Supporting Information

Treutlein et al. 10.1073/pnas.1403244111

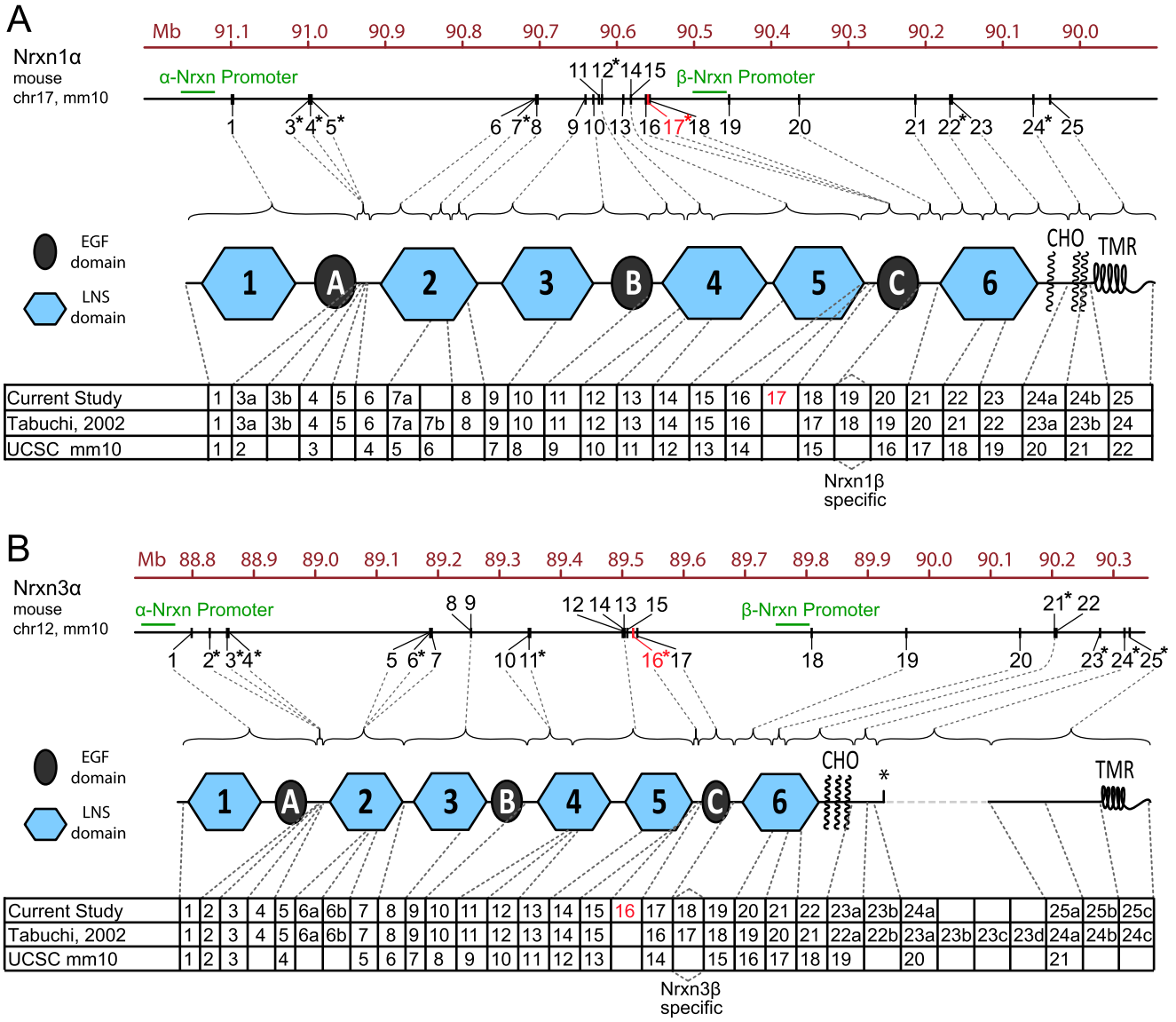
SI Materials and Methods

Library Preparation and Pacific Biosciences Sequencing. Reverse transcription (RT) was performed for each neurexin separately in a total volume of 20 μL using 1 μg of mouse prefrontal cortex total RNA, the SuperScript III Reverse Transcriptase System (Life Technologies), and 0.5 μM primer complementary to the last exon of the respective neurexin gene [Nrxn1α and Nrxn1β, exon 25 (24 in ref. 1): 5′-ACGTGGTACGACCCTTCATC-3′; Nrxn2β, exon 24 (23 in ref. 1): 5′-GGAGAATGAGGATGCAGAGC-3′; Nrxn3α and Nrxn3β, exon 25 (24 in ref. 1): 5′-CCTCGTCTCGTCCACTTGAT-3′]. The resulting single-stranded cDNA was amplified by 30 PCR cycles in a total volume of 50 μL using 2 μL RT product as template, the Platinum Taq DNA Polymerase High Fidelity System (Life Technologies), as well as a primer pair targeting the first and last exons of the respective neurexins (forward primers: Nrxn1α: 5′-ACCAGGCCGTATGCGACT-3′; Nrxn1β: 5′-TGGCCCTGATCTGGATAGTC-3′; Nrxn2β: 5′-CCACCACTTCCACAGCAAG-3′; Nrxn3α: 5′-AGCAGTTGGCACTTCCTCAT-3′; Nrxn3β: 5′-CACTCTCAGCAGGAACACCA-3′; reverse primers are identical to the RT primers). PCR products were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter) at a volume ratio of 0.6:1 (magnetic beads:PCR product) for all β-neurexins and 0.45:1 for all α-neurexins. cDNA was quantified using the Qubit HS dsDNA Kit (Life Technologies) and quality was assessed using the Agilent 2100 bioanalyzer (High Sensitivity DNA Kit).

SMRT bell sequencing libraries were prepared using Pacific Biosciences DNA Template Prep Kit 2.0 (001-540-835) according to the 2- or 5-kb template preparation and sequencing protocol provided by Pacific Biosciences, with a minimum of 1 μg and 500 ng of cDNA input into each library preparation for α-neurexins and β-neurexins, respectively. SMRT bell templates were bound to polymerases using DNA/Polymerase Binding Kit XL 1.0 (100-150-800) and v2 primers. Sequencing was carried out on the Pacific Biosciences real-time sequencer using C2 sequencing reagents with 90-min movies. Each α-neurexin was sequenced on six SMRT cells, yielding a total of 151,776 raw reads for Nrxn1α and 198,505 raw reads for Nrxn3α. β-Neurexins were sequenced on two or three SMRT cells, each yielding a total of 56,449 raw reads for Nrxn1β, 73,637 raw reads for Nrxn2β, and 92,537 raw reads for Nrxn3β.

Processing, Alignment, and Analysis of Pacific Biosciences Sequencing Data. Subread filtering was performed using Pacific Biosciences SMRT analysis software (v1.3.3). For all β-neurexins, circular consensus (CCS) reads were constructed from molecules that the DNA polymerase passed at least twice (reads containing four times the hairpin adapter sequence), whereas CCS reads for α-neurexins were constructed from molecules that the polymerase fully passed at least once. Mapping of CCS reads to the mouse genome (mm10) was carried out using STAR 2.2.0 (2) with the following settings:–seedSearchLmax 30–seedSearchStartLmax 30–outFilterMismatchNmax 100–seedPerReadNmax 100000–seedPerWindowNmax 100–alignTranscriptsPerReadNmax 100000–alignTranscriptsPerWindowNmax 10000–outSAMattributes All–outSAMstrandField intronMotif. Aligned reads that did not contain the first and last exons (amplification primer binding sites) of the respective neurexin were excluded from the analysis because they are likely to be products of mispriming or fragmentation. Subsequent analysis was performed in Python (www.python.org/). The STAR output file Aligned.out.sam was parsed for the genomic positions of splice junctions of each CCS read, and the output file SJ.out.tab, listing the genomic positions and abundances of all splice junctions detected in the CCS reads, was annotated based on nomenclature used in previous studies of neurexin gene structure (1). Splice junction positions of each CCS read were intersected with the annotated SJ.out.tab as a *.bed file using the Python package pybedtools to annotate the exon structure of each transcript. Finally, a binary splice matrix was created, which shows for each CCS read the presence and absence of each exon. All further analysis as well as the generation of data figures was done using R (3).

1. Tabuchi K, Südhof TC (2002) Structure and evolution of neurexin genes: Insight into the mechanism of alternative splicing. *Genomics* 79(6):849–859.
2. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.
3. R Development Core Team (2009) R: A Language and Environment for Statistical Computing (R Found Stat Comput, Vienna).



Current Study	1	3a	3b	4	5	6	7a		8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24a	24b	25		
Tabuchi, 2002	1	3a	3b	4	5	6	7a	7b	8	9	10	11	12	13	14	15	16		17	18	19	20	21	22	23a	23b	24		
UCSC mm10	1	2		3		4	5	6		7	8	9	10	11	12	13	14		15		16	17	18	19	20	21	22		

Current Study	1	2	3	4	5	6a	6b	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23a	23b	24a			25a	25b	25c
Tabuchi, 2002	1	2	3	4	5	6a	6b	7	8	9	10	11	12	13	14	15		16	17	18	19	20	21	22a	22b	23a	23b	23c	23d	24a	24b	24c
UCSC mm10	1	2	3		4			5	6	7	8	9	10	11	12	13		14		15	16	17	18	19		20				21		

Fig. S1. Gene and protein domain structures as well as exon annotations of *Nrxn1α* and *Nrxn3α*. Gene and protein domain structures are shown for mouse (*A*) *Nrxn1α* and (*B*) *Nrxn3α* together with a comparison of exon annotations used by this study, a previous study (1), and the UCSC genome browser (mm10) (http://genome.ucsc.edu/). The gene diagrams depict the genomic positions of exons and introns on the respective chromosome. The positions of α- and β-neurexin promoters are indicated. In the gene structure diagrams, exons that are subject to canonical alternative splicing are marked by asterisks. Protein domain structures are shown for the α-neurexins [light blue hexagons, LNS (*l*aminin-α, *n*eurexin, *s*ex hormone-binding globulin) domains; black ovals, EGF-like domains; CHO, O-linked sugar modifications; TMR, transmembrane region]. Dotted gray lines indicate which exon encodes the part of the respective protein structure. In the Nrxn3 protein domain structure, the asterisk denotes a stop codon encoded by one of the alternatively spliced exons; this stop codon could potentially cause production of secreted forms of neurexin-3.