Published in final edited form as:

Neurochem Int. 2014 November ; 77: 86–93. doi:10.1016/j.neuint.2014.06.006.

Expression of transcribed ultraconserved regions of genome in rat cerebral cortex

Suresh L. Mehta^a, Ashutosh Dharap^{a,b}, and Raghu Vemuganti^{a,*}

^aDepartment of Neurological Surgery, University of Wisconsin, Madison, WI, USA

^bTheoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract

Emerging evidence indicates that 481 regions of the genome (>200 bp) that actively transcribe noncoding RNAs shows 100% homology between humans, rats and mice. These transcribed ultraconserved regions (T-UCRs) are thought to control the essential regulatory functions basic for life in rodents and mammals. Using microarray analysis, we presently show that 107 T-UCRs are actively expressed in adult rat cerebral cortex. They are grouped into intragenic (61) and intergenic (46) based on their genic location. Interestingly, 10 T-UCRs are expressed at unusually high levels in cerebral cortex. Additionally, many T-UCRs also showed cogenic expression. We further analyzed the correlation of intragenic T-UCRs with their host protein coding genes. Surprisingly, most of the expressed intragenic T-UCRs (54 out of 61) displayed a negative correlation with their host gene expression. T-UCRs are thought to control the splicing and transcription of the protein-coding genes that host them and flank them. Bioinformatics analysis indicated that the protein products of majority of these genes are nuclear in localization, share protein domains and are involved in the regulation of diverse biological and molecular functions including metabolism, development, cell cycle, binding and transcription factor regulation. In conclusion, this is the first study to shows that many T-UCRs are expressed in rodent brain and they might play a role in physiological brain functions.

Keywords

Brain; Non-coding RNA; lncRNA; Transcription; Gene expression

1. Introduction

Humans and rodents were separated phylogenetically >80 million years ago and their genomes continuously evolved over generations (Prat et al., 2009). As mammalian genome is known to mutate at a rate of $\sim 2.5 \times 10^{-8}$ per base per generation (Kumar and

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding author. Address: Department of Neurological Surgery, Mail code CSC-8660, 600 Highland Ave, Madison, WI 53792, USA. Tel.: +1 608 263 4055; fax: +1 608 263 1728., Vemuganti@neurosurgery.wisc.edu (R. Vemuganti).

Conflicts of interest

Authors declare no conflict of interest.

Subramanian, 2002; Nachman and Crowell, 2000), there are millions of base differences in the genetic code between the mammals and rodents. However, 481 regions longer than 200 bp each (range: 200–779 bp) remained conserved with 100% homology between the orthologous regions of human, mouse and rat genomes (Bejerano et al., 2004b). All these regions are known to actively transcribe RNAs. Their sequences have been published previously (Bejerano et al., 2004b) and they were subsequently renamed transcribed ultraconserved regions (T-UCRs) (Calin et al., 2007). The functional significance of these T-UCRs is not known but as they resisted mutations for such a long time, they are thought to play functions essential for life in mammals (Bejerano et al., 2004b). Although, the RNAs transcribed by T-UCRs are long and non-coding, they are distinct from long noncoding RNAs (lncRNAs) as T-UCRs are absolutely conserved while lncRNAs are least conserved among non-coding RNAs (Babak et al., 2005; Derrien et al., 2012; Johnsson et al., 2014).

The extreme degree of the genetic conservation of T-UCRs for many millions of years is thought to be related to functional evolutionary constraints such as negative or purifying selection that points towards an essential role for these elements (McLean and Bejerano, 2008). Furthermore, the remarkable depletion in human segmental duplications and copy number variants of T-UCRs suggest a disruption of their normal copy number fitness (Derti et al., 2006). The T-UCRs are both intragenic and intergenic and are often located at fragile sites and genomic regions involved in various pathologies including different types of cancer (Braconi et al., 2011; Calin et al., 2007; Hudson et al., 2013; Pauli et al., 2011; Sana et al., 2012; Scaruffi, 2011). Although the functional importance of T-UCRs is debatable, recent reports have suggested that they may serve roles such as chromatin regulation, acting as long-range enhancers of flanking genes, regulating splicing and epigenetic modifications, transcriptional co-activation, RNA scaffolding and protein scaffolding (Baira et al., 2008). The T-UCRs are also thought to control the expression of their host genes and/or the up-/down-stream genes (Bejerano et al., 2004b).

In the present study, we analyzed the expression profile of T-UCRs in adult rat cerebral cortex and conducted bioinformatics analysis to evaluate the Gene Ontology (GO) networks, subcellular localization and the biological functions of the nearest upstream and downstream genes and the host genes of the expressed T-UCRs. To further understand their functional significance, we correlated the expression of the intragenic T-UCRs with the expression of their host genes.

2. Materials and methods

2.1. RNA extraction and microarray analysis

These procedures are same as those published previously (Dharap et al., 2012). All the animal procedures were approved by the Research Animal Resources and Care Committee of the University of Wisconsin-Madison and the animals were cared 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). Total RNA was extracted from the cerebral cortex of adult, male rats (280–300 g; Charles River, Wilmington, MA, USA) ($n = 3$) using the TRIzol reagent (Invitrogen 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 samples were linearly amplified, labeled with Cy3-dCTP, purified by RNeasy Mini Kit (Qiagen USA), fragmented and hybridized to microarrays containing probes for T-UCRs (Arraystar USA) as described earlier (Dharap et al., 2012). We used 3 arrays to hybridize 3 RNA samples. 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) built into the GeneSpring GX V11.0 software. 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 call in all samples were chosen for further analysis.

2.2. Bioinformatics

The genomic coordinates of expressed T-UCRs obtained from UCSC genome browser assembly Baylor 3.4 (<http://genome.ucsc.edu>) were converted to RGSC 5.0/rn5 by UCSC's Batch Coordinate Conversion tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>). The coordinates of host and nearest up- and down-stream genes of the T-UCRs and the full length genomic sequences of host genes were queried for the possible exonic, intronic and exonic-intronic overlap using BLAT tool from UCSC Genome Browser and the sequences were cross-checked with the Ensembl Genome Browser (<http://useast.ensembl.org/index.html>). The co-expression and shared functional protein domains of the protein products of the host genes and nearest up- and down-stream genes of the T-UCRs were identified with GeneMANIA (<http://www.genemania.org/>) as described earlier (Zuberi et al., 2013). Functional coupling analysis for the subcellular localization of the protein products of the host genes and nearest up- and down-stream genes of the T-UCRs was analyzed with FunCoup (<http://funcoup.sbc.su.se/search/>) by the sub-network selection parameters set stringently (confidence threshold of 0.5, expansion depth of 1 and nodes per expansion depth are 20) and with expansion algorithm (Alexeyenko et al., 2012). Co-localization was weighted by the specificity of the localization, where specific localizations get a higher weight and unspecific localizations get a lower weight (Alexeyenko et al., 2012). The Protein ANalysis Through Evolutionary Relationships classification system (PANTHER, v. 9.0; <http://www.pantherdb.org/>) was used for high-throughput analysis of molecular function and biological process networks of the up/down stream and host genes of the T-UCRs (Mi et al., 2013). The functional classification was analyzed using the ontology of the terms generated with InterProScan and GO curation.

2.3. Real-time PCR validation

Real-time PCR was conducted to validate the microarray expression pattern of T-UCRs. Briefly, total RNA was extracted from the cerebral cortex of three adult, male rats using the mirVana™ miRNA Isolation Kit (Invitrogen USA). Each sample was reverse transcribed using Universal cDNA synthesis kit (Exiqon USA) and amplified (1 cycle at 50 °C for 2 min, 1 cycle at 95 °C for 10 min and 40 cycles at 95 °C for 10 s and 60 °C for 1 min) with primers specific for T-UCRs uc.188 (forward 5'-ACCTGCAAGCACCATCAA-3' and reverse 5'-CGTGGTGTGTCTATTTGTAGGA-3') and uc.419 (forward 5'-TACGGCTTCTGCTACGACTA-3' and reverse 5'-CTGCCTACATCCGGGTAAAG-3') using ExiLent SYBR® Green master mix, (Exiqon) in ABI 7000 Sequence Detection System (Applied Biosystems). 18S rRNA was used as an internal control. For each sample, real-time PCR was conducted twice.

3. Results

3.1. T-UCR expression in cerebral cortex

In the adult rat cerebral cortex, 107 (61 intragenic and 46 intergenic) T-UCRs were observed to be expressed (Fig. 1A) and their expression level ranged from 1 to 75,000 units (normalized -4.62 to 8.59). Of those, 10 were expressed at a very high level (10,001–75,000 units), 9 were expressed at a high level (2001–10,000 units), 17 were expressed at a medium level (501–2000 units), 29 were expressed at a low level (201–500 units) and 42 were expressed at a very low level (<200 units) (Fig. 1A). Interestingly, expression of 6 T-UCRs (uc.465, uc.46, uc.154, uc.471, uc.141 and uc.420) accounted for ~73% of the cumulative units of the 61 intragenic T-UCRs (Fig. 1B). Similarly, expression of 4 T-UCRs (uc.84, uc.188, uc.45 and uc.75) accounted for ~91% of the cumulative units of expression of the 46 intergenic T-UCRs (Fig. 1B). Using real time PCR, we confirmed that uc.188 was expressed in adult rat ~6-fold higher than uc.419 (Fig. 2). The 107 T-UCRs expressed in the cerebral cortex are observed to be distributed unevenly over chromosomes from which they transcribe. While 6 chromosomes (1, 3, 5, 6, 10 and X) showed an average of 11.5 T-UCRs/chromosome, 6 chromosomes (11, 15, 16, 19 and 20) showed an average of one T-UCRs/chromosome (Supplementary Table 1). The chromosome 12 showed no T-UCRs (Supplementary Table 1). The top 15 highly expressed intragenic and the top 15 highly expressed intergenic T-UCRs in rat cerebral cortex are shown in Tables 1 and 2. The remaining 77 T-UCRs (46 intragenic and 31 intergenic) are shown in Supplementary Tables 2 and 3. Of the 107 T-UCRs expressed in rat cerebral cortex, 49 (24 intragenic and 25 intergenic) are located in relatively protein-coding gene deserts (there are no protein-coding genes within 100,000 bases upstream or downstream except for the host genes in case of intragenic T-UCRs) (Tables 1 and 2; Supplementary Tables 2 and 3). Whereas, 19 T-UCRs (8 intragenic and 11 intergenic) are in the close vicinity of protein-coding genes (either an upstream or a downstream protein-coding gene present within 10,000 bases from the T-UCR) (Tables 1 and 2; Supplementary Tables 2 and 3). Eleven intragenic T-UCRs (uc.305/uc.308, uc.255/uc.256, uc.460/uc.462/uc.465, uc.418/uc.419, uc.208/uc.209) and 8 intergenic T-UCRs (uc.293/uc.298, uc.85/uc.86, uc.18/uc.19 and uc.6/uc.10) expressed in the cerebral cortex are observed to be co-genic (Supplementary Tables 4 and 5).

3.2. Correlation of intragenic T-UCRs with their host genes

Of the 61 intragenic T-UCRs expressed in the cerebral cortex, 34 are located in exons or exon-overlapping regions, 25 are located in introns or intron-overlapping regions and 2 are located at the exon–intron junctions (Fig. 3A). Forty-three of them are coded by the sense strand while 18 are coded by the antisense strand (Fig. 3A). The microarray platform we used also contained probes for 14,077 protein coding RNAs including the 32 host genes of the intragenic T-UCRs. The correlation expression analysis of the intragenic T-UCRs to the expression of 32 host genes of the intragenic T-UCRs in rat cerebral cortex revealed that only 7 of these intragenic T-UCRs showed a positive correlation with the expression of their host genes, while the remaining 25 showed a negative correlation (Fig. 3B). GeneMANIA analysis showed that many of the protein products of the host genes and the nearest up- and down-stream genes of the T-UCRs expressed in rat cerebral cortex are coexpressed (Supplementary Fig. 1) and share the protein domains (Supplementary Fig. 2). FunCoup analysis of the protein products of the host genes and nearest up- and down-stream genes of the intragenic T-UCRs expressed in rat cerebral cortex are nuclear while a limited number of them are also localized in the endoplasmic reticulum, mitochondria and on the membrane (Fig. 4).

3.3. Functional annotation of T-UCR expression networks in brain

T-UCRs have been identified previously based upon their genomic location of in or around protein-coding genes related to differentiation and splicing (Bejerano et al., 2004b). To explore the putative shared functions of the intragenic and intergenic T-UCRs expressed in cerebral cortex, we explored the functional significance of the networks of the nearest upstream/downstream protein coding genes and the host genes (in case of intragenic T-UCRs) using PANTHER Gene List Analysis algorithm. The putative common functions of each group were annotated in relation to biological process and molecular function (Fig. 5). Both the intragenic and intergenic T-UCRs expressed in the rat cerebral cortex are associated with the same set of biological functions (Fig. 5A). However, the intragenic T-UCRs are predominantly associated with metabolism, development and cell cycle while the intergenic T-UCRs are predominantly associated with cellular process and cell communication (Fig. 5A). Both intragenic and intergenic T-UCRs are associated with the same molecular functions (binding, transcription regulation and catalytic activity are the top categories) (Fig. 5B).

4. Discussion

As they code for proteins, the mRNAs are considered important for controlling the physiological functions under normal and disease conditions. However, <2% of the RNAs encoded by mammalian genome are mRNAs while the remaining ~98% transcriptional output is noncoding (nc) RNAs of various classes. Recent studies indicate that ncRNAs have important regulatory roles which are indispensable for the cellular homeostasis (Marchese and Huarte, 2013; Vemuganti, 2013). Many types of ncRNAs are <200 nucleotides in length (Vemuganti, 2013). However, genome also encodes several lncRNAs (>200 bp in length), which regulate molecular functions like scaffolding of the chromatin modifying proteins and transcriptional modulation, protein targeting to genomic loci and epigenetic silencing (Gupta

et al., 2010; Hung et al., 2011; Pasmant et al., 2007; Rinn and Chang, 2012; Tsai et al., 2010; Yu et al., 2008). Unlike smaller ncRNAs such as microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs), the lncRNAs surprisingly lack sequence conservation among species (Babak et al., 2005; Derrien et al., 2012; Johnsson et al., 2014). Although >200 bp each, the T-UCRs are distinct and unlike lncRNAs show an extreme sequence conservation between the orthologous regions of human, mouse and rat genomes (Bejerano et al., 2004b; Calin et al., 2007). As the T-UCRs resisted mutations during evolution, these regions of the genome and the RNAs they encode are thought to play essential life functions such as tissue-specific regulation of gene expression (Ferdin et al., 2013; Pennacchio et al., 2006). Interestingly, the host genes and the nearest up- and down-stream genes of many T-UCRs are known to be involved in transcription and RNA splicing (Bejerano et al., 2004a,b, 2006). We currently observed that the majority of the nearest up- and down-stream genes and the host genes of the T-UCRs expressed in cerebral cortex are nuclear in localization indicating that they might participate in transcriptional events.

CNS is known to be very active in ncRNA expression, but no studies to date analyzed T-UCRs detection in the rodent brain. The present study shows that rat cerebral cortex is an active site of transcription for T-UCRs and many of their up- and down-stream genes and host genes are those involved in important biological and molecular functions. While the functional significance of this high activity of T-UCRs in brain is not yet known, we observed that the expression of 25 of the 32 intragenic T-UCRs are inverse in relation to their host gene expression. This indicates that T-UCRs might alter the expression of the host genes or vice versa in brain. In addition to this mutual regulation, other factors like transcriptional factors, epigenetic modulators and RNA degradation might regulate these transcripts. Expression of 7 intragenic T-UCRs expressed in brain were observed to be positively correlated with the expression of their respective host genes and interestingly all of them showed a sense exon overlap, suggesting a sharing conserved feature of the host gene rather than coding from an independent transcriptional unit (Mestdagh et al., 2010). A further interesting observation we made is that 70% (43 of the 61) of the intragenic T-UCRs expressed in rat cerebral cortex are transcribed in sense direction to the host gene indicating the possibility of shared promoters. Network analysis of the genes that flank and host the T-UCRs showed that many of them are coexpressed and share protein domains indicating an interactive mechanism of control exerted by the regulatory elements in brain.

To understand the functional significance of T-UCRs in cerebral cortex, we conducted GO analysis of the nearest protein-coding genes that flank them on either side as well as host genes (in case of intragenic T-UCRs). Surprisingly, the top biological functional categories of the intragenic and intergenic T-UCR associated genes are different. While the top 2 categories of intergenic T-UCRs are cellular process and cell communication, the top 2 categories of intragenic T-UCRs are metabolism and development. This indicates a divergence of function of the T-UCRs based on their genomic localization. However, both classes of T-UCR associated genes showed binding and transcriptional regulatory activity as the top molecular functions indicating that the T-UCRs might have stayed ultraconserved under elevated levels of purifying selection during the evolution (Ovcharenko et al., 2005). This is further supported by the observation of low density of synteny breakpoints in the gene deserts (many of them harbor the intergenic T-UCRs) compared to the gene-rich

regions of the genome indicating preclusion of rearrangement of regulatory elements from the host and nearby genes (Ovcharenko et al., 2005).

The protein-coding genes that flank and host the T-UCRs in cerebral cortex are also observed to be associated with other functions like cell cycle, transport, cell adhesion, apoptosis, enzyme regulatory activity, catalytic activity and structural molecular activity indicating a strong possibility that their extreme conservation might be evolved to regulate the functions important for cell survival. The extreme conservation might also be a protective mechanism towards a lack of tolerance to change to maintain the essential *in vivo* functionality of higher organisms (Poulin et al., 2005). However, mice that underwent deletion of 2 large gene deserts harboring some intergenic T-UCRs were observed to be viable and indistinguishable from their wild-type littermates (Nobrega et al., 2004). Furthermore, individual genetic deletion of T-UCRs uc.248, uc.329, uc.467 and uc.482 also resulted in viable mice with no observable phenotypic changes (Ahituv et al., 2007). This intriguing evidence suggests a lack of function for some of the T-UCRs. However, we cannot generalize that all T-UCRs are not functional as many recent studies showed altered T-UCR expression profiles in different types of cancers which might be associated with carcinogenesis in neuroblastoma, leukemia, hepatic, prostate and colorectal cancers (Braconi et al., 2011; Calin et al., 2007; Hudson et al., 2013; Mestdagh et al., 2010; Pauli et al., 2011; Sana et al., 2012; Scaruffi, 2011). Interestingly, many T-UCRs are also shown to be induced under HIF-1 regulation when tumor cells were subjected to hypoxia (Ferdin et al., 2013). Cerebral hypoxia and ischemia are known to rapidly alter the expression of ncRNAs including miRNAs, piRNAs and lncRNAs with significant impact on the post-stroke functional outcome (Dharap et al., 2009; Dharap et al., 2011, 2012). However, it is not yet known if cerebral hypoxia or ischemia influences T-UCRs and if yes whether they play any role in mediating the ischemic secondary brain damage.

In conclusion, the present results show that several T-UCRs are expressed in adult rat brain. Furthermore, extensive bioinformatics analysis of the expressed T-UCRs indicate a role in diverse biological and molecular processes including metabolism, development, cell cycle, cell communication and transcriptional control essential for homeostatic brain function. Additional studies are needed to infer the mechanisms that control the expressions of T-UCRs in normal and diseased brain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The study was supported by NIH Grants NS061071 and NS074444.

References

- Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 2007; 5:1906–1911.
- Alexeyenko A, Schmitt T, Tjarnberg A, Guala D, Frings O, Sonnhammer ELL. Comparative interactomics with Funcoup 2.0. *Nucleic Acids Res.* 2012; 40:D821–D828. [PubMed: 22110034]

- Babak T, Blencowe BJ, Hughes TR. A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription. *BMC Genomics*. 2005; 6:104. [PubMed: 16083503]
- Baira E, Greshock J, Coukos G, Zhang L. Ultraconserved elements genomics, function and disease. *RNA Biol*. 2008; 5:132–134. [PubMed: 18708752]
- Bejerano G, Haussler D, Blanchette M. Into the heart of darkness: large-scale clustering of human non-coding DNA. *Bioinformatics*. 2004a; 20:40–48. [PubMed: 14693806]
- Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*. 2006; 441:87–90. [PubMed: 16625209]
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science*. 2004b; 304:1321–1325. [PubMed: 15131266]
- Braconi C, Valeri N, Kogure T, Gasparini P, Huang NY, Nuovo GJ, Terracciano L, Croce CM, Patel T. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci USA*. 2011; 108:786–791. [PubMed: 21187392]
- Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu XP, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*. 2007; 12:215–229. [PubMed: 17785203]
- 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. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012; 22:1775–1789. [PubMed: 22955988]
- Derti A, Roth FP, Church GM, Wu CT. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nat Genet*. 2006; 38:1216–1220. [PubMed: 16998490]
- Dharap A, Bowen K, Place R, Li LC, Vemuganti R. Transient focal ischemia induces extensive temporal changes in rat cerebral MicroRNAome. *J Cereb Blood Flow Metab*. 2009; 29:675–687. [PubMed: 19142192]
- Dharap A, Nakka VP, Vemuganti R. Altered expression of PIWI RNA in the rat brain after transient focal ischemia. *Stroke*. 2011; 42:1105–1109. [PubMed: 21311060]
- Dharap A, Nakka VP, Vemuganti R. Effect of focal ischemia on long noncoding RNAs. *Stroke*. 2012; 43:2800–2802. [PubMed: 22949471]
- Ferdin J, Nishida N, Wu X, Nicoloso MS, Shah MY, Devlin C, Ling H, Shimizu M, Kumar K, Cortez MA, Ferracin M, Bi Y, Yang D, Czerniak B, Zhang W, Schmittgen TD, Voorhoeve MP, Reginato MJ, Negrini M, Davuluri RV, Kunej T, Ivan M, Calin GA. HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death Differ*. 2013; 20:1675–1687. [PubMed: 24037088]
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang YL, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:U1071–U1148.
- Hudson RS, Yi M, Volfovsky N, Prueitt RL, Esposito D, Volinia S, Liu CG, Schetter AJ, Van Roosbroeck K, Stephens RM, Calin GA, Croce CM, Ambs S. Transcription signatures encoded by ultraconserved genomic regions in human prostate cancer. *Mol Cancer*. 2013; 12. [PubMed: 23409748]
- Hung T, Wang YL, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerod A, Borresen-Dale AL, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011; 43:U621–U629.

- Johnsson P, Lipovich L, Grander D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*. 2014; 1840:1063–1071. [PubMed: 24184936]
- Kumar S, Subramanian S. Mutation rates in mammalian genomes. *Proc Natl Acad Sci USA*. 2002; 99:803–808. [PubMed: 11792858]
- Marchese FP, Huarte M. Long non-coding RNAs and chromatin modifiers. Their place in the epigenetic code. *Epigenetics*. 2013;9.
- McLean C, Bejerano G. Dispensability of mammalian DNA. *Genome Res*. 2008; 18:1743–1751. [PubMed: 18832441]
- Mestdagh P, Fredlund E, Pattyn F, Rihani A, Van Maerken T, Vermeulen J, Kumps C, Menten B, De Preter K, Schramm A, Schulte J, Noguera R, Schleiermacher G, Janoueix-Lerosey I, Laureys G, Powel R, Nittner D, Marine JC, Ringner M, Speleman F, Vandesompele J. An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. *Oncogene*. 2010; 29:3583–3592. [PubMed: 20383195]
- Mi HY, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013; 8:1551–1566. [PubMed: 23868073]
- Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000; 156:297–304. [PubMed: 10978293]
- Nobrega MA, Zhu YW, Plajzer-Frick I, Afzal V, Rubin EM. Megabase deletions of gene deserts result in viable mice. *Nature*. 2004; 431:988–993. [PubMed: 15496924]
- Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L. Evolution and functional classification of vertebrate gene deserts. *Genome Res*. 2005; 15:137–145. [PubMed: 15590943]
- Pasman E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res*. 2007; 67:3963–3969. [PubMed: 17440112]
- Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet*. 2011; 12:136–149. [PubMed: 21245830]
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 2006; 444:499–502. [PubMed: 17086198]
- Poulin F, Nobrega MA, Plajzer-Frick I, Holt A, Afzal V, Rubin EM, Pennacchio LA. In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics*. 2005; 85:774–781. [PubMed: 15885503]
- Prat Y, Fromer M, Linial N, Linial M. Codon usage is associated with the evolutionary age of genes in metazoan genomes. *BMC Evol Biol*. 2009; 9:285. [PubMed: 19995431]
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012; 81(81): 145–166. [PubMed: 22663078]
- Sana J, Hankeova S, Svoboda M, Kiss I, Vyzula R, Slaby O. Expression levels of transcribed ultraconserved regions uc. 73 and uc.388 are altered in colorectal cancer. *Oncology*. 2012; 82:114–118. [PubMed: 22328099]
- Scaruffi P. The transcribed-ultraconserved regions: a novel class of long noncoding RNAs involved in cancer susceptibility. *Sci World J*. 2011; 11:340–352.
- Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010; 329:689–693. [PubMed: 20616235]
- Vemuganti R. All's well that transcribes well: non-coding RNAs and post-stroke brain damage. *Neurochem Int*. 2013; 63:438–449. [PubMed: 23954844]
- Yu WQ, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui HM. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*. 2008; 451:U202–U210.
- Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, Morris Q. GeneMANIA prediction server 2013 update. *Nucleic Acids Res*. 2013; 41:W115–W122. [PubMed: 23794635]

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.neuint.2014.06.006>.

Fig. 1.

In adult rat cerebral cortex, 107 T-UCRs were observed to be expressed. Of those, 61 are intragenic (left pie-chart of A) and 46 are intergenic (right pie-chart of A). On an arbitrary scale of 1–75,000 units, 34% of them are expressed at medium to very high levels. Six T-UCRs accounted for 73% of the total expression units of intragenic and 4 T-UCRs accounted for 91% of the total expression units of intergenic T-UCRs (B).

Fig. 2.

Representative real-time PCR amplification plots of uc.188, uc.419 and 18s rRNA. PCR was conducted with 3 samples in duplicate. The values of different brain samples were observed to be similar (>5% variation). The real-time PCR show that uc.188 expression was ~6-fold higher than uc.419 in cerebral cortex of adult rat. Microarray analysis also showed higher expression of uc.188 compared to uc.419.

Fig. 3.

Genomic locations of the intragenic T-UCR genes indicate that they can be transcribed either in sense or antisense direction to their host genes (A). The intragenic T-UCR genes mainly overlap with the exons of the host genes (A). We correlated the expression levels of 32 intragenic T-UCRs with their host genes. Only 7 T-UCRs showed a positive correlation with the expression of their host genes, while the remaining 25 showed a negative correlation (B). The bars represent the T-UCRs and the circles in the line graph represent the host genes (B). The T-UCRs with positive correlation to their host genes are shown in red (B).

Fig. 4.

FunCoup analysis showed that nucleus is the major site of location of the protein products of the host genes and nearest up- and down-stream genes of the T-UCRs expressed in rat cerebral cortex. The blue nodes indicate mRNA co-expression whereas red nodes represent subcellular co-localization.

Fig. 5.

GO analysis was conducted for the nearest upstream/downstream protein coding genes and the host genes (in case of intragenic T-UCRs) using PANTHER algorithm. Metabolism is the top biological function associated with intragenic T-UCRs while cellular process is the top biological processes associated with intergenic T-UCRs (A). Binding, transcription regulation and catalytic activity are the top molecular functional categories for both intergenic and intragenic T-UCRs (B).

Table 1

Top 15 intragenic T-UCRs expressed in rat cerebral cortex.

T-UCR			Host gene		Upstream gene		Downstream gene	
Name	Length	Expression	Genomic location		Name	Distance	Name	Distance
uc.465 ^c	310	8.28	chrX:63074891–63075200	Pola1	Arx	104,949	Peytlb	272,119
uc.46 ^b	217	7.72	chr13:100671444–100671660	Hmrnpu	Cox20	6296	Efcab2	72,554
uc.154 ^d	203	6.98	chr2:48238548–48238750	Tnpo1	Fcho2	26,610	Pcd2	493,827
uc.471 ^b	239	6.71	chrX:11199647–11199885	Ddx3x	Nyx	182,584	Usp9x	108,530
uc.141 ^a	295	6.66	chr14:61286542–61286836	Dhx15	Sod3	209,497	Kctd8	728,045
uc.420 ^b	233	6.05	chr10:94727631–94727863	Ddx5	Polg2	2672	Cep95	6199
uc.144 ^b	205	5.66	chr14:11146299–11146503	Hmrpdl	Enoph1	4270	Hmrpd	53,788
uc.360 ^c	287	5.65	chr6:76539046–76539332	Novo2	Stxbp6	1637,295	Foxg1	2993,523
uc.455 ^a	245	5.35	chr3:159432679–159432923	Rbm39	Phf20	15,167	LOC 100363776	31,497
uc.472 ^b	202	4.47	chrX:10954072–10954273	Cask	Gpr34	125,008	Nyx	41,754
uc.481 ^b	204	4.45	chrX:128711894–128712097	Stag2	Xiap	168,167	Sh2d1a	266,721
uc.50 ^b	222	3.74	chr6:2861050–2861271	Srsf7	Galm	23,415	Ttc39d	2386
uc.419 ^b	289	3.60	chr10:74763575–74763863	Srsf1	Cuedc1	153,007	Dynll2	65,992
uc.255 ^b	232	3.53	chr5:113452493–113452724	Elavl2	Dmrta1	1202,681	Zfp352	819,005
uc.371 ^a	296	3.14	chr6:85932384–85932679	Ralgapal	Aldoa2	47,514	Brms1l	274,799

Sixty-one intragenic T-UCRs were expressed in rat cerebral cortex. This table shows top 15 of those. The remaining 46 are shown in the Supplementary Table 2.

^a Sense-intron-overlap;

^b sense-exon-overlap;

^c antisense-intron-overlap;

^d antisense-exon-overlap. The genomic locations and the distances (to and from T-UCR) are given in base pairs. The expression level is given as mean quantile normalized value ($n = 3$ /group; $SD < 15\%$ in each case). T-UCR, transcribed ultraconserved region; uc, ultraconserved; Aldoat2, Aldolase 1 A retrogene 2 (NM_001013943); Arx, aristaless related homeobox (NM_00100174); Brms1l, breast cancer metastasis-suppressor 1-like (NM_001106731); Cask, calcium/calmodulin-dependent serine protein kinase (MAGUK family) (NM_022184); Cep95, centrosomal protein 95 kDa (NM_001013862); Cox20, cyclooxxygenase-2 chaperone homolog (NM_001105976); Cuedc1, CUE domain containing 1 (NM_001013971); Ddx3x, DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked (NM_00108246); Ddx5, DEAD (Asp-Glu-Ala-Asp) box helicase 5 (NM_001007613); Ddx15, DEAH (Asp-Glu-Ala-His) box polypeptide 15 (NM_001191597); Dmrtal, DMRT-like family A1 (NM_001107945); Dynl12 (dynein light chain LC8-type 2 (NM_080697); Efcab2, EF-hand calcium binding domain 2 (NM_001105977); Elavl2, ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B) (NM_173309);

Enoph1, enolase-phosphatase 1 (NM_001009391); Echo2, FCH domain only 2 (NM_001191632); Foxg1, forkhead box G1 (NM_012560); Galm, galactose mutarotase (aldose 1-epimerase) (NM_001007704); Gpr34, G protein-coupled receptor 34 (NM_001024925); Hnmpu, heterogeneous nuclear ribonucleoprotein U (NM_057139); Hnmpd, heterogeneous nuclear ribonucleoprotein D (NM_024404); Hnmpd, Hnmpd-like (NM_001033696); Kctd8, potassium channel tetramerisation domain containing 8 (NM_001100172); LOC100363776, RNA-binding region containing protein 2-like (NM_001177904); Nova2, neuro-oncological ventral antigen 2 (NM_001100541); Nyx, nyctalopin (NM_001100967); Pcytlb, phosphate cytidylyltransferase 1, choline, beta (NM_173151); Phf20, PHD finger protein 20 (NM_001107795); Pola1, polymerase (DNA directed), alpha 1 (NM_053479); Polg2, polymerase (DNA directed), gamma 2, accessory subunit (NM_001107060); Ptcid2, pentatricopeptide repeat domain 2 (NM_001107648); Ralgapa1, Ral GTPase activating protein, alpha subunit 1 (catalytic) (NM_020083); Rbm39, RNA binding motif protein 39 (NM_001013207); Sh2d1a, SH2 domain containing 1A (NM_001109313); Sod3, superoxide dismutase 3, extracellular (NM_012880); Srsf1, serine/arginine-rich splicing factor 1 (NM_001109552); Srsf7, serine/arginine-rich splicing factor 7 (NM_001039035); Stag2, stromal antigen 2 (NM_001173507); Sixbp6, syntaxin binding protein 6 (amisy) (NM_001191872); Tnpo1, transportin 1 (NM_001100692); Ttc39d, tetratricopeptide repeat domain 39D (NM_001191906); Usp9x, ubiquitin specific peptidase 9, X-linked (NM_001135923); Xiap, X-linked inhibitor of apoptosis (NM_022231); Zfp352, zinc finger protein 352 (NM_001109357).

Table 2

Top 15 Intergenic T-UCRs expressed in rat cerebral cortex.

T-UCR			Upstream gene		Downstream gene		
Name	Length	Expression	Genomic location	Name	Distance	Name	Distance
uc.84	209	8.59	chr3:48196541-48196749	Nr4a2	7287	Gpd2	98,679
uc.188	215	7.91	chr17:21565232-21565446	Atxn1	5961	Gmpr	3285
uc.45	203	7.67	chr13:100670536-100670	Cox20	5388	Hnmpu	73
uc.75	236	6.17	chr3:35062735-35062970	Gtdcl	54456	Zeb2	939
uc.184	230	3.57	chr10:15880498-15880727	RGD 1311343	25,633	Cpeb4	1386
uc.408	252	3.06	chr19:53585358-53585609	Pnfbp1	579,480	Glg1	225,473
uc.199	256	2.29	chr5:42106693-42106948	Pou3f2	781,297	Mms22.1	1366,650
uc.367	298	2.21	chr6:84022519-84022816	Akap6	476,943	Egln3	570,078
uc.101	254	2.08	chr3:66119619-66119872	Cdca7	478,158	Olal	155,555
uc.193	319	1.94	chr8:95599848-95600166	Snx14	15,900	Syncrip	2615
uc.448	232	1.88	chr1:95133694-95133925	Tshz3	1237,400	Uri1	301,921
uc.18	238	1.68	chr5:140297857-140298094	Tnem53	376,393	Dmap1	25,120
uc.86	340	1.39	chr3:48699085-48699424	Gpd2	272,519	Galnt5	442,431
uc.432	211	1.36	chr18:18268390-18268600	Celf4	437,880	Pik3c3	4426,342
uc.217	221	1.28	chr14:99295138-99295358	Sec61g	221,054	Cld	816,246

Forty-six intergenic T-UCRs were expressed in rat cerebral cortex. The table shows top 15 of those. The remaining 31 are shown in the Supplementary Table 3. The genomic location and the distance (to and from T-UCR) are given in base pairs. The expression level is given as mean quantile normalized value ($n = 3$ /group; $SD < 15\%$ in each case). T-UCR, transcribed ultraconserved region; uc, ultraconserved; Akap6, A kinase (PRKA) anchor protein 6 (NM_022618); Atxn1, ataxin 1 (NM_012726); Cld, C1D nuclear receptor co-repressor (NM_001106021); Cdca7, cell division cycle associated 7 (NM_001025693); Celf4, CUGBP, Elav-like family member 4 (NM_001107400); Cox20, Cox2 chaperone homolog (NM_001105976); Cpeb4, cytoplasmic polyadenylation element binding protein 4 (NM_001106992); Dmap1, DNA methyltransferase 1-associated protein 1 (NM_0010315006); Egln3, EGL nine homolog 3 (*C. elegans*) (NM_019371); Galnt5, UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5) (NM_031796); Glg1, golgi glycoprotein 1 (NM_017211); Gmpr, guanosine monophosphate reductase (NM_057188); Gpd2, glycerol-3-phosphate dehydrogenase 2, mitochondrial (NM_012736); Gtdcl, glycosyltransferase-like domain containing 1 (NM_001024274); Hnmpu, heterogeneous nuclear ribonucleoprotein U (NM_057139); Mms22.1, MMS22-like, DNA repair protein (NM_001135780); Nr4a2, nuclear receptor subfamily 4, group A, member 2 (NM_019328); Olal, Olg-like ATPase 1 (NM_001033927); Pik3c3, phosphoinositide-3-kinase, class 3 (NM_022958); Pnfbp1, polyamine modulated factor 1 binding protein 1 (NM_134393); Pou3f2, POU class 3 homeobox 2 (NM_172085); RGD1311343, similar to RIKEN cDNA 4930524B15 (NM_001108270); Sec61g, gamma subunit (NM_001135020); Snx14, sorting nexin 14 (NM_001108174); Syncrip, synaptotagmin binding, cytoplasmic RNA interacting protein (NM_001047916); Tnem53, transmembrane protein 53 (NM_001107964); Tshz3, teashirt zinc finger homeobox 3 (NM_001107506); Uri1, prefoldin-like chaperone (NM_001107507); Zeb2, zinc finger E-box binding homeobox 2 (NM_001033701).