multiple sequence alignment of NL proteins was generated by PROMALS (24), which uses secondary structure predictions to improve alignment quality. Weakly similar N- and C-terminal segments and gapped positions (gap fraction >0.1) were removed from the alignment. A maximum-likelihood tree was built by using the MOLPHY package version 2.3 (25) with a JTT amino acid substitution model (26). The local estimates of bootstrap percentages were obtained by the RELL method (27), as implemented in the program ProtML of MOLPHY.

Antibodies. Rabbit antibodies were generated against different NL isoforms: 19C, raised to recombinant protein containing residues Y762 → V925 of mouse NL4* (cross-reacts with all NL isoforms despite low sequence similarity); 169C, raised to the NL2 synthetic peptide MB0703 (N-CNTAYGRVRGVRRELN-C) coupled to keyhole limpet hemocyanin; 639B, raised to recombinant protein containing residues Y687 → V805 of NL3. Antibodies 19C and 639B were affinity-purified. All other antibodies were described previously or obtained from commercial sources.

Miscellaneous. Neuronal and cell cultures were done as described (18). All expression plasmids were CMV promoter-based vectors. Transfections of COS cells, HEK-293T cells, and neurons, immunocytochemistry experiments, and cell-surface labeling assays were carried out as described (18). Similarly, electrophoresis and immunoblotting analyses were performed by using standard methods. Please see *SI Methods* for a more detailed description of these methods.

ACKNOWLEDGMENTS. We thank I. Kornblum and E. Borowicz for technical support. This work was supported by National Institute of Mental Health Grant R37 MH52804-08 (to T.C.S.) and a Simons Foundation grant (to T.C.S.). M.F.B. was partially supported by a postdoctoral fellowship from the Swiss National Science Foundation, S.M. by a fellowship from the German Research Association, and A.A.B. by a fellowship from the Canadian Institutes of Health Research.

1. Ichtchenko K, *et al.* (1995) Neuroligin 1: A splice site-specific ligand for β-neurexins. *Cell* 81:435–443.
2. Ichtchenko K, Nguyen T, Südhof TC (1996) Structures, alternative splicing, and neurexin binding of multiple neuroligins. *J Biol Chem* 271:2676–2682.
3. Ushkaryov YA, Petrenko AG, Geppert M, Südhof TC (1992) Neurexins: Synaptic cell surface proteins related to the α-latrotoxin receptor and laminin. *Science* 257:50–56.
4. Nguyen T, Südhof TC (1997) Binding properties of neuroligin 1 and neurexin 1β reveal function as heterophilic cell adhesion molecules. *J Biol Chem* 272:26032–26039.
5. Scheiffele P, Fan J, Choih J, Fetter R, Serafini T (2000) Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* 101:657–669.
6. Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM (2004) Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell* 119:1013–1026.
7. Nam CI, Chen L (2005) Postsynaptic assembly induced by neurexin-neuroligin interaction and neurotransmitter. *Proc Natl Acad Sci USA* 102:6137–6142.
8. Missler M, *et al.* (2003) α-Neurexins couple Ca^{2+} channels to synaptic vesicle exocytosis. *Nature* 423:939–948.
9. Varoqueaux F, *et al.* (2006) Neuroligins determine synapse maturation and function. *Neuron* 51:741–754.
10. Bolliger MF, Frei K, Winterhalter KH, Gloor SM (2001) Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression. *Biochem J* 356:581–588.
11. Song JY, Ichtchenko K, Südhof TC, Brose N (1999) Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *Proc Natl Acad Sci USA* 96:1100–1105.
12. Chubykin AA, *et al.* (2007) Activity-dependent validation of excitatory versus inhibitory synapses by neuroligin-1 versus neuroligin-2. *Neuron* 54:919–931.
13. Varoqueaux F, Jamain S, Brose N (2004) Neuroligin 2 is exclusively localized to inhibitory synapses. *Eur J Cell Biol* 83:449–456.
14. Jamain S, *et al.* (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 34:27–29.
15. Laumonnier F, *et al.* (2004) X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 74:552–557.
16. Yan J, *et al.* (2005) Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry* 10:329–332.
17. Zoghbi HY (2003) Postnatal neurodevelopmental disorders: Meeting at the synapse? *Science* 302:826–830.
18. Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC (2005) A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α- and β-neurexins. *Neuron* 48:229–236.
19. Araç D, *et al.* (2007) Structures of neuroligin-1 and the neuroligin-1/neurexin-1β complex reveal specific protein–protein and protein–Ca^{2+} interactions. *Neuron* 56:992–1003.
20. Chen X, Liu H, Shim AHR, Focia PJ, He X (2008) Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. *Nat Struct Mol Biol* 15:50–56.
21. Fabrichny IP, *et al.* (2007) Structural analysis of the synaptic protein neuroligin and its β-neurexin complex: Determinants for folding and cell adhesion. *Neuron* 56:979–991.
22. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
23. Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. *Adv Appl Math* 12:337–357.
24. Pei J, Grishin NV (2007) PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* 23:802–808.
25. Adachi J, Hasegawa M (1996) MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs* (The Institute of Statistical Mathematics, Tokyo), No. 28.
26. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282.
27. Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol* 31:151–160.