Supplemental Figure Legends

Figure S1. Intronic sequences are detectable in dendritic mRNA using microarray analysis and Illumina sequencing. (A) Heatmap of microarray signals for 92 intronic sequence probes (rows) for three independent rat hippocampal dendrite populations (columns). Data are presented as normalized log-fold intensity difference of dendritic RNA compared to a null reference sample. Selected groups of introns are labeled with gene abbreviations. Red indicates high signal, green indicates low. Signal range indicator displayed at bottom. (B) Gene features detected by Illumina sequencing of three independent rat hippocampal dendrite populations. Pooled exonic (E) and individual intronic (i) regions are labeled for each bar graph for each gene (see Table S2 for full annotation). Scale indicates log2 transformed counts of sequencing hits summed over all samples. (C) PCR products generated from dendritic cDNA templates for intronic sequences. Primers sets used are the same as those for microarray positives with inconclusive Illumina sequencing read evidence. Bands were subcloned and confirmed by sequencing. 1KB Plus DNA ladder shown for size estimates. STX1Ai1 and STX1Ai9 failed to yield PCR products and were subsequently removed as candidate retained introns.

Figure S2. Exon and intron directed antisense *in situ* hybridization results with sense strand controls. Antisense *in situ* hybridization results from Figure 2 are presented along with MAP2 immunostaining (insets) to confirm neuronal area within fields. Antisense

probe results (left) are presented with corresponding sense strand probe results (right) for each target exon and intron. (A) CAMK2Be3, (B) CAMK2Bi3, (C) FMR1e1, (D) FMR1i1, (E) GABRG3e5, (F) GABRG3i5, (G) GRIK1e1, (H) GRIK1i1.

Figure S3. *In situ* hybridization results with calculation paths for EGFP (A) and FMR1i1ID1 (B) transfected hippocampal neurons. Paths are shown against MAP2 immunostaining (left panels) along with zoomed regions (dashed lines) shown at higher magnitude (bottom) for individual dendrites. Blue text indicates transfected DNA construct, white text indicates *in situ* probe sequence. (C) Intronic B2 element sequence does not confer dendritic localization to reporter gene mRNA. *In situ* hybridizations with antisense biotinylated EGFP riboprobe on primary hippocampal neurons transfected with FMR1i1B2-EGFP. Blue text indicates transfected DNA construct, white text indicates *in situ* probe sequence. Inset represents MAP2 immunostaining. (D) Immunostaining results to confirm translation of EGFP after transfection with ID-EGFP fusion constructs. Immunofluorescence with antibody to GFP on primary hippocampal neurons transfected with FMR1i1ID1-EGFP. Blue text indicates transfected DNA construct, white text indicates antibody target. All scale bars = 20μm.

Figure S4. Intronic ID element sequences disrupt dendritic localization patterns of endogenous mRNA. *In situ* hybridizations with antisense biotinylated intron riboprobes on primary hippocampal neurons transfected with pEGFP-N1, GABRG3i5ID2-EGFP (A), and GRIK1i1ID4-EGFP (B) constructs. Graphs at right represent *in situ* signal $\Delta F/F$

against distance from soma for ID-EGFP constructs versus pEGFP-N1 (C) Additional ID element cross-competition *in situ* hybridization results. *In situ* hybridizations with antisense biotinylated intron riboprobes on primary hippocampal neurons transfected with pEGFP-N1, CAMK2Bi3ID1-EGFP, FMR1i1ID1-EGFP, GABRG3i5ID2-EGFP, and GRIK1i1ID4-EGFP constructs. Blue text indicates transfected DNA construct, white text indicates *in situ* probe sequence. Insets represent MAP2 immunostaining. Scale bars = 20µm.

Supplemental Table Legends

Table S1. Comparison of microarray probe signals (rows) when hybridized with dendritic RNA versus background random hexamer RNAs. Mean normalized signal over three array replicates are reported for dendritic RNA and over two replicates for hexamer RNA. Individual t tests were performed per probe (3 d.o.f.) to test that the mean intensity for dendritic RNA is greater than the signal obtained for random RNA. Probes are ranked by increasing t test p value.

Table S2. High-confidence sequence reads aligning to exons and introns from genes of interest. All reported exonic reads align uniquely to the spliced gene sequence (except for STX1A, which appears in two different genomic locations in the rn4 genome assembly). Intronic counts represent only genome-unique paired reads, with at least one read of an end pair anchored in non-repetitive sequence, and not overlapping

predicted coding regions as evidenced by homology or computational models. Reads overlapping splice junctions were found for Introns marked with an asterisk (*): Adcy4 i11-e12, e12-i12, i12-e13; Gria3 e11-i11, i13-e14, i15-e16; Gria4 e13-i13; Grik1 i16-e17.

Table S3. ID elements found in candidate genes. All ID elements listed occur in the plus orientation relative to the host gene, whose strand is indicated in the "Strand" column. Sequencing evidence for each ID element is listed as "intron" = unique paired sequence reads are found in the containing intron; or "cis" = unique paired sequence reads span the ID element itself.

Table S4. Dendrite RNA pools are enriched in sequence reads aligning to ID elements. Sequence reads are listed for the three dendritic sequencing replicates. Number of reads with significant sequence similarity (by BLAST) to the ID element/BC1 5' hairpin, the BC1 3' domain, and the B2 SINE element are shown. Approximate genomic instance totals for each class of element (ID and B2) were based on RepeatMasker annotations and calculated by dividing the summed length of all genomic instances of each element by the typical length of the element (e.g., 74 nts for ID hairpins).

Table S5. Top-scoring RNA hairpins in retained introns with microRNA-like characteristics. For each hairpin, the following measures are shown: minimum free energy of the predicted structure (kcal/mol), minimum-free energy z-score value, self-containment index value, minimum e-value from a search against all miRNA RFAM

models, and the miRNA prediction results using CSHMM, MiPred, and Triplet-SVM classifiers (1 = miRNA).

Table S1

Symbol	Intron	Mean dendritic normalized signal	Mean background normalized signal	Difference	t statistic	p value
CAMK2B	1	-0.055	-0.624	0.569	10.160	0.001
STX1A	1	1.720	0.929	0.791	7.157	0.003
SNCA	2	1.125	0.547	0.578	6.448	0.004
GABRG3	5	0.034	-0.576	0.609	6.383	0.004
GABRD	8	0.640	0.437	0.203	5.047	0.008
CACN1B	18	0.210	-0.068	0.279	4.249	0.012
ADCY4	11	0.320	0.158	0.162	3.841	0.016
CAMK2B	3	0.053	-0.383	0.436	3.152	0.026
FMR1	1	0.735	0.559	0.176	2.527	0.043
CDH1	15	-0.596	-1.105	0.510	2.419	0.047
GRIN1	8	0.802	0.653	0.149	2.218	0.057
GRIK1	1	-0.250	-0.870	0.620	2.167	0.059
STX1A	9	1.403	1.000	0.403	2.163	0.060
MAP2	4	-0.151	-2.648	2.497	2.087	0.064
STX1B2	1	0.611	0.468	0.143	1.967	0.072
CAMK2G	9	0.345	0.232	0.113	1.583	0.106
GABRA4	1	0.444	0.278	0.167	1.577	0.106
ADCY6	2	1.079	0.846	0.233	1.567	0.108
CAMK2B	15	0.682	0.225	0.456	1.523	0.113
GABRA4	8	0.338	0.041	0.297	1.448	0.122
GRIA3	1	0.378	-0.037	0.415	1.432	0.124
CACN1B	45	0.327	0.065	0.262	1.309	0.141
GRIA3	14	-0.058	-0.217	0.159	1.080	0.180
SNCB	2	0.317	0.204	0.113	1.039	0.188
GABBR1	6	-0.026	-0.347	0.321	0.989	0.198
ALB	6	1.697	1.488	0.209	0.936	0.209
CACNA1H	5	0.002	-0.130	0.132	0.626	0.288
ADCY5	1	0.246	0.220	0.026	0.612	0.292
KCND2	1	-0.214	-0.305	0.091	0.575	0.303
ADCY2	1	0.615	0.422	0.193	0.490	0.329
SNCG	4	-0.207	-0.241	0.034	0.414	0.353
ADCY3	3	0.845	0.713	0.132	0.365	0.370
CREB1	2	-0.236	-0.315	0.078	0.296	0.393
GABBR1	21	0.273	0.175	0.098	0.286	0.397
ADCY4	2	0.854	0.808	0.046	0.284	0.398
ALB	1	0.066	0.038	0.028	0.274	0.401
ADCY2	20	-0.365	-0.378	0.013	0.257	0.407
GRIN1	1	0.654	0.564	0.091	0.221	0.420
GABRD	1	0.133	0.113	0.020	0.105	0.461
STX1B2	7	0.534	0.545	-0.010	-0.146	0.554
ADCY6	1	0.235	0.280	-0.045	-0.177	0.565
CACNA1H	33	-0.642	-0.592	-0.051	-0.341	0.622
FMR1	7	-0.280	-0.180	-0.100	-0.388	0.638
SNCG	1	0.279	0.327	-0.048	-0.433	0.653
ADCY5	3	0.468	0.495	-0.027	-0.567	0.695
ADCY5	20	0.396	0.438	-0.042	-0.682	0.728
KCND2	4	0.091	0.121	-0.030	-0.902	0.783
SNCB	5	0.820	0.958	-0.138	-1.016	0.808
CREB1	6	0.319	0.808	-0.488	-1.148	0.833
STX1A	3	1.303	1.590	-0.287	-1.350	0.865
GABRG3	9	-0.448	-0.248	-0.200	-1.374	0.868
APP	17	0.113	0.346	-0.234	-1.443	0.878
ADCY4	24	0.132	0.279	-0.147	-1.559	0.892
GRIA4	16	0.094	0.303	-0.209	-1.563	0.892

GRIA4	4	-0.637	-0.378	-0.260	-1.571	0.893
CDH1	3	-0.105	0.073	-0.178	-1.677	0.904
STX1B2	4	0.023	0.305	-0.282	-1.712	0.907
APP	6	-0.062	0.034	-0.096	-1.795	0.915
KCNA2	1	0.271	0.486	-0.214	-2.080	0.936
GABRG3	2	-0.620	-0.473	-0.147	-2.086	0.936
GRM7	1	-0.059	0.405	-0.465	-2.608	0.960
ADCY6	21	-0.644	-0.021	-0.623	-2.672	0.962
CACN1B	1	0.304	0.594	-0.290	-2.884	0.968
ADCY3	2	0.644	1.236	-0.591	-3.008	0.971
CAMK2D	18	-0.609	0.049	-0.658	-3.047	0.972
GRIK1	3	-0.036	0.206	-0.242	-3.654	0.982
CDH1	1	0.321	0.706	-0.385	-4.716	0.991
MAP2	3	0.331	0.713	-0.382	-5.537	0.994
GRIN1	19	0.309	0.933	-0.624	-5.874	0.995
GRIA4	2	-0.350	-0.015	-0.335	-7.342	0.997
ADCY2	4	-1.013	-1.970	0.957	3.855	0.015
CREB1	8	-1.253	-1.846	0.594	3.252	0.024
CAMK2A	10	-0.936	-1.482	0.546	1.916	0.076
MAP2	12	-0.859	-1.006	0.147	1.435	0.123
GRIK1	16	-0.787	-1.064	0.277	1.176	0.162
GABRA4	7	-0.878	-1.106	0.228	1.129	0.170
GABRD	2	-0.853	-1.090	0.237	1.111	0.174
CAMK2A	3	-0.736	-0.879	0.143	0.792	0.243
CACNA1H	1	-0.892	-1.013	0.122	0.756	0.252
CAMK2A	1	-0.864	-0.922	0.058	0.583	0.300
GABBR1	11	-0.756	-0.655	-0.101	-0.238	0.586
SNCA	4	-0.641	-0.570	-0.071	-0.445	0.657
CAMK2G	18	-0.654	-0.513	-0.141	-0.774	0.752
APP	1	-1.849	-1.426	-0.423	-0.926	0.789
CAMK2G	2	-1.095	-0.711	-0.384	-1.493	0.884
ADCY3	21	-1.028	-0.500	-0.529	-1.758	0.912
CAMK2D	4	-1.157	-0.643	-0.514	-2.420	0.953
FMR1	15	-1.040	-0.585	-0.454	-3.242	0.976
KCND2	2	-1.387	-0.703	-0.683	-3.389	0.979
ALB	14	-0.960	-0.333	-0.626	-3.552	0.981
GRIA3	4	-1.402	-0.950	-0.452	-3.619	0.982
CAMK2D	1	-1.314	-0.558	-0.756	-4.257	0.988

Table S2

Symbol	Feat.	D1	D2	D3	Feat.	D1	D2	D3	Feat.	D1	D2	D3	Feat.	D1	D2	D3
ADCY2	Exonic	284	0	3887	i2	68	4	0	i14	0	0	10	i18	0	0	2
ADCY3	Exonic	202	218	198	i2	0	0	18								
ADCY4	Exonic	402	0	0	i11*	402	0	2	i12*	402	0	0				
ADCY5	Exonic	68	2949	506	i17	0	0	6								
ADCY6	Exonic	1135	1831	294												
ALB	Exonic	1	0	0												
APP	Exonic	16045	37321	10240	i1	26	0	0	i2	0	18	0	i5	0	0	78
	i6	26	0	0	i8	196	0	2	i13	264	0	524	i15	0	0	50
CACNA1B	Exonic	1	9	56	i31	24	0	0								
CACNA1H	Exonic	0	23	35	i9	2	0	0								
CAMK2A	Exonic	47	492	328												
CAMK2B	Exonic	2856	21311	14927	i1	28	0	0								
CAMK2D	Exonic	6241	968	6784	i2	2	0	32	i3	34	0	60	i11	0	0	14
	i18	94	0	0												
CAMK2G	Exonic	706	822	1306	i9	0	38	0	i13	36	0	266	i18	0	0	6
CREB1	Exonic	0	0	19	i1	20	0	56								
FMR1	Exonic	920	156	640	i1	2	0	0	i7	0	0	20	i12	2	0	0
GABBR1	Exonic	670	13559	5077	i4	1346	8	1658	i14	8	0	0				
GABRA4	Exonic	4792	5	3753	i8	0	4	0								
GABRG3	Exonic	0	0	0	i2	0	46	0	i3	0	0	22	i5	4	0	30
	i9	0	0	18												
GRIA3	Exonic	4087	34	691	i2	6	0	0	i11*	6	0	0	i12	0	296	4
	i13*	188	0	0	i15*	0	0	280								
GRIA4	Exonic	0	11	67	i3	474	0	8	i4	92	94	2	i7	0	0	16
	i13*	0	2	0												
GRIK1	Exonic	4	769	45	i1	76	192	0	i7	0	0	32	i16*	0	0	44
GRIN1	Exonic	63	259	176	i8	2	0	0								
GRM7	Exonic	4	61	352	i1	22	0	144	i5	27	8	34	i6	1	0	0
KCND2	Exonic	90	580	1353	i1	8	0	2								
MAP2	Exonic	21380	10062	11896	i4	0	0	12	i7	0	0	2	i9	0	0	2
	i10	2	0	6												
SNCA	Exonic	7028	10337	27965	i4	0	0	14								
SNCB	Exonic	539	1687	2512												
STX1A	Exonic	13	625	759												
STX1B2	Exonic	119	1774	1453												

Table S3

Symbol	Intron	Coord	Strand	ID	Seq. evidence
ADCY2	i1	chr17:4557075-4557148	+	ADCY2i1ID1	
ADCY2	i13	chr17:4923799-4923872	+	ADCY2i13ID1	intron
ADCY2	i14	chr17:4931390-4931463	+	ADCY2i14ID1	intron
ADCY2	i3	chr17:4679992-4680065	+	ADCY2i3ID1	
ADCY2	i3	chr17:4721539-4721612	+	ADCY2i3ID2	
ADCY2	i3	chr17:4820243-4820316	+	ADCY2i3ID3	
ADCY4	i21	chr15:33932416-33932489	-	ADCY4i21ID1	
ADCY5	i1	chr11:67398142-67398213	-	ADCY5i1ID1	cis
ADCY5	i3	chr11:67338681-67338754	-	ADCY5i3ID1	
APP	i13	chr11:24481873-24481946	-	APPi13DI1	cis
APP	i14	chr11:24474905-24474978	-	APPi14ID1	
CACNA1B	i31	chr3:2915346-2915419	-	CACNA1Bi31ID1	intron
CAMK2A	i2	chr18:56897536-56897608	+	CAMK2Ai2ID1	
CAMK2B	i1	chr14:86712517-86712590	-	CAMK2Bi1ID1	intron
CAMK2B	i3	chr14:86667060-86667132	-	CAMK2Bi3ID1	
CAMK2D	i2	chr2:223879849-223879922	+	CAMK2Di2ID1	intron
CAMK2D	i4	chr2:224012159-224012231	+	CAMK2Di4ID1	intron
CAMK2D	i7	chr2:224038648-224038720	+	CAMK2Di7ID1	
CAMK2G	i18	chr15:3784885-3784955	+	CAMK2Gi18ID1	cis
CAMK2G	i2	chr15:3748334-3748403	+	CAMK2Gi2ID1	
CDH1	i2	chr19:36456777-36456850	+	CDH1i2ID1	intron
CDH1	i2	chr19:36488291-36488364	+	CDH1i2ID2	intron
FMR1	i1	chrX:154759680-154759747	+	FMR1i1ID1	cis
FMR1	i2	chrX:154764980-154765052	+	FMR1i2ID1	
GABRA4	i2	chr14:39058639-39058711	+	GABRA4i2ID1	
GABRG3	i3	chr1:108690876-108690948	-	GABRG3i3ID1	intron
GABRG3	i5	chr1:108287957-108288025	-	GABRG3i5ID1	intron
GABRG3	i5	chr1:108285219-108285291	-	GABRG3i5ID2	intron
GABRG3	i5	chr1:108273028-108273101	-	GABRG3i5ID3	intron
GABRG3	i5	chr1:108394161-108394234	-	GABRG3i5ID4	intron
GABRG3	i6	chr1:108223713-108223786	-	GABRG3i6ID1	
GRIA3	i2	chrX:3683750-3683823	-	GRIA3i2ID1	intron
GRIA3	i4	chrX:3569721-3569788	-	GRIA3i4ID1	
GRIA4	i14	chr8:985632-985705	-	GRIA4i14ID1	intron
GRIA4	i14	chr8:971449-971519	-	GRIA4i14ID2	intron
GRIA4	i4	chr8:1236057-1236129	-	GRIA4i4ID1	cis
GRIA4	i6	chr8:1088993-1089066	-	GRIAi6ID1	
GRIA4	i7	chr8:1055625-1055698	-	GRIA4i7ID1	intron
GRIK1	i1	chr11:27943131-27943204	-	GRIK1i1ID1	intron
GRIK1	i1	chr11:27913623-27913697	-	GRIK1i1ID2	intron
GRIK1	i1	chr11:28081118-28081190	-	GRIK1i1ID3	intron
GRIK1	i1	chr11:28044738-28044811	-	GRIK1i1ID4	intron
GRIK1	i1	chr11:28098620-28098693	-	GRIK1i1ID5	intron

GRIK1	i16	chr11:27709224-27709296	-	GRIK1i16ID1	intron
GRIK1	i6	chr11:27814404-27814476	-	GRIK1i6ID1	
GRIN1	i4	chr3:3471203-3471276	-	GRIN1i4ID1	
GRIN1	i8	chr3:3462386-3462457	-	GRIN1i8ID1	intron
GRIN1	i8	chr3:3461694-3461767	-	GRIN1i8ID2	intron
GRM7	i1	chr4:146624192-146624265	+	GRM7i1ID1	intron
GRM7	i1	chr4:146736661-146736734	+	GRM7i1ID2	intron
GRM7	i1	chr4:146746191-146746264	+	GRM7i1ID3	intron
GRM7	i1	chr4:146760671-146760744	+	GRM7i1ID4	intron
GRM7	i1	chr4:146937318-146937390	+	GRM7i1ID5	intron
GRM7	i1	chr4:146951304-146951377	+	GRM7i1ID6	intron
GRM7	i1	chr4:146433109-146433182	+	GRM7i1ID7	intron
GRM7	i2	chr4:146972743-146972815	+	GRM7i2ID1	
GRM7	i5	chr4:147097472-147097544	+	GRM7i5ID1	intron
GRM7	i6	chr4:147204899-147204971	+	GRM7i6ID1	intron
GRM7	i6	chr4:147222137-147222210	+	GRM7i6ID2	intron
KCND2	i1	chr4:47841144-47841217	+	KCND2i1ID1	intron
KCND2	i1	chr4:47972539-47972610	+	KCND2i1ID2	intron
KCND2	i1	chr4:47998363-47998436	+	KCND2i1ID3	intron
KCND2	i1	chr4:48009814-48009886	+	KCND2i1ID4	intron
KCND2	i1	chr4:47642184-47642253	+	KCND2i1ID6	intron
MAP2	i10	chr9:65246459-65246531	+	MAP2i10ID1	intron
MAP2	i10	chr9:65249684-65249754	+	MAP2i10ID2	cis
MAP2	i11	chr9:65253115-65253188	+	MAP2i11ID1	
SNCA	i2	chr4:89713754-89713825	-	SNCAi2ID1	
SNCG	i3	chr16:10028853-10028926	-	SNCGi3ID1	
STX1B2	i4	chr1:187098603-187098671	-	STX1B1i4ID1	

Table S4

Dendrite sample	**D1**	**D2**	**D3**
Total number of reads (single)	25,923,420	25,428,726	21,647,526
Num. reads aligning:			
ID hairpin (74 nt)	101,981	109,919	75,227
BC1 3’ domain (78 nt)	0	0	0
B2 5’ portion (74nt)	11,793	8,603	11,143
Num. aligned reads per genomic instance			
ID (161,321 genomic instances)	0.623	0.681	0.466
B2 (352,447 genomic instances)	0.033	0.024	0.032
Fold enrichment ID > B2	18.9	27.9	14.7

Table S5

Symbol	i	Genomic coord	MFE	MFEZ	SC	RFAM	CSHMM	MiPred	Triplet
GABRG3	i5	chr1:108,421,636-108,421,725	-44.3	-6.59	0.99	2.7E-8	1	1	1
GRIK1	i1	chr11:28,068,838-28,068,904	-27.4	-5.96	0.98	3.1E-3	1	1	1
FMR1	i1	chrX:154,761,229-154,761,378	-41.5	-3.55	0.92	1.5E-3	1	0	0
CAMK2B	i3	chr14:86,668,207-86,668,312	-54.3	-3.43	0.95	8.1E-3	1	0	0

Supplemental Text

PCR confirmation of intronic loci presence in dendrites. As an additional means of confirming the presence of retained intronic sequences within dendritically localized transcripts, dendritic cDNA was used as template material for PCR reactions. Select primer pairs used to generate sequences for microarray printing were used to amplify sequences from cDNA made from isolated dendrites. Bands were detected using this approach, subcloned and sequence verified as the intronic target sequences for all introns queried (Figure S1C), except for those in STX1A. The PCR products confirm that the target intronic sequences are part of the transcripts bearing the coding sequence for target genes, as the 5' primer lies within the upstream exon in all cases.

Verifying candidate intronic retention loci. To confirm that the candidate retained introns were not unannotated alternate exons or overlapping genes, we analyzed each retained intronic locus using publicly available annotations and base composition properties. Some of the candidate intronic loci were in fact found to have high sequence similarity to annotated exons in related species using RefSeq and Ensembl, or were spanned by N-SCAN or Genescan predicted exons; these were removed from consideration. However, a vast majority of detected CIRTs showed no detectable conservation or gene prediction signals in the retained intronic region. We also analyzed the distributions of open-reading frame (ORF) lengths among candidate CIRTs of interest and further eliminated candidates that had unusually long stop codon-free

regions in any reading frame, using a cutoff empirically determined by the stop codon characteristics of known coding sequence. In all, only six introns were eliminated from consideration on the basis of any of these characteristics: ADCY2i21, ADCY6i1, and GRIA4i10 due to annotated coding evidence; and ADCY5i1, ADCY5i8, and CAMK2Di13 due to long ORFs.

We also sought to ensure that the observed intronic read coverage was in fact due to selective intron sequence retention and not due to RNA sample contamination. First, we found that reads mapping to intergenic regions are generally not present in the sequencing data. For our candidate CIRTS, we found a highly significant enrichment of intronic read coverage as compared to neighboring intergenic regions of equal length ($p < 2.2E-16$ by the Binomial Proportion Test), indicating that read alignments within the gene boundaries do not arise from genomic DNA contamination.

We also reasoned that if pre-mRNAs are present in the RNA extracts, all introns in a gene should be represented roughly proportionally in the sequencing data; however, among the genes with strong evidence of exonic sequence coverage (at least 20 reads aligning to exons), only 15 percent of the introns show evidence for retention. There is a high degree of variability in the number of introns with retention evidence across our genes of interest, which is not positively correlated with the number of introns in the gene ($p = 0.93$, Pearson's correlation). Similarly, there is no correlation between intron position in the gene architecture and retention, as roughly equal numbers of first and last introns were retained (five versus seven), and there is no significant rank correlation between degree of read coverage per intron and intron

position in the gene for any of the candidate CIRTs ($0.080 < p < 0.99$, Spearman's correlation).

Measuring the prevalence of ID elements in dendritic RNA pools. We analyzed the entire pool of sequencing data from the dendrite samples and found a pronounced enrichment in reads that contain ID-element sequence. Compared to transposable elements belonging to the SINE B2 element family, which occur approximately 2.2 times more frequently than the ID element in the rat genome as annotated by RepeatMasker, ID elements occur on average 9.4 times more often in the sequence reads, implying a 20-fold enrichment in ID element sequence in the dendritic RNA pools (Table S4). These ID-related sequence reads do not arise from BC1 RNAs, since the comparably-sized BC1 3' domain does not appear at all in the sequencing data. These results suggest that a large number of dendritically localized transcripts contain ID element sequences.

***In situ* hybrization control experiments.** For *in situ* hybridization studies, all cultures were immunostained with antibodies to MAP2 protein to confirm dendrito-somatic regions of neurons, and sense strand probes were used as nucleotide compositional controls (Figure S2). Additionally, we performed controls recommended by the Allen Mouse Brain Atlas (http://mouse.brain-map.org/documentation/index.html). We used different detection systems (DAB and Quantum Dot), tested for background signal in ISH's performed without probes, and repeated the ISH studies on distinct cultures from

different dates of harvesting and different litters (data not shown).

To ensure that exogenous transcript localization patterns were not the result of association with translational machinery in the cytoplasm, we also looked for EGFP protein expression in cells transfected with the ID-EGFP constructs. We found that the ID-containing constructs engage the translational machinery, producing detectable levels of EGFP distributed throughout the cytoplasm (Figure S3D).

Quantifying differences in dendritic *in situ* hybridization signals between experimental and control cells. For each cell under consideration, signal intensity was measured along the length of the dendrites, out to a distance of approximately 50 µm from the soma, and normalized to set the highest intensity pixel to 1. The RNA intensities were pooled in 8 µm intervals and paired t-tests were carried out to assess the significance of the difference between the test probe and control EGFP probe within each interval. Since the resulting set of t-tests may not be independent due to shared residuals from a gradient-like generating process along the dendrites, we carried out a conservative Bonferroni correction for non-independent multiple tests. All test probes showed significantly greater signal intensity along the length of the dendrites compared to the EGFP control using a Fisher's combined statistic for Bonferroni-corrected t-test p-values from each interval. Fisher's combined p statistics and probabilities of the Fisher's combined p value are as follows -- CAMK2Bi3ID1: 71.99, p = 1.36E-10; FMRi1ID1: 89.02, p = 7.63E-14; GABRG3i5ID2: 74.19, p = 5.24E-11; GRIK1i1ID4: 74.19, p = 5.24E-11). Similar calculations were performed for the competition and

cross-competition assays. For the competition assays, *in situ* signal for intron-containing transcripts in the absence of competitor construct was compared to signal in the presence of competitor constructs, yielding the following Fisher's statistics and p values – CAMK2Bi3: 64.04, p = 4.10E-9; FMR1i1: 67.42, p = 9.69E-10; GABRG3i5: 64.36, p = 3.58E-9; GRIK1i1: 51.55, p = 7.46E-7. For the cross-competition assays, *in situ* signal was again compared in the absence and presence of competitor constructs, yielding the following Fisher's statistics and p values – CAMK2Bi3 vs FMR1ID1: 44.51, p = 1.25E-5; FMR1i1 vs CAMK2Bi3ID1: 78.03, p = 9.8E-12.

Diffusion modeling of dendritic *in situ* intensities. In order to compare our *in situ* results to a non-specific diffusion model, we tested the differential gradient by fitting the entire probe intensity curve to a negative hyperbolic function of the form $I = c - \frac{sd}{g + d}$ where *I* represents probe intensity and *d* represents distance from soma. The parameters *c* and *s* represent translation and scale of the curve with *(c – s)* forming the asymptote of the curve. The parameter *g* represents the steepness of the curve; i.e., the steepness of the gradient, and is therefore the parameter of interest. The ISH signals for control EGFP probes (n=8) and test probes (n=10, each) were fitted using a nonlinear least-squares fitting procedure (R statistical package). For the EGFP control, the 95% confidence interval for the parameter *g* is 1.63+/-0.241, while for the test probes, the intervals are CAMK2B = 2.32+/-0.420, FMR1 = 4.94+/-1.133, GABRG3 = 3.56+/-0.499, GRIK1 = 4.96+/-0.866. Thus, EGFP forms a significantly steeper gradient along the dendrites than any of the four quantified test probe ISHs, suggesting more active

transport of the mRNA corresponding to the test probes. It should be noted that the parameter estimate *g* provides an assessment of the RNA gradient controlling for the expression levels (computationally rescaling the expression levels results in the same *g* estimate; data not shown). In effect, fitting a hyperbolic curve and then testing the steepness parameter establishes the spatial pattern as a self-control that is invariant of expression levels or probe specific effects.

ID element mutations block dendritic targeting capacity. We introduced two separate mutations in wild-type ID element sequences to verify that the BC1-like targeting motifs conferred targeting capacity: deletion of the uracil at position 22, which eliminates a bulge in the predicted secondary structure; and truncation of the basal stem by deleting seven nucleotides from each end of the ID element (positions 1-7 and 68-74) leaving the rest of the ID element and its flanking intronic sequence intact. Both of these mutations in 5' BC1 significantly disrupted its dendritic localization in neurons (Muslimov et al., 2006), and we see similar results when these same mutations are incorporated into intronic ID elements, as dendritic localization of exogenous ID-EGFP transcripts is inhibited (Figure 2B). Both mutations decreased transgene mRNA targeting to dendrites by nearly the same extent.

Supplemental Experimental Procedures

Microarray preparation Thirty-three dendritic genes were selected for investigation.

RefSeq gene models were used, as shown in the following table:

RefSeq ID	Symbol	Description	Strand	Chr	Start	End
NM_031007	ADCY2	ADENYLATE CYCLASE 2	+	17	4543490	5040433
NM_130779	ADCY3	ADENYLATE CYCLASE 3	+	6	27118400	27202275
NM_019285	ADCY4	ADENYLATE CYCLASE 4	-	15	33930534	33946315
NM_022600	ADCY5	ADENYLATE CYCLASE 5	-	11	67290968	67437468
NM_012821	ADCY6	ADENYLATE CYCLASE 6	-	7	137339933	137360020
NM_134326	ALB	ALBUMIN	-	14	19126965	19142199
NM_019288	APP	AMYLOID BETA (A4) PRECURSOR PROTEIN	-	11	24457855	24693851
NM_147141	CACNA1B	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, N TYPE, ALPHA 1B SUBUNIT	-	3	2873391	3039747
NM_153814	CACNA1H	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, T TYPE, ALPHA 1H SUBUNIT	-	10	14621372	14679051
NM_012920	CAMK2A	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE II ALPHA SUBUNIT	+	18	56879247	56948537
NM_021739	CAMK2B	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE II BETA SUBUNIT	-	14	86634690	86721261
NM_012519	CAMK2D	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE II, DELTA	+	2	223840650	224108082
NM_133605	CAMK2G	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE II GAMMA	+	15	3729433	3786057
NM_031334	CDH1	CADHERIN 1	+	19	36442693	36512091
NM_031017	CREB1	CAMP RESPONSIVE ELEMENT BINDING PROTEIN 1	+	9	63170785	63234725
NM_052804	FMR1	FRAGILE X MENTAL RETARDATION SYNDROME 1 HOMOLOG	+	X	154756031	154793782
NM_031028	GABBR1	GAMMA-AMINOBUTYRIC ACID (GABA) B RECEPTOR 1	-	20	1553313	1582398
NM_080587	GABRA4	GAMMA-AMINOBUTYRIC ACID (GABA-A) RECEPTOR, SUBUNIT ALPHA 4	+	14	39047461	39122526
NM_017289	GABRD	GAMMA-AMINOBUTYRIC ACID A RECEPTOR, DELTA	-	5	172203065	172214960
NM_024370	GABRG3	GAMMA-AMINOBUTYRIC ACID (GABA) A RECEPTOR, SUBUNIT GAMMA 3	-	1	108189311	108821051
NM_032990	GRIA3	GLUTAMATE RECEPTOR, IONOTROPIC, AMPA3 (ALPHA 3)	-	X	3454606	3719276
NM_017263	GRIA4	GLUTAMATE RECEPTOR, IONOTROPIC, 4	-	8	957190	1438021
NM_017241	GRIK1	GLUTAMATE RECEPTOR, IONOTROPIC, KAINATE 1	-	11	27703875	28106450
NM_017010	GRIN1	GLUTAMATE RECEPTOR, IONOTROPIC, N-METHYL D-ASPARTATE 1	-	3	3453784	3480381
NM_031040	GRM7	GLUTAMATE RECEPTOR, METABOTROPIC 7	+	4	146332578	147270224
NM_012970	KCNA2	POTASSIUM VOLTAGE-GATED CHANNEL, SHAKER-RELATED SUBFAMILY, MEMBER 2	+	2	202560175	202564305
NM_031730	KCND2	POTASSIUM VOLTAGE GATED CHANNEL, SHAL-RELATED FAMILY,	+	4	47541787	48047906

		MEMBER 2				
NM_013066	MAP2	MICROTUBULE-ASSOCIATED PROTEIN 2	+	9	65174379	65255995
NM_019169	SNCA	SYNUCLEIN, ALPHA	-	4	89613731	89722807
NM_080777	SNCB	SYNUCLEIN, BETA	+	17	15907598	15915704
NM_031688	SNCG	SYNUCLEIN, GAMMA	-	16	10025979	10030513
NM_053788	STX1A	SYNTAXIN 1A (BRAIN)	-	12	22737113	22765064
NM_012700	STX1B2	SYNTAXIN 1B2	-	1	187089182	187108643

Fragments were amplified using forty rounds of PCR with an annealing temperature of 50°C. The template used was rat genomic DNA isolated from rat liver. 1ug of each of 96 PCR products were submitted to the University of Pennsylvania Microarray Facility for printing on Corning UltraGap slides. These samples were dried and resuspended in 10ul of Corning Spotting Buffer. 1nl of each sample was then denatured and printed in each spot on individual slides and cross-linked using ultra-violet light for immobilization. All PCR products were subcloned and confirmed by sequencing prior to array printing. The following probe sequences were printed:

RefSeq ID	Symbol	Intron	Intron length	Probe length	Sense primer	Antisense primer
NM_031007	ADCY2	1	15927	417	CTGCTAGCCGTCTTCTTC	TTTGCCAAAGCAGCCTGT
NM_031007	ADCY2	4	4857	412	TCTGTATGTCAAACGACA	TGGAATGCACTGGCTTCA
NM_031007	ADCY2	20	1700	486	CTGAATGAGATCATTGCT	TGGCATTTGCCAGAGACC
NM_130779	ADCY3	2	37523	449	GAGCTAGAAGGGATGCAG	GGGCAGACAGGATCTCCA
NM_130779	ADCY3	3	26352	480	ATGAATCTGGAGGAGCAG	ACCCATTCACCCCACACA
NM_130779	ADCY3	21	399	396	GAATCCACAGGGGTCATG	GCCAGGATGGTAGGAAAGAA
NM_019285	ADCY4	2	1452	387	CTGCCCGCCGTCGCCTGG	AGCCCCACATTCCCAAAT
NM_019285	ADCY4	11	2612	424	AGTCCTGCATCCACATCC	GCCCTGCCCTAGCTTCAT
NM_019285	ADCY4	24	279	287	ACTCGTTCAACAACTTCC	TGGGGAGGCATGCTAGAG
NM_022600	ADCY5	1	90496	403	AACGCCCAGGACCAGTTT	TTCAGGAGGGGATGGCTA
NM_022600	ADCY5	3	15753	451	GATTTACATCCAGAAACA	GTGGAGGCGAAGTGTGCT
NM_022600	ADCY5	20	1031	389	GACAGCACTGGGGTGCCT	GGCCAGGAGAACTGCAAG
NM_012821	ADCY6	1	5953	471	CGTGGCCCGATCCGGTTC	ACAAGCGGCCACTGTTCT
NM_012821	ADCY6	2	2717	389	AACAATGGTGACCCCTTC	GGCCATGGTGGGTTACAA
NM_012821	ADCY6	21	193	215	GACAGCACAGGAGTTCCT	TGAGAGGGGAGCCTGCTT
NM_134326	ALB	1	703	498	TTTCCAGGGGTGTGTTTC	GCATTCAAACCCAAGACG
NM_134326	ALB	6	1518	489	GAGATTTGGAGAGAGAGC	AAGGCCCTGGGTTCAGTC
NM_134326	ALB	14	1061	381	CTTAGCCTAAACACATCA	GGTTTCAGCTTCATTCAGAGG
NM_019288	APP	1	52131	405	TGGCCGCCTGGACGGTTC	TGAGACACTGCGCTTTCG
NM_019288	APP	6	35426	490	CCACTGAGTCTGTGGAGG	CCAGCCCAGTGAATGACC
NM_019288	APP	17	5628	424	TACACATCCATCCATCAT	CAGGGCCCAAACAAAACA
NM_147141	CACNA1B	1	880	466	CAAATATGCTAAGCGCAT	CCTGCAACCACACAGCAC
NM_147141	CACNA1B	18	6129	406	ATGTCTGCTGCCAACATC	GGCTGGGCTGCACTAGAA
NM_147141	CACNA1B	45	476	405	GCGGCACTAGAGCCAGGA	AGGCAAATGGAGGGGAAG

NM_153814	CACNA1H	1	34322	477	CAGCTGGTGCCTCCGACT	TCCGCGGAGAACTAAGGA
NM_153814	CACNA1H	5	1331	427	CCGATGCTTCCTGGACAG	CTCCACCCAGGCTCTCCT
NM_153814	CACNA1H	33	669	406	CTGAGTCTCTCACGGATA	GGACCCAGGTCACCTCCT
NM_012920	CAMK2A	1	15908	454	AGAGTACCAGCTCTTCGA	GCGGGGGTGAAGTCTCTC
NM_012920	CAMK2A	3	14909	429	GAGGGGCACCACTACCTT	TCCCATGTGCTCCTGTCC
NM_012920	CAMK2A	10	4347	429	TGGACTTTCATCGATTCT	ACTGGGCATCTGGGATGA
NM_021739	CAMK2B	1	35619	411	CGAGTACCAGCTATACGA	GGCTAGGACCGACCCTTC
NM_021739	CAMK2B	3	8295	424	TCTGCCGCCTGCTGAAGC	CATGGCTACCAGGGAAGG
NM_021739	CAMK2B	17	6677	392	ACACAACCATAGAGGACG	ATCACTGGCAGGGCTCAG
NM_012519	CAMK2D	1	1579	468	CGAGTATCAGCTCTTCGA	TCGGCATCGATAAACAAGG
NM_012519	CAMK2D	4	43278	491	AGAGGGCTTCCATTACTT	AGCGCTAGGCAGTCATGC
NM_012519	CAMK2D	18	1157	468	TCGTTCGGGGTCCCCAAC	CCCCCTTGTTCTCTTTCCA
NM_133605	CAMK2G	2	7736	411	TATGCCGACTTCTGAAAC	AGTGCGCAGGACACTTCC
NM_133605	CAMK2G	9	3744	440	GACCAGGCTCTCAAGCAC	GAGGTCCTCGGAACTCAGC
NM_133605	CAMK2G	18	1505	444	TGGATTTCCATAAGTTTT	TCGATGTCCAGGGTGTCC
NM_031334	CDH1	1	750	419	AGCTTCTCCGCGCTCCTG	TACGGTCCCTCCCTCACA
NM_031334	CDH1	3	3930	405	GGCCACCACCACCACCGG	GGGGGTTGTCGTCAGAGA
NM_031334	CDH1	15	1726	390	CCTGATGAAATCGGGAAC	GAGCTGGCCGATTCAAGA
NM_031017	CREB1	2	7008	393	GCCCAGCCACAGATTGCC	ACACCTTCAGGTCCCAAAAA
NM_031017	CREB1	6	3999	438	CGATTTACCAAACTAGCA	CATGCAGGCACACACACA
NM_031017	CREB1	8	21931	481	AAAGAGAGAGGTTCGTCT	TGCATAAAAATGCAAAAGAAAG
NM_052804	FMR1	1	8022	466	GTGCGGGGCTCCAATGGC	AGCCCAAGACTCCGAACC
NM_052804	FMR1	7	2108	457	ACTGAGAAATGAAGAAGCCAGT	GCCTCTGGCAGGAATCTG
NM_052804	FMR1	15	4185	439	AAAGGAAGAACAACTGAT	GCTCTTAGTAAAGGTGCCTGGA
NM_031028	GABBR1	6	2933	467	AGCTCAAGCTTATCCACC	CAGAATCAGGCAGCCTTACC
NM_031028	GABBR1	11	5346	461	CCTGCCAACACCCGAAGC	TGAAGCAACCCCCTGAAG
NM_031028	GABBR1	21	105	296	ATCACTCTGGTTGTGCTC	TGTTTGTCACCGCACACC
NM_080587	GABRA4	1	424	399	CCTCCTGCACGTCCTGTG	ACAGGTGGAAAGGGATTGC
NM_080587	GABRA4	7	8278	434	AGTCTGTTCCAGCCAGAA	GCAGTCGGCTGAGACTTTG
NM_080587	GABRA4	8	42313	424	AAGGAAAAACATACAGAA	GGGTCATTCTGGCTGAGG
NM_017289	GABRD	1	7764	474	GCTGTGCACGCAGCCGCA	ACTCACGCAAGGGCAGAC
NM_017289	GABRD	2	309	326	ACGCCCGAAACTTCCGAC	GGCCTGCTAGCGTTCGT
NM_017289	GABRD	8	273	301	AAGGTCAAGGTCACGAAG	CAGAGAGTGGGGCAGGAG
NM_024370	GABRG3	1	4902	401	CCTCTGCCTGTTCTCGGG	CTGGGTCCGGAGGTTTCT
NM_024370	GABRG3	5	188354	469	CGCATGCCTGCCCTTTGA	TGGTCTCCTTGTGGAATGAA
NM_024370	GABRG3	9	5156	436	CCTGATCGAATAAGCCTT	AGACGTATGGGCTGGAGAAG
NM_032990	GRIA3	1	1150	447	ACGGAGGATTCCCCAACA	CCCAGGTTATGCGTCCTG
NM_032990	GRIA3	4	28342	411	GTCGAAAGGATTAACACA	CCTGTCAGCCCTGCCTAA
NM_032990	GRIA3	14	2560	402	GAATGTGGAGCCAAGGAC	CTGTCACAGCTGGATCTTTTT
NM_017263	GRIA4	2	1151	489	TGGGAGCCTTTCCAAGCA	ACTGGCAAGGCAACAAGG
NM_017263	GRIA4	4	151900	398	GTTTCGTCTTCCTGTATG	CACAAAGCCTGTTTGCTGAG
NM_017263	GRIA4	16	2697	417	AAGTCCAGGGCAGAGGCG	CAATTTTAGGGAAATCATGCTG
NM_017241	GRIK1	1	243354	466	CTCAGACCTCCCCTCAAG	GGCTCTGGCATGCTTTTC
NM_017241	GRIK1	3	17589	447	CAGTGACGGTGGTGTATG	CCCTTCCCCAAGGAATGT
NM_017241	GRIK1	16	18626	417	AAATCACGGAAGAACAAT	TGTATCCTGAAGGGCTGGT
NM_017010	GRIN1	1	1359	414	TCAGTGTGTGAGGACCTC	CTGCTTGCTTTGGCACCT
NM_017010	GRIN1	8	4252	421	CGAGGATACCAGATGTCC	TGCAAATCAGGCTGAGGA
NM_017010	GRIN1	19	1911	396	TTCAAGAGACGTAGGTCC	AGGGGCAACCTGGGATAG
NM_031040	GRM7	1	618843	473	ATGGTAGCCAACATCTTG	TGGCTTAGGCATGGGAAG
NM_012970	KCNA2	1	1402	447	ACATGGAGGCTCGGGTAC	CCCAGCCAAGTGGAAAAA
NM_031730	KCND2	1	486424	492	GTACACCATCGTCACCAT	GGCAAACCTAGACTTCCAAACA
NM_031730	KCND2	2	11469	501	CAACGAGCGGACAAACGA	GCCCGATTTCTAAGGGATAAA
NM_031730	KCND2	4	2607	437	CACCTGCTTCACTGCCTG	CCCATCCCTCCTTTGCTT
NM_013066	MAP2	3	1945	485	ACAAGGATCAGCCTGCAG	TGGGAAATGGCAAGGCTA
NM_013066	MAP2	4	10939	461	CATCAGAACAAACAGCTG	AGGGCTGGGGGAAGTTCT
NM_013066	MAP2	12	480	408	CACCATGTACCTGGAGGT	CCCTCCTTCCCGTTTGAG
NM_019169	SNCA	2	10738	399	GGAAGACAAAAGAGGGCG	TCTTTGGGCCTGCACTTC
NM_019169	SNCA	4	89732	469	GGTTTTGTCAAGAAGGAC	AGCCCAGTTTGCAAAATCAT
NM_080777	SNCB	2	2922	439	AGAAGACCAAGGAAGGCG	TCATACCCGGGCTTCAGA
NM_080777	SNCB	5	379	401	GAAGGGGAAAGTTACGAG	AGGGAGGGGTGGAACAGA
NM_031688	SNCG	1	762	404	AGAAGACCAAGGAAGGGG	AGGTGGGGTCCTGTAGCC
NM_031688	SNCG	4	237	215	GAGGCCAAAGAGCAAGAG	ATTCCCAGGCCCACAACT
NM_053788	STX1A	1	12905	502	ATGAAGGACCGAACCCAG	TGATGAGGAGCCCCAAGA

NM_053788	STX1A	3	3564	536	CATCCTGGCCTCCCCGAA	GGTAGCCTGCTTGGTCAGA
NM_053788	STX1A	9	828	427	GCCGTCAAGTACCAGAGC	GGCGTGGCTCAGTGGTAG
NM_012700	STX1B2	1	8126	407	ATGAAGGATCGGACTCAG	CAGTAACCACCCCCTCCA
NM_012700	STX1B2	4	3284	479	CGGCAAACAAGGTCCGGT	CATTTGCTGAGCCCACCT
NM_012700	STX1B2	7	2455	473	AGCGGGAAGCTGGCCATC	GGGCTGTGGGTCATTGTC

Following three rounds of aRNA (Miyashiro et al., 1994), labeled single stranded cDNA was generated by incorporation of amino-allyl labeled dUTP and conjugation with Cy3. Labeled material from dendrites was hybridized to our custom microarrays and screened for positives.

Slides were blocked (pre-hybridized) at 42°C for 3 hours in 1% bovine serum albumin (BSA), 1% sodium dodecyl sulphate (SDS), and 3X saline-sodium citrate (SSC). Hybridization was carried out in Corning slide chambers for sixteen hours at 42°C in a 25% percent formamide, 0.1% SDS, 4X SSC buffer with human Cot-1 DNA, single stranded (SS) poly dA and poly dT DNA, yeast transfer RNA (tRNA) and T7-oligo dT primer as blocking agents. Slides were washed two times for five minutes at room temperature (RT) in 2X SSC, 0.1% SDS, two times for five minutes at 42°C in 0.2X SSC, 0.1% SDS, and two times for five minutes at RT in 0.2X SSC. Slides were scanned using an Axon Instruments GenePix 4200 series scanner provided by the University of Pennsylvania Microarray Facility, and analyzed with GenePix 6.0 software.

Raw signal intensities obtained for each array were background subtracted, logged, then median centered and interquartile range normalized. For each intron microarray spot, we compared the signal intensities across the three dendritic RNA sample replicates against corresponding signal from two arrays hybridized to random hexamer sequence as a compositional control. Introns for which spot intensity on the

experimental arrays was higher than signal intensity on the control arrays, at a significance level of $p < 0.1$ by t-test, were classified as candidates for retention.

Illumina sequencing: Specific read coverage (Tables S2) for our 33 genes of interest was performed using Bowtie (Langmead et al., 2009) version 0.11.3, allowing up to three nucleotide differences per 50-nt read to the reference Rat Genome v. 3.4 (Gibbs et al., 2004). Only paired-end reads genome-wide uniquely aligning to intronic loci, in which at least one end of the pair overlaps non-RepeatMasker-annotated repetitive sequence, were considered. Due to the repeat-rich nature of intronic sequence, this policy will tend to underestimate actual read coverage. Exonic read coverage was estimated by using genome-wide unique single-end read alignments to RefSeq-defined mature mRNA sequences, plus up to 2000 nts of repeat-masked sequence downstream the annotated 3' end of the gene to account for alternate UTR isoforms that are supported by homology evidence. Intergenic read coverage was determined by counting paired reads in non-repeat-overlapping transcription-free regions as annotated by GenBank, RefSeq, Ensembl gene models; as well as genomic alignment of expressed sequence tags and homologous RefSeq transcripts by the UCSC Genome Browser.

***In situ* hybridization and imaging**: Antisense digoxigenin or biotin-labeled probes were produced as runoff transcripts from plasmid DNAs that were digested at a site downstream of the region to be transcribed. Primary rat hippocampal neurons were

fixed for 15 minutes in 4% paraformaldehyde, washed in PBS and permeabolized with 0.3% TritonX-100. Cells were prehybridized at 42C with 50% formamide, 1X Denhardt's solution, 4X SSC, 10mM DTT, 0.1% CHAPS, 0.1% Tween-20, 500ug/ml yeast tRNA, 500ug/ml salmon sperm DNA. *In situ* hybridization was performed at 42C with 10ng/ul (for EGFP probes) or 20ng/ul (for intron probes) probe in prehybridization buffer with additional 8% Dextran sulfate. Rabbit anti-MAP2 antibody was added to cells after probe hybridization followed by goat anti-rabbit antibody and streptavidin conjugated to Qdot 525 and Qdot 605 molecules for imaging. The samples were visualized by confocal microscopy. The emission wavelengths for each fluorescent dye were selectively collected by specific spectral ranges of dyes with either slit width (Olympus fluoview 1000, 60x N.A.1.2 or 20x N.A.0.7) or Meta detector (Zeiss 510 meta, 40x N.A 1.0). The collected images were minimally processed in Metamorph image analysis software and extracted information in regions of interest was transferred to Excel. The images were background subtracted and scaled 0 to 2000 in 12bit bit depth unless indicated in text.

Oligo probe sequences for splice junction *in situ* hybridization were as follows:

RefSeq ID	Symbol	Junction	Sequence
NM_032990	GRIA3	e11-i11	GATATTTCTCCAAGGTTTGTTTGTG
NM_017263	GRIA4	e13-i13	AGAATTCTTCAGAGTAAGTCACAGGA
NM_017241	GRIK1	i16-e17	TGTACTCTCCAGTGTCTCTCTTTCA
NM_017241	GRIK1	i16-altexon	TTGGTGTAGAGAAGGTTTGTTCATT

ID element targeting and competition. Intron-derived ID elements were studied using PCR products with primers designed to amplify selected ID elements plus flanking

sequence from retained intron regions. Rat genomic DNA was used as the template to produce amplicons of 137-152 nts in length each roughly centered on the 74-nt ID element. Products were cloned into pEGFP-N1 expression vectors (CMV promoter driven), with the ID region placed upstream of the EGFP coding sequence, to generate ID-EGFP transcripts upon transfection into primary rat hippocampal neurons. *In situ* hybridization was performed on the transfected cells 48 hours post-transfection with probes targeted to the EGFP portion of the sequence. pEGFP-N1 transfected cells were used as a control for ID-independent RNA localization

Sequence elements used for the targeting and competition studies were as follows:

RefSeq ID	Symbol	Intron	Element_name	Sense Primer	Antisense Primer
NM_021739	CAMK2B	3	CAMK2Bi3ID1long	GGTTGGGTCGCTTGCAGAAG	CCTGAAGGCAATGCTCCAAG
NM_021739	CAMK2B	3	CAMK2Bi3ID1short	TGTTAAAAAATGGCATCCAC	TAGCGGATATGTGGACGCCA
NM_052804	FMR1	1	FMR1i1ID1long	TCTTTGTGTATTGCAATTTGGG	ATTTCCAAGGTGGGGTGTTA
NM_052804	FMR1	1	FMR1i1ID1short	CTTTAGAAATTTTTAAGCCG	TTTAAATTCTATGTAGGC
NM_024370	GABRG3	5	GABRG3i5ID2long	AGGACCCACTTCCTCATGTGCTGT	GGTGGCTCTGGGCTCTGTTT
NM_024370	GABRG3	5	GABRG3i5ID2short	CACATAAGAGACCATGACAC	CGGCTTGGGGTGTCATGTTC
NM_017241	GRIK1	1	GRIK1i1ID4long	TGGCCTCCAGGAAACTAGTA	ATGACCAAGCCCATCTAACA
NM_017241	GRIK1	1	GRIK1i1ID4short	AAGAAACTAAACTTGAGGGG	AAGCATTTCTGAAAATAGTC
NM_052804	FMR1	1	FMR1i1B2	TGAAACCTTGGGTTTGAATCAT	CATATCTAACTAGGGGTGAGGAAA

For the IGOR calculations, quantification paths are manually drawn tracing 3 dendrites of selected cells based on MAP2 immunostaining. Path origins were chosen at the somal end of the dendritic process. Generated paths are 11 pixels wide (4.4µm). The average signal intensity along the paths was computed for the *in situ* hybridization channel. These average intensities were normalized to the maximum signal along the path. The average of the normalized values was computed for each cell and then plotted against the distance from the path origin, using Graphpad Prism.

Plots of calculated *in situ* signal in dendrites are presented as $\Delta F/F$ – i.e., relative

difference in fluorescence signal in ID-EGFP fusion constructs compared to controls – against distance from the cell soma in microns. Signal from each ID-EGFP transfection experiment is binned to find average intensity values across 8μm distances. The differences from binned experimental signal values relative to control signal values are calculated to generate mean and SEM values to distances of 48um in each plot.

Computational sequence analysis: Pairwise NCBI BLAST (Altschul et al., 1990) was run on each pair of intronic sequences represented on the microarray using an e-value cutoff of 1e-10, and results were clustered (single-linkage) based on overlapping gene coordinates. Individual clusters were annotated for presence of repetitive elements, including the ID element, using RepeatMasker (Smit, 1996-2004).

Repeat-masked intron sequence was computationally folded using Vienna RNAplfold 1.7 (Bernhart et al., 2006; Bompfunewerer et al., 2008; Hofacker, 2003) using the default settings. Using a custom Python script, we segmented the base pairs into stable structures of length >50nts, such that at least one base pair had probability > 0.85 in the ensemble. The minimum free energy z-score was calculated on each hairpin candidate using the standard protocol (Clote et al., 2005) as implemented in http://clavius.bc.edu/clotelab/RNAdinucleotideShuffle/dinucleotideShuffle.html. The best MFEZ value per intron was extracted for subsequent statistical analysis.

Three miRNA classification programs were run using default parameters on stable (MFEZ < -2) hairpins from introns of interest -- CSHMM (Agarwal et al., 2010), MiPred (Jiang et al., 2007), and Triplet-SVM (Xue et al., 2005). RFAM covariance

models for known miRNAs were also used to score each candidate using the Infernal 1.0.2 cmsearch program (Gardner et al., 2009; Nawrocki et al., 2009), which returns an e-value and bit score for a query sequence against a given miRNA family model. Self containment index was calculated as described in (Lee and Kim, 2008) and indicates structural robustness on a [0.0, 1.0] scale, with a score of 0.9 or better typical of miRNA precursors.

Supplemental References

Agarwal, S., Vaz, C., Bhattacharya, A., and Srinivasan, A. (2010). Prediction of novel precursor miRNAs using a context-sensitive hidden Markov model (CSHMM). BMC Bioinformatics *11 Suppl 1*, S29.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Bernhart, S.H., Hofacker, I.L., and Stadler, P.F. (2006). Local RNA base pairing probabilities in large sequences. Bioinformatics *22*, 614-615.

Bompfunewerer, A.F., Backofen, R., Bernhart, S.H., Hertel, J., Hofacker, I.L., Stadler, P.F., and Will, S. (2008). Variations on RNA folding and alignment: lessons from Benasque. J Math Biol *56*, 129-144.

Clote, P., Ferre, F., Kranakis, E., and Krizanc, D. (2005). Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. RNA *11*, 578-591.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R., and Bateman, A. (2009). Rfam: updates to the RNA families database. Nucleic Acids Res *37*, D136-140.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E.*, et al.* (2004). Genome sequence of

the Brown Norway rat yields insights into mammalian evolution. Nature *428*, 493-521.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res *31*, 3429-3431.

Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X., and Lu, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res *35*, W339-344.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lee, M.T., and Kim, J. (2008). Self containment, a property of modular RNA structures, distinguishes microRNAs. PLoS Comput Biol *4*, e1000150.

Miyashiro, K., Dichter, M., and Eberwine, J. (1994). On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. Proceedings of the National Academy of Sciences of the United States of America *91*, 10800-10804.

Muslimov, I.A., Iacoangeli, A., Brosius, J., and Tiedge, H. (2006). Spatial codes in dendritic BC1 RNA. The Journal of cell biology *175*, 427-439.

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335-1337.

Smit, A., Hubley, R & Green, P (1996-2004). RepeatMasker Open-3.0. .

Xue, C., Li, F., He, T., Liu, G.P., Li, Y., and Zhang, X. (2005). Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics *6*, 310.

Supplemental Figure 1

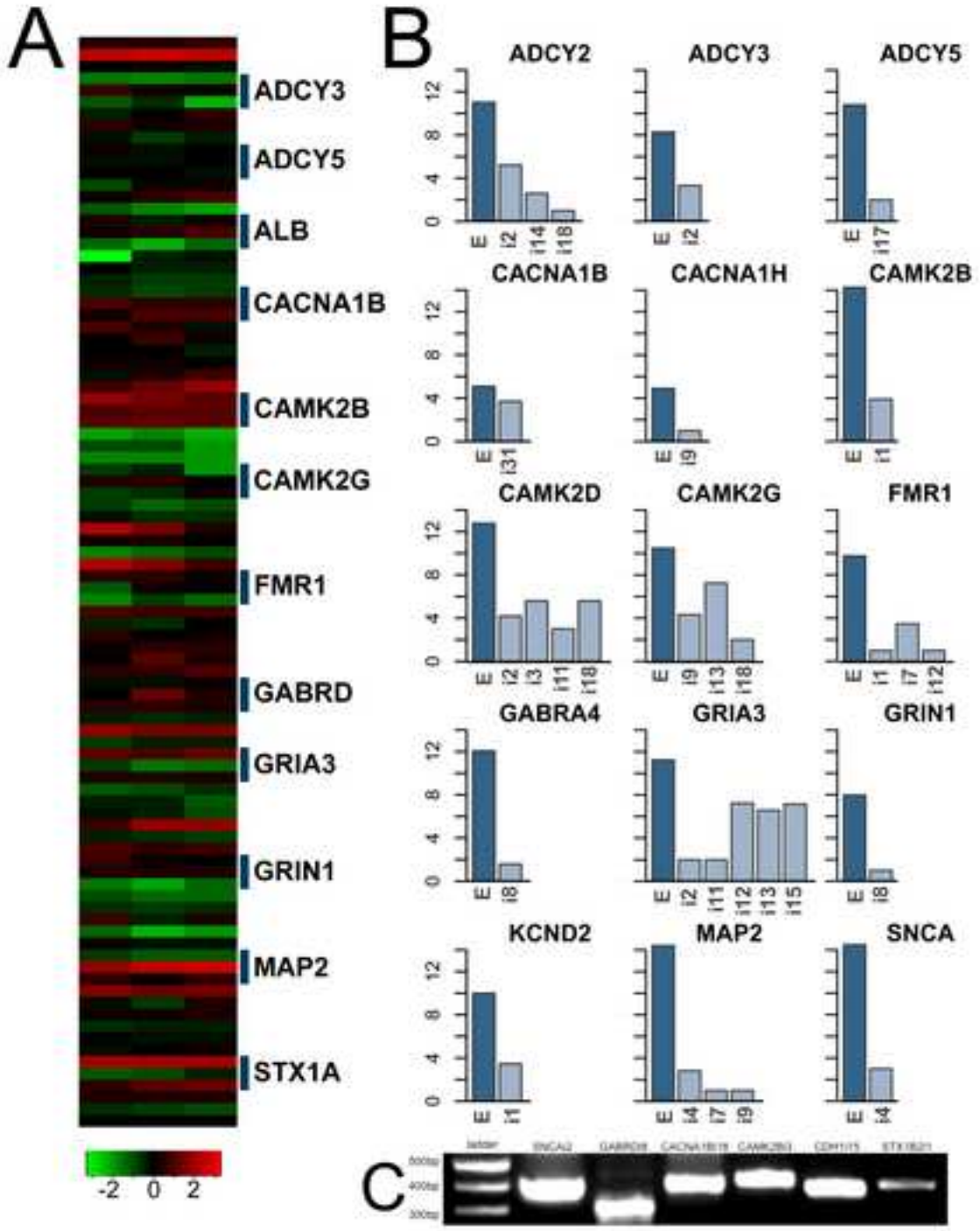

Figure S1

Supplemental Figure 2

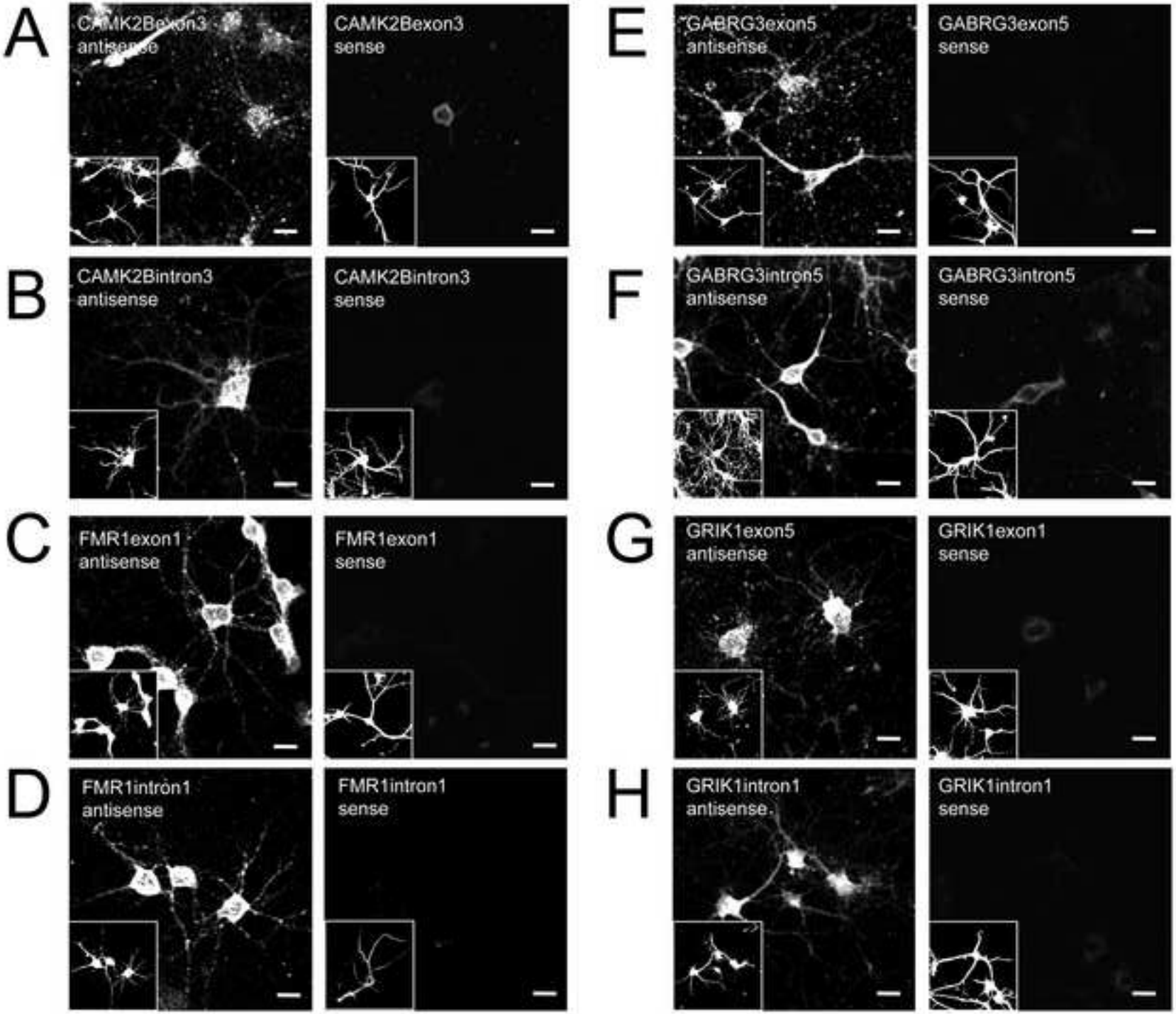

Figure S2

Supplemental Figure 3

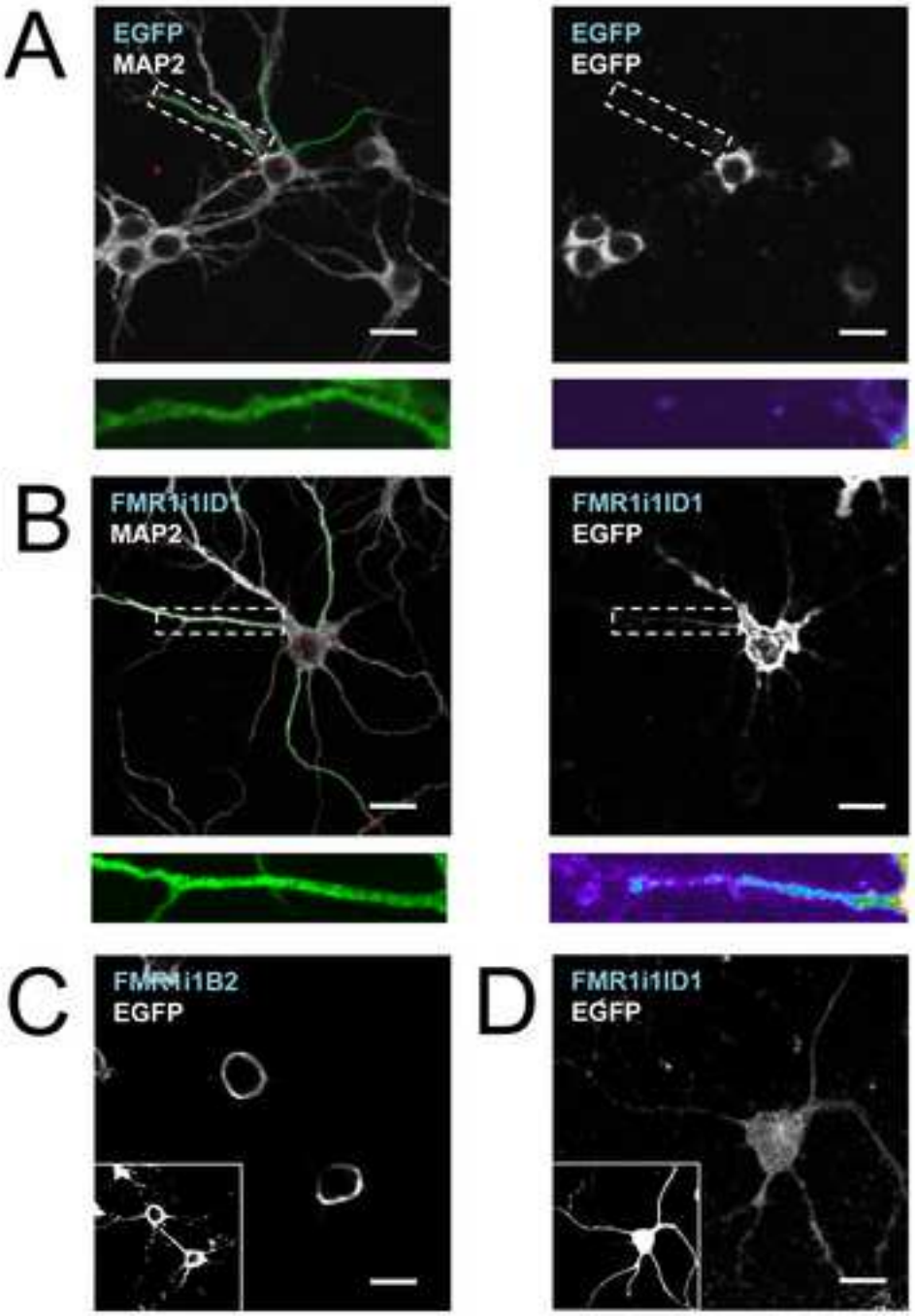

Figure S3

Supplemental Figure 4

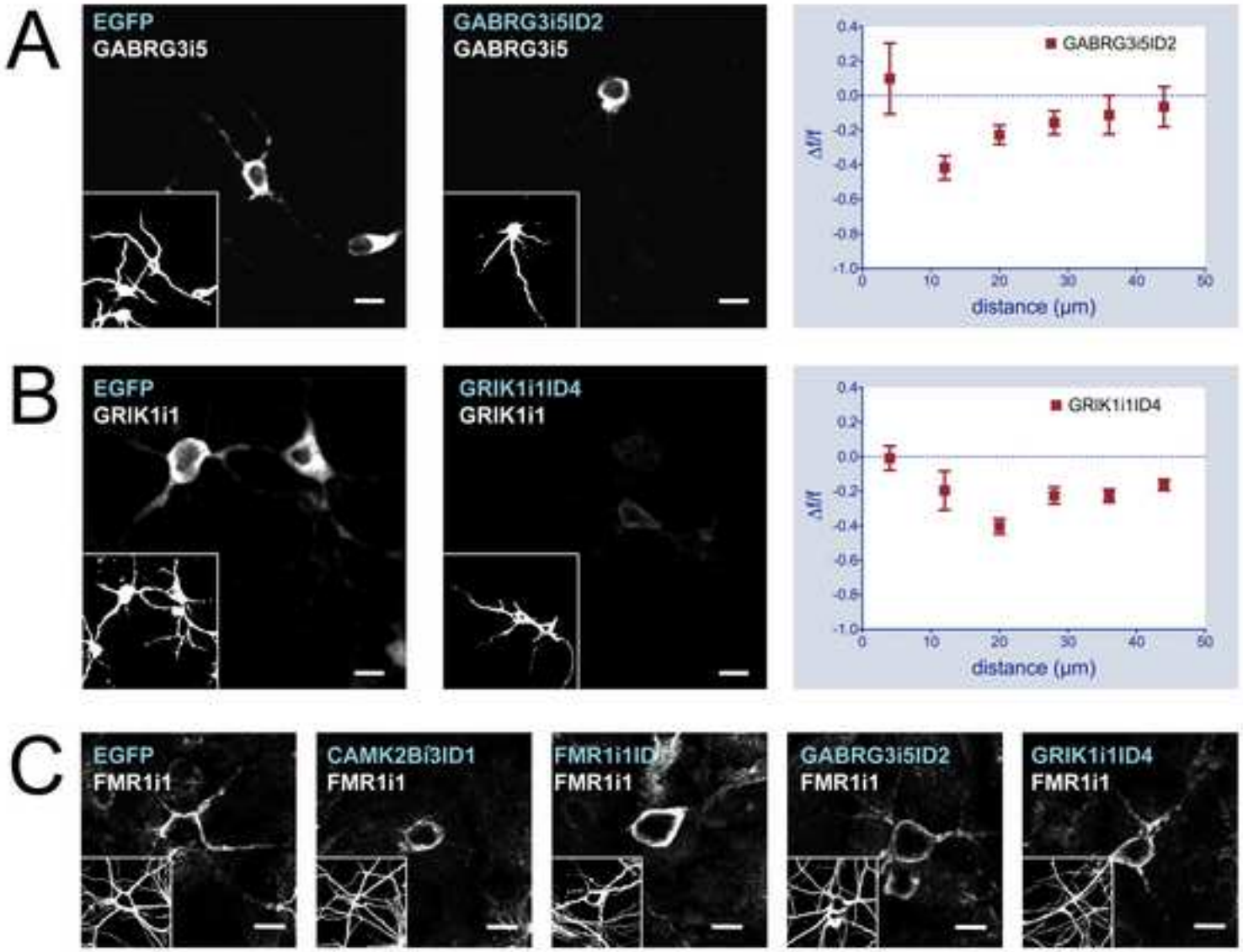

Figure S4