ONLINE SUPPLEMENT

Long noncoding RNAs altered temporally after focal cerebral ischemia display exon mimicry with protein-coding RNAs

Ashutosh Dharap, PhD, Venkata Prasuja Nakka, PhD and Raghu Vemuganti, PhD

Department of Neurological Surgery, University of Wisconsin, Madison, WI, USA

Detailed methods

Focal ischemia: Adult, male, spontaneously hypertensive rats (SHR; 280-320g; Charles River, Wilmington, MA) used in these studies were cared for in accordance with the *Guide for the Care and Use of Laboratory Animals*, U.S. Department of Health and Human Services Publication number 86-23 (revised 1986). The Research Animal Resources and Care Committee of the University of Wisconsin-Madison approved the surgical procedures. Transient MCAO was induced with an intraluminal suture. In brief, a rat was anesthetized with halothane, placed in a stereotaxic frame fitted with a nose cone with 2% isoflurane anesthesia. A craniotomy (4 mm in diameter, 2-4 mm lateral and 1-2 mm caudal to bregma) was performed with extreme care over the MCA territory using a trephine. The dura was left intact and a laser Doppler flow-meter probe (model PD-434; Vasamedics, LLC, St Paul, MN, USA) was placed on the surface of the ipsilateral cortex and fixed to the periosteum with a 4-0 silk suture. The probe was connected to a laser flowmeter device (Laserflow blood perfusion monitor BPM 403A; TSI Inc., St Paul, MN) for continuous monitoring of regional cerebral blood flow (rCBF). The left femoral artery was cannulated for continuous monitoring of arterial blood pressure and to obtain the measurements of pH, Pao2, Paco2, hemoglobin and blood glucose concentration (i-STAT; Sensor Devices, Waukesha, WI). The rectal temperature was controlled at 37.0 ± 0.5°C during surgery with a feedback-regulated heating pad. After a midline skin incision, the left external carotid artery (ECA) was exposed, and its branches were coagulated. A 3-0 surgical monofilament nylon suture, blunted at the end, was introduced into the ECA lumen and gently advanced to the internal carotid artery until rCBF was reduced to 10 to 16% of the baseline (recorded by laser Doppler flowmeter). After a 1h occlusion, the suture was withdrawn to restore the blood flow (confirmed by laser Doppler). Rats were euthanized at 3h, 6h and 12h of reperfusion. After suturing the wound, the rat was allowed to recover from anesthesia and returned to the cage with *ad libitum* access to food and water. During the MCAO, Pao2 (100 to 200 mm Hg) and Paco2 (30 to 40 mm Hg) were maintained at physiological levels. Sham-operated rats served as control.

lncRNA microarray: Rats subjected to transient MCAO were sacrificed at various reperfusion periods (3h, 6h and 12h; n = 3 at each time point). Three sham-operated rats served as control. Total RNA was extracted from the ipsilateral cortex of each rat using the mirVana total RNA extraction kit (Ambion USA). Sample labeling and array hybridization were performed according to the Agilent One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technology USA). Briefly, RNA from each sample was linearly amplified, labeled with Cy3-dCTP, purified by RNAeasy Mini Kit (Qiagen USA), fragmented and hybridized to an Arraystar lncRNA expression microarray.

Microarray data analysis: The arrays were scanned with the Agilent DNA Microarray Scanner and analyzed with the Agilent Feature Extraction software (version 10.5.1.1). The array quality was confirmed by checking the spot centroids at 4 corners of the array and the spatial distribution of the population and non-uniformity outliers distributed across the array. The level and the shape of the signal distribution were confirmed by negative control stats (average and SD of the net signals; mean signal-scanner offset and background-subtracted signals). The array quality was also confirmed by correcting for local background inliers and checking reproducibility statistics (percent coefficient of variation replicated probes). A transcript was considered detectable if the signal intensity was 3 times the maximal background signal and the spot coefficient of variation (SD/signal intensity) was <0.5. The expression data files obtained by the Agilent Feature Extraction Software were imported into the GeneSpring GX V11.0 software, data sets from different arrays were quantile normalized and the transcripts that obtained a present in all samples were chosen for further analysis. Differentially expressed transcripts were identified by fold-change screening with a threshold of ≥2-fold. Statistically significant differences between the groups were identified by the statistical measures built in the GeneSpring based on the *t*-test

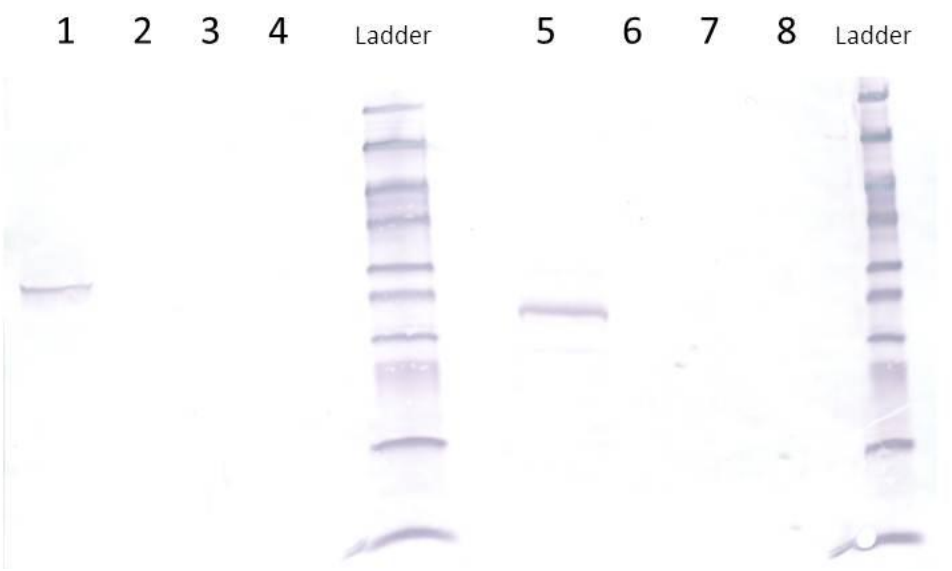
p-value value method with a high stringency (fold change cutoff of >2 and a probability value of <0.001 to decrease false-positives).

lncRNA sequence analysis: The annotated, full-length lncRNA sequences were obtained from the RefSeq database. For each sequence, a BLAT query was conducted in the UCSC genome browser against the rat Baylor 3.4/rn4 genome assembly to identify the genomic loci and sequence conservation with known genes. The browser is able to identify sequences that show homology to the exons of annotated protein-coding genes.

Promoter analysis: Promoters of the stroke-responsive lncRNAs that showed >90% sequence homology to protein-coding genes were analyzed by entering the genomic locus of each lncRNA into the UCSC genome browser and scanning the nearest upstream promoter on the same strand (within 10 kb from transcription start site [TSS]). Annotated promoters for the corresponding protein-coding genes were also obtained from the UCSC genome browser. In each case, a 1 Kb sequence (upstream from TSS) was analyzed for TF binding sites using the 'Common TFs' algorithm of the Genomatix Pattern Search & Analysis tool (Genomatix Software Suite GmbH) to identify TF matrices that were common to all the 48 lncRNA promoters scanned. The TF matrices that are common to paired lncRNA/protein-coding gene promoters were identified. Matrix scores were assigned based on conserved nucleotide matches of the query sequence to the defined matrix using a matrix similarity score of >0.75.

In vitro transcription and translation: Three lncRNAs (XR_005672 that showed 4 ORFs with homology to Eno1, XR_007365 that showed 3 ORFs with homology to Eef1a1 and XR_005605 that showed 2 ORFs with homology to RpL3) were selected for this experiment. These lncRNAs were selected based on the following criteria (1) >1kb in length (putative peptides are >10 Kda) and (2) show homology to separate classes of proteins. The DNA sequences corresponding to lncRNAs XR_005672, XR_007365 and XR_005605 were cloned into the pIDTBlue plasmid (Integrated DNA Technologies USA) downstream to a T7 promoter. The plasmids were transcribed using the T7 mMessage Machine kit (Ambion USA) and ~5 μg RNA was used for *in vitro* translation using the Retic Lysate IVT kit (Ambion USA). We also used an alternate method of coupled transcription/translation reactions directly from ~400 ng circular or linearized plasmids using the T7 TNT Quick Coupled reticulocyte lysate system (Promega USA). The translation reactions were supplemented with biotinylated tRNA complexes containing lysine. After the translation reaction, 10 μl aliquot was electrophoresed on a 4-20% SDS-PAGE gel, blotted onto nitrocellulose and the protein bands were visualized with Transcend non-radioactive detection system (Promega USA) that detects the biotinylated lysine residues. We also conducted coupled transcription/translation reactions directly from the plasmids. Multiple reactions were conducted to include linearized as well as circular plasmids, tweaked conditions for capped and uncapped RNA and varying incubation durations. Although RNA was successfully transcribed from all 3 lncRNA plasmids, neither produced any detectable protein/peptide bands at any of the predicted sizes (based on ORFs). Since RNA 5' capping can influence translation, we adjusted the reactions for capped or uncapped RNA, but still the RNAs derived from lncRNA genes failed to translate any protein products. Both the positive controls (luciferase plasmid and pTRI-Xef RNA) showed robust protein bands indicating successful translation reactions.

Supplementary Figure 1: Immunoblot showing the protein products of the *in vitro* translation reactions



Three lncRNAs (XR_005672 that showed 4 ORFs with homology to Eno1, XR_007365 that showed 3 ORFs with homology to Eef1a1 and XR_005605 that showed 2 ORFs with homology to RpL3) were translated *in vitro* and blotted onto a nitrocellulose membrane and the biotinylated lysine residues incorporated into proteins were detected by using a biotin/streptavidin colorimetric system. The samples from lanes 1-4 were translated directly from the plasmid using a coupled transcription/translation system. The samples from lanes 5-8 were translated from RNA that was synthesized from the plasmids using an *in vitro* transcription system. Lanes: 1 - positive control luciferase plasmid; 2 – XR_005672; 3 – XR_007365; 4 – XR_005605; 5 – Positive control pTRI-Xef RNA; 6 – XR_005672; 7 – XR_007365; 8 – XR_005605. Although the positive controls showed detectable bands, no bands were observed for the 3 lncRNAs tested.

Supplementary Figure 2: Graphical representation of the lncRNA ORFs and the ORFs of the corresponding homologous protein-coding genes.

The ORFs for each transcript were obtained from the NCBI ORF Finder online tool. The putative peptides from the lncRNAs were queried against the NCBI protein database using BLAST. The green bars represent those ORFs in the lncRNA that encode peptides with homology to the corresponding annotated protein. The remaining ORFs did not show homology to any known protein. The start and stop codons for each ORF are presented in bold.

Supplementary Table 1: lncRNA expression profile in normal rat cortex

Hybridization signal	Expression level	# of transcripts	% of transcripts
<100	Negligible	2,729	32.8
101 to 1,000	Low	1,419	17.1
1,001 to 5,000	Low to medium	670	8.1
5,001 to 20,000	Medium	529	8.4
20,001 to 50,000	Medium to high	289	6.4
50,001 to 100,000	High	124	1.5
100,001 to 670,000	Very high	98	1.1

Each value is the average of n = 3 lncRNA arrays (3 different rats). The total number of lncRNAs represented on the lncRNA microarray is 8,314. Of those, 5,858 obtained a present call in all 3 arrays. Between samples, the variation in the number of lncRNAs expressed was <10%. The expression level (hybridization signal) of individual lncRNAs was within 20% between different samples.

Supplementary Table 2: Other lncRNAs altered after stroke

lncRNA	3h	6h	12h	Chr	lncRNA	3h	6h	12h	Chr
XR_007499	9.3	6.5	9.6	13	uc.88-	0.4	0.4	0.5	3
XR_007501	8.1	14.8	13.9	17	MRAK013677	0.4	0.4	0.4	4
MRAK154943	7.8	12.9	10.7	M	uc.426+	0.4	0.6	0.5	18
XR_007384	7.7	12.1	14.3	5	MRAK078930	0.4	0.4	0.4	1
XR_006772	7.6	13.4	17.0	1	MRAK086453	0.4	0.5	0.5	7
MRAK160242	7.3	6.9	7.5	9	MRAK048867	0.4	0.4	0.5	13
AJ131848	7.2	5.0	3.3	20	MRAK008871	0.4	0.4	0.6	6
MRAK139013	7.1	11.9	10.8	4	XR_005532	0.4	0.4	0.5	3
XR_005800	7.1	15.8	15.4	11	MRAK044009	0.4	0.5	0.5	9
S77494	6.9	5.0	6.6	18	AJ646872	0.5	0.6	0.4	2*
AF034247	6.9	18.7	19.7	2	MRAK162515	0.5	0.5	0.5	3
XR_005722	6.8	13.4	16.1	1	MRAK037342	0.5	0.5	0.4	11
MRuc008ymd	6.4	10.5	7.9	14	XR_009186	0.5	0.4	0.3	7
MRAK132400	6.4	13.4	13.5	15	L20992	0.5	0.5	0.5	15
XR_007539	6.4	14.6	13.2	5	MRAK157642	0.5	0.4	0.3	4
XR_008814	6.2	22.7	12.6	7	MRAK046990	0.5	0.4	0.4	9
XR_005592	6.2	16.5	8.8	16	XR_008822	0.5	0.5	0.5	13
MRuc008qku	6.1	10.0	8.8	2	MRAK090145	0.5	0.5	0.5	12
MRuc007ifd	6.0	3.3	6.2	14	AB190259	0.5	0.4	0.4	16
MRAK032862	5.9	9.4	10.4	8	MRAK081202	0.5	0.5	0.3	13
XR_005904	5.8	12.9	13.7	16	BC158785	0.5	0.5	0.4	16
MRAK084456	5.7	11.1	9.1	16	MRBC062110	0.5	0.4	0.3	1
XR_005821	5.7	11.8	10.3	X	XR_007475	0.5	0.4	0.5	8
MRAK162509	5.6	5.2	4.1	3	AF468011	0.5	0.4	0.4	2
MRAK137543	5.6	10.0	9.7	6	MRAK048764	0.5	0.4	0.4	14
XR_009002	8.0	19.1	15.5	7	MRuc009hnl	0.4	0.4	0.3	1
XR_007148	7.5	25.2	17.4	17	MRAK042330	0.4	0.3	0.4	20
XR_007625	7.5	20.6	20.8	4	uc.212+	0.4	0.3	0.3	4
XR_006737	7.3	25.7	24.3	1	BC091360	0.4	0.2	0.1	1
MRAK081997	7.1	23.4	16.5	2	L11024	0.4	0.3	0.3	15
MRAK005097	9.3	19.4	14.8	15	MRAK045028	0.3	0.5	0.4	7
XR_006504	8.8	28.8	23.6	19	U78146	0.4	0.4	0.4	10
XR_006541	8.4	23.3	19.9	9	MRAK088349	0.4	0.3	0.5	X
XR_008674	8.4	16.5	13.2	5	uc.18-	0.4	0.4	0.6	5
XR_007247	8.3	18.0	17.8	15	uc.226+	0.4	0.3	0.3	4

We presented the data for the 30 lncRNAs altered maximally after focal ischemia in Table 1. This Supplementary Table 2 shows 70 more lncRNAs that were also altered significantly at all 3 time points of reperfusion after transient MCAO evaluated compared to sham. The values given are mean fold increases in the ischemic groups over sham group that are statistically significant (≥2-fold, $p<0.05$;

ANOVA). 3h, 6h and 12h are the reperfusion time points after a 1h transient MCAO. For each group, 3 RNA samples were hybridized to 3 microarrays. For each time point, 3 RNA samples were hybridized to 3 microarrays. The lncRNA IDs are listed according to the Refseq annotations. The chromosome number (Chr) indicates the chromosome from which the lncRNA is transcribed. The symbol * refers to Chr2_random in the UCSC genome browser, which indicates a sequence on chromosome 2 that is not in a finished state but whose location on the chromosome is known.

Supplementary Table 3: Stroke-induced lncRNAs that show exon mimicry with protein-coding genes

LNC ID	Mimicked Gene	Chr# of lncRNA gene	Chr# of mRNA gene	Sequence Similarity (%)
XR_009313	Hsp90aa1	1	6	92.8
XR_006073	Kpna2	6	10	93.5
XR_009002	Srsf7	7	6	95.2
XR_007499	Tuba1 (a & c)	13	7	92.6
XR_006222	Rp17	X	5	92.5
XR_006541	Hnrnpa3	9	3	94.7
XR_007384	RpL21	5	12	96.3
XR_006772	Rps8	1	5	90.1
XR_007148	Ppp2ca	17	10	90.9
XR_005722	Actg1	1	10	96.1
XR_007539	Hnrpa1	5	7	95.3
XR_005592	Ppia	16	14	95.8
XR_005821	Ppid	X	2	95.6
XR_005972	Tipr1	5	13	92.7
XR_006599	Rbms2	18	7	93.9
XR_007647	Hnrpa1	X	7	94.8
XR_005513	Rps12	7	1	92.5
XR_007206	Rps7	8	6	91.5
XR_006232	RpL23a	6	10	92.8
XR_005605	RpL3	X	7	92.7
XR_006148	RpL9	2	14	93.2
XR_008505	Spcs2	4	1	94.4
XR_009130	RpL23a	3	10	91.7
XR_005754	Rps7	16	6	90.6
XR_006375	BTF3	X	2	94.0
XR_008876	RpL27a	1	1	95.9
MRAK034983	Sav1	4	6	96.6
XR_009536	Pgk1	11	X	91.9
XR_006528	Rps2	13	10	93.1
XR_007963	RpL31	7	9	96.0
XR_007321	Rps27a	18	14	92.2
XR_009459	Ppid	10	2	95.8
XR_007355	Rps27a	16	14	94.3
XR_005604	RpL15	11	15	94.0
XR_005515	RpL15	X	15	90.8
XR_009094	Nme2	15	10	91.5
XR_006647	Pomp	6	12	91.7
XR_005917	RpL24	15	11	93.7

XR_008103	H3f3b	2	13	92.6
XR_007270	RpL10	9	X	93.5
XR_006417	Chmp5	12	5	95.7
XR_006271	RpL31	4	9	94.2
XR_006345	Actr2	14	14	95.5
XR_005813	Ywhaz	X	7	92.0
XR_007302	Tceb2	11	10	90.4
XR_008016	RpL12	8	3	91.8
XR_005647	Rps10	2	20	94.5
XR_006948	Rp16	1	12	92.4
XR_008143	Prdx2	1	19	93.2
XR_007404	Rp19	3	14	94.9
XR_006373	Pmpcb	6	4	91.8
XR_005936	Hsp90aa1	13	6	94.6
XR_005957	Rps9	X	1	92.9
XR_008555	Rps24	14	16	95.5
XR_005823	Rps2	1	10	90.2
XR_005672	Eno1	8	5	92.2
XR_008939	RpL27	12	10	93.4
XR_006550	Rps6	4	5	92.9
XR_006417	Chmp5	12	5	95.7
XR_005821	Ppid	X	2	95.6
XR_008555	Rps24	14	16	95.5
XR_006345	Actr2	14	14	95.5
XR_007365	eEF1A1	2	8	95.4
XR_007539	Hnrpa1	5	7	95.3

Table 2 of the manuscript lists the lncRNAs with highest percent sequence similarity to the protein-coding genes. This supplementary table lists other lncRNAs that also showed significant exon mimicry. The exon mimicry was determined using the BLAT program of the UCSC genome browser. All lncRNAs in this table showed >90% sequence similarity to the exons of the mimicked genes.

Supplementary Table 4: Stroke-induced lncRNAs showing common exon mimicry targets

	Fold Change						
ProbeName	**3h**	**6h**	**12h**	**Chr#**	**Mimicked Gene**	**Sequence Similarity (%)**	**Chr# of Mimicked Gene**
XR_005604	3.8	7.4	6.2	chr11	RpL15	94.0	15
XR_005515	3.6	7.4	6.2	chrX	RpL15	90.8	15
XR_007321	4.0	8.7	7.8	chr18	Rps27a	92.2	14
XR_007355	3.8	7.7	6.2	chr16	Rps27a	94.3	14
XR_009083	10.8	22.1	21.9	chr14	RpL21	95.9	12
XR_007384	7.7	12.1	14.3	chr5	RpL21	96.3	12
XR_006232	4.9	9.4	10.1	chr6	RpL23a	92.8	10
XR_009130	4.6	10.7	10.9	chr3	RpL23a	91.7	10
XR_006528	4.3	10.3	8.4	chr13	Rps2	93.1	10
XR_005823	2.3	3.9	3.5	chr1	Rps2	90.2	10
XR_007963	4.2	6.8	7.1	chr7	RpL31	96.0	9
XR_006271	3.0	4.2	4.7	chr4	RpL31	94.2	9
XR_007539	6.4	14.6	13.2	chr5	Hnrpa1	95.3	7
XR_007647	5.1	12.3	9.9	chrX	Hnrpa1	94.8	7
XR_009313	25.2	53.7	45.7	chr1	Hspca/Hsp90aa1	92.8	6
XR_005936	2.5	2.0	2.4	chr13	Hspca/Hsp90aa1	94.6	6
XR_007206	5.1	10.4	9.5	chr8	Rps7	91.5	6
XR_005754	4.6	9.3	7.7	chr16	Rps7	90.6	6
XR_005821	5.7	11.8	10.3	chrX	Ppid	95.6	2
XR_009459	3.9	7.2	7.3	chr10	Ppid	95.8	2

Pairs of lncRNAs showed homology to the same protein-coding genes. Sequence comparisons were derived using the BLAT program of the UCSC genome browser. The expression patterns for most of these lncRNA pairs were similar at all 3 time points of reperfusion indicating that their expression might be controlled by common regulatory elements. A majority of the mimicked protein-coding genes code for ribosomal proteins.

Table 5: Second tier of the TF families that showed binding sites in the lncRNA and protein-coding gene promoters

Family	# of lncRNA promoters bound	# of sites in the lncRNA promoters	# of protein-coding gene promoters overrepresented
ABDB	45	181	14
GATA	46	174	12
IRFF	45	113	12
EVI1	46	192	10

The top 6 TF families in this category were shown in Table 3 and the next 4 are shown in this Supplementary Table.

Supplementary Table 6: Full list of HOMF and ETSF Transcription Factors

Human and Murine ETS1 factors (ETSF)	Homeodomain Transcription Factors (HOMF)
c-Ets-1(p54)	BarH-like homeobox 1
Ets homologous factor	BarH-like homeobox 2
Ets - family member ELF-2 (NERF1a)	BARX homeobox 1
E74-like factor 3 (epithelial-specific ets domain transcription factor)	BARX homeobox 2
E74-like factor 4	Barx2, homeobox transcription factor that preferentially binds to paired TAAT motifs
E74-like factor 5	Brain specific homeobox
ETS-like gene 1 (ELK-1)	Hematopoietically expressed homeobox, proline-rich homeodomain protein
ETS-domain protein (SRF accessory protein 1)	H6 family homeobox 1 / NKX5-3
ETS-domain protein (SRF accessory protein 2)	H6 homeodomain HMX3/Nkx5.1 transcription factor
v-ets erythroblastosis virus E26 oncogene homolog	Hmx3/Nkx5-1 homeodomain transcription factor
c-Ets-1 binding site	Hmx2/Nkx5-2 homeodomain transcription factor
c-Ets-2 binding site	H6 homeodomain HMX3/Nkx5.1 transcription factor
Ets variant 1	Ladybird homeobox 2
Ets variant 3	Homeodomain proteins MSX-1 and MSX-2
Ets variant 4	Muscle-segment homeobox 1, msh homeobox 1
Ets variant 5	Muscle segment homeo box 2, homologue of Drosophila (HOX 8)
Ets variant 6	Muscle-segment homeobox 2, msh homeobox 2
ETS family member FLI	Homeobox, msh-like 3
GABP: GA binding protein	Homeobox containing germ cell-specific transcription factor NOBOX
GA binding protein transcription factor, alpha	NOBOX oogenesis homeobox
GA repeat binding protein, beta 1)	T-cell leukemia homeobox 1
Nuclear respiratory factor 2	T-cell leukemia, homeobox 2
Prostate-derived Ets factor	--
SAM pointed domain containing ets transcription factor	--
SPI-1 proto-oncogene; hematopoietic transcription factor PU.1	--
Spi-B transcription factor (Spi-1/PU.1 related)	--
Spi-C transcription factor (Spi-1/PU.1 related)	--

The transcription factors that comprise the HOMF and ETSF matrices were used to compile the binding scores for each family on the lncRNA and protein-coding gene promoters.

Supplementary Table 7: Full names and summaries of the protein-coding genes that are mimicked by the lncRNAs

Symbol	Official Full Name	Summary
Actg1	actin, gamma 1	The beta and gamma actins co-exist in most cell types as components of the cytoskeleton, and as mediators of internal cell motility. A cytoplasmic actin found in non-muscle cells. Mutations in this gene are associated with DFNA20/26, a subtype of autosomal dominant non-syndromic sensorineural progressive hearing loss.
Actr2	ARP2 actin-related protein 2 homolog (yeast)	The specific function of this gene has not yet been determined; however, the protein it encodes is known to be a major constituent of the ARP2/3 complex. This complex is located at the cell surface and is essential to cell shape and motility through lamellipodial actin assembly and protrusion.
Btf3	basic transcription factor 3	This protein forms a stable complex with RNA polymerase IIB and is required for transcriptional initiation. This gene has multiple pseudogenes.
Chmp5	charged multivesicular body protein 5	CHMP5 belongs to the chromatin-modifying protein/charged multivesicular body protein (CHMP) family. These proteins are components of ESCRT-III (endosomal sorting complex required for transport III), a complex involved in degradation of surface receptor proteins and formation of endocytic multivesicular bodies (MVBs).
Eef1a1	eukaryotic translation elongation factor 1 alpha 1	A member of the EF-1 alpha family that tethers aminoacyl-tRNA to the ribosome during peptide synthesis.
Eno1	enolase 1, (alpha)	This gene encodes alpha-enolase, one of three enolase isoenzymes found in mammals. Each isoenzyme is a homodimer composed of 2 alpha, 2 gamma, or 2 beta subunits, and functions as a glycolytic enzyme. Alpha-enolase in addition, functions as a structural lens protein (tau-crystallin) in the monomeric form. Alternative splicing of this gene results in a shorter isoform that has been shown to bind to the c-myc promoter and function as a tumor suppressor. Several pseudogenes have been identified, including one on the long arm of chromosome 1.
H3f3b	H3 histone, family 3B	histone that may be involved in brain development.
Hnrnpa3	heterogeneous nuclear ribonucleoprotein A3	An mRNA binding protein involved in cytoplasmic trafficking of messages to specific sites in the cell.
Hnrpa1	heterogeneous nuclear ribonucleoprotein A1	Involved in the packaging of pre-mRNA into heterologous nuclear ribonucleoprotein particles and in the transport of poly-A mRNA from the nucleus to the cytoplasm; may modulate splice site selection.
Hspca	heat shock protein 90, alpha (cytosolic), class A	Molecular chaperone; involved in sequestering damaged proteins and in ATP-dependent folding of proteins.

	member 1	
Kpna2	karyopherin alpha 2	receptor which binds to Glut2; participates in nuclear import.
Nme2	non-metastatic cells 2, protein	kinase involved in synthesis of nucleoside triphosphates.
Pgk1	phosphoglycerate kinase 1	kinase enzyme that is important for phosphoprotein glycolysis.
Pmpcb	peptidase (mitochondrial processing) beta	A beta subunit of mitochondrial processing peptidase, which catalyzes cleavage of the amino terminal leader peptide of nuclear encoded mitochondrial proteins.
Pomp	proteasome maturation protein	The protein encoded by this gene is a molecular chaperone that binds 20S preproteasome components and is essential for 20S proteasome formation. The 20S proteasome is the proteolytically active component of the 26S proteasome complex. The encoded protein is degraded before the maturation of the 20S proteasome is complete. A variant in the 5' UTR of this gene has been associated with KLICK syndrome, a rare skin disorder.
Ppia	peptidylprolyl isomerase A (cyclophilin A)	May play a role in protein folding or intracellular protein transport.
Ppid	peptidylprolyl isomerase D	This gene encodes a 370 aa protein that is found in the cytoplasm.
Ppp2ca	protein phosphatase 2, catalytic subunit, alpha isoform	Catalyzes the removal of serine- and threonine- bound phosphate groups.
Prdx2	peroxiredoxin 2	A protective enzyme against oxidative damage by cellular reactive sulfur species.
Rbms2	RNA binding motif, single stranded interacting protein 2	The protein encoded by this gene is a member of a small family of proteins which bind single stranded DNA/RNA. The RBMS proteins have been implicated in such diverse functions as DNA replication, gene transcription, cell cycle progression and apoptosis.
Rp17	ribosomal protein L7	A component of the 60s subunit of the ribosome; important for protein synthesis.
RpL10	ribosomal protein L10	A 60S ribosomal subunit protein.
RpL12	ribosomal protein L12	A 60S ribosomal subunit protein.
RpL15	ribosomal protein L15	60S ribosomal subunit protein.
RpL19	ribosomal protein L19	A ribosomal protein.
RpL21	ribosomal protein L21	A ribosomal subunit protein.
RpL23a	ribosomal protein L23a	60 S ribosomal subunit protein.
RpL24	ribosomal protein L24	60S ribosomal subunit protein.
RpL27	ribosomal protein L27	A ribosomal protein.

RpL27a	ribosomal protein L27a	60 S ribosomal subunit protein.
RpL3	ribosomal protein L3	May act as a structural constituent of the ribosome.
RpL31	ribosomal protein L31	A ribosomal protein subunit.
RpL9	ribosomal protein L9	A ribosomal protein.
Rps10	ribosomal protein S10	A ribosomal protein.
Rps12	ribosomal protein S12	40 S ribosomal subunit protein.
Rps16	ribosomal protein S16	A ribosomal subunit protein.
Rps2	Rps2	40 S ribosomal subunit protein.
Rps24	ribosomal protein S24	A ribosomal protein subunit.
Rps27a	ribosomal protein S27a	A ribosomal protein subunit.
Rps6	ribosomal protein S6	A ribosomal protein with many phosphorylated seryl residues.
Rps7	ribosomal protein S7	40 S ribosomal subunit protein.
Rps8	ribosomal protein S8	A ribosomal subunit protein.
Sav1	salvador homolog 1 (Drosophila)	WW domain-containing proteins are found in all eukaryotes and play an important role in the regulation of a wide variety of cellular functions such as protein degradation, transcription, and RNA splicing. This gene encodes a protein which contains 2 WW domains and a coiled-coil region.
Spcs2	signal peptidase complex subunit 2 homolog (S. cerevisiae)	No official summary available on NCBI. UniProt summary: Component of the microsomal signal peptidase complex which removes signal peptides from nascent proteins as they are translocated into the lumen of the endoplasmic reticulum.
Srsf7	serine/arginine-rich splicing factor 7	The protein encoded by this gene is a member of the serine/arginine (SR)-rich family of pre-mRNA splicing factors, which constitute part of the spliceosome. The RS domain is rich in serine and arginine residues and facilitates interaction between different SR splicing factors. In addition to being critical for mRNA splicing, the SR proteins have also been shown to be involved in mRNA export from the nucleus and in translation.
Tceb2	transcription elongation factor B (SIII), polypeptide 2	A positive regulatory subunit of the general transcription factor SIII heterotrimer; acts to increast the rate of RNA polymerase II transcription; regulates RNA polymerase II elongation complex activity.
Tiprl	TIP41, TOR	TIPRL is an inhibitory regulator of protein phosphatase-2A (PP2A).

	signaling pathway regulator-like (S. cerevisiae)	
Tuba1 (a & c)	tubulin, alpha 1A	Forms heterodimers with beta-tubulin; acts as a structural component of the microtubule cytoskeleton; plays a role in microtubule-based processes.
Ywhaz	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	A 30kDa component of the mitochondrial import stimulation factor, a protein complex that facilitates the import of in vitro synthesized precursor proteins into mitochondria.

The symbols, official full names and summaries were obtained from the NCBI database. If the summary for a gene was unavailable in the rat database, the summary of the human gene homolog was retrieved.

Supplementary Table 8: Conserved ORFs in the lncRNA sequences that show exon mimicry to protein-coding genes

LncRNA	Protein-Coding Gene	Conserved ORF
XR_007365	Eef1a1	64-381 Frame +1
		514-1107 Frame +1
		1274-1453 Frame +2
XR_006073	Kpna2	138-458 Frame +3
		862-1296 Frame +1
		1247-1426 Frame +2
XR_009083	RpL21	156-491 Frame +3
XR_007499	Tuba1 (a & c)	1-657 Frame +1
		651-1394 Frame +3
XR_006222	Rp17	1-363 Frame +1
		385-714 Frame +1
XR_006541	Hnrnpa3	190-456 Frame +1
XR_009002	Srsf7	106-657 Frame +1
XR_007384	RpL21	47-343 Frame +2
		394-528 Frame +1
XR_006772	Rps8	No conserved
XR_007148	Ppp2ca	323-616 Frame +2
		579-890 Frame +3
XR_005722	Actg1	2 56-1024 Frame +2
		894-1154 Frame -3
		961-1176 Frame +1
XR_007539	Hnrpa1	190-450 Frame +1
XR_005592	Ppia	40-498 Frame +1
		344-532 Frame +2
XR_005821	Ppid	1-225 Frame +1
		241-750 Frame +1
		855-1058 Frame +3
XR_005972	Tiprl	524-2218 Frame +2
XR_006599	Rbms2	1-465 Frame +1
		347-673 Frame +2
XR_007647	Hnrpa1	142-606 Frame +1
XR_005513	Rps12	1-150 Frame +1
XR_007206	Rps7	1 1-453 Frame +1
XR_006232	RpL23a	17-454 Frame +2
XR_005605	RpL3	29-655 Frame +2
		657-935 Frame +3
XR_006148	RpL9	1-345 Frame +1
XR_007963	RpL31	27-410 Frame +3
XR_007321	Rps27a	1-345 Frame +1
XR_009459	Ppid	1-234 Frame +1
		391-945 Frame +1
		1008-1262 Frame +3
XR_007355	Rps27a	23-232 Frame +2
		226-462 Frame +1
XR_005604	RpL15	28-516 Frame +1
		504-638 Frame +3
XR_005515	RpL15	78-323 Frame +3
		530-697 Frame +2
XR_009094	Nme2	125-472 Frame +2
XR_006647	Pomp	1-471 Frame +1
		167-424 Frame +2
XR_005917	RpL24	1-345 Frame +1
XR_008103	H3f3b	49-540 Frame +1
XR_007270	RpL10	14-640 Frame +2
XR_006417	Chmp5	155-466 Frame +2
XR_006271	RpL31	156-374 Frame +3
XR_006345	Actr2	1-444 Frame +1
		456-617 Frame +3
		463-1167 Frame +1
XR_005813	Ywhaz	1-798 Frame +1
		746-1009 Frame +2
		1115-1228 Frame +2
XR_007302	Tceb2	509-655 Frame +2
XR_008016	RpL12	1-456 Frame +1
XR_005647	Rps10	135-413 Frame +3
XR_006948	Rps16	2-511 Frame +2
XR_008143	Prdx2	1-129 Frame +1
XR_007404	RpL9	1-159 Frame +1
		199-576 Frame +1
XR_006373	Pmpcb	609-755 Frame +3
		1024-1554 Frame +1
XR_005936	Hspca	351-512 Frame +3
		533-898 Frame +2
		1103-1765 Frame +2

		342-577 Frame +3			1820-2191 Frame +2
XR_008505	Spcs2	119-256 Frame +2	XR_005957	Rps9	440-583 Frame +2
		335-640 Frame +2	XR_008555	Rps24	29-439 Frame +2
XR_009130	RpL23a	216-335 Frame -3	XR_005823	Rps2	1-309 Frame +1
XR_005754	Rps7	3-209 Frame +3			575-874 Frame +2
XR_006375	Btf3	158-475 Frame +2	XR_009394	RpL19	26-235 Frame +2
XR_008876	RpL27a	1-315 Frame +1			251-616 Frame +2
MRAK034983	Sav1	No conserved	XR_005672	Eno1	1-354 Frame +1
XR_009536	Pgk1	19-330 Frame +1			351-512 Frame +3
		514-765 Frame +1			590-859 Frame +2
		793-1236 Frame +1			678-1286 Frame +3
XR_006528	Rps2	1-279 Frame +1	XR_008939	RpL27	1-351 Frame +1
		300-413 Frame +3	XR_006550	Rps6	1-408 Frame +1
		608-733 Frame +2			386-745 Frame +2

The lncRNA sequences were analyzed using the NCBI ORF scanner to locate putative ORFs. Each lncRNA contained multiple putative ORFs. Each ORF was further analyzed using BLAST against the NCBI protein database to identify those that show conservation with annotated protein sequences. 97% of the lncRNAs that showed exon mimicry displayed ≥1 ORFs that are conserved with the corresponding protein-coding genes.