The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease

Thomas C. Roberts¹,², Kevin V. Morris²,³ and Matthew J. A. Wood¹

¹Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK
²Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
³School of Biotechnology and Biomedical Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia

Long non-coding RNAs (lncRNAs) are transcripts with low protein-coding potential that represent a large proportion of the transcriptional output of the cell. Many lncRNAs exhibit features indicative of functionality including tissue-restricted expression, localization to distinct subcellular structures, regulated expression and evolutionary conservation. Some lncRNAs have been shown to associate with chromatin-modifying activities and transcription factors, suggesting that a common mode of action may be to guide protein complexes to target genomic loci. However, the functions (if any) of the vast majority of lncRNA transcripts are currently unknown, and the subject of investigation. Here, we consider the putative role(s) of lncRNAs in neurodevelopment and brain function with an emphasis on the epigenetic regulation of gene expression. Associations of lncRNAs with neurodevelopmental/neuropsychiatric disorders, neurodegeneration and brain cancers are also discussed.

1. Introduction

It is now clear that the majority of the mammalian genome produces RNA transcripts despite only approximately 1% of the DNA sequence encoding proteins (a phenomenon known as pervasive transcription) [1]. The majority of loci produce a forest of interlaced [2] and overlapping [3] transcripts in both sense and antisense orientations [4,5]. Complementary results have been observed using multiple transcriptomics methodologies (i.e. RNA-seq [6], RNA tiling arrays [3,7–11], sequencing of full-length cDNA libraries [2,12], high-throughput rapid amplification of cDNA ends (RACE) [7] and sequencing of CAGE tags [2]) suggesting that the observed transcription is real and not a technical artefact or background genomic DNA/pre-mRNA.

Long non-coding RNAs (lncRNAs) are RNA transcripts more than 200 nucleotides in length that do not encode proteins. lncRNA transcripts are generally ‘mRNA-like’ [13,14] as they are frequently transcribed by RNA polymerase II, contain canonical splice sites (GU/AG), have similar intron/exon lengths to mRNAs, exhibit alternative splicing, may be polyadenylated or non-polyadenylated and associate with the same types of histone modification as protein-coding genes [14–16]. In contrast to mRNAs, a large fraction of lncRNAs (42% of lncRNAs in the GENCODE v7 catalogue) consist of only two exons [16]. lncRNAs generally exhibit low coding potential and are devoid of extended open reading frames (ORFs). Putative lncRNA ORFs have also been shown to be of similar quality to ORFs found in random genomic sequence [16], lack the pattern of cross-species mutation accumulation typical of protein-coding sequence [17] and show little similarity with ORFs of recently evolved proteins [18]. lncRNAs are associated with ribosomes (as are other non-coding RNAs and non-coding regions of mRNAs) but are distinct from coding transcripts in that they lack a characteristic ribosome drop-off.
signature found at the 3'-end of bona fide ORFs [18], suggesting that the majority are not translated into proteins. Nevertheless, some lncRNAs may encode short peptide sequences [19–21].

A major challenge in biology is to decode the genomic language that governs the architecture and function of the central nervous system (CNS). The mammalian CNS arguably represents the most complex system within all of biology. Not only does it comprise hundreds of billions of cells of neuronal and glial origin, but this complexity is amplified by the hundreds of trillions of synaptic interactions between these cells. Establishing this intricate cellular architecture during neurodevelopment and maintaining it effectively during adult life with appropriate adaptation and learning is a significant undertaking. It is highly likely that the cells of the CNS take advantage of all the subtleties of genomic evolution in order to achieve these complex cellular behaviours. Here, we explore the possible roles of lncRNAs as critical genomic regulators within the brain.

2. Are long non-coding RNAs functional?

The degree to which non-coding transcription is functional is currently a matter of debate, with some arguing that the majority is simply noise resulting from stochastic promoter firing or illegitimate transcripts arising from ‘promiscuous’ promoters. A study by van Bakel et al. [22] argued in favour of the transcriptional noise hypothesis and showed that the majority of non-coding RNA transcripts are associated with known genes. These conclusions have been vigorously opposed by others who have suggested that association of lncRNA transcripts with protein-coding loci is consistent with pervasive transcription, and point to insufficient sequencing depth in the van Bakel study [1]. Furthermore, a contradictory finding that the majority of lncRNAs are independent transcriptional units, was reported by the ENCODE consortium [16].

By contrast, many studies point to a functional role for non-coding transcription in the general case. Firstly, lncRNA genes are expressed in a tissue-specific manner. Investigation of the transcriptional landscape of multiple human cell lines found that 29% of lncRNAs were expressed specifically in a single cell type, while only 10% were expressed in all cell types (in stark contrast to protein-coding genes for which the numbers were 7% and 53%, respectively) [6]. Furthermore, among the most differentially expressed lncRNAs, approximately 40% are expressed specifically in the brain [16]. Using *in situ* hybridization data in mouse brain sections taken from the Allen Brain Atlas, Mercer et al. [23] found that most lncRNAs are associated with distinct neuroanatomical loci. For example, expression of the lncRNA *AK037594* was found only in the dentate gyrus and CA1–3 regions of the hippocampus. Similarly, *MIAT* (*Gomafu*) a nuclear-localized lncRNA is expressed only in differentiating neural progenitors and a subset of postmitotic neurons [13].

A study by Ponjavic et al. [24] found that the genomic loci of lncRNAs expressed in the developing brain were preferentially located in the vicinity of protein-coding genes that are (i) highly expressed in brain, (ii) involved in transcriptional regulation or (iii) involved in CNS development. Furthermore, analysis of a subset of these lncRNA–protein-coding gene pairs revealed co-expression in the same specific brain regions consistent with cross-talk between coding and non-coding transcripts arising from the same loci [24]. Similarly, many overlapping sense mRNA–antisense lncRNA pairs are co-expressed and specifically localized to synaptoneurosomes (specialized structures enriched at the pinched-off dendritic spines of pyramidal neurons) in the adult mouse forebrain [25]. Some of these mRNAs have known roles in synaptogenesis (e.g. *BC1*, *Camk2a*, *Dag1*) or have been implicated in Alzheimer’s disease (AD) pathophysiology (e.g. *Bace1* and *App*). Additionally, many lncRNAs are localized to specific subcellular compartments or to subnuclear structures [13,26]. For example, the lncRNA *Ntab* was found to be expressed only in the developing and adult rat CNS and transported to processes distal from the cell soma [27].

Targeted sequencing of cDNAs eluted from tiling arrays has revealed a plethora of low abundance transcripts (some originating from so-called gene deserts) exhibiting well-defined exon–exon boundaries, indicative of high-fidelity lncRNA splicing [28]. Interestingly, a recent study by Tilgner *et al.* [29] showed that the efficiency of lncRNA splicing is significantly lower than for mRNAs, and that many lncRNA transcripts (including well-studied functional examples such as *Airn* and *KCNQ1OT1*) remain unspliced.

Secondly, lncRNAs exhibit signs of regulated expression [30]. For example, enhancer-derived lncRNAs are differentially expressed in an activity-dependent manner in neuronal cultures [31]. Similarly, 174 lncRNAs were differentially expressed during the 16-day differentiation of mouse embryonic stem (ES) cells to embryoid bodies [32]. Four lncRNAs (including *Miat*) also showed dynamic patterns of expression following retinoic acid-induced neuronal differentiation in a separate study in mouse ES cells [33]. The observation that pluripotency factors, such as *Oct4* and *Nanog*, can bind to the promoters of lncRNA genes and modulate their transcription suggests that lncRNAs constitute an important component of the genetic circuitry that regulates the balance between maintenance of pluripotency and lineage commitment (figure 1). RNA interference (RNAi) knockdown and overexpression of two of these lncRNA transcripts led to alterations in Nanog and Oct4 expression, and promoted the adoption of lineage-specific differentiation programmes [33].

A separate study identified *Sox2OT* as a lncRNA gene that is dynamically expressed during neural cell differentiation [34]. *Sox2OT* encodes a sense-orientation transcript that overlaps with the pluripotency-associated transcription factor (TF) *Sox2* (sex determining region Y-box 2). The genomic proximity of *Sox2OT* and *Sox2* suggested a possible regulatory role for *Sox2OT* in the maintenance of pluripotency, which was recently confirmed experimentally [35].

While the precise processing, tissue-specificity, sub-cellular localization and differential expression of lncRNA transcripts have been used as arguments for functionality, it could be equally argued that many of these observations are also consistent with lncRNAs being the product of noisy transcription. In this case, lncRNA expression might be explained as a result of low-level TF binding and RNA polymerase engagement [36]. Given that the expression of TFs and other gene regulatory mechanisms are differentially active in specific tissues/cell types and during changes in cellular metabolism, this could give rise to patterns of
tissue-specific or apparently regulated lncRNA noise. However, the observation that many lncRNA transcripts are localized to distinct subcellular compartments is more difficult to dismiss as noise [23]. In a recent study, Sauvageau et al. [37] developed 18 transgenic knockout mice strains in order to investigate possible lncRNA functions. These researchers focused on a subclass of lncRNAs called long intergenic non-coding RNAs (lincRNAs). lincRNAs are biochemically indistinct from other lncRNAs but differ in their genomic organization as they reside in the space between genes. As lincRNAs do not overlap with protein-coding genes, functions can unambiguously be ascribed to the non-coding transcript rather than as indirect effects on neighbouring protein-coding genes. lincRNAs targeted for knockout were replaced with a \( \text{lacZ} \) expression cassette such that transcription from each lincRNA loci was maintained. As a result, any phenotypes observed in the knockout mice can be attributed to the lincRNA sequence, rather than as sequence-independent effects mediated by the act of transcription itself. Of the 18 lincRNA knockout strains, three lncRNA knockout strains (\( \text{Fendrr} \), \( \text{Peril} \) and \( \text{Mdgt} \)) had perinatal and postnatal lethal phenotypes indicating critical roles for these transcripts in development. Another strain knocked out for linc-\( \text{Brn1b} \) showed a reduction in the number of intermediate progenitor cells in the subventricular zone, suggesting that this lncRNA plays a key role in the developing cortex [23].

Importantly, linc-\( \text{Brn1b} \) showed many features consistent with the results of Mercer et al. [23,38]. linc-\( \text{Brn1b} \) expression is primarily restricted to specific brain regions (i.e. telencephalon, ventricular zone and subventricular zone), is predominantly nuclear localized in cultured neural progenitor cells derived from the cerebral cortex and shows...
spatio-temporally regulated patterns of expression during cortical development [37]. These findings lend support to the attribution of potential function on the basis of tissue-specific and regulated expression patterns. While the results presented by Sauvageau et al. are highly encouraging (especially given the inability to find essential functions for IncRNAs in other studies [39,40]) many more knockout studies are required to demonstrate further functions for IncRNAs in vivo. Thirdly, IncRNA genes show evidence of being under evolutionary constraint (although generally to a lesser extent than for protein-coding genes). The exons of IncRNA genes show a tendency to have lower base substitution rates than their corresponding intronic regions, indicative of evolutionary conservation [17,41]. Similarly, IncRNA exons show other signs of conservation such as enrichment for phastCons elements and indel-purified sequence [42,43]. Additionally, the promoter regions and splice sites of IncRNA genes are conserved at rates higher than would be expected by chance [42,44]. A separate study found that while conservation of IncRNA genes was low when looking at the full-length transcript, the degree of conservation became much higher when transcripts were analysed in 50 nucleotide windows [45]. This is consistent with short conserved functional sequences residing within longer transcripts that are generally under less evolutionary constraint. Furthermore, conservation of RNA secondary structural motifs within IncRNA genes unambiguously points to functions for their RNA gene products [46–50]. RNA structure may be critical to the functionality of many IncRNAs, whereas the primary base sequence may be less important. As a result, conservation analyses which fail to take into account the preservation of RNA secondary structure motifs despite changes to the primary base sequence will tend to underestimate the degree of actual IncRNA conservation.

Although conventional metrics of evolutionary constraint suggest that IncRNAs are under selective pressure, these findings should be treated with a degree of caution. Annotation of a transcript as definitively non-coding is not trivial, and so it is possible that a substantial number of protein-coding transcripts have been misclassified as IncRNAs. Such an eventuality would ‘contaminate’ the pool of so-called IncRNA genes with conserved sequence and bias estimations of IncRNA conservation [14]. Similarly, estimates of IncRNA functionality based on conservation may be skewed as a result overlap with protein-coding genes or other conserved DNA elements (such as enhancers).

While primary sequence conservation of IncRNA genes may be limited, some show other signs of being under evolutionary constraint, such as positional conservation [5]. For example, the IncRNA MALAT1 is syntenically conserved across a wide variety of organisms [39]. Similarly, 68 IncRNAs derived from pseudogene loci [51] showed positional conservation between human and at least two other mammals [52]. Analysis of IncRNAs in the GENCODE v7 catalogue showed that approximately 30% of IncRNAs are specific to the primates and therefore lack evolutionary conservation outside that lineage [16]. Importantly, while evolutionary conservation is indicative of function, lack of conservation does not necessarily imply lack of function [45]. For example, several IncRNAs (i.e. Xist and Airn) with well-established epigenetic regulatory roles are poorly conserved between human and mouse at the primary sequence level [53,54]. Additionally, rapidly evolving IncRNAs that are lineage-specific likely represent recent evolutionary innovations. One such primate-specific IncRNA gene HAR1F (human accelerated region 1F) is expressed in Cajal–Retzius neurons of the neocortex [55]. Interestingly, despite considerable sequence changes, the expression pattern of HAR1F in developing cortex is highly conserved between humans and cynomolgus macaques, suggesting that HAR1F expression is functionally significant.

In some cases, IncRNAs may have sequence-independent functions, whereby the act of their transcription alone may regulate expression of neighbouring genes (a phenomenon called transcriptional interference [56] or promoter occlusion [57]). As a result, the nucleotide sequence of the IncRNA may be inconsequential with respect to its functionality and therefore not subject to evolutionary constraint [58–61]. In support of this, Derrien et al. showed that IncRNA promoters are generally more conserved (at a level similar to protein-coding exons) than IncRNA exons [16], suggesting that the transcription of many IncRNAs is more important than the IncRNA sequence itself.

3. What are the functions of long non-coding RNAs?

To date, IncRNAs have been implicated in a wide variety of processes including modulation of splicing [62,63], organellar formation [26,64], telomere function [65], post-transcriptional gene regulation [66–69], sequestration of signalling proteins [70], generation of small RNAs (e.g. nucleolar RNAs (snoRNAs), microRNAs (miRNAs) and endogenous small interfering RNAs) [71–73], competition for miRNA binding [74–76] and regulation of protein localization [77]. A major function of IncRNAs appears to be in the epigenetic regulation of gene transcription and, as such, IncRNAs have been implicated in practically every epigenetic process: X-chromosome dosage compensation [78,79], mono-allelic expression of imprinted genes [80–82], control of chromatin macro structure [83], direction of genomic loci to distinct nuclear sub-substructures [84] and lineage commitment/cell fate determination [33,85,86].

Epigenetics is the study of heritable traits that are not encoded in the primary DNA sequence itself, but rather in the patterns of covalent alteration of DNA nucleobases (e.g. cytosine methylation) and histone protein post-translational modification (e.g. the histone code) [87–91]. Epigenetic modifications regulate the accessibility of the genome to the transcriptional machinery [92] and are thus important controllers of gene expression. As such, it has been proposed that IncRNAs act as ‘analogue–digital convertors’ [93] capable of facilitating the flow of information between proteins and nucleic acids. The structural plasticity of RNA enables the simultaneous binding of IncRNAs to proteins by forming secondary structure motifs (i.e. analogue interactions), and to nucleic acids through Watson–Crick, Hoogsteen and reverse Hoogsteen base pairings (i.e. digital interactions). IncRNAs may consist of multiple binding modules and are therefore, in theory, capable of bringing together any cellular component [94–96]. Specifically, IncRNAs act to direct epigenetic modifying complexes and TFs to specific chromatin loci. The observation that IncRNAs tend to be
enriched in nuclear extracts [6,15], and more specifically in the chromatin fraction [15,16] (whereas coding transcripts are primarily cytoplasmic), is consistent with specific interactions of lncRNAs with genomic DNA (figure 2).

Two landmark studies used RNA-immunoprecipitation methodologies in order to systematically identify lncRNAs which bind to chromatin-modifying proteins. Khalil et al. [97] performed RIP-chip using antibodies against PRC2, SMCX and CoREST (a general transcriptional co-repressor which acts to repress neural-specific genes) in order to precipitate and analyse bound lncRNA transcripts. These epigenetic modifier complexes were found to associate with 38% of the approximately 1100 lncRNA genes featured on the microarray chips. Furthermore, there was little overlap between the lncRNA-binding partners for each protein complex, suggesting that each complex binds a distinct repertoire of lncRNAs [97]. By contrast, very few mRNAs (approx. 2% of those featured on the arrays) associated with PRC2, suggesting that PRC2 binding is an lncRNA-specific phenomenon. Similar results were obtained by Zhao et al. [98] who immunoprecipitated Ezh2 (the component of PRC2 which tri- methylates H3K27 in order to induced transcriptional silencing) and identified approximately 9000 bound transcripts by RNA sequencing. Subsequently, a plethora of other epigenetic modifier complexes were shown to associate with lncRNAs including PRC1, Cbx1, Cbx3, Tip60/P400, Setd8, ESET, and Suv39h1, Jarid1b, Jarid1c, HDAC1 and YY1 [85]. While the majority of studies have identified lncRNAs that bind to repressive epigenetic modifying complexes, associations with activating complexes have also been observed [32,97,99]. For example, the lncRNAs Eexl as and Hoxb5/6as (which show concordant expression with their overlapping sense-orientation protein-coding genes during mouse ES cell differentiation) immunoprecipitated with the H3K4 trimethylase Mll1, suggesting that they may be cis positive regulators [32].

In several cases, lncRNAs have been shown to be composed of distinct protein or nucleic-acid-binding modules, and this has been proposed as a general mode of lncRNA function [94,100]. Modular binding of proteins allows for the activities of multiple epigenetic modifier complexes to be directed to specific genomic loci in a coordinated manner. The best described example of this is the lncRNA HOTAIR which is a trans negative regulator of the HOXC cluster [101,102] and other loci [103]. The HOTAIR transcript acts as a scaffold for PRC2 and a complex of LSD1/CoREST/REST (Repressor Element 1-Silencing TF) at its 5' and 3' termini, respectively [101,104]. As a result, HOTAIR coordinates the H3K27 trimethylase and H3K4 demethylase activities of these protein complexes in order to facilitate gene silencing at specific target loci.

RNAi screening loss-of-function studies targeting lncRNAs in mouse ES cells have shown that many non-coding transcripts act to control pluripotency and differentiation [33,85,105]. Interestingly, knockdown of lncRNA generally resulted in comparable numbers of up- and downregulated transcripts. Given that the majority of studies have focused on lncRNAs with gene silencing functions, this observation suggests that gene-activating lncRNAs may be of equal importance and that many positive regulators of gene expression remain to be discovered. Knockdown of many lncRNAs produced gene expression changes associated with a loss of pluripotency and the adoption of early differentiation lineages (including neuroectoderm). lncRNA knockdown did not, in general, affect neighbouring genomic loci, suggesting that the primary mode of gene regulation is in trans rather than in cis [85]. Similar results were obtained by Khalil et al. [97], who showed that RNAi knockdown of PRC2-associated lncRNAs
resulted in activation of polycomb targets while not significantly affecting lncRNA-neighbouring genes, again indicative of trans regulation.

An alternative and complementary approach, termed ‘guilt-by-association’ has been used to infer lncRNA functions. Firstly, lncRNAs and protein-coding genes are clustered according to the degree of correlation between their expression patterns. The degree of association of each lncRNA with each gene ontology term is determined and biclustering used to identify groups of lncRNAs associated with specific functions [17]. Similarly, Liao et al. [106] identified probable functions (including neuronal development) for 340 lncRNAs based on coding/non-coding gene co-expression networks.

4. Long non-coding RNAs are involved in neural development and brain function

Multiple studies have implicated non-coding RNAs in brain development and function. Here, we focus only on lncRNAs, although small non-coding RNAs, such as miRNAs, are also important and have been discussed elsewhere [107,108]. Dynamic expression of lncRNAs has been observed in human-induced-pluripotent stem cells (iPSCs) [109] and human ES cells [105] during neuronal differentiation using RNA-seq and custom microarray, respectively. Neurogenesis-associated lncRNAs were found to associate directly with SUZ12 (a component of the polycomb repressive complex 2, PRC2), REST (discussed below) and SOX2 (a pluripotency-associated TF) suggesting that lncRNAs may act as guides for these proteins. Importantly, knockdown of these lncRNAs by RNAi resulted in impaired neuronal differentiation, suggesting that lncRNAs are critical regulators of neurogenesis [105]. A landmark study by Lipovich et al. [110] measured lncRNA expression in surgically resected in vivo human neocortical samples. Analysis of a range of samples from patients of different ages identified eight lncRNAs which showed strong statistical associations with aging and, by extension, brain development [110]. The majority of these lncRNAs were antisense to neighbouring protein-coding genes, suggesting possible gene regulatory functions. Interestingly, these lncRNAs also exhibited features consistent with recent evolutionary origins, including anthropopid-specific exons and mRNA processing sites which reside within primate-specific sequence [110]. Taken together, these findings implicate lncRNAs in the development of the human brain.

Similar results have also been observed in mouse cells where lncRNAs have been shown to control neuronal-glial cell fate decisions. Using custom microarray analysis of both coding and non-coding transcripts, Mercer et al. [38] identified lncRNAs that were differentially expressed between mouse embryonic forebrain-derived neural stem cells (NSCs), bipotent GABAergic neuron/oligodendrocyte cells and the various stages of terminally differentiated neurons and glia [38]. For example, the lncRNAs Neat1 and Neat2 (Malat1) were downregulated in the bipotent precursor cells but upregulated in differentiated neuronal and glial cells. Treatment of oligodendrocyte progenitor cells with the histone deacetylase (HDAC) inhibitor trichostatin A (known to suppress the maturation of oligodendrocyte precursors and induce a more neuronal-like pattern of gene expression) also affected expression of lncRNAs, suggesting that their expression is under HDAC control [38]. A separate study identified the lncRNA Nkx2.2AS, which is a natural antisense transcript overlapping the TF gene Nkx2.2, as a further regulator of oligodendrocyte differentiation [111]. Overexpression of Nkx2.2AS induced differentiation and resulted in an increase in Nkx2.2 mRNA expression.

lncRNAs have also been implicated in the differentiation of other types of CNS tissue. Photoreceptors are specialized neurons in the retina which facilitate vision through the process of phototransduction [112]. The lncRNA TUG1 (which is highly expressed in brain) has been shown to be required for photoreceptor differentiation, although the mechanism of action has not yet been identified [86].

REST is a TF that represses expression of genes involved in neurogenesis and neuronal function in non-neural and immature neural cell types [113]. REST is therefore a key player in maintaining pluripotency and regulating neurogenesis. Johnson et al. showed that two brain-restricted lncRNAs are repressed by REST in NSCs [114] and the HAR1F/R lncRNA locus (discussed in §2) in a separate study [115]. Similarly, ROR1 (also known as CoREST) is another protein that acts to repress expression of neuronal genes [116]. RIP-chip analysis using antibodies against ROR1 identified 63 associated lncRNAs, many of which were also found to bind PR2, suggesting that non-coding transcripts may play a key role in neural cell differentiation [97].

Imprinting is an epigenetic process by which certain genes are expressed in a parent-allele specific manner. A common theme in epigenetic imprinting is the reciprocal allelic expression of an imprinted gene and an imprinted non-coding RNA cis regulator. One of the most well-understood examples of this is the lncRNA Airn (also known as Air) which mediates epigenetic silencing of the Igf2r/Slc22a2/Slc22a3 locus on chromosome 17 [82]. Airn encodes an antisense transcript which overlaps with the Igf2r gene but not Slc22a2 or Slc22a3. Airn is expressed only from the parental allele, leading to epigenetic silencing of the parental Igf2r/Slc22a2/Slc22a3 locus in cis. Conversely, on the maternal allele, Airn is itself silenced by a reciprocal imprinting process and expression of the maternal Igf2r/Slc22a2/Slc22a3 locus is unhindered. Airn-mediated silencing occurs by at least two different mechanisms. In the case of Igf2r, transcription alone is sufficient to induce silencing [117]. Conversely, Slc22a3 silencing is dependent upon Airn-dependent recruitment of EHMT2 (a H3K9 histone methylase also known as G9a) [118]. In the majority of tissues, Igf2r is expressed only from the maternal allele, whereas Airn is expressed only from the paternal allele. However, this pattern of reciprocal allelic expression is not observed in brain where Igf2r is expressed in a biallelic manner as a result of neuron-specific relaxation of Airn-mediated imprinting [119].

Dlx genes encode homeodomain proteins that play key roles in the regulation of neuronal differentiation and migration [120,121]. The lncRNA Evf2 is transcribed from an ultraconserved region between the Dlx5 and Dlx6 protein-coding genes and is a direct target of SHH (Sonic hedgehog), a master regulator of vertebrate CNS development. Evf2 RNA forms a stable complex with Dlx4 protein and enhances its transcriptional activation functionality in C17 NSCs [122]. In a follow-up study, Evf2 was shown to act via both cis- and trans-acting mechanisms to recruit both Dlx and Mecp2 (methyl CpG-binding protein 2) to the
Dbx5/6 ultraconserved region in the ventral forebrain [123]. Transgenic mice deficient in Evf2 transcription exhibited an imbalance in gene expression that led to a decrease in the number of GABAergic interneurons in the postnatal hippocampus, thereby illustrating the importance of this IncRNA in the patterning of the brain [123].

*Malat1* is one of the most well-studied IncRNAs. It is well conserved, highly abundant and expressed in a wide range of tissues [39]. In the brain, *Malat1* is expressed at high levels in neurons and low levels in glia and astrocytes, suggesting an important neuronal function [124]. Genes affected by antisense oligonucleotide-mediated *Malat1* depletion were enriched for gene ontology terms associated with synaptic function and dendrite development. Knockdown of *Malat1* in primary hippocampal neuron cultures resulted in reduced synaptic density, whereas *Malat1* overexpression showed the opposite effect. Changes in the expression of *Nlgn1* and *SynCAM1* were observed upon *Malat1* knockdown, suggesting that *Malat1* regulates synaptogenesis by modulating the expression of genes in synapse formation [124].

5. Long non-coding RNAs and neurodegeneration

The human genome overwhelmingly (approx. 99%) consists of non-protein-coding sequence and it is therefore not surprising that the majority of mutations identified by genome-wide association studies (GWAS) occur in non-coding regions [125,126]. As such, a number of neurodegenerative disorders are known to be caused by mutations in IncRNA genes.

Perhaps the clearest example is spinocerebellar ataxia type 8 (SCA8) which is caused by a CTG triplet expansion in the brain-expressed *ATXN8OS* gene which is an antisense IncRNA transcript that partially overlaps with its neighbouring protein-coding gene, *KLHL1* [127]. Although the aetiology of the disease is not well understood, the microsatellite expansion in the antisense transcript is believed to interfere with its endogenous role in regulating *KLHL1* expression [128]. Microsatellite expansions in non-coding regions are also known to cause toxic RNA gain-of-function pathologies (such as in myotonic dystrophy) by sequestering factors involved in alternative splicing such as MBNL and CELF [129].

In 2011, a hexanucleotide (GGGGCC) repeat expansion in a protein-coding gene, *C9ORF72* (chromosome 9 ORF 72) was identified as the first causative mutation for both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia [130,131]. Since this landmark discovery, non-coding transcripts have now also been identified at the *C9ORF72* locus. The *C9ORF72* repeat expansion region undergoes bidirectional transcription [132]. Antisense *C9ORF72* transcripts are elevated in the brains of ALS patients with both sense and antisense transcripts forming nuclear RNA foci [132,133]. The importance of the antisense *C9ORF72* transcript is exemplified by the observation that targeted degradation of the corresponding sense transcript using antisense oligonucleotides is insufficient to correct the disease-associated gene expression signature in patient-derived fibroblasts [134]. These findings would be consistent with a toxic RNA-type cellular pathology, although the reality may be more complex as both sense and antisense transcripts produce dipeptide repeat proteins by repeat-associated non-ATG translation [132,135,136].

By interrogating published microarray gene expression data from Huntington’s disease (HD) patient caudate nucleus [137], Johnson [138] was able to identify lncRNAs with a HD-specific pattern of differential expression. Three novel lncRNAs were elevated in HD brains in addition to TUG1 and NEAT1 (which were upregulated in HD) and MEG3 and Dgcr5 (which were downregulated). The role of these lncRNAs in HD pathophysiology is currently unknown, although the observation that MEG3 and TUG1 associate with PRC2 suggests that they may act as epigenetic regulators which induce disease-specific gene expression signatures [97]. Similarly, a separate study found that the expression of lincRNAs originating from the HAR1F/R locus was repressed in the striatum of post-mortem HD brains [115]. Using a similar data mining approach, Michelbaugh et al. [139] found that the *Miat*, *MEG3*, *NEAT1* and *NEAT2* lincRNAs were all upregulated in the post-mortem dissected nucleus accumbens of heroin users, suggesting a possible role for lncRNAs in addictive behaviours.

BACE1 (β-site amyloid precursor protein cleaving enzyme 1, also known as β-secretase) is an enzyme central to the pathology of AD. BACE1 catalyses the cleavage of amyloid precursor protein to generate β-amyloid peptides which aggregate to form plaques [140]. Studies by Faghihi and co-workers identified a conserved antisense transcript overlapping (BACE1-AS) at the BACE1 locus [66,141]. BACE1-AS is concurrently expressed with BACE1 sense mRNA and acts as a feed-forward positive regulator of BACE1 expression. Expression of BACE1-AS was also found to be elevated in the hippocampus, superior frontal gyrus and entorhinal cortex in post-mortem AD brain tissue [66]. The mechanism of BACE1 regulation by BACE1-AS was subsequently shown to be via the formation of an RNA duplex between the overlapping transcripts which masks the binding site for miR-485–5p, thereby relieving miRNA-mediated gene silencing [67].

6. Long non-coding RNAs and neurodevelopmental/neuropsychiatric disorders

The non-coding RNA BC200 is restricted to brain tissue (specifically to the neurite outgrowths of neurons) and its expression gradually declines with aging. However, BC200 expression is elevated in the brains of AD patients and mislocalized to the neuronal cell bodies rather than to dendritic spines [142]. The molecular function of BC200 appears to be in the regulation of neuronal protein translation and so it may contribute to amyloid plaque formation and subsequent AD [143,144]. The murine homologue of BC200, BC1, was targeted in a transgenic knockout model. Interestingly, BC1 knockout mice showed no obvious phenotype in a laboratory cage environment. However, when introduced into a controlled ‘natural outdoor’ environment, these mice showed signs of increased anxiety and reduced survival [145]. This study demonstrates that non-coding transcripts may exert subtle effects on complex behaviour and raises the intriguing possibility that lncRNAs may be involved in the pathogenesis of neurodevelopmental and neuropsychiatric diseases with poorly understood aetiologies.

Autism spectrum disorder (ASD) refers to a heterogeneous group of neurodevelopmental disorders that are
characterized by defects in social interactions, communication and repetitive stereotyped behaviours. Although ASD is known to have a strong genetic basis, its pathophysiology is poorly understood [146]. Microarray analysis of human post-mortem brain tissue (prefrontal cortex and cerebellum) from ASD patients and unaffected controls identified 222 differentially expressed IncRNAs which were enriched at protein-coding gene loci associated with brain development. Interestingly, the ASD brains were more transcriptionally homogeneous than the controls, both in terms of mRNA and IncRNA expression [147]. Similarly, interrogation of publicly available RNA-seq data identified overlapping antisense IncRNAs at 38 protein-coding loci associated with ASD. Furthermore, one of these antisense transcripts, SYNAGPI1-AS, was found to be up-regulated in the ASD post-mortem prefrontal cortex and superior temporal gyrus [148].

A study by Kerin et al. [149] identified a single nucleotide polymorphism (SNP) associated with ASD at a non-coding locus in a GWAS. This locus was found to encode an IncRNA (MSNP1AS) antisense to a processed pseudogene of moesin (M5SNP1) which shows no evidence of being transcribed in the sense orientation. The SNP-containing MSNP1AS transcript was shown to be elevated in post-mortem brain tissue (temporal cerebral cortex) of ASD patients and regulated expression of Moesin protein (a known regulator of nuclear architecture [150]) in human cells [149], suggesting a possible role in ASD pathophysiology.

Fragment X syndrome (FXS) and fragile X tremor ataxia syndrome (FXTAS) are intellectual disabilities caused by expansions of a CGG repeat in the 5’-UTR of the FMR1 protein-coding gene [151]. Normal individuals typically carry 5–54 repeats, whereas 55–200 repeats (so-called premutation alleles) lead to FXTAS, and more than 200 repeats lead to FXS. As in the case of SCAB, FMR1 has an upstream partially overlapping antisense transcript, FMR4 (also known as FMR1-AS1), which is presumably driven by a bidirectional promoter. In FXS, the repeat expansion region becomes hypermethylated and transcription of the gene products in both orientations is diminished. siRNA-mediated knockdown of either FMR1 or FMR4 did not affect the expression of each transcript’s antisense partner, suggesting that FMR4 is not a regulator of FMR1. Instead, FMR4 knockdown was shown to promote apoptosis, suggesting that its endogenous function is as an RNA anti-apoptotic signal [152]. By contrast, FMR1 and FMR4/FMR1-AS are upregulated in carriers of premutation alleles (i.e. FXTAS) [153]. The application of a high-throughput sequencing RACE methodology to the FMR1 locus identified a further two IncRNA transcripts, FMR5 and FMR6 (in sense and antisense orientations, respectively). Analysis of post-mortem brain tissues from carriers of both full and premutation alleles showed that expression of FMR6 was suppressed in both cases relative to wild-type controls [154]. Furthermore, a recent study identified a role for CTCF in regulating bidirectional transcription of FMR1 through chromatin structure [155]. Consequently, the variable patterns of antisense RNA expression at the FMR1 gene as a result of different repeat region lengths have been proposed as an explanation for the different clinical features of FXS and FXTAS, despite both syndromes being caused by CGG expansions [153].

Several IncRNAs have been implicated in the pathogenesis of schizophrenia (SZ). The IncRNA MIAT is downregulated upon neuronal activation [156]. Investigation of SZ patient post-mortem brain tissue (superior temporal gyrus) found that MIAT was downregulated. The MIAT transcripts directly interact with the splicing factors QKI and SRSF1 and loss of MIAT expression results in global changes in alternative splicing similar to those observed for other SZ-associated genes (i.e. DISC1) [156,157].

7. Long non-coding RNAs and brain cancers

Multiple studies have identified IncRNAs involved in cancer [102,158]. For example, IncRNAs have been shown to be direct targets of p53, including Inc-p21, PANDA, TUG1 and Pint [159,160]. The IncRNA ANRIL is implicated in melanoma-neural system tumour [161] and has been shown to interact with PRC2 in order to epigenetically silence the p15 tumour suppressor [162].

The IncRNA CRNDE is highly upregulated in gliomas [163] and in iPSCs undergoing neuronal differentiation [109]. CRNDE shares a bidirectional promoter with the IRX5 gene (which is involved in neurogenesis) and the two genes exhibit concordant expression patterns. CRNDE binds to CoREST [97] and in the human adult brain CRNDE is predominantly expressed in the basal ganglia, thalamus, cerebellum and surrounding structures [164]. Conversely, the imprinted IncRNA gene, MEG3, is a brain-specific tumour suppressor that suppresses cell growth, promotes p53-mediated apoptosis and is lost in pituitary tumours [165,166], and in meningiomas [167].

8. Conclusion

In summary, the proposition that IncRNAs are functional is supported by the following: (i) specific spatial and temporal expression patterns, (ii) high-fidelity transcript processing, (iii) differential expression during cellular processes, (iv) evolutionary conservation, (v) extinction of mouse models, (vi) RNAi loss-of-function screens, (vii) guilt-by-association co-expression studies, (viii) interactions with chromatin-modifying proteins and TFs, (ix) correlation with disease pathophysiology and (x) focused studies demonstrating function in specific cases. The relatively low abundance and tissue-restricted expression of IncRNA transcripts suggest that they function as subtle regulators in the determination of cell fate and identity, rather than in the execution of housekeeping functions. The preponderance of evidence suggests that IncRNAs constitute a previously under-appreciated component of cellular metabolism that together with TFs, chromatin remodelling complexes and miRNAs, regulates differential gene expression. Given that the degree of organismal complexity scales with the amount of non-coding DNA sequences [168], it is tempting to speculate that the increase in regulatory complexity afforded by the interplay of IncRNAs and protein-coding genes may be responsible for the differences in cognitive abilities between humans and other animals [169].

The importance of IncRNAs in the brain is exemplified by their involvement in the maintenance of pluripotency, neuroectodermal differentiation, neuronal-glial cell fate determination, neuron-specific relaxation of epigenetic imprinting, repression of neural genes in non-neural cells, brain tissue patterning and synaptogenesis. Given that epigenetic mechanisms underlie memory formation, it is likely that IncRNAs may also be involved in this process [170]. The involvement of IncRNAs in neurodegenerative, neurodevelopmental and neuropsychiatric disorders, and in brain cancers further underlines their
importance in CNS development and function. lncRNAs may themselves drive or mediate the disease pathophysiology (as in the case of ATXN8OS and FMR4), or they may regulate the expression of disease-associated genes (as in the case of BACE1-AS). As a result, lncRNAs are promising novel targets for therapeutic intervention [171,172].

References

27. Sauvageau M et al. 2013 Multiple knockout mouse models reveal lncRNAs are required for life and brain development. elife 2, e01749. (doi:10.7554/ elife.01749)
29. Zhang B et al. 2012 The IncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult.
et al


133. Mizielinska S, Lashley T, Norona FE, Clayton EL, Lagier-Tourenne C.

134. Lagier-Tourenne C.


silencing of p15(INK4B) tumor suppressor gene.


