LINC00507 Is Specifically Expressed in the Primate Cortex and Has Age-Dependent Expression Patterns

Mills, J.D.; Ward, M.; Chen, B.J.; Iyer, A.M.; Aronica, E.; Janitz, M.

DOI
10.1007/s12031-016-0745-4

Publication date
2016

Document Version
Final published version

Published in
Journal of molecular neuroscience

J Mol Neurosci (2016) 59:431–439
DOI 10.1007/s12031-016-0745-4

LINC00507 Is Specifically Expressed in the Primate Cortex and Has Age-Dependent Expression Patterns

James D. Mills[1,3] · **Melanie Ward**[1] · **Bei Jun Chen**[1] · **Anand M. Iyer**[2] · **Eleonora Aronica**[2] · **Michael Janitz**[1]

Received: 19 February 2016 / Accepted: 22 March 2016 / Published online: 8 April 2016

Abstract Over the past decade, there has been an increase in the appreciation of the role of non-coding RNA in the development of organism phenotype. It is possible to divide the non-coding elements of the transcriptome into three categories: short non-coding RNAs, circular RNAs and long non-coding RNAs. Long non-coding RNAs are those transcripts that are greater than 200 nts in length and lack any significant open reading frames that produce proteins greater then 100 amino acids. Long intervening non-coding RNAs (lincRNAs) are a subclass of long non-coding RNAs. In contrast to protein coding RNAs, lincRNAs are expressed in a more tissue- and species-specific manner. In particular, many lincRNAs are only conserved amongst higher primates. This coupled with the propensity of many lincRNAs to be expressed in the brain, suggests that they are in fact one of the major drivers of organism complexity. We analysed 39 lincRNAs that are expressed in the frontal cortex and identified *LINC00507* as being expressed in a cortex-specific manner in non-human primates and humans. The expression patterns of *LINC00507* appear to be age-dependent, suggesting it may be involved in brain development of higher primates. Moreover, the analysis of *LINC00507* potential to bind ribosomes revealed that this previously identified non-coding transcript may harbour a micropeptide.

Keywords *LINC00507* · Micropeptide · lincRNAs · RNA-Seq · Non-coding RNA · Cerebral cortex · Brain

Electronic supplementary material The online version of this article (doi:10.1007/s12031-016-0745-4) contains supplementary material, which is available to authorized users.

✉ Michael Janitz
m.janitz@unsw.edu.au

[1] School of Biotechnology and Biomolecular Sciences, University of New South Wales, NSW, Sydney 2052, Australia

[2] Department of (Neuro) Pathology, Academic Medical Center and Swammerdam Institute for Life Sciences, Center for Neuorscience, University of Amsterdam, Amsterdam, The Netherlands

[3] Present address: Department of (Neuro) Pathology, Academic Medical Center and Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, Amsterdam, The Netherlands

Introduction

Long non-coding RNAs (lncRNAs) are typically described as RNA molecules that are longer than 200 nucleotides and lack protein-coding potential (Mattick and Rinn 2015). Other than length, there does not appear to be any common sequence or structural features shared between lncRNAs. It is possible to split lncRNAs into various subclasses based on their orientation and relationship to surrounding genes. These subclasses include antisense, intronic, intervening, bi-directional and overlapping (Mattick and Rinn 2015). Long intervening non-coding RNAs (lincRNAs) are those lncRNAs that are transcribed from loci positioned between other transcriptionally active regions of the genome. LincRNAs have a diverse range of functions. It has been revealed that lincRNAs play a role in embryonic stem cell physiology (Guttman et al. 2011), in brain function and development (Sauvageau et al. 2013; Mills et al. 2015a), and are associated with chromatin-modifying complexes and thus gene expression (Khalil et al. 2009).

LincRNAs appear to be highly tissue-specific, and in particular, the central nervous system demonstrates high tissue-specific expression patterns (Ulitsky et al. 2011). A study of over 8000 human lincRNAs demonstrated

strikingly tissue-specific expression when compared to protein-coding genes (Cabili et al. 2011). While only 19 % of protein-coding genes demonstrate tissue-specific expression patterns, 78 % of lincRNAs appear to be expressed in a tissue-specific manner (Cabili et al. 2011). Further, while maintaining tissue-specific expression patterns, lincRNAs also appear to have high levels of species-specific transcription (Ward et al. 2015). A comparative transcriptome study of human and macaque prefrontal cortex (PFC) found that while 13,722 of the 14,745 protein-coding genes expressed in human were also expressed in macaque, only 514 of 1061 lincRNAs were expressed in macaque (He et al. 2014).

LincRNAs show very little conservation in their sequences and they evolve rapidly, especially when compared to protein-coding genes (Ulitsky et al. 2011; Kutter et al. 2012; Necsulea et al. 2014). While approximately 98 % of protein-coding genes are conserved across primates, only 30 % of lncRNAs are conserved amongst primates (Derrien et al. 2012). This is further supported by a study, which demonstrated that 70 % of lincRNAs were not conserved in species with greater than 50 million years of evolutionary divergence (Hezroni et al. 2015). Even amongst closely related species, traditional sequence conservation patterns do not hold. Amongst the set of lincRNAs across mouse and rat, nearly half have been gained or lost since divergence from the last common ancestor (Kutter et al. 2012).

Further, lincRNAs appear to have different evolutionary constraints to protein-coding genes. While protein-coding genes need to maintain an extensive open reading frame (ORF) to preserve function, this is not the case for lincRNAs. It has been suggested that the evolution of lincRNAs is driven by the conservation of secondary structures. Many lincRNAs with differing sequences are able to bind to the same protein, indicating that secondary structure is more important to function than primary sequence (Khalil et al. 2009; Guttman et al. 2011). Conversely, sequence conservation is not always equivalent to transcriptional conservation. Two lincRNAs, surrounding the sex-determining region Y-box21 (*SOX21*), *linc-SOX21-B* and *linc-SOX21-C*, are only expressed in humans. However, both lincRNAs have high sequence conservation. *Linc-SOX21-B* overlaps a sequence that is highly conserved across vertebrates, and *linc-SOX21-C* has a conserved sequence across mammals (Hezroni et al. 2015). This suggests that the traditional methods of detecting functionality based on identifying regions of sequence conservation may not apply to lincRNAs.

It has recently been discovered that some lincRNAs encode small proteins, known as micropeptides. Micropeptides are shorter than 100 amino acids; they are not cleaved from a larger precursor protein and lack an N-terminal signalling sequence (Galindo et al. 2007; Hashimoto et al. 2008). Recently, it was shown that *LINC00948* harbours a short 128 nucleotide ORF that is translated into a 46-amino acid micropeptide known as myoregulin (MLN) (Anderson et al. 2015). The micropeptide appears to be involved in the regulation of muscle performance. While it is not yet known how many lincRNAs conceal micropeptides, it is possible that some lincRNAs may carry out function both as an RNA molecule and as a translated micropeptide.

It is thought that one of the major evolutionary drivers of the complexity of the human brain is the non-coding transcriptome (Haygood et al. 2010). Indeed, the human brain and, in particular, the frontal cortex, appears to be a rich source of lincRNAs (Mills et al. 2013). It is possible that lincRNAs with their high levels of tissue-specific and species-specific expression may be involved in developing the higher cognitive function seen in the frontal cortex of the humans and other primates. In the current study, we analysed 39 lincRNAs that are expressed in the frontal cortex and identified *LINC00507*, first described within the Mammalian Gene Collection Program (Strausberg et al. 2002), as being expressed in a cortex-specific manner in non-human primates and humans. To this end, we transcriptionally characterised *LINC00507*, followed by a comprehensive analysis of micropeptide-coding potential.

Materials and Methods

Tissue-Specific Expression Patterns of lincRNAs

The tissue-specific expression patterns of lincRNAs were assessed using the Genotype-Tissue Expression (GTEx) project (http://www.gtexportal.org/home/). The GTEx database contains RNA-Seq data from multiple tissues derived from multiple donors (Mele et al. 2015). This allows the expression and splicing patterns of all annotated human genes to be tracked across different tissue types.

Conservation of LINC00507

Sequence conservation of *LINC00507* was assessed using the UCSC Genome Browser website (http://genome.ucsc.edu/index.html). The UCSC genome browser contains the reference sequences and working draft assemblies of a large collection of genomes from a variety of organisms. Using the UCSC Genome Browser, a BLAST-like alignment tool (BLAT) search was performed to identify the

level of sequence conservation of *LINC00507* across chimpanzee, baboon, rhesus macaque, mouse lemur, mouse and dog.

RNA-Sequencing Resources

A variety of publicly available resources were utilised in order to analyse the lincRNAs. RNA-Seq data of primate tissues was downloaded from the Nonhuman Primate Reference Transcriptome Resource (NHPRTR, http://nhprtr.org/index.html). All sequencing data was strand-specific, ribosomal RNA (rRNA)-depleted with paired-end reads of 100 nucleotides length. Cell-type specific expression levels were evaluated using published RNA-Seq data (Darmanis et al. 2015). These reads were paired-end and 75 nucleotides in length.

All RNA-Seq data was analysed using the Tuxedo suite (Trapnell et al. 2012). The first reads were mapped to the appropriate reference genome using Tophat2, with the default settings for paired-end data. The following reference genomes were used for read alignment; *hg38* for RNA-Seq data derived from human tissue, *panTro4* for analysis of chimpanzee RNA-Seq data, *papAnu2* for baboon RNA-Seq data and *rheMac3* for rhesus macaque RNA-Seq data. Cufflinks was used to assemble the reads generated by Tophat2 into transcripts and quantifies the abundance of each transcript. For this study, transcript abundance is represented as fragments per kilobase of exon per million fragments mapped (fpkm). A transcript was considered expressed if it had an fpkm greater than 1.

Determination of RNA Secondary Structure and Protein Coding Potential

The Open Reading Frame Finder tool from NCBI (http://www.ncbi.nlm.nih.gov/projects/gorf/) was used to detect putative open reading frames in isoforms of *LINC00507*. RNAfold from the ViennaRNA web suite (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was then used to predict the secondary structure of *LINC00507* using the default settings. RNAfold predicts secondary structures based on minimum free energy (Zuker and Stiegler 1981). RNA secondary structures were then manually inspected for location of ORFs.

Ribo-Seq profiles of human brain tissue were acquired from the GWIPS-viz browser (http://gwips.ucc.ie/). Ribo-Seq is a next-generation sequencing tool designed to demonstrate the binding to a ribosome on a particular RNA sequence. Firstly, RNAs which are not protected by ribosomes are digested. The remaining RNA fragments are then sequenced and can be analysed in the same manner as standard RNA-Seq data (Ingolia et al. 2009). The ribosome binding profile of *LINC00507* was analysed to determine the binding of a ribosome at any of the open reading frames that were flanked by RNA open structures. The Ribo-seq and messenger RNA (mRNA) RNA-Seq data used in this study were from control brain samples (Gonzalez et al. 2014).

RT-qPCR

For analysis of expression of *LINC00507* across different age groups, total RNA was extracted from cortical tissue of 18 samples ranging in age from 22 gestational weeks to 52 years (Table S1). Tissue was provided from the archives of the Department of (Neuro) Pathology of the Academic Medical Center, University of Amsterdam. Tissue was obtained and used in accordance with the Declaration of Helsinki and the AMC Research Code provided by the Medical Ethics Committee. Total RNA was isolated using the miRNeasy Mini Kit or RNeasy Lipid Tissue Mini Kit (Qiagen Benelux, Venlo, The Netherlands) according to manufacturer's instructions. The concentration and purity of RNA samples were determined at 260/280 nm using a Nanodrop spectrophotometer (Ocean Optics, Dunedin, FL, USA).

First total RNA were reverse-transcribed into cDNA using oligo(dT) primers. Five nanomoles of oligo(dT) primer were annealed to 2 μg total RNA in a total volume of 25 μL, by incubation at 72 °C for 10 min, and cooled to 4 °C. Reverse transcription was performed by the addition of 25 μL RT-mix, containing First Strand Buffer (Invitrogen-Life Technologies, The Netherlands), 2 mM dNTPs (Pharmacia, Germany), 30 U RNase inhibitor (Roche Applied Science, Indianapolis, IN, USA) and 400 U M-MLV reverse transcriptase (Invitrogen-Life Technologies, The Netherlands). The total reaction mix (50 μL) was incubated at 37 °C for 60 min, heated to 95 °C for 10 min and stored at −20 °C until use.

For quantitative PCR (qPCR), a master mix was prepared on ice, containing, per sample, 1 μL cDNA, 2.5 μL of FastStart Reaction Mix SYBR Green I (Roche Applied Science, Indianapolis, IN, USA) and 0.4 μM of both reverse and forward primers. The final volume of the reaction was 6 μL. qPCR was carried out on a LightCycler 480 (Roche Applied Science, Germany). For qPCR experiments, the sequence of the forward primer used to amplify *LINC00507* was AGTTTCACCTGCCTGCACAT, and the sequence of the reverse primer was GTTTGTTCACCTCTGCGCTC. The forward and reverse primers used to amplify proteasome (prosome, macropain) subunit, beta type 4 (*PSMB4*), were TCAGTCCTCGGCGT and TAAGTTC and CTGATCATGTGGGCAATATCC, respectively. *PSMB4* was utilised as a reference gene to normalise expression data. The

cycling conditions were as follows: initial denaturation at 95 °C for 5 min, followed by 45 cycles of denaturation at 95 °C for 5 s, annealing at 60 °C for 10 s and extension at 72 °C for 15 s. The fluorescent product was measured by a single acquisition mode at 72 °C after each cycle. For distinguishing specific from non-specific products and primer dimers, a melting curve was obtained after amplification. The qPCR output was analysed using the software LinRegPCR (Ramakers et al. 2003); the data was passed to R (http://www.r-project.org/) for statistical testing, where Mann-Whitney-Wilcox tests were performed.

Results

LINC00507 Is Expressed Exclusively in the Brain Amongst Human Tissues

Using RNA-Seq, 39 lincRNAs were found to have an expression level higher than 5 fpkm in the superior frontal gyrus of the healthy human brain (Mills et al., unpublished data). Each of these 39 lincRNAs was assessed for the expression patterns across 53 different tissue types using the GTEx database. Amongst these 39 lincRNAs, only *LINC00320* and *LINC00507* were identified as brain-specific. While *LINC00320* appeared to be expressed throughout the brain, significant expression of *LINC00507* was only detected in the cortex. Only basal levels of expression for *LINC00507* were detected in the cerebellum, hippocampus and the amygdala (Supp. Figure 1). This suggests that the expression of *LINC00507* is limited to the brain and more specifically to the cortex.

LINC00507 is alternatively spliced to produce three splice variants: *LINC00507-001*, *LINC00507-002* and *LINC00507-003* (Ensembl IDs) (Supp. Figure 2). Amongst the regions of the brain in which *LINC00507* was expressed, the isoform predominantly expressed in the brain was *LINC00507-002* (Fig. 1). This isoform is 1930 bps long and consists of three exons.

Conservation of LINC00507 Sequence and Expression Across Primates

Sequence conservation of *LINC00507-002* was assessed across chimpanzee, baboon, rhesus macaque, mouse lemur, mouse and dog (Fig. 2). All three exons of *LINC00507-002* were highly conserved in chimpanzee, baboon and rhesus macaque. The presence of exon 1 was detected in mouse lemur. No reliable sequence homologue of any exon was detected in mouse or dog.

As sequence conservation is not always indicative of sequence expression, the expression of *LINC00507* was assessed in the cerebellum and frontal cortex of chimpanzee, baboon and rhesus macaque (Fig. 3). *LINC00507* maintained its cortex-specific expression pattern across non-human primates; expression was detected in the frontal cortex, but not in the cerebellum.

Open Reading Frame Prediction

A prediction of ORFs for the most highly expressed isoform in the human frontal cortex, *LINC00507-002*, was performed to assess the presence of potential protein-coding sequences. ORF frames less than 100 nucleotides in length were excluded. In total eight ORFs were detected in *LINC00507-002* (Table 1). Each of the detected ORFs was between 100 and 200 nucleotides in length.

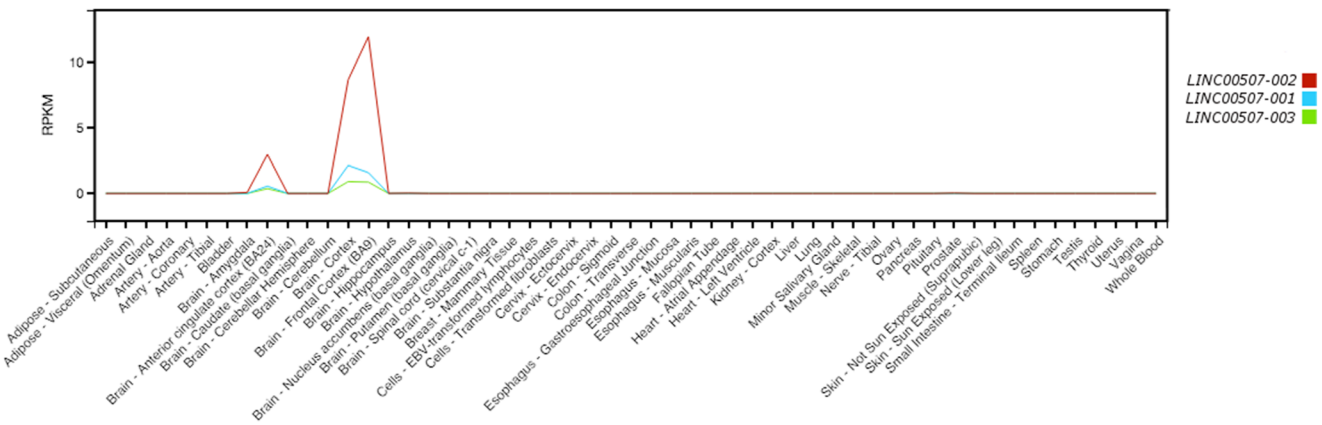

Fig. 1 Isoform-specific expression of *LINC00507*. The breakdown of the isoform specific expression across tissues. The absolute expression value of each splice variant is given. In all of the tissues where *LINC00507* was expressed, *LINC00507-002* was the dominant isoform. Expression (*y*-axis) is in reads per kilobase of transcript per million mapped reads (RPKM). The figure was adapted from the GTEx database

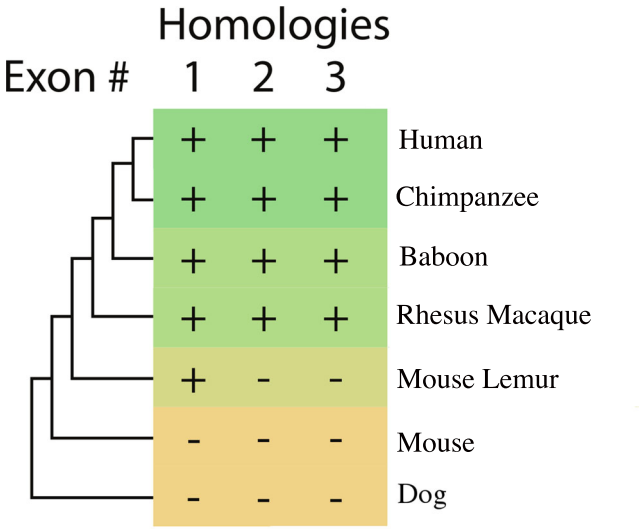

Fig. 2 Sequence conservation of *LINC00507-002* across vertebrate species. The sequence of all three exons was conserved across chimpanzee, baboon and rhesus macaque. The first exon was detected in mouse lemur. No exons were detected for mouse or dog. *Plus sign* represents sequence detected; *minus sign* represents no sequence detected

RNA Secondary Structure Prediction

Both protein-coding and non-coding RNA can be folded to form secondary structures. The secondary structure of RNA can determine how it interacts with other molecules present in the cell. A common feature of many RNA protein-coding regions is that the transcription start site and stop codon are flanked by single-stranded RNA (ssRNA) secondary structures (Wan et al. 2014). Therefore, in order for an open reading frame to be a considered a candidate for translation, it should be flanked by ssRNA structures on both the 5′ and 3′ ends. Secondary structure analysis revealed that the ORF located in frame 1 (809–981), frame 2 (431–556) and frame 3 (99–218) (Table 1) was flanked by regions of single-stranded RNA (Fig. 4).

Protein-Coding Potential Based on Ribo-Seq Data

The Ribo-Seq profiles, derived from publicly available data sets (see Materials and Methods for details), of apolipoprotein E (*APOE*) and *LINC00507* were analysed. *APOE* is an abundantly expressed protein-coding gene. Therefore, it was used as a positive control to show Ribo-Seq data output for transcripts being actively translated (Fig. 5a). Strong peaks along the protein-coding regions of *APOE* indicate that these regions were bound to the ribosomes. Further, the peaks of the Ribo-Seq data were of a similar height of the mRNA RNA-Seq data, demonstrating that the majority of the *APOE* transcripts were also translated.

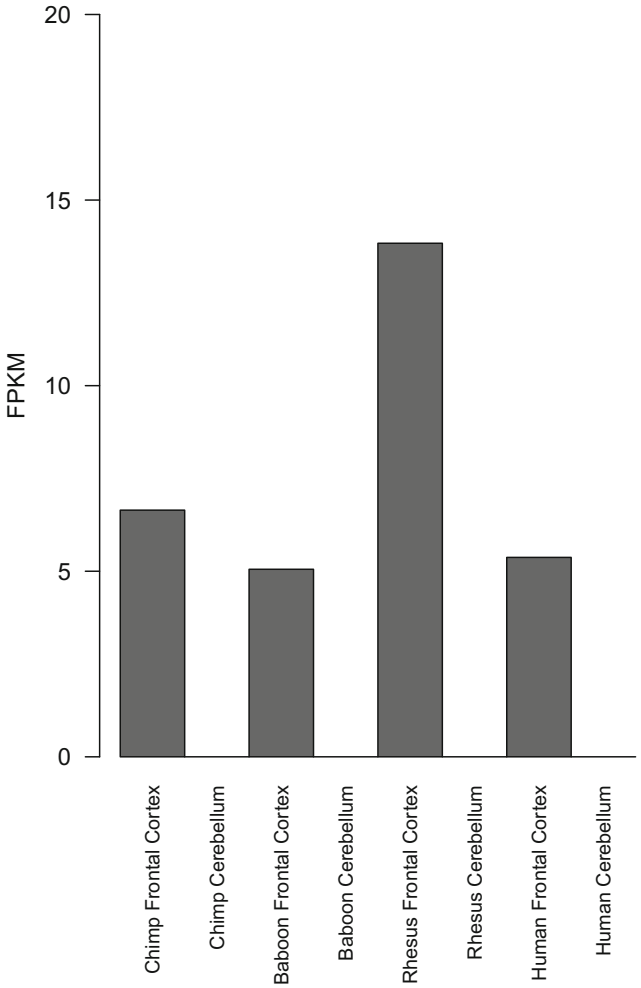

Fig. 3 Expression of *LINC00507* in primates. The expression of *LINC00507* in the frontal cortex and cerebellum of selected primates is shown. Similarly to humans, there is expression in the frontal cortex, while no expression is detected in the cerebellum

For *LINC00507-002*, there was a marked peak in the Ribo-Seq data corresponding to short sequence reads aligning to the ORF 99–218 bp in exon 1 (Fig. 5b). The peak had approximately the same height as the mRNA

Table 1 Location of potential ORFs in *LINC00507*

Frame	Open reading frame location	Length (nts)
+1	808–981	174
+1	1696–1836	141
+2	431–556	126
+2	722–829	108
+2	1340–1510	171
+3	99–218	120
+3	561–725	165
+3	1569–1688	120

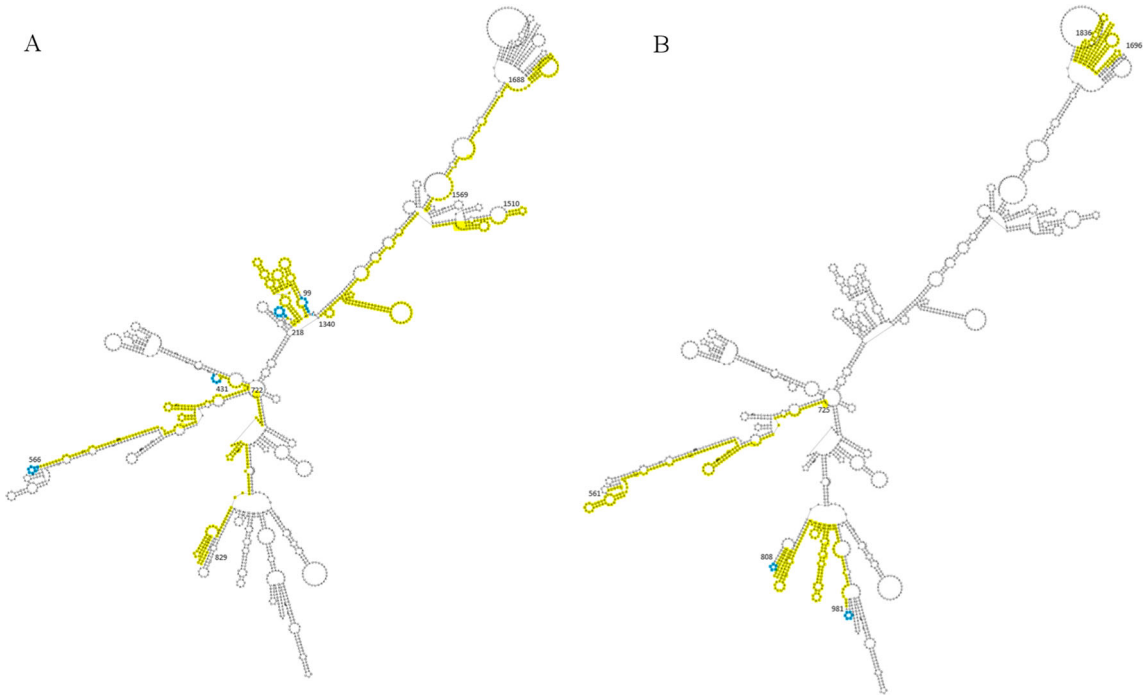

Fig. 4 Predicted RNA secondary structure of *LINC00507-002*. Predicted ORFs are highlighted in *yellow*. *Blue highlighted areas* represent open single-stranded RNA structures. The *numbers* represent the start and end of the respective ORFs, based on nucleotide count from the 5′ end of the transcript. **a** The positioning of ORFs located between nucleotides 99–218, 431–566, 722–829, 1340–1510 and 1569–1688 is shown. **b** The positioning of ORFs located between nucleotides 561–725, 808–981 and 1696–1836 is shown. Of interest are those ORFs located at 99–218, 431–556 and 809–981 as these ORFs are flanked by single-stranded RNA structures

RNA-Seq data, indicating a large proportion of the *LINC-00507-002* was bound by the ribosomes. As this region also corresponds to an ORF that is flanked by single stranded RNA secondary structures, there is strong evidence that it encodes for a micropeptide. The predicted micropeptide would be 40 amino acids in length.

Fig. 5 Ribo-Seq profiles of *APOE* and *LINC00507* genes. The *red graph* represents the Ribo-Seq data, the *green graph* represents the mRNA RNA-Seq data. The *red lines* denote the intron exon arrangements of the transcript, where *filled boxes* correspond to exons. **a** For *APOE* a large proportion of the transcribed *APOE* appears be bound to the ribosomes. **b** There appears to be higher levels of transcription of the first and last exon of *LINC00507*. There is a large peak on the Ribo-Seq reads mapping to exon 1 of *LINC00507*. This region corresponds to the ORF located between nucleotides 99–218. This ORF was predicted to be flanked by single-stranded RNA after RNA secondary structure formation

Expression Across Individual Cell Types

Many brain cells often demonstrate distinct gene expression patterns. Hence, single-cell RNA-Seq data sets derived from individual cell types, isolated from human brain tissue, were analysed. The cell types included astrocytes, endothelial cells, microglia, neurons, oligondendrocytes and oligodendrocyte precursor cells (OPC) (Darmanis et al. 2015). No expression of *LINC00507* was detected in any of these isolated cells.

Expression of LINC00507 Across Age Ranges

In order to further elucidate the expression patterns of *LINC00507*, RT-qPCR was carried out on 18 samples of varying ages extracted from the cortex of the human brain (Table S1). The samples were split into three categories based on age; gestational–1 year (gestational/infants), 10–18 years (pre-adolescent/adolescent) and 25+ years (adult) (Fig. 6).

Amongst the gestational/infants group, the expression of *LINC00507* was almost not detectable. In contrast, when compared to the gestational/infants group, *LINC00507* was up-regulated 48-fold (p-value < 0.005) and 32-fold (p-value < 0.005) in the pre-adolescent/adolescent and adult groups, respectively. The highest level of expression was seen in the pre-adolescent/adolescent group. Between the pre-adolescent/adolescent and adult groups, there was a non-significant 1.5-fold downregulation. Interestingly the highest level of variation in expression was seen in the pre-adolescent/adolescent group.

Discussion

There is a large body of evidence to suggest lincRNAs are transcribed in a highly tissue-specific manner (Cabili et al. 2011; Ulitsky et al. 2011; Mills et al. 2015a, b). As presented here, transcriptional analysis of *LINC00507* across a number of different tissue types adds to this body evidence. The expression of *LINC00507* was limited to the cortex of the brain, with no expression seen in other peripheral tissues and other brain regions. Interestingly, it was not possible to detect the expression of *LINC00507* in the major cell types that are constituents of human brain tissue. This may suggest a role for *LINC00507* in cell-to-cell communication. In purified cell populations, it is possible that the transcription profile of a cell may be altered, resulting in no expression of *LINC00507*. A similar phenomenon has also been noted to occur for *LINC00320* (Mills et al. 2015b). *LINC00320* is highly expressed in the superior frontal gyrus; however, no expression could be detected in the following immortalised cell lines: neurons, oligodendrocytes, foetal and mature astrocytes. This suggests that the expression of lincRNAs may be a function of their cellular context. The implication for the brain would be that the complete network of different cell types is required for the detection of the complete repertoire of lincRNAs expressed by a cell. This might be also a reason we were able to detect *LINC00507* in entire cortical tissue but not in individual cells.

The evaluation of the sequence of *LINC00507-002* demonstrated conservation in higher primates, where its cortex-specific expression appears to be preserved. Consequently, *LINC00507-002* may be a primate- and cortex-specific lincRNA. This may also suggest that *LINC00507* could possibly be involved in the higher cognitive function seen in the cortex of primates.

Interestingly, while only basal levels of *LINC00507* could be detected in cortical brain tissue taken from gestational and infant patients, a significant up-regulation was seen in cortical tissue extracted from pre-adolescent/adolescent and adults. Further, the highest level of expression and variation in expression of *LINC00507* was seen amongst those in the pre-

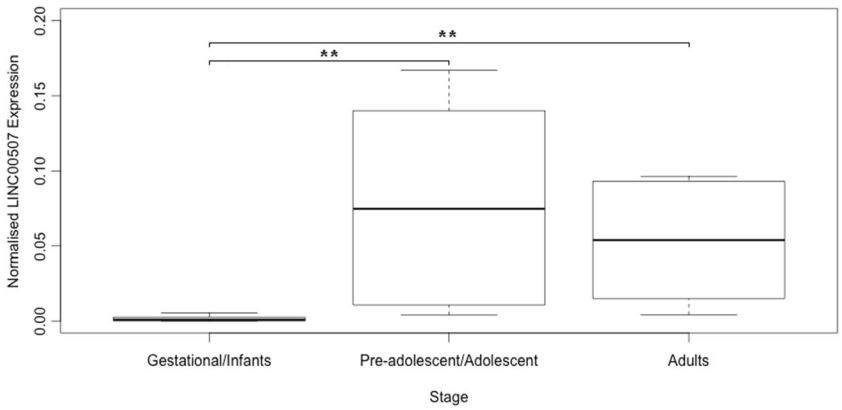

Fig. 6 Boxplot of *LINC00507* expression in different age groups. The expression values of *LINC00507* appear to be dependent on age. Only basal levels of *LINC00507* expression were detected the gestational/infant group. *LINC00507* was up-regulated 48-fold (p value < 0.005 indicated by double asterisk) in pre-adolescent/adolescent group when compared to the gestational/infant group and 32-fold (p value < 0.005 indicated by double asterisk) when adults where compared to the gestational/infant group. The highest levels of expression were seen in the pre-adolescent/adolescent group. Expression levels were normalised to the reference gene *PSMB4*

adolescent/adolescent group. It appears that at some point in human brain development, *LINC00507* gene expression is switched on, thus indicating a possible role of *LINC00507* in the development of the cerebral cortex. Evidence has shown that neuron development continues from adolescence into early adulthood (Huttenlocher and Dabholkar 1997; Petanjek et al. 2011). Synaptic connections in the cerebral cortex of primates are initially overproduced and then are pruned during puberty (Huttenlocher and Dabholkar 1997). Age appears to have a significant effect on a number of metabolites in the brain. For example, out of a hundred metabolites detected in the superior frontal gyrus of the prefrontal cortex and lateral part of the cerebellar cortex, 88 % show significant concentration changes with age (Fu et al. 2011). Currently, very little specific information is available on age-dependent lincRNAs. It has been shown that 409 lincRNAs demonstrated significant expression level changes with age (He et al. 2014). Two hundred seven of these lincRNAs also showed age-related expression levels in macaque (Derrien et al. 2012). This indicates that these lincRNAs have evolved recently and perhaps function in development of the primate brain. In future studies, a greater range and number of age samples could be used to confirm the age-specific dependent expression of *LINC00507*. Further, attempts could be made to correlate *LINC00507* with developmental milestones or intelligence levels.

The report presented here provides strong evidence that *LINC00507-002* harbours micropeptide-coding potential. This is based on three lines of evidence: (i) the presence of an ORF, (ii) the ORF is flanked by an open RNA secondary structure and (iii) the ORF region is bound to the ribosomes. Nevertheless, the presence of an ORF alone is not enough evidence for micropeptide translation, as majority of lincRNAs contain multiple ORF simply by chance. There is a tendency for the start sites and stop sites of ORFs in protein-coding RNAs to be more single-stranded then other regions of the RNA transcript (Wan et al. 2014). This principle can be applied to identify micropeptides in lncRNAs, and indeed the micropeptide MLN is flanked by regions of single-stranded RNA (Anderson et al. 2015). Finally, to be translated, the RNA must be bound to the ribosome. Here, an ORF in *LINC00507-002*, which was flanked by single-stranded RNA regions, was found to be bound to ribosomes. This strongly suggests that this ORF is translated.

Further investigation is required to confirm whether *LINC00507-002* does indeed harbour a micropeptide. Mass spectrometry could be used to detect the micropeptide. If the presence of a micropeptide is confirmed, a knockdown or over-expression study of *LINC00507-002* may elucidate the role that it plays in the brain. A knockout of the micropeptide MLN in mice revealed its role in Ca^{2+} handling and skeletal muscle performance. If the presence of a micropeptide is not found, it is foreseeable that *LINC00507-002*, by engaging the ribosomes, is performing a regulatory function by blocking the translation of other RNA molecules.

It has been shown here that *LINC00507* is a primate- and cortex-specific lincRNA that has age-dependent expression patterns. Thus, *LINC00507* may play an important role in the development of the brain and, in particular, primate-specific intelligence. Overall, *LINC00507* poses an interesting target for further investigation into the role of lincRNAs in primate brain biology.

Acknowledgments The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171) and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by supplements to University of Miami grants DA006227 and DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 and MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina–Chapel Hill (MH090936 and MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). This research was supported by the Framework Programme FP7/2007-2013 under the project EPISTOP (grant agreement no: 602391 to EA) and (NeuroGeM grant 733051052 to EA and JDM).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160:595–606

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25:1915–1927

Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR (2015) A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci U S A 112:7285–7290

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J,

Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789

Fu X, Giavalisco P, Liu X, Catchpole G, Fu N, Ning ZB, Guo S, Yan Z, Somel M, Paabo S, Zeng R, Willmitzer L, Khaitovich P (2011) Rapid metabolic evolution in human prefrontal cortex. Proc Natl Acad Sci U S A 108:6181–6186

Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP (2007) Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol 5:e106

Gonzalez C, Sims JS, Hornstein N, Mela A, Garcia F, Lei L, Gass DA, Amendolara B, Bruce JN, Canoll P, Sims PA (2014) Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J Neurosci 34:10924–10936

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477: 295–300

Hashimoto Y, Kondo T, Kageyama Y (2008) Lilliputians get into the limelight: novel class of small peptide genes in morphogenesis. Develop Growth Differ 50(Suppl 1):S269–S276

Haygood R, Babbitt CC, Fedrigo O, Wray GA (2010) Contrasts between adaptive coding and noncoding changes during human evolution. Proc Natl Acad Sci U S A 107:7853–7857

He Z, Bammann H, Han D, Xie G, Khaitovich P (2014) Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. Rna 20:1103–1111

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep 11: 1110–1122

Huttenlocher PR, Dabholkar AS (1997) Regional differences in synaptogenesis in human cerebral cortex. J Comp Neurol 387:167–178

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106:11667–11672

Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, Odom DT, Marques AC (2012) Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet 8: e1002841

Mattick JS, Rinn JL (2015) Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 22:5–7

Mele M, PG F, Reverter F, DS DL, Monlong J, Sammeth M, TR Y, JM G, DD P, TJ S, Johnson R, AV S, Djebali S, Niarchou A, GT C, FA W, Lappalainen T, Calvo M, Getz G, ET D, KG A, Guigo R (2015) Human genomics. The human transcriptome across tissues and individuals. Science 348:660–665

Mills JD, Kavanagh T, Kim WS, Chen BJ, Waters PD, Halliday GM, Janitz M (2015a) High expression of long intervening non-coding RNA OLMALINC in the human cortical white matter is associated with regulation of oligodendrocyte maturation. Mol Brain 8:2

Mills JD, Chen J, Kim WS, Waters PD, Prabowo AS, Aronica E, Halliday GM, Janitz M (2015b) Long intervening non-coding RNA 00320 is human brain-specific and highly expressed in the cortical white matter. Neurogenetics 16:201–213

Mills JD, Kavanagh T, Kim WS, Chen BJ, Kawahara Y, Halliday GM, Janitz M (2013) Unique transcriptome patterns of the white and grey matter corroborate structural and functional heterogeneity in the human frontal lobe. PLoS One 8:e78480

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505:635–640

Petanjek Z, Judas M, Simic G, Rasin MR, Uylings HB, Rakic P, Kostovic I (2011) Extraordinary neoteny of synaptic spines in the human prefrontal cortex. Proc Natl Acad Sci U S A 108:13281–13286

Ramakers C, Ruijter JM, Deprez RHL, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Let 339:62–66

Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, Swerdel MR, D'Ecclessis MF, Moore JC, Lai V, Gong G, Yancopoulos GD, Frendewey D, Kellis M, Hart RP, Valenzuela DM, Arlotta P, Rinn JL (2013) Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife 2:e01749

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T, Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P, Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH, Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW, Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J, Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW, Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U, Smailus DE, Schnerch A, Schein JE, Jones SJ, Marra MA, Team MGCP (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A 99: 16899–16903

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562–578

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147:1537–1550

Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, Chang HY (2014) Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505:706–709

Ward M, McEwan C, Mills JD, Janitz M (2015) Conservation and tissue-specific transcription patterns of long noncoding RNAs. J Hum Transcriptome 1:2–9

Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res 9:133–148