Cerebral Cortex June 2014;24:1451–1459
doi:10.1093/cercor/bhs414
Advance Access publication February 1, 2013

Developmental Changes in the Transcriptome of Human Cerebral Cortex Tissue: Long Noncoding RNA Transcripts

Leonard Lipovich[1], Adi L. Tarca[1], Juan Cai[1], Hui Jia[1], Harry T. Chugani[2], Kirstin N. Sterner[3], Lawrence I. Grossman[1], Monica Uddin[1], Patrick R. Hof[4], Chet C. Sherwood[5], Christopher W. Kuzawa[6], Morris Goodman[1†] and Derek E. Wildman[1]

[1]Center for Molecular Medicine and Genetics, [2]Department of Pediatrics, Wayne State University, Detroit, MI 48201, USA, [3]Department of Anthropology, University of Oregon, Eugene, OR 97403, USA, [4]Fishberg Department of Neuroscience and Friedman Brain Institute, Mount Sinai School of Medicine, New York, NY 10029, USA, [5]Department of Anthropology, The George Washington University, Washington, DC 20052, USA and [6]Department of Anthropology, Northwestern University, Evanston, IL 60208, USA

[†]Deceased 15 November 2010.

Address correspondence to Leonard Lipovich, Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, Detroit, MI 48201, USA. Email: llipovich@med.wayne.edu

The human neocortex is characterized by protracted developmental intervals of synaptogenesis and myelination, which allow for an extended period of learning. The molecular basis of these and other postnatal developmental changes in the human cerebral cortex remain incompletely understood. Recently, a new large class of mammalian genes, encoding nonmessenger, long nonprotein-coding ribonucleic acid (lncRNA) molecules has been discovered. Although their function remains uncertain, numerous lncRNAs have primate-specific sequences and/or show evidence of rapid, lineage-specific evolution, making them potentially relevant to the evolution of unique human neural properties. To examine the hypothesis that lncRNA expression varies with age, potentially paralleling known developmental trends in synaptogenesis, myelination, and energetics, we quantified levels of nearly 6000 lncRNAs in 36 surgically resected human neocortical samples (primarily derived from temporal cortex) spanning infancy to adulthood. Our analysis identified 8 lncRNA genes with distinct developmental expression patterns. These lncRNA genes contained anthropoid-specific exons, as well as splice sites and polyadenylation signals that resided in primate-specific sequences. To our knowledge, our study is the first to describe developmental expression profiles of lncRNA in surgically resected in vivo human brain tissue. Future analysis of the functional relevance of these transcripts to neural development and energy metabolism is warranted.

Keywords: gene expression, human, in vivo, long noncoding RNA, neocortex

Introduction

Anatomical and functional changes during postnatal life contribute to neuronal plasticity in human infancy and childhood. In particular, exuberant synaptic proliferation and subsequent pruning take place during childhood and into early adulthood (Huttenlocher and Dabholkar 1997; Petanjek et al. 2011). In addition, functional connectivity of MRI-measured long-range networks is not adult-like until early adolescence, and myelination of axonal connections persists until after puberty (Power et al. 2010; Lebel and Beaulieu 2011). The elongated period of neurodevelopment in humans, furthermore, is energetically costly, accounting roughly for a two-fold increase in mass-specific metabolic rate of cortical tissue measured in mid-childhood compared with uptake rates in the same regions in human adults (Chugani et al. 1987).

Previous microarray studies of mostly protein-coding transcripts in human brain revealed that the majority of the expression variance is age-dependent, relating gene expression directly to profound developmental changes in the prefrontal cortex (Weickert et al. 2009). A metabolomic survey of human and nonhuman primate prefrontal cortex revealed that the majority of brain metabolites exhibit age-related changes (Fu et al. 2011), a finding that suggests a need to better understand age-dependent brain gene expression that may be responsible for these metabolic patterns. These studies have provided the first characterizations of gene expression of protein-coding transcripts in the human neocortex as well as brain metabolism in humans and nonhuman primates.

In addition to its ~20 000 protein-coding genes, the human genome encodes nearly 23 000 transcriptional units of mostly unknown function that lack protein-coding potential (Carninci and Hayashizaki 2007). Long noncoding RNA (lncRNA) transcriptional units, which are distinct from other prominent types of functional noncoding RNA (e.g., micro or small RNA), are defined as RNAs that do not have open reading frames greater than 100 amino acids in size, and which also do not have any smaller open reading frames similar to known proteins in public databases (Lipovich et al. 2010). Although the functions of most lncRNAs remain largely uncharacterized, functional roles have been and are still being established for a growing list of lncRNAs.

The recently discovered roles of mammalian lncRNAs are heterogeneous and diverse, revealing the central importance of these molecules in cellular function, development, and disease. In particular, telomerase is a ribonucleoprotein complex, consisting of a protein reverse transcriptase and a highly conserved telomerase lncRNA. This lncRNA serves as the template for telomeric DNA synthesis by the reverse transcriptase (Blackburn et al. 2006). Other lncRNAs serve as key components of nuclear architecture at nuclear speckles and paraspeckles (Wilusz et al. 2009). LncRNAs are central to X chromosome inactivation, which is mediated by the lncRNA XIST (Spencer et al. 2011). Numerous autosomal lncRNAs, such as H19, a direct target of c-Myc, are imprinted and integral in growth, development, and oncogenesis (Keniry et al. 2012). LncRNA importance in early development is well established: notably, specific lncRNAs mediate the maintenance of pluripotency, as well as early cell fate decisions, in the mouse (Sheik Mohamed et al. 2010). Subsequently, similar functions

of long intergenic RNAs (lincRNAs) in human pluripotency were demonstrated (Guttman et al. 2011). Hundreds of these RNAs, including HOTAIR, a central regulator of homeobox transcription factor expression, function through direct interaction with the polycomb repressor complex 2, a key epigenetic modifier (Khalil et al. 2009). Additional lncRNAs, including GAS5 (a ribo-mimic of the glucocorticoid receptor binding site DNA sequence), function as endogenous riboregulators in posttranscriptional, rather than epigenetic, contexts (Bond et al. 2009; Kino et al. 2010).

Besides these global properties, recent work suggests an important role for lncRNAs specifically in neuronal function and, potentially, cortical development. MALAT-1, an lncRNA essential to nuclear architecture, for instance, has been shown in vitro to regulate mRNA levels of synaptic genes (Bernard et al. 2010). Synaptic functions for additional lncRNAs are suggested by the colocalization in synaptosomes of lncRNA-mRNA antisense pairs (Smalheiser et al. 2008). Recently identified lncRNAs with neuronal functions, including MIAT (Ishii et al. 2006), similarly speak to the potentially important role of lncRNA in human neurons (reviewed in Lipovich et al. 2010). Bidirectional transcription of regulatory lncRNAs is a hallmark of protein-coding loci involved in microsatellite expansion disorders of the central nervous system, such as Fragile X Mental Retardation 1, whereas BACE1-AS, another lncRNA, directly contributes to Alzheimer's disease pathogenesis (Pastori and Wahlestedt 2012). These studies summarily point to the need for additional research into the role of lncRNA in regulating gene expression in the developing human cerebral cortex.

As a step toward this goal, we report here the first measurement of lncRNA expression profiles in human neocortical samples collected during brain surgery. These samples are from 36 individuals spanning in age from infancy to adulthood, thus allowing identification of lncRNA transcripts that show age-linked variation in expression to identify potential mechanisms of postnatal changes in neuronal and synaptic function and underlying metabolic processes. This focus on lncRNA complements our recent parallel examination of protein-coding gene expression patterns in these samples (Sterner et al. 2012). To examine lncRNA transcripts, we used a unique custom microarray containing 8 probes for each of 5586 human lncRNA genes selected from our human lncRNA catalog (Jia et al. 2010), and statistical techniques capable of detecting transcripts that vary in expression as linear or curvilinear functions of age while accounting for multiple comparisons.

Materials and Methods

Human in vivo brain sample procurement from curative surgeries and selection for microarray analysis

Samples used for this study were obtained from surgically resected tissue originally removed from patients receiving treatment for neurological pathologies including epilepsy. Our study was designed after these surgeries were completed and did not influence any of the clinical decisions. As part of these surgical procedures, a small amount of healthy tissue was removed to ensure full excision of the unhealthy tissue. Tissues were flash-frozen in dry ice and stored at −80°C. Tissue samples were homogenized in TRI Reagent (Applied Biosytems/Ambion, Austin, TX, USA). We extracted RNA from tissues by using either the MagMax-96 kit (Applied Biosystems/Ambion) or a commercial version of the TRIzol protocol (Invitrogen, Carlsbad, CA, USA). After initial isolation, we further purified RNA, by either the TURBO DNase treatment (Applied Biosytems/Ambion), or the RNeasy kit used in conjunction with the RNase-Free DNase Set (Qiagen, Valencia, CA, USA). We assessed the concentration and the quality of the isolated DNA-free RNA samples, on a Nanodrop 1000 instrument (Thermo Scientific, Wilmington, DE) and the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), respectively. We proceeded with further analysis only on samples which had ABS 260/280 ratios greater than 1.7 with no evidence of degradation. The study was performed in accordance with all applicable guidelines (Wayne State University IRB approval: HIC # 071608MP4X).

The 36 human neocortex samples used in this analysis were drawn from roughly comparable numbers of males and females, and represented an age range from 11 months to 47 years. Most, but not all, of the samples were from the temporal lobe (Table 1; Sterner et al. 2012). Specifically, 29 were obtained from the temporal lobe and 7 from other regions. In hopes of maximizing sample size, we evaluated whether model results would vary depending on whether the analyses were limited to the temporal lobe samples versus pooling all samples. We found that pooling samples and limiting analyses to temporal lobe samples yielded similar results. Consequently, we report here the results of pooled analyses (see detailed discussion in Supplementary Analyses). Samples were obtained primarily from epilepsy patients, and were derived from electrophysiologically least-active (nonspiking) regions as defined by 3-dimensional presurgery whole-brain electrode profiling.

LncRNA Discovery and Microarray Design

Discovery and construction of our human genomewide lncRNA dataset were previously reported (Jia et al. 2010). We used our

Table 1

Demographic and phenotypic characteristics of the 36 human in vivo brain samples used in the developmental time-course microarray analysis

Sample name	Age (years)	Sex	Race/ ethnicity	Year of tissue collection	Cortical region
R420	8.75	M	W	2007	Temporal
R421	11.58	M	A	1999	Occipital
R422	14.25	M	W	2008	Frontal
R425	6.17	M	W	2008	Temporal
R426	0.92	F	A	2006	Frontal
R430	6	F	W	2001	Temporal
R431	15.58	F	B	2005	Temporal
R432	3	M	W	1994	Temporal
R433	10.92	M	W	2004	Temporal
R435	7.17	M	W	2004	Inferior occipital
R436	10.17	M	W	2008	Temporal
R437	10.33	M	W	2004	Temporal
R438	11.08	M	W	2002	Temporal
R439	11	F	B	2002	Frontal
R442	8.58	M	W	2002	Temporal
R445	6.25	M	W	1998	Frontal
R446	8.08	F	W	2002	Temporal
R447	2.58	M	W	2002	Temporal
R449	8.17	F	W	1997	Temporal
R451	42.5	F	W	1997	Temporal
R452	8	M	N/A	1996	Temporal
R454	3.42	F	N/A	1996	Temporal
R455	47.83	M	W	1996	Temporal
R458	10.5	F	W	2003	Temporal
R460	17.67	M	W	2003	Temporal
R462	43.5	F	W	2000	Temporal
R473	36.92	M	W	1999	Temporal
R475	3.33	F	W	1999	Temporal
R476	3.5	M	Hispanic	2003	Temporal
R477	15.92	M	N/A	1999	Temporal
R481	13.67	F	N/A	1999	Temporal
R487	7.42	M	W	1999	Posterior Temporal
R490	0.92	F	W	1997	Temporal
R491	2.58	M	W	1994	Parietal
R492	24.33	F	W	1996	Temporal
R493	7.33	M	N/A	1996	Temporal

M, male; F, female; A, Asian; B, Black; W, White; N/A, not available.

lncRNA catalog to design, test, validate, and implement custom lncRNA expression analysis microarrays. We selected the Agilent high-density platform because it has the advantages of genomewide screening, extremely low background signals, and reproducibility that are critical for us to expand our ability to detect both common and unique gene expression changes in the human neocortex during development. Our choice of a custom microarray platform was motivated by cost advantages relative to RNAseq, and by the known shortcomings of commercial microarray platforms, which underrepresent and misrepresent genomically complex loci, including those encoding lncRNA genes (Orlov et al. 2007; Jia et al. 2010).

Of the 6736 unique lncRNA genes in our dataset, 5586 are represented (by 8 probes per unique gene) on our custom human lncRNA microarray. The remaining lncRNAs failed Agilent EArray probe design or RepeatMasker-based and/or BLAT-based postdesign checks of genomic specificity, and were therefore not pursued further. Only one reference transcript per lncRNA gene was selected for probe design on the array. The majority of lncRNAs in our dataset, and hence on our custom microarray, are not redundant by genomic position relative to 2 major public datasets of "lincRNAs" and other long ncRNAs (Fig. 1*A*). In addition to the lncRNA genes, 161 known protein-coding control genes (6 housekeeping genes, 111 known genes expressed in the human neocortex (Beaumont et al. 2012), and 44 known brain-expressed genes from a postmortem study of human drug abusers (Albertson et al. 2006) were represented on the microarray as positive controls, since detecting the expression of these genes in our human brain samples would help validate our microarray platform as suitable for brain lncRNA profiling as well.

Eight 60-mer probes per gene, unambiguously mapping by BLAT to a single genomic location and free of interspersed and simple repeats, were designed using the Agilent Technologies OpenGenomics eArray interface for the reference transcript of each lncRNA and for each known-gene control mRNA (Supplementary Dataset accession number GSE43639). Each probe had a different sequence. All 8 probe sequences, for each gene, matched the lncRNA or control target sequence. The probes were designed with a bias toward the 3′ end in order to maximize the efficiency of hybridization to copies of a reverse-transcribed template, because reverse transcriptase synthesizes first-strand cDNA in a 3′-to-5′ direction relative to the RNA template and may not always reach the 5′end of the template. After first-strand cDNA synthesis (antisense to the original lncRNA or mRNA) and filling of the cDNA second-strand, antisense cRNA was synthesized and labeled. The sense-orientation probes hybridize to the antisense cRNA (Fig. 1*B*). Microarrays were custom-manufactured in the 1 × 244k (244 000 features per slide) format (Agilent Technologies), with 5 randomly placed on-chip replicates for each of the 8 probes per gene, a design aspect that aimed to enhance array performance through the provision of multiple data points to gauge the extent of within-array technical variation and to increase technical reproducibility. A 1-color (Alexa 555 green) microarray experiment was performed separately on each of the 36 samples.

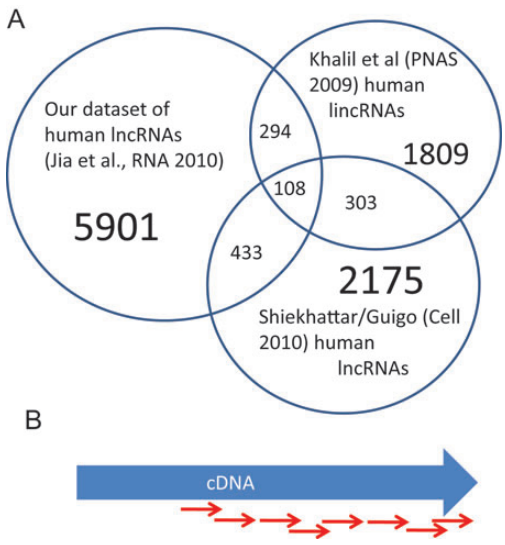

Figure 1. Genome-wide lncRNA custom microarray analysis of the human in vivo brain during the developmental time-course from early childhood to adulthood: experimental design and results. (*A*) Venn diagram view of the {5901 + 433 + 108 + 294 = } 6736 lncRNA genes in our lncRNA dataset (Jia et al. 2010) relative to 2 major public datasets of "lincRNAs" and other long ncRNAs (Ørom et al. 2010; Khalil et al. 2009). Source: Lipovich et al. (2012), Supplementary File S4, Supplementary Exhibit S4.2. Permission granted by the Genetics Society of America. (*B*) The relationship between the 8 distinct oligonucleotide probes that we designed for each gene (red) and the cDNA of the gene whose expression is being profiled (blue). Arrows indicate the biological direction of transcription of the gene.

The on-chip replicates allowed us to confirm that microarray noise, as inferred from expression level variability among the 5 replicates for each probe, was low. The median coefficient of variation of the probes was 2.6%, with approximately 75% of the probes having a coefficient of variation less than 6% (Supplementary Analyses).

Statistical Analysis: Detection of Transcripts and Identification of Age Trends in Expression

Data preprocessing was performed with Agilent Feature Extraction software. We defined a gene (lncRNA or protein-coding control) to be detectable (present) if at least one on-chip replicate (out of the 5 on-chip replicates) of at least 1 probe (out of the 8 probes per gene) was detectable by the Agilent Feature Extraction "well above background" (gIsWellAboveBg = 1) criterion in at least half (15) of our 31 samples with age less than 18 years shown in Table 1. Samples from patients older than 18 were not included in further analysis, since datapoints after 18 years were too sparse (only 5 samples covering the 18-48 years interval) to allow nonlinear modeling of the data. The reason we required only one probe to be detectable is that probes were designed from public flcDNA clones, none of which are from cortical cDNA libraries. These cDNA clones do not necessarily represent the structure (5′, 3′ends, splicing) of actual transcripts from the same locus in the brain where alternative initiation, splicing, and/or polyadenylation may all occur.

We applied a linear model to fit the log-transformed expression level of each probe as a second degree polynomial function of age. This function, which concurrently adjusted for sex of each individual, is represented by the equation:

$$y_i = a_0 + a_1 \cdot \mathrm{AGE}i + a_2 \cdot \mathrm{AGE}_i^2 + a_3 \cdot \mathrm{SEX}_i + a_4 \cdot D_i + \varepsilon_i.$$

In the above equation, i denotes the ith of the available samples, AGE indicates age of the patient in years, and SEX indicates the patient's gender.

Nominal P-values based on F-tests were calculated for each gene representing the probability that both coefficients related to age (a1 and a2) are null. We applied the False Discovery Rate procedure (Benjamini and Hochberg 1995) in order to correct nominal P-values for multiple testing. These P-values quantify the evidence for association between the microarray probe expression level and the individual's age, while adjusting for sex. Then, we sorted genes as a function of these nominal P-values. We defined all genes with pFDR <0.25 as being significantly associated with age. Our model assumes a quadratic relationship between gene expression levels, on a log scale, and age. We developed this model as a compromise between 2 conflicting needs: The necessity to fit nonlinear trends, and the need to minimize the number of parameters in the model in order to ensure adequate power.

Classification of Developmental Patterns

The quadratic relationship between gene expression and age allowed a limited number of profile types to be observed as determined by the sign and the magnitude of the a1 and a2 coefficients in the equation above for each probe. Individual probes fell into 1 of 8 profile classes derived from the fit of their expression as a function of age:

1. Linear up: No inflection point in the curve and the coefficient a1 is positive.
2. Linear down: No inflection point in the curve and the coefficient a1 is negative.
3. u-shape: Minimum value on the curve is lower than both the fitted values at age = 0 and 18 by at least 1.25-fold.
4. n-shape: Maximum fitted value on the curve is higher than both the values at age = 0 and 18 by at least 1.25-fold.
5. Up then flat: There is an inflection point but the expression at age 18 does not drop enough to be a "n-shape".
6. Flat then down: There is an inflection point but the expression at age 0 does not drop enough to be a "n-shape".
7. Down then flat: There is an inflection point but the expression at age 18 does not rise enough to be a "u-shape".
8. Flat then up: There is an inflection point but the expression at age 0 does not rise enough to be a "u-shape".

Results

We first examined the extent to which the lncRNA genes and the protein-coding control genes from our microarray were reproducibly present in the 36 human in vivo surgically excised brain samples. When detectability was defined using minimally conservative criteria (see Materials and Methods), 3952 (71%) of the lncRNA genes and 147 (91%) of known brain-expressed protein-coding control genes were detectable (Supplementary File S1). We then investigated which of these detectable lncRNA genes were also differentially expressed in an age-dependent manner across the ages represented by our samples. We defined a gene to be differentially expressed if at least 1 of its 8 probes significantly fit 1 of our 8 curve-type differential expression trends, adjusting significance for multiple comparisons (see Methods).

Using these methods, we identified 32 individual lncRNA probes with evidence of age-dependent differential expression (Supplementary File S2). Manual annotation of the corresponding 11 transcriptional units in the UCSC Genome Browser revealed that 3 of the lncRNA transcriptional units localized in close (<10 kb) proximity to, and in the same transcriptional orientation as, extended 3′-end untranslated regions (3′UTRs) of nearby protein-coding genes. Because these 3 differentially expressed loci showed evidence for protein-coding activity and thus may not correspond to lncRNA genes, we did not pursue them further. The underlying lncRNA annotation behind our array is still changing, and it is theoretically possible that some of our lncRNAs will be re-classified by projects such as Gencode (Harrow et al. 2006) into protein-coding transcripts, just as some formerly protein-coding transcripts will be re-annotated as new lncRNAs that are not on this array; our determination of these transcripts as lncRNAs was based on the best available protein-coding capacity data at the time when this custom array was designed. The 8 remaining loci identified as following age trends (specific age-dependence trends for each microarray probe of each lncRNA locus; Supplementary File S3) represented bona fide lncRNA genes (Table 2), and all 8 were represented by full-length cDNA clones (corresponding to full-length mRNA molecules) from the NCBI database GenBank. We performed extensive manual annotation of the 8 lncRNAs from Table 2 by visual curation of the 46-species MultiZ alignment, within the context of each lncRNA gene's genomic structure, in the UCSC Genome Browser (http://genome.ucsc.edu). We analyzed all splice donors (GT-), splice acceptors (-AG), and canonical polyadenylation signals (AnTAAA) of these lncRNA genes for interspecies

Table 2
The 8 age-dependent differentially expressed lncRNA genes inferred from the developmental time-course microarray analysis of the human in vivo brain samples

lncRNA GenBank accession number	Number of probes on our custom microarray (out of 8) with significant age-dependent differential expression	Annotation	Conservation of lncRNA gene structure (splice sites and polyadenylation signals)
AK093356	8	Spliced. Public cDNA/EST GenBank data supports expression in brain (medulla). Antisense to UBE2D3; divergent, head-to-head transcript from an internal promoter of UBE2D3.	Single intron: GT- conserved in eutherians. -AG is in an ERV-L LTR, primate-specific repeat insertion. ATTAAA polyA is in the same primate-specific repeat.
BC037918	3	Spliced. Public cDNA/EST GenBank data supports expression in brain (hypothalamus). Genomic span: 432 kb. Broad Inst lincRNA support.	Single intron: GT- is in an ERV1 LTR present only in the great apes; -AG is nonrepeat but conserved in only a subset of eutherians. No canonical polyA signal.
BC080570	3	Unspliced. Antisense to an exon of FUT2. Supported by several ESTs.	No canonical polyA signal. Hence, could not assess gene structure conservation.
BC062470	2	Spliced. Near, but distinct from, PIGP.	Single AT-AC intron. AT- and -AC are present only in the great apes. No repeats. ATTAAA polyA signal is conserved in eutherians.
AK123971	3	Unspliced. Near, but distinct from, SSH1.	AATAAA polyA signal is only in human, great apes, and baboon. Rhesus, no alignment. Marmoset, diverged from consensus. Sequence is not a repeat but is absent in prosimians and all nonprimates.
CR626620	1	Spliced. Repetitive sequences are prominent in exons. Excellent cDNA/EST and Broad Inst lincRNA support. Shares a bidirectional promoter with NUP50. Public cDNA/EST GenBank data supports expression in fetal brain, retina, and placenta.	The CR626620 isoform has 2 introns. Intron 1 –AG and intron 2 GT- are within the same primate-specific FLAM_C Alu element. Intron 2: -AG is within an ERV-L LTR (MLT1J) present in some but not all eutherians. No canonical polyA signal.
BC004397	1	Spliced. Repetitive sequences are prominent in exons. Public cDNA/EST support. Antisense to AFAP1, and is hence known as AFAP1-AS1 in public databases.	BC043614 (upstream full-length cDNA) has 1 intron. GT- is within an ERV-L LTR (THE1A) repeat insertion that is present only in the great apes and old world monkeys. –AG is not in a repeat, but is primate-specific (substitutions abolish this splice acceptor beyond the prosimians). ATTAAA polyA signal is conserved in some but not all eutherians.
BC050676	1	Spliced. Repetitive sequences are prominent in exons. 1st 2 exons are shared with C7ORF49 coding gene, but transcript is noncoding. 3rd exon is unique, intronic antisense to AGBL3 gene.	Intron 1 GT-AG, intron 2 GT- splice junctions conserved in eutherians. -AG is completely primate-specific, in L1 repeat. AATAAA polyA signal is also primate-specific, in a different, L3, repeat.

conservation in this genomewide multispecies alignment. We determined that the majority (7 of the 8) of our age-dependent differentially expressed in vivo human neocortex lncRNAs depend on at least 1 primate-specific gene structure feature (a splice site and/or a polyA signal) for their existence. These results are summarized in the "Conservation of lncRNA gene structure" column of Table 2.

It is well established in gene expression literature (e.g., Johnston et al. 2007) that microarray-based measurements of gene expression must be validated by a microarray-independent method, such as quantitative real-time PCR (qRTPCR), that measures the expression of the same genes in the same series of samples. If the qRTPCR measurements closely parallel the microarray measurements, then the validation is successful; we have used this approach in our previous studies of gene expression in the human brain (e.g., Michelhaugh et al. 2011). We chose the Taqman platform (Applied Biosystems) for qRTPCR. We considered which specific probes were responsible for the differential expression of the selected lncRNA gene (BC037918) on our microarray, and designed custom Taqman primers-and-probe combinations to target only the region of the selected transcript that overlapped with the differentially expressed probes, and not any other regions of the transcript. This allowed us to evaluate gene expression fold change between different age timepoints using 2 independent methods (microarray and qRTPCR), and to determine whether the methods agreed with one another. The Pearson's correlation coefficient calculated between the array-derived and PCR-derived expression values from the same samples was 0.61, $P = 3 \times 10^{-7}$ (Fig. 2*A*), and the curve shape and inflection point derived from the PCR data recapitulated those derived from the array data (Fig. 2*B* and 2*C*). Accordingly, for this particular lncRNA, quantitative real-time PCR validates age-dependent lncRNA differential expression inferred by custom microarray analysis.

Discussion

In contrast to recent studies of the human brain transcriptome that utilized postmortem samples (e.g., Fu et al. 2011; Kang et al. 2011), our samples, obtained in vivo during brain surgeries, comprise a unique and underutilized resource for studying gene expression. By applying our custom microarray platform to this resource, we identified 8 lncRNA genes that followed distinct, statistically significant age trends. The fact that only 8 of 5586 lncRNAs demonstrated these trends implicates these particular RNAs, which we have manually ascertained to be bona fide noncoding transcripts, as likely contributing to age-dependent neocortical function. To our knowledge, these findings are the first to point to possible roles of long, nonprotein-coding transcripts in specifically developmental, age-dependent regulation of brain gene expression.

Mammalian genomes are characterized by abundant occurrence of a genomic phenomenon known as Complex Loci (Engström et al. 2006), defined as genes near, or overlapping, other genes. Bidirectional gene arrangements at these complex loci include sense-antisense gene pairs where genes are encoded by opposite DNA strands within the same locus, as well as bidirectional promoters shared by paired genes that are transcribed in opposite orientations. Seven of our 8 age-dependent lncRNAs resided in close genomic proximity to, or in a sense-antisense pair with, known protein-coding genes. In view of the extensive evidence for functional regulation by lncRNAs of genomic-neighbor protein-coding genes, including in sense–antisense pairs (Lipovich et al. 2010), a possible regulatory role of these lncRNAs of their nearby and overlapping protein-coding genes warrants additional investigation.

Profiling multiple probes per gene on our custom microarray adds a level of reliability compared with conventional microarrays that use one probe per gene because we can check whether only some, or most, or all probes representing the same gene follow consistent age-dependent expression patterns. Our confidence in age-dependent expression of the gene is increased when we see a majority of, or all, probes for that gene showing the same age-dependent expression pattern.

Here, we focus on 2 lncRNA genes exhibiting the greatest between-probe consistency in expression patterns. The lncRNA gene whose reference cDNA sequence corresponds to NCBI GenBank accession # AK093356 had all 8 probes differentially expressed, and # BC037918 had 3 of 8 differentially expressed. Both lncRNA genes showed prior evidence of human brain expression as determined from the tissue origin information available in public databases (GenBank repository of DNA sequences; http://www.ncbi.nlm.nih.gov/entrez) The present analysis provides the first evidence of neocortical brain transcript localization of these 2 novel human lncRNA genes, and also shows that they follow expression patterns that vary with stage of development. It is notable that these lncRNAs have previously been shown, in public GenBank data, to be expressed in noncortical regions of the brain (AK093356 in medulla, BC037918 in hypothalamus). We extend upon the previous public GenBank data by demonstrating neocortical expression of the lncRNAs, which raises the possibility that these lncRNAs have functions in multiple regions of the brain.

In light of the potential functional role of these transcripts in human brain, we characterized the 2 lncRNAs at the sequence level to clarify their likely functional relevance and time of evolutionary emergence. The origins of many repetitive sequences in genomic DNA have been traced to known evolutionary timepoints using the molecular clock, thus rendering these sequences useful as evolutionary markers. In particular, Alu elements and other repeats, such as certain specific subclasses of endogenous retrovirus (ERV) sequences, are useful markers of primate specificity because these sequences first appeared after the origin of primates and are not shared by other mammals (Batzer and Deininger 2002). If an exon of a gene contains a primate-specific repetitive element, the exon likely either arose or became modified during primate evolution. In this respect, it is notable that the lncRNA gene AK093356 has an ERV-containing terminal exon. Although ERV repeats, as a class, are not unique to primates, this particular ERV insertion clearly corresponds to sequence that is only found in primates (red rectangle, Fig. 3*A*), suggesting that the origin of this exon may postdate the emergence of the primate lineage. Interestingly, 4 of our 8 differentially expressed lncRNA genes had ERV insertions that contributed splice sites or polyadenylation signals.

AK093356 also shares a CpG-island bidirectional promoter with the ubiquitin-conjugating enzyme E2D 3 (UBE2D3) gene (Fig. 3*A*). This is interesting because bidirectional promoters often join conserved protein-coding genes with less

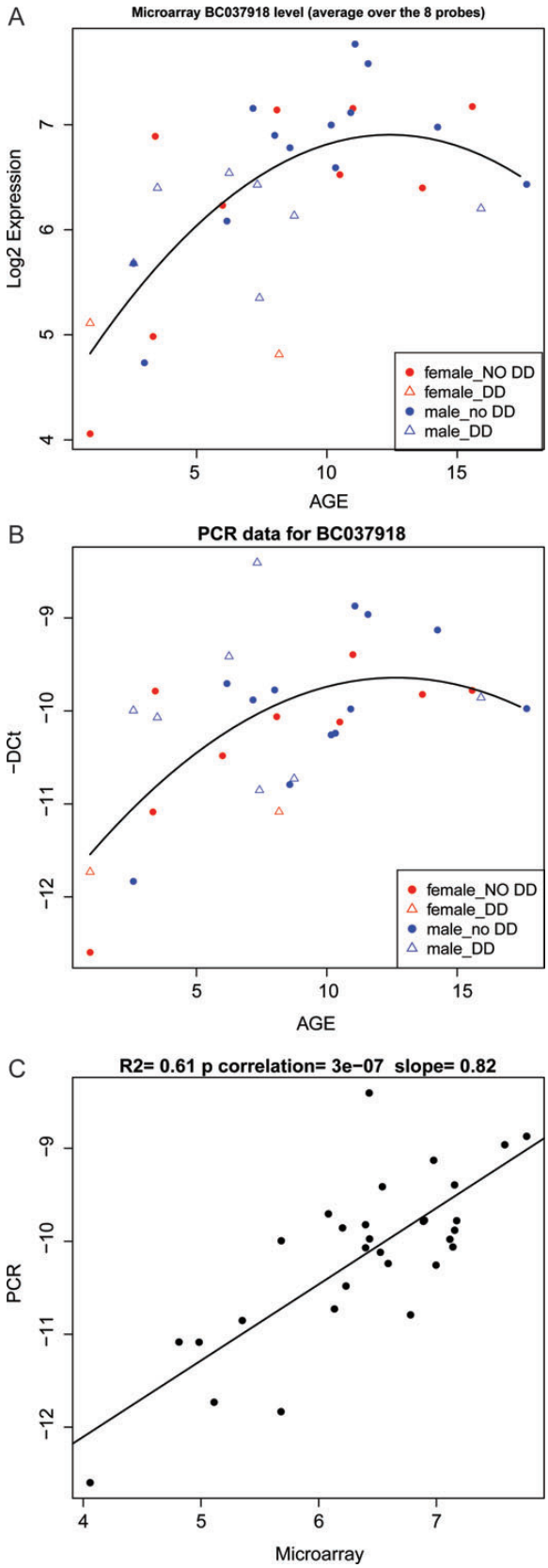

Figure 2. Validation of custom lncRNA microarray results by quantitative real-time PCR for the lncRNA BC037918. (*A*) Plot of microarray and quantitative real-time PCR results for the human surgically resected brain samples, showing Pearson's correlation coefficient. (*B*) Microarray-derived quadratic-fit curve of expression levels on age, based on all 8 microarray probes. (Each probe had a different sequence, and there were also 5 on-chip replicates for each of the right probes, for every lncRNA and control gene.). (*C*) PCR-derived quadratic-fit curve of expression levels on age recapitulates the microarray-derived curve in Figure 2*b*.

conserved lncRNA genes, and the AK093356-UBE2D3 bidirectional promoter is an example of this class.

Many primate-specific genes, including lncRNAs, have acquired novel functions in recent evolution (Tay et al. 2009). Although representing only circumstantial evidence, it is intriguing that the lncRNA gene AK093356, which contains primate-specific repeats in its exons and is also not conserved in mouse, follows an age-dependent brain expression pattern that roughly mirrors derived age changes in human neuronal functions, such as synaptogenesis and related changes in neuronal metabolism (Chugani et al. 1987). Indeed, all 8 probes corresponding to AK093356, our highest-scoring lncRNA, have bell-shaped differential expression curves that share a common inflection point at ~11 years of age (representative probe: Fig. 3*B*). The second highest-ranking lncRNA gene (BC037918) displays a remarkably similar age-dependent expression profile and age of inflection (Figs 2*B* and 3*B*). This pattern of expression corresponds roughly with age trends in many aspects of brain development, including changes in synapse density and shifts in mass-specific energy uptake as reflected in PET-based estimates of cerebral glucose uptake (Chugani et al. 1987; Huttenlocher and Dabholkar 1997).

Although the developmental trend in expression at these 2 loci is notable, the genomic context of these RNAs provides only limited clues to their function. AK093356 may cis-regulate the UBE2D3 gene through the shared bidirectional promoter, but our analyses exclude such a role. We have interrogated all lncRNAs from Table 2 for coexpression with the protein-coding neighbor genes mentioned in that table, using our companion dataset where all protein-coding genes had been profiled in the same samples (Sterner et al. 2012), and did not observe significant coexpression (Supplementary Analyses). The potential functions of standalone lncRNAs such as BC037918 are even more obscure. This transcript is expressed from an apparent gene desert, a chromosomal region far from any known protein-coding genes. Determining the function of these RNAs would require additional approaches, including reverse genetics, which can determine lncRNA function by turning off (using RNA interference) or on (using overexpression) an lncRNA in a human neuronal cell line or primary neuronal culture, and by then screening the cells for phenotypic changes. A complementary approach would be direct identification of protein-complexes that bind to lncRNAs. Using an lncRNA as a bait to trap interacting proteins, which can then be characterized by mass spectrometry, has also proved to be an effective method of assigning cellular functions to novel lncRNAs (Willingham et al. 2005).

Of our 8 age-dependent differentially expressed lncRNAs, 7 were informative for UCSC Genome Browser annotation, whereas 1 lacked any splice sites or canonical polyadenylation signals. All 7 had at least 1 gene structure element (a splice donor, splice acceptor, or polyadenylation signal) that was demonstrably primate-specific, either harbored in a primate-specific (e.g., Alu) repetitive element class or present in single-copy sequences that diverged beyond the GT-, -AG, or AnTAAA consensus in nonprimates. The existence of these

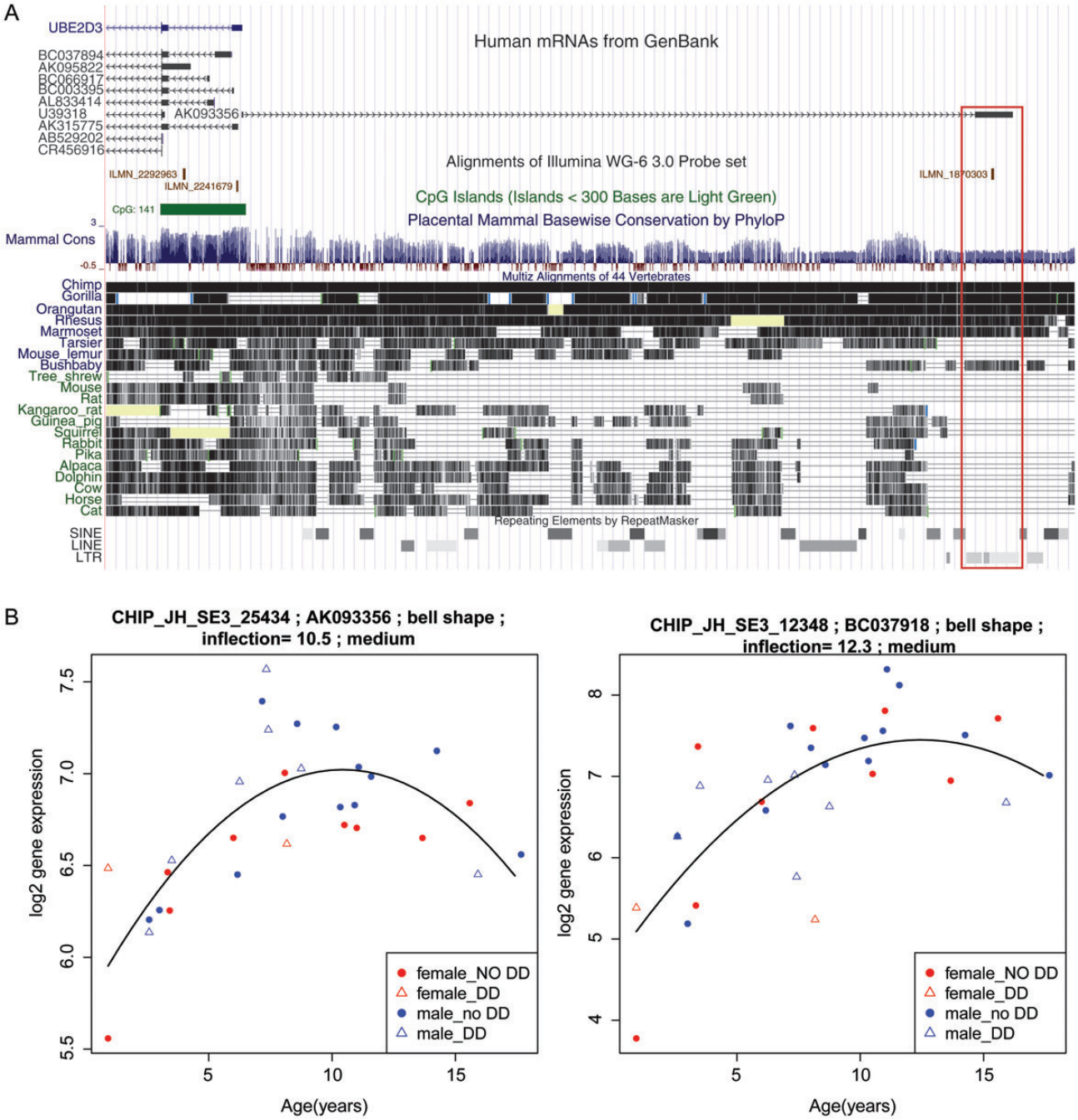

Figure 3. UCSC Genome Browser views and expression level plots for selected age-dependent differentially expressed lncRNA genes. (*A*) UCSC Genome Browser (Kent et al. 2002) view of the lncRNA AK093356. Red box shows the exon that is completely contained within a primate-specific repetitive element (ERV-L LTR unalignable to any species outside of primates). (*B*) Expression plots show $\log_2$ transformed and normalized lncRNA expression levels as a function of age for selected genes. The curve (black) is a quadratic fit of the expression levels on age. A representative probe of AK093356, left (CHIP_JH_SE3_25434; AK093356; bell-shape; inflection = 10.5; medium) and a representative probe for BC037918, right (CHIP_JH_SE3_12348; BC037918; bell-shape; inflection = 12.3; medium) are shown.

primate-specific gene structure features in these differentially expressed lncRNAs suggests that gene structure differences may exist between primate and nonprimate orthologs of these genes. Whether these structure differences relate to the developmental expression differences, which were observed by us in humans, still remains to be determined. However, rapid evolution of lncRNA genes even within primates, with sequence differences resulting in human–chimpanzee distinctions of an orthologous lncRNA gene's transcript secondary structure, has already been described (Pollard et al. 2006). Therefore, these 7 of our 8 lncRNAs, whose rare and significant differential expression is consistent with developmental function in the human brain, should also be considered as candidates for harboring functional interspecies structure differences.

Here, we analyzed a limited number of samples with a relatively coarse temporal resolution, and we also lacked information on developmental tempo among study subjects. Thus, we cannot rule out the possibility that individuals vary in developmental tempo independent of calendar age, which

might bias the results presented here, or limit our ability to detect developmentally based changes in gene expression. Ideally, future analyses should include additional samples to allow finer temporal resolution and standardize individual transcription data to an objective measure of the developmental status of each study participant rather than calendar age.

Despite these issues of sample availability and phenotyping constraints, we here describe statistically significant developmental changes in expression of 8 long noncoding RNA genes measured in human cortical tissue. The 2 genes that follow most consistent age-related patterns rise to peak expression in mid-childhood, when many neuronal processes are active in human brain development. We also find evidence for primate-specific origins of these loci, consistent with their relatively recent evolutionary origin. However, our analyses are incapable of establishing the regulatory or other functional roles of these genes, which await future work using complementary methods. Our findings underscore the need for research to consider the contribution of the extensive noncoding regions of the human genome to brain development. The genes identified as following significant developmental trends in our age series are an excellent starting point for future investigations in this important area.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

This work was supported by the National Science Foundation (grant numbers BCS 0827546, BCS 0827531, BCS 0550209, and DBI 0965741) and the James S. McDonnell Foundation (grants 22002078 and 220020293).

Notes

We thank Amy Weckle and Lucie Grégoire for RNA isolation. *Conflict of Interest*: None declared.

References

Albertson DN, Schmidt CJ, Kapatos G, Bannon MJ. 2006. Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse. Neuropsychopharmacology. 31:2304–2312.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet. 3:370–379.

Beaumont TL, Yao B, Shah A, Kapatos G, Loeb JA. 2012. Layer-specific CREB target gene induction in human neocortical epilepsy. J Neurosci. 32:14389–14401.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc Ser B (Methodol). 57:289–300.

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F et al. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. 29:3082–3093.

Blackburn EH, Greider CW, Szostak JW. 2006. Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging. Nat Med. 12:1133–1138.

Bond AM, Vangompel MJ, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD. 2009. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci. 12:1020–1027.

Carninci P, Hayashizaki Y. 2007. Noncoding RNA transcription beyond annotated genes. Curr. Opin. Genet. Dev. 17:139–144.

Chugani HT, Phelps ME, Mazziotta JC. 1987. Positron emission tomography study of human brain functional development. Ann Neurol. 22:487–497.

Engström PG, Suzuki H, Ninomiya N, Akalin A, Sessa L, Lavorgna G, Brozzi A, Luzi L, Tan SL, Yang L et al. 2006. Complex loci in human and mouse genomes. PLoS Genet. 2(4):e47.

Fu X, Giavalisco P, Liu X, Catchpole G, Fu N, Ning ZB, Guo S, Yan Z, Somel M, Pääbo S et al. 2011. Rapid metabolic evolution in human prefrontal cortex. Proc Natl Acad Sci USA. 108:6181–6186.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al. 2011. LincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 477:295–300.

Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Lagarde J, Gilbert JG, Storey R, Swarbreck D et al. 2006. GENCODE: producing a reference annotation for ENCODE. Genome Biol. 7(S4):1–9.

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

Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, Takahashi A, Miyamoto Y, Ikegawa S, Kamatani N, Hori M et al. 2006. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. J Hum Genet. 51:1087–1099.

Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L. 2010. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. RNA. 16:1478–1487.

Johnston DS, Jelinsky SA, Zhi Y, Finger JN, Kopf GS, Wright WW. 2007. Identification of testis-specific male contraceptive targets: insights from transcriptional profiling of the cycle of the rat seminiferous epithelium and purified testicular cells. Ann N Y Acad Sci. 1120:36–46.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G et al. 2011. Spatio-temporal transcriptome of the human brain. Nature. 478:483–489.

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W. 2012. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol. 14:659–665.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res. 12:996–1006.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin modifying complexes and affect gene expression. Proc Natl Acad Sci USA. 106:11667–11672.

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. 2010. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal. 3:RA8.

Lebel C, Beaulieu C. 2011. Longitudinal development of human brain wiring continues from childhood into adulthood. J Neurosci. 31:10937–10947.

Lipovich L, Dachet F, Cai J, Balan K, Bagla S, Jia H, Loeb JA. 2012. Activity-dependent human brain coding/non-coding gene regulatory networks. Genetics. doi:10.1534/genetics.112.145128 (e-publication ahead of print).

Lipovich L, Johnson R, Lin CY. 2010. MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. Biochim Biophys Acta. 1799:597–615.

Michelhaugh SK, Lipovich L, Blythe J, Jia H, Kapatos G, Bannon MJ. 2011. Mining affymetrix microarray data for long non-coding RNAs: altered expression in the nucleus accumbens of heroin abusers. J Neurochem. 116:459–466.

Orlov YL, Zhou J, Lipovich L, Shahab A, Kuznetsov VA. 2007. Quality assessment of the affymetrix U133A&B probesets by target sequence mapping and expression data analysis. In Silico Biol. 7:241–260.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell. 143:46–58.

Pastori C, Wahlestedt C. 2012. Involvement of long noncoding RNAs in diseases affecting the central nervous system. RNA Biol. 9: (e-publication ahead of print).

Petanjek Z, Judaš M, Šimic G, Rasin MR, Uylings HB, Rakic P, Kostovic I. 2011. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. Proc Natl Acad Sci USA. 108: 13281–13286.

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 443:167–172.

Power JD, Fair DA, Schlaggar BL, Petersen SE. 2010. The development of human functional brain networks. Neuron. 67:735–748.

Sheik Mohamed J, Gaughwin PM, Lim B, Robson P, Lipovich L. 2010. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA. 16:324–337.

Smalheiser NR, Lugli G, Torvik VI, Mise N, Ikeda R, Abe K. 2008. Natural antisense transcripts are co-expressed with sense mRNAs in synaptoneurosomes of adult mouse forebrain. Neurosci Res. 62:236–239.

Spencer RJ, del Rosario BC, Pinter SF, Lessing D, Sadreyev RI, Lee JT. 2011. A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. Genetics. 189:441–454.

Sterner KN, Weckle A, Chugani HT, Tarca AL, Sherwood CC, Hof PR, Kuzawa CW, Boddy AM, Abbas A, Raaum RL et al. 2012. Dynamic gene expression in the human cerebral cortex distinguishes children from adults. PLoS ONE. 7(5):e37714.

Tay SK, Blythe J, Lipovich L. 2009. Global discovery of primate-specific genes in the human genome. Proc Natl Acad Sci USA. 106:12019–12024.

Weickert CS, Elashoff M, Richards AB, Sinclair D, Bahn S, Paabo S, Khaitovich P, Webster MJ. 2009. Transcriptome analysis of male-female differences in prefrontal cortical development. Mol Psychiatry. 14:558–561.

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG. 2005. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science. 309: 1570–1573.

Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23:1494–1504.