



HAR1: an insight into lncRNA genetic evolution

Ella Waters^{*,‡,1} , Perla Pucci^{‡,1,2} , Mark Hirst^{‡,1} , Simon Chapman^{‡,1} , Yuzhuo Wang^{‡,3} , Francesco Crea^{‡,1}  & Christopher J Heath^{‡,1} 

¹School of Life, Health & Chemical Sciences, The Open University, Milton Keynes, MK7 6AA, UK

²Division of Cellular & Molecular Pathology, Department of Pathology, University of Cambridge, Cambridge, CB2 0QQ, UK

³Department of Urologic Sciences, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

*Author for correspondence: ella.waters@open.ac.uk

‡Authors contributed equally

Long noncoding RNAs (lncRNAs) have a wide range of functions in health and disease, but many remain uncharacterized because of their complex expression patterns and structures. The genetic loci encoding lncRNAs can be subject to accelerated evolutionary changes within the human lineage. *HAR1* is a region that has a significantly altered sequence compared to other primates and is a component of two overlapping lncRNA loci, *HAR1A* and *HAR1B*. Although the functions of these lncRNAs are unknown, they have been associated with neurological disorders and cancer. Here, we explore the current state of understanding of evolution in human lncRNA genes, using the *HAR1* locus as the case study.

First draft submitted: 22 February 2021; Accepted for publication: 1 October 2021; Published online: 22 October 2021

Keywords: DNA conservation • genetic evolution • *HAR1* • lncRNAs • molecular structure

Introduction to long noncoding RNAs

Genomics has advanced considerably within the last decade. Substantial research has shown that approximately 99% of the transcriptome consists of nonprotein-coding regions [1,2]. An abundant class of these noncoding RNAs (ncRNAs) are long noncoding RNAs (lncRNAs) [3]. lncRNA transcripts are defined over 200 nucleotides in length and usually lack significant open reading frames but can be capped at the 5' end, spliced and polyadenylated [4]. The DNA sequences from which ncRNAs, including lncRNAs, are transcribed were originally believed to be 'junk' DNA [5], but lncRNAs have a wide range of biological functions and have been found in some cases to be differentially expressed between healthy and diseased cells [6–9].

lncRNA research is rapidly producing novel developments and knowledge with a relevant impact in molecular biology and medicine. To simplify this new knowledge and aid the understanding of lncRNA functions and characterization, lncRNAs are classified depending on their genomic location, subcellular location, structure and function.

lncRNAs can be broadly split into five categories depending on their genomic location, relative to immediately neighboring protein-coding genes: sense, antisense, bidirectional, intronic and intergenic [2,10] (Figure 1). Sense lncRNAs are transcript variants of a known protein-coding gene that overlap with their sense sequence, whereas antisense lncRNAs are transcribed from, and overlap with, the antisense strand of a protein-coding gene. Bidirectional lncRNA transcripts are head-to-head with a protein-coding gene on the opposite strand of DNA, and intronic lncRNAs overlap within the introns of a protein-coding gene. Intergenic lncRNAs are located between protein-coding genes.

The identification of alterations in lncRNA expression has prompted the characterization of their roles in important biological processes. A peculiarity of human lncRNAs is that they are often cell-, tissue- and disease-specific [11]. Because of their specificity within the cell, their functions can be indicated by their subcellular location. Nuclear lncRNA functions include transcriptional regulation such as enhancer lncRNAs, epigenetic and chromatin structure regulation [12] and alternative splicing [13]. Cytoplasmic lncRNAs function as miRNA sponges, to be involved in modifying mRNA stability [14], to promote proteasomal degradation [15], and cytoskeletal

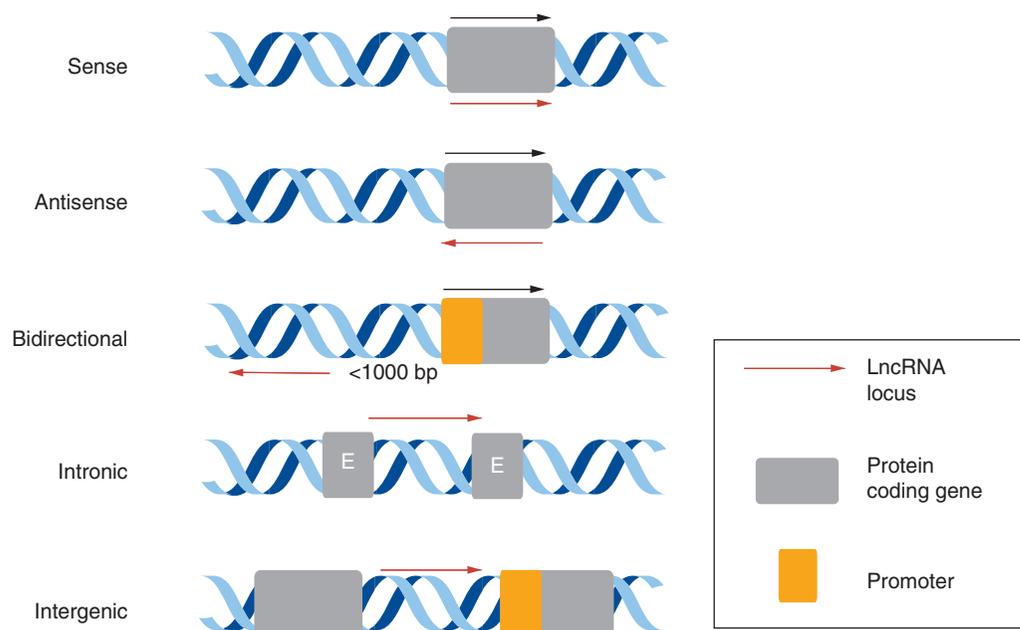


Figure 1. Long noncoding RNAs (red arrow) categorized as sense, antisense, bidirectional, intronic, or intergenic, in relation to protein coding genes (grey boxes).

E: Exon; LncRNA: Long noncoding RNA.

arrangement [16,17]. In addition to these examples of functions, some lncRNAs have also recently been found to encode for biologically active, short peptides called micropeptides [18,19], despite their noncoding name.

lncRNAs can also be characterized depending on their structure and function, interacting with other types of RNA, DNA and proteins:

- Signals regulate transcription in response to a stimulus;
- Decoys present binding sites to regulate the availability of regulatory factors, such as transcription factors and miRNAs;
- Scaffolds allow the assembly of multiple-component complexes including a RNP by providing structural domains, which can result in transcriptional activation or repression;
- Guides direct RNPs to their target genes;
- Enhancers are regions which can interact with promoters by recruiting RNA polymerase, thereby influencing the expression of transcription factors tethered to this region by other lncRNAs [20].

Depending on the site of these functions, lncRNAs can be described as acting in *cis* or *trans* [21,22]. *Cis* acting lncRNAs function at their loci and often regulate the expression of nearby genes. *Trans* acting lncRNAs function away from their site of transcription, such as regulating the expression of genes with loci on different chromosomes.

Despite the increasing volume of evidence showing that lncRNAs are involved in the regulation of gene expression and are implicated in numerous different human diseases, to date, only a relatively small percentage of lncRNA structures and functions have been elucidated [23,24]. This is likely due to the multifaceted expression patterns of lncRNAs, their complex molecular structures and relatively lower cellular abundance in physiological conditions [25], although there are specific patterns of lncRNA upregulation in cancer and other diseases.

Understanding the evolutionary constraints of lncRNAs and their degree of sequence conservation in humans may provide insights into their molecular interactions and functional roles, especially for those lncRNAs with genetic regions that have undergone accelerated evolution in the human lineage [26]. This review explores some of the current understanding of lncRNA evolution, using examples of key lncRNAs within the *HARI* locus to further analyze the relationship between their evolution and biological functions in health and disease.

LncRNA origin

LncRNA loci are often evolutionarily poorly conserved in their sequence among closely related organisms that share a common ancestor with humans. This phenomenon suggests a high gene turnover rate within the human lineage. This sequence turnover can refer to the lncRNA sequences diverging through mutation but retaining functionality, or to the process of functional lncRNAs originating from nonfunctional lncRNAs [27].

Novel lncRNAs can occur via *de novo* formation [28], although emerging evidence shows that lncRNAs may also be formed via lncRNA duplication or via the genetic rearrangements of protein-coding gene [29–31]. *De novo* evolution of lncRNAs can occur in regions not previously transcriptionally active, that become transcribed, such as transposable element activity. Transposable elements could influence the tissue specificity of lncRNAs [32]. The authors of the study also showed that a promoter region, which was originated from the transposable element family *LIPA2*, could act as a promoter for lncRNAs that are specifically expressed in the placenta [32]. This suggested that functional lncRNAs can originate from a nonfunctional transposable element, with functional lncRNAs contributing to a selectable trait persisting.

The phenomenon of lncRNAs formation via gene duplication of other lncRNAs is relatively understudied in comparison to protein-coding genes, although the duplication of a gene is the main process by which new protein-coding genes are created [33]. LncRNAs can form multiple lncRNA genes in this way, for example, a family of long intergenic ncRNAs (lincRNAs) originated from another lincRNA sequence *FAM230C* by segmental duplication [29]. LncRNAs can also emerge by assembly from the material of protein-coding sequences that are in the process of pseudogenization [34]. A relatively small number of lncRNAs originate from protein-coding sequences in this way [31]. The authors of the study also reported that despite losing their protein-coding abilities, these functional lncRNAs had elements that were conserved for millions of years.

LncRNA evolution & conservation

Conserved regions are sequences that remain unchanged or very similar in either different species or the same species, throughout generations in evolution. In lncRNAs, conserved sequences can vary in length, and have been shown in several algorithms to be of functional importance [35]. In contrast to protein-coding genes, the primary structures of multiple lncRNAs have poor evolutionary sequence conservation overall [36–38].

Large evolutionary studies of lncRNAs are instrumental in developing our understanding of regulatory gene networks, and therefore determining lncRNA functions [36]. This RNA-sequencing study of eight organs in 11 different species uncovered key information, such as the strong conservation of lncRNA promoter sequences for younger and older lncRNAs, comparable to protein-coding gene promoters [36]. This suggests stronger selective constraints at the transcriptional level. Furthermore, lncRNAs older than 90 million years (Myr) have a higher exonic sequence conservation, similar to that of protein-coding exons. Older lncRNAs may have more detectable homologous sequences within evolution, whereas the newer sequences may have evolved more rapidly, therefore evading detection with the current algorithms. The authors also confirmed that lncRNA transcription has diverged more rapidly over time compared with protein-coding genes, but tissue specificity is maintained.

A key advance in this area was made by the development of a software package called *slncky*, which identifies lncRNAs from RNA-sequencing data, and so improves the accuracy of lncRNA analysis [37]. Previous approaches have had difficulty in classifying conserved lncRNAs, as these were often mistaken for protein-coding genes or pseudogenes [37]. Owing to its increased sensitivity, *slncky* enables a more specific classification of lncRNAs and facilitates the study of lncRNA evolution, conservation and potential functions. The software identifies a syntenic region for a lncRNA, a region with the gene locus on the same chromosomal location in different species, the sequence and transcript conservation are then characterized. *Slucky* can also identify evolutionary properties, as shown by the software discovering specific selection on two distinct classes of ancestral intergenic lncRNAs. 232 lncRNAs were analyzed, finding that about 20% of these only have constraint on the act of transcription, with limited conservation on the actual transcript sequence. Approximately 80% of the lncRNAs analyzed have a strong purifying selection, therefore preserving and stabilizing the transcript in evolution. The latter 80% of lncRNAs could be considered as being part of a different class of lincRNAs. Further developing and utilizing software such as *slucky* to reveal more information on lncRNAs by evolutionary analysis, and further characterize lncRNAs, will assist in clarifying more of their functional roles.

More recently, comparative transcriptomics has been used to analyze lncRNAs at different developmental stages of mouse, rat and chicken, using tissues from the brain, testes, liver and kidney [38]. The study reported that the lncRNAs that are expressed during embryonic development, particularly in the brain and kidney, had higher levels

of functional constraint. As the most conserved sequences were located in the promoters of lncRNAs, rather than their exons, these lncRNAs were suggested to have RNA sequence-independent biological functions. This means the process of transcription influences gene function, rather than the lncRNA product being functional itself. Many of lncRNAs that act with this kind of crosstalk are *cis* regulators [39]. This type of lncRNA functionality along with their specific expression patterns are significant findings and are a potential focus for future studies of lncRNAs.

With several previous studies being focussed on protein-coding gene or shorter RNA characterization, these studies represent some of the lncRNA-specific research over the past decade. Although these studies indicate that it may be difficult to predict the functions of lncRNAs based on their sequence similarities alone, as lncRNAs have species-specific expression patterns, their syntenic conservation and expression conservation within regulatory regions are the main types of conservation aiding the understanding of lncRNA functional regions. These studies and those alike are essential for characterizing understudied lncRNAs and identifying regions of interest and likely functional regions within their sequences.

Rapid evolution of noncoding regions

While increasing the understanding of lncRNA conservation and evolution, it has also been discovered that there are multiple regions within the human genome that have undergone rapid sequence and structural changes within their recent evolution. As previously mentioned, in contrast to protein-coding genes, the primary structures of multiple lncRNAs have poor evolutionary sequence conservation, although they may have conserved genomic locations [40]. As lncRNAs are transcripts with a long sequence length, they can fold into complex secondary and tertiary structures. This folding can increase their structural stability and conservation [41], which may explain why lncRNAs with rapidly evolving sequences still have a functional role, although few lncRNA structures have been mapped and characterized yet. Understanding the high-order structure of a lncRNA may also enable the identification of binding sites and specific motifs that are important for predicting and understanding function, more effectively than considering the primary structure in isolation [42].

Rapid structural changes and structural versatility are likely the largest contributors to the wide range of human specific lncRNA functions that have evolved, in addition to their evolutionary origin. Examples of lncRNA structural motifs influencing their functions include conserved pseudoknots, helical secondary structures connected by loops, forming in the lncRNA *MEG3* tertiary structure allowing *p53* interaction [43]; G-quadruplex secondary structure, rich in guanine, in the lncRNA *GSEC* being important for colon cancer cell migration [44] and conserved protein binding elements and domains essential for the mechanism of action of the lncRNA *HOTAIR* predicted from its secondary structure [45]. These few examples demonstrate the importance of understanding the structure of a lncRNA.

Case study: *HAR1*

HARs are short (~260 base pairs) noncoding genomic regions that have undergone an accelerated rate of nucleotide substitutions and deletions, specifically in the human lineage, compared to their chimpanzee orthologous sequences [46]. 49 HARs with significantly high substitution rates were originally identified [46]. Of these, the most accelerated evolution was observed at the *HAR1* locus, which has undergone 18 substitutions within its 118 base pairs in the human genome, compared to the expected 0.27 substitutions, as the divergence of the *Homo* and *Pan* genera [47]. This noncoding *HAR1* sequence resides in a shared region within two divergently transcribed lncRNAs named after *HAR1*, *HAR1A* and *HAR1B*, in chromosome region 20q13.33 (Figure 2).

The high-order structure of the short *HAR1* region has been studied in detail. The secondary structure of *HAR1* was modeled to be a stable cloverleaf [48,49]. In these studies, the human *HAR1* structure was compared to the orthologous *HAR1* sequence in chimpanzees; the latter was predicted to fold into an unstable and extended hairpin. The structural stability of the *HAR1* helix in humans likely arises due to the length of the region, the transition from weaker AT to stronger GC base-pairing occurring through its 18 substitutions, increasing the base stacking energy within the helix. A recent study has further analyzed the secondary structure of human *HAR1*, identifying a bulge within the structure as a potential binding site for molecular interaction [50].

There is a close similarity between human and chimpanzee genomes because of the relatively short amount of time available for mutations to accumulate; an approximate 5–7 million-year history of separate evolution. The *HAR1* region is more ancient, dating back at least 310 Myr to a shared ancestor with chicken [47]. In contrast to the 18 altered bases between human and chimpanzee *HAR1* genes, the chimpanzee and chicken genes differ by only two bases out of 118 (Figure 3). Furthermore, it has been suggested that the human-specific substitutions in *HAR1*

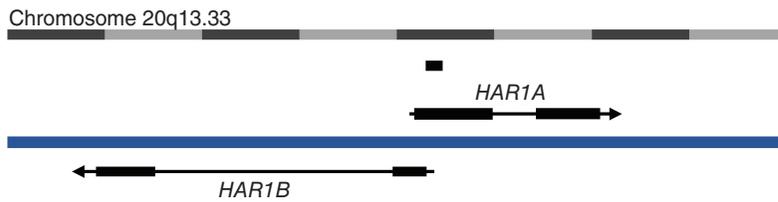


Figure 2. Simplified schematic of the HAR1 loci. The long noncoding RNA HAR1A is on the sense strand, and the long noncoding RNA HAR1B is on the antisense strand on chromosome 20. These divergently transcribed long noncoding RNAs have a region of overlap, containing the ‘accelerated’ 118bp HAR1. There is also a miscellaneous RNA overlapping with HAR1A.

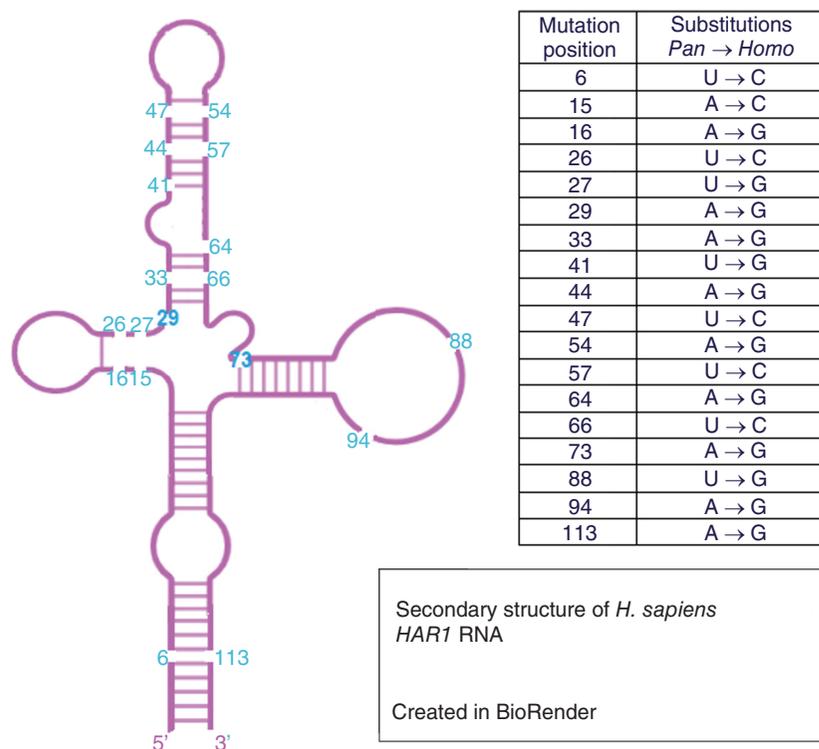


Figure 3. The human HAR1 secondary structure, including the mutations (numbers in blue) that have rapidly occurred between chimpanzee and human. Figure created using BioRender.com.

may have occurred in the human lineage within the last 1 Myr [47].

A computational model was developed to understand how the human HAR1 secondary structure might have evolved, comparing the chimpanzee (ancestral), the *Denisovan* and the modern human HAR1 secondary structures [51]. The computational model reconstructed the statistically most likely order of the 18 substitutions. The likely final mutation in HAR1 human evolution thus far is the variant resulting in the stabilization of the lowest stem. This stem has been recreated in the modern human, as it was weakened in the evolution of ancestral HAR1 to *Denisovan* [51]. These analyses, along with the predicted stability of the HAR1 helix structure having rapidly increased in humans, suggest that the structure of HAR1 is highly relevant to its function.

Although the secondary structure and nucleotide changes within the 118 base pair region HAR1 region has been mapped, little is known about the overall HAR1A and HAR1B structures, and any other functionally relevant sequences within these lncRNAs. As previously explained, identifying other short sequences, mapping their structures and validating in the laboratory is important for understanding how a lncRNA is functional, as it

has been recently shown that short syntenic sequences within lncRNAs are the functional elements, rather than the whole lncRNA sequence. A tool which may be relevant for *HARIA* and *HARIB* is the lncLOOM algorithm [35], which is designed to identify biologically relevant short sequences, 6–12 nucleotides long, that are deeply conserved within lncRNAs and other elements. These short sequences were compared across several species in lncLOOM, including humans, mice, opossums, chickens and zebrafish, with 18 different species compared for their analysis of the *Cyano* lncRNA. These short sequences were identified in syntenic regions as having been conserved over large evolutionary distances because the order of nucleotides was identical across different species. This synteny may be important for function; therefore, these highly conserved motifs are likely the functional elements within the lncRNA. Many of these motifs were found to be regions of the lncRNA that contribute to the binding sites for RNA binding proteins and smaller RNAs [35].

Another study has found two lncRNAs, *APOLO* and *UPAT*, one in plants and the other in humans, respectively, have very different sequences but are involved in the same pathways and protein interactions [52]. As these lncRNAs are seemingly evolutionary unrelated in terms of sequence, it would be interesting to identify whether they share common short motifs or if their 3D structures are similar. This study demonstrates how lncRNA evolution is seemingly unique, and further study is required to increase the understanding of lncRNA functionality. As lncLOOM, or similar tools, can be used to identify functionally relevant short sequences within lncRNAs, it would be interesting to identify whether the *HARI* sequence includes one of these elements, with this being highly evolutionary accelerated, or if there are functionally conserved elements elsewhere in *HARIA* and *HARIB* that are more relevant to their functions in health and disease.

Identifying the tertiary interactions indicated by the secondary structure will help in establishing the mechanism of action of the lncRNAs overlapping with this accelerated region: *HARIA* and *HARIB*. Furthermore, it is necessary to identify whether short syntenic elements within the *HARIA* and *HARIB* sequences are responsible for a function or can be linked to a mechanism of action. These studies can, in turn, further identify the possible roles of *HARIA* and *HARIB* in health and disease.

HARI lncRNA loci in human health & disease

As explored in this review, the extent of a lncRNA's evolutionary conservation is determined by several different factors, including the mechanism of its origin, its age, genomic environment and the structure and sequences of the lncRNA itself. Many of these factors, especially the selection of functionally conserved regions through long evolutionary distances, and the folding of the lncRNA structures, may contribute to these genes have functional roles in multiple diseases. To corroborate these models, we discuss recent evidence, showing the highly evolved *HARI* loci may have a functional role in neurodevelopment, neurological disorders and cancer.

HARI loci in neurodevelopment

HARIA expression was identified specifically in Cajal-Retzius cells in the developing human neocortex and the dorsal telencephalon, between the 7th and 9th gestational weeks, and in the subpial granular layer up to 19 gestational weeks [47]. *HARIB* expression was shown to be more widespread in the brain. The main function of Cajal-Retzius cells is the organization of neurons during neocortical development [53]. Cortical lamination is the organization of cells into six excitatory layers (I–VI). This cell lamination is achieved by the release of signals, such as the extracellular glycoprotein RELN [54]. As the expression of *HARIA* overlaps with the expression of *RELN* in Cajal-Retzius cells between the critical development period of 17th–19th gestational weeks, it is necessary to understand their possible functional relationship, and to discern the role of *HARIA* and *HARIB* in nervous system development.

HARI loci in neurological & psychiatric disorders

Owing to the preferential expression of *HARI* in the human cortex [47], it is not surprising that emerging evidence has linked this locus to neurological and psychiatric disorders. A key study has shown that the *HARI* locus may have a functional role in Huntington's disease [55]. Huntington's disease is an autosomal-dominant neurological disorder, with typical symptoms including involuntary movements, depression and dementia [56]. It is caused by a CAG trinucleotide expansion in the *HTT* gene [57], that results in the relocation of the transcriptional repressor, *REST*. The *REST* gene represses thousands of genes, many which are neuron specific [58].

The *HARI* locus was shown to be transcriptionally regulated by the REST protein in Huntington's disease in cell lines and human tissue [55]. Firstly, the authors identified three REST binding sites in the vicinity of

HAR1. Two of these contain canonical RE1 binding motifs that are specifically found in promoter regions [59]. Consistent with a transcriptional silencing effect of *REST*, quantitative PCR analysis demonstrated increased levels of *HARIA* and *HARIB* mRNAs when *REST* was silenced *in vitro* [55]. The authors also analyzed human brain tissue, detecting a significantly lower level of *HARIA* and *HARIB* transcripts in the striatum of Huntington's disease patient tissue, compared to normal brain. The striatum has a significant role in motor control, and it is the main region causing neurodegeneration in Huntington's disease [60]. The normal mechanisms of action of the *HARIA* and *HARIB* lncRNAs, likely regulated by *REST*, are poorly understood, but their dysregulation in Huntington's disease striatum suggests they may contribute to abnormal cellular function and phenotype.

HARIA and *HARIB* have also been associated with Alzheimer's disease. A machine-learning based model called Laplacian Regularized Least Squares for lncRNA–Disease Association (LRLSLDA) predicts novel lncRNA–disease associations, where the lncRNA may have a functional role [61]. LRLSLDA associated *HARIA* and *HARIB* to Alzheimer's disease [62]. This model is based on the assumption that similar diseases have associations with functionally similar lncRNAs. These parameters may be a limitation of the model, and *in vitro* confirmation of this prediction is required.

The role of *HARIA* was investigated in schizophrenia, however, a significant association was not identified within the 285-patient sample [63]. Nevertheless, the authors did find that a CCCC GC haplotype combined of six single nucleotide polymorphisms covered the *HARIA* region completely and was significantly associated with auditory hallucinations in 221 psychiatric regions [63]. As hallucinations can be as a result of erroneous neuronal connections forming in brain development, the activity of *HARIA* may be altered, increasing the chance of psychosis, but this requires further validation.

These studies indicate that *HARIA* and *HARIB* may have a role in neurological and psychiatric disorders, but more evidence is required. However, there is increasing recognition for the important role that lncRNAs have in Huntington's disease and other neurological disorders [64–66].

HAR1 loci in cancer

Similar to neurological and psychiatric disorders, the *HAR1* locus has been implicated in multiple cancers, including gliomas. Bioinformatics analysis showed significant downregulation of *HARIA* in diffuse glioma samples [67]. The authors found that lower *HARIA* expression resulted in significantly worse survival of diffuse glioma patients. Upregulation of *HARIA* increased the survival rates of patients who underwent radiotherapy and chemotherapy, demonstrating a tumor suppressor role. Another study showed that *HARIA* downregulation also resulted in worse survival of diffuse glioma patients, specifically with the prognostic marker IDH mutant [68]. These results suggest that the downregulation of *HARIA* may be a prognostic biomarker for diffuse gliomas, though the mechanisms of action and cellular function of this lncRNA are still unknown.

In another study using microarray analysis, *HARIB* was one of nine lncRNAs in a signature pattern of gene expression, which were found to be significantly associated with glioblastoma survival [69]. The other eight lncRNAs are *AC078883.3*, *AC104653.1*, *RP11-944L7.4*, *RP4-635E18.7*, *RP5-1172N10.2*, *TP73-AS1*, *SAPCD1-AS1* and *HOTAIR*. These lncRNAs together could be a potential biomarker for diagnosis and prediction of survival. Further research into the lncRNAs biological functions and mechanism of action is also required.

In addition to gliomas, *HARIA* and *HARIB* are dysregulated in other types of cancer. A recent study identified *HARIB* as a predictive biomarker in bone and soft-tissue sarcoma cell lines [70]. *HARIB* has differential expression in these cell lines, and siRNA silencing of *HARIB* resulted in an increased resistance to a therapeutic inhibitor, pazopanib. *HARIB* was upregulated in clear cell renal cell carcinoma within a signature containing *MIR155HG*, *PVT1* and *TCL6*, and was correlated with poor overall survival [71]. *HARIB* may also have a role in human parathyroid tumors, as silencing of the tumor suppressor *MEN1* upregulated *HARIB* [72]. The study reported that *HARIB* silencing increased *SOX2* and *NANOG* levels in primary cell lines. Overexpression of these genes can result in cancer hallmarks such as metastasis. The extent of the role of *HARIB* in this cancer requires further investigation. A study in breast cancer showed that the upregulation of *HARIA* within a signature containing eight other lncRNAs (*LINC00310*, *LINC00323*, *LINC00574*, *LINC00704*, *LINC00705*, *ARRDC1-AS1*, *FAM74A3* and *UMODL1-AS1*) predicted cancer reoccurrence [73]. For six of these lncRNAs together, including *HARIA*, the alteration was significantly higher within the advanced breast cancer group. *HARIA* has also been associated with recurrence free survival in papillary thyroid cancer [74]. A different study found that low expression of *HARIA* and *HARIB* in patients with hepatocellular carcinoma, compared to liver cirrhosis and chronic hepatitis B cases, was significantly associated with poor prognosis, advanced histological grade and progressive TNM stage [75].

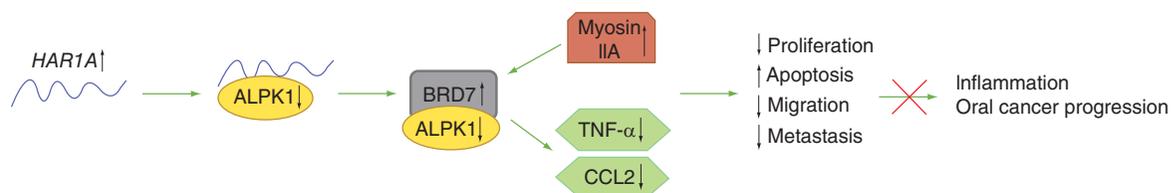


Figure 4. The mechanism of action of *HARI1* in oral cancer cell lines, demonstrating *HARI1* as a tumor suppressor. When *HARI1* is upregulated, ALPK1 levels reduce. This results in increased levels of BRD7 and myosin IIA, resulting in tumor suppression. Decreased levels of TNF- α and CCL2 with *HARI1* upregulation suggest *HARI1* has an anti-inflammatory role.

Another recent publication has analyzed the downstream mechanism of action and functional role of *HARI1* in oral cancer progression [76]. *HARI1* was found to bind to the oncogene *ALPK1* in the nucleus of an oral squamous cell carcinoma cell line, SAS, with *HARI1* and *ALPK1* having an inverse correlation. The discovery of *HARI1* being primarily localised in the nucleus is important because of lncRNAs functional specificity depending on their subcellular localization. *HARI1* knockdown increased the protein levels of pro-inflammatory cytokines TNF- α and CCL2, with *ALPK1* knockdown decreasing these levels protein levels. Therefore, it is proposed that *HARI1* may suppress a ALPK1/BRD7/myosin IIA pathway (Figure 4), as *HARI1* knockdown may cause the ALPK1 protein to translocate into the nucleus, bind and downregulate BRD7, resulting in inflammation and oral cancer progression [76]. Further to a possibly anti-inflammatory role, the function of *HARI1* was also analyzed in oral cancer cell lines. *HARI1* silencing increases oral cancer cell proliferation, migration and promote apoptosis, further implicating *HARI1* as being a tumor suppressor. *HARI1* silencing may also increase metastasis, investigated using epithelial-mesenchymal transition (EMT) markers. Mesenchymal markers N-cadherin, fibronectin, vimentin and slug were upregulated with *HARI1* silencing, and this did not upregulate the epithelial marker E-cadherin. These results further implicate *HARI1* as having a tumor suppressor role [76]. Although this study is one of the first identifying *HARI1* downstream interacting partners, lncRNAs are often cell and disease specific, therefore further investigation into this mechanism of action and other interacting partners are needed in different cancers and *in vivo* studies.

These cancer studies show that *HARI1* and *HARI2* expression can be oncogenic in some cancers [71,73] and tumor suppressors in other cancers [67,75,76], and may be used as a prognostic biomarker to indicate the stage of tumor progression. They also suggest a functional role of these lncRNAs in human cancers, especially gliomas and oral carcinomas, but this requires further research in different types of cancer. As with neurological disorders, there is also an increasing focus for identifying the functional role of lncRNAs in cancer [77–79].

Conclusion

HARI1 and *HARI2* are lncRNAs at the HARI1 loci which have been implicated in health and disease. As explained within this review, the functions of *HARI1* and *HARI2* in disease and the upstream mechanisms of action regulating these lncRNAs is currently unknown. An investigation into *REST* regulating *HARI1* and *HARI2* in diseases other than Huntington's disease is required. Furthering our knowledge of evolutionary accelerated regions, through investigating their sequences and structures, will enhance our understanding of *HARI1* loci functions in human disease. However, functionally conserved regions elsewhere in the *HARI1* and *HARI2* sequences may also contribute to their lncRNA function.

Future perspective

lncRNA regions comprise a significant part of our genome, owing to the abundance of studies identifying novel lncRNAs over the last decade. The more that lncRNA structures and sequences are analyzed, the more is shown that they have a wide range of functions and expression patterns within a cell, tissue, disease or species. Owing to their specificity, lncRNAs are recently emerging as promising biomarkers and therapeutic targets [80].

As lncRNAs typically have low sequence conservation and no protein-coding ability, their functionality was previously questioned. Studies are now emerging to further understand lncRNAs roles in health and disease. Many lncRNAs have been found to contain regions of high sequence conservation which may be responsible for their function and interactions with other molecules [81]. Other lncRNA regions, such as *HARI1* and *HARI2* in the *HARI1* loci, have undergone rapid evolution, specifically within the human lineage. This high rate of sequence

turnover is possibly the reason the structure and function of these lncRNAs are likely unique in humans. The *in silico* development of lncRNA-specific algorithms has enabled more data to be gathered regarding the degree of lncRNA sequence conservation and their structures [35–38].

lncRNAs that are expressed in the developing brain and kidney, during embryonic development, had higher levels of functional constraint and therefore may function in *cis* from the process of transcription rather than from the lncRNA itself [38]. This is a significant finding and a potential focus for future studies of lncRNAs. For example, if *HARIA*, which was previously shown to be expressed early in brain development [47], conforms to these findings, *HARIA* may have evolved to have higher levels of functional constraints within humans, and therefore may have a possible active function within neocortical development. Using algorithms such as lncLOOM may allow evolutionary conserved motifs to be identified within the lncRNA sequence that are responsible for the function [35], and tools may also identify other regions that have undergone rapid sequence changes, like *HAR1*.

As discussed in this review, the lncRNAs *HARIA* and *HARIB* have been shown to have significantly altered expression and association in various diseases including neurodevelopment, Huntington's disease [55], and cancers including gliomas [67] and oral cancer [76], although their functional roles and upstream mechanisms of action are unknown, so their novelty presents a wide scope for study. As *HARIA* is co-expressed with *RELN*, and *RELN* is responsible for the lamination of cortical neurons [54], *HARIA* may have a functional role in migration and possibly regulates *RELN*. It has previously been demonstrated that cortical layers III, V and VI of Huntington's disease patients experience loss of neurons, especially as the disease progresses [82]. Downregulation of *HARIA* in Huntington's disease, transcriptionally regulated by *REST*, may contribute to the architecture of layering neurons and this loss of cells. Furthermore, tumor cell invasion was identified in deeper layers (IV–VI) of the cerebral cortex in malignant gliomas [83]. *HARIA* was identified to be downregulated in more aggressive gliomas [67], meaning *HARIA* may have a role in cancer cell invasion and migration here. The extent of their regulatory relationship with *REST*, likely located upstream of *HARIA* and *HARIB*, will be interesting to further establish. This potential migratory role requires laboratory characterization *in vitro* and *in vivo* relevant to these diseases, using techniques such as transwell migration assays, and Boyden chamber assays.

HARIA was identified to also have a tumor suppressor role in oral cancer [76]. This publication demonstrated that *HARIA* and the protein kinase *ALPK1* may bind in monocytes, and *HARIA*, possibly indirectly, inhibits this gene in oral cancer cell lines. *HARIA* knockdown resulted in increased cell viability, increased migration, and reduced cell apoptosis [76]. *HARIA* silencing resulted in oral cancer progression and inflammation. The domains of *HARIA* binding *ALPK1* have not yet been identified. It would be interesting to understand whether the *HAR1* evolutionary accelerated region is included in this, or whether there are unknown functional motifs. Another key result was the identification of *HARIA* in the nucleus [76]. In oral cancer, *HARIA* may function in *cis* to epigenetically regulate *ALPK1*, possibly resulting in inflammation-promoted tumorigenesis. These findings suggest that *HARIA* may have a number of different biological roles within health and disease including migration, inflammation and tumorigenesis, dependent on the cell and tissue type.

As *HARIB* is less characterized in comparison to *HARIA*, *HARIB* requires further laboratory characterization to analyze the subcellular location in the appropriate cell line, identify a phenotypic effect and lncRNA function through gain-of-function and loss-of-function experiments comparing cells with/without lncRNA expression, and find possible binding partners and interacting-macromolecules to discover a mechanism of action. Direct binding partners can be investigated using techniques such as RNA pull-down assays and mass spectrometry [84], followed by luciferase reporter assays and chromatin precipitation [85]. *HARIA* also requires further characterization in other cancers, such as gliomas, and neurodegenerative disorders. The secondary structures of these lncRNAs also require mapping. As well as *in vitro* and *in vivo* characterization of these, understanding the tertiary structure of *HARIA* will aid in locating binding sites and other motifs of functional interest. The RNA structure of the lncRNA *Xist* was found to vary, likely depending on the experiment [86]. Structural and wet lab characterization of lncRNAs should be a consistent set of experiments to reduce variability between laboratories.

As shown by the *HAR1* loci, further knowledge of the evolutionary history of lncRNAs, including their evolved or evolving structures and degree of sequence conservation within the human lineage, can guide their molecular and functional characterization, with the final aim to identify their roles in health and disease. Methods to further understand the roles of lncRNAs such as *HARIA* and *HARIB* that have aberrant expression in health and disease will provide insights into improving diagnostics and therapeutic options.

Executive summary

Introduction to long noncoding RNAs

- Long noncoding RNAs (lncRNAs) are nonprotein coding regions over 200 nucleotides in length. lncRNAs are often cell- and disease-specific, and their functions can be inferred by their subcellular location.
- lncRNAs are involved in several human diseases, but the majority of lncRNAs are still structurally and functionally uncharacterized.

lncRNA origin, evolution & conservation

- lncRNA evolutionary conservation can be determined by the mechanism of origin, age, genomic environment and structure.
- Numerous studies have indicated that many lncRNAs have poor evolutionary sequence conservation; therefore, functionality may be conserved via specific sequence motifs or their folded structures.

Rapid evolution of lncRNAs

- The genetic loci encoding lncRNAs can be subject to an accelerated rate of nucleotide substitutions within the human lineage.

Case study: *HAR1*

- An example of a region that has undergone accelerated evolution is the *HAR1*. The *HAR1* sequence resides within two overlapping but divergently transcribed lncRNAs, *HAR1A* and *HAR1B*.
- The *HAR1* secondary structure has been mapped in *Homo* and *Pan*. This structure stabilizes in humans, suggesting the structure is relevant for *HAR1* function.

HAR1 lncRNA loci in human health & disease

- *HAR1A* is expressed in the brain during neurodevelopment. The function of *HAR1A* in neurodevelopment is currently unknown.
- The *HAR1* locus is involved in multiple neurological disorders. In Huntington's disease, *HAR1A* and *HAR1B* are transcriptionally regulated by *REST*.
- The *HAR1* locus is also implicated in multiple cancers, including tumor suppressor roles in gliomas and oral cancer. *HAR1A* is present in the nucleus. *HAR1A* was also shown to have an anti-inflammatory role.

Future perspective

- Further studies are required to identify the genes regulating the *HAR1* loci in cancer and neurological disorders.
- Understanding lncRNA structures and their evolutionary conservation will aid in identifying more lncRNA functions and their mechanisms of action in health and disease, and therefore improve therapeutic options.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Forrest ARR, Kawaji HFANTOM Consortium and the RIKEN PMI and CLST (DGT) A promoter-level mammalian expression atlas. *Nature* 507(7493), 462–470 (2014).
2. Harrow J, Frankish A, Gonzalez JM *et al.* GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22(9), 1760–1774 (2012).
3. Fang S, Zhang L, Guo J *et al.* NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46(Database issue), D308–D314 (2018).
4. Guttman M, Amit I, Garber M *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235), 223–227 (2009).
5. Ohno S. So much “junk” DNA in our genome. *Brookhaven Symp. Biol.* 23, 366–370 (1972).
6. Carlevaro-Fita J, Lanzós A, Feuerbach L *et al.* Cancer lncRNA census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* 3(1), 1–16 (2020).

7. Jiang MC, Ni JJ, Cui WY, Wang BY, Zhuo W. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am. J. Cancer Res.* 9(7), 1354–1366 (2019).
8. Li L, Zhuang Y, Zhao X, Li X. Long non-coding RNA in neuronal development and neurological disorders. *Front. Genet.* 9 (2019).
9. Shi C, Zhang L, Qin C. Long non-coding RNAs in brain development, synaptic biology, and Alzheimer's disease. *Brain Res. Bull.* 132, 160–169 (2017).
10. Derrien T, Johnson R, Bussotti G *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22(9), 1775–1789 (2012).
11. Brunner AL, Beck AH, Edris B *et al.* Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* 13(8), R75 (2012).
12. Modali SD, Parekh VI, Kebebew E, Agarwal SK. Epigenetic regulation of the lncRNA MEG3 and its target c-MET in pancreatic neuroendocrine tumors. *Mol. Endocrinol.* 29(2), 224–237 (2015).
13. Gonzalez I, Munita R, Agirre E *et al.* A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat. Struct. Mol. Biol.* 22(5), 370–376 (2015).
14. He X, Zheng Y, Zhang Y *et al.* Long non-coding RNA AK058003, as a precursor of miR-15a, interacts with HuR to inhibit the expression of γ -synuclein in hepatocellular carcinoma cells. *Oncotarget.* 8(6), 9451–9465 (2016).
15. Li Z, Hou P, Fan D *et al.* The degradation of EZH2 mediated by lncRNA ANCR attenuated the invasion and metastasis of breast cancer. *Cell Death Differ.* 24(1), 59–71 (2017).
16. Chen R, Kong P, Zhang F *et al.* EZH2-mediated α -actin methylation needs lncRNA TUG1, and promotes the cortex cytoskeleton formation in VSMCs. *Gene* 616, 52–57 (2017).
17. Tang Y, He Y, Zhang P *et al.* LncRNAs regulate the cytoskeleton and related Rho/ROCK signaling in cancer metastasis. *Mol. Cancer* 17, 77 (2018).
18. Guo B, Wu S, Zhu X *et al.* Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 39(1), e102190 (2020).
19. Niu L, Lou F, Sun Y *et al.* A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. *Sci. Adv.* 6(21), eaaz2059 (2020).
20. Fang Y, Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics, Proteomics & Bioinformatics.* 14(1), 42–54 (2016).
21. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21(2), 102–117 (2020).
22. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172(3), 393–407 (2018).
23. Jones AN, Sattler M. Challenges and perspectives for structural biology of lncRNAs – the example of the Xist lncRNA A-repeats. *J. Mol. Cell Biol.* 11(10), 845–859 (2019).
24. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22(2), 96–118 (2021).
25. Cabili MN, Trapnell C, Goff L *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25(18), 1915–1927 (2011).
26. Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* 22(1), 1–5 (2006).
27. Ponting CP, Nellåker C, Meader S. Rapid turnover of functional sequence in human and other genomes. *Annu. Rev. Genomics Hum. Genet.* 12, 275–299 (2011).
28. Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 30(10), 439–452 (2014).
29. Delias N. Formation of a family of long intergenic noncoding RNA genes with an embedded translocation breakpoint motif in human chromosomal low copy repeats of 22q11.2 – some surprises and questions. *Noncoding RNA* 4(3), 16 (2018).
30. Elisaphenko EA, Kolesnikov NN, Shevchenko AI *et al.* A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE.* 3(6), e2521 (2008).
31. Hezroni H, Ben-Tov Perry R, Meir Z, Housman G, Lubelsky Y, Ulitsky I. A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Bio.* 18(1), 162 (2017).
32. Chishima T, Iwakiri J, Hamada M. Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs. *Genes (Basel)* 9(1), (2018).
33. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4(11), 865–875 (2003).
34. Liu WH, Tsai ZTY, Tsai HK. Comparative genomic analyses highlight the contribution of pseudogenized protein-coding genes to human lincRNAs. *BMC Genomics* 18 (2017).

35. Ross CJ, Rom A, Spinrad A, Gelbard-Solodkin D, Degani N, Ulitsky I. Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biol.* 22(1), 29 (2021).
- **Discusses a new algorithm identifying long noncoding RNA (lncRNA) functional regions.**
36. Necseulea A, Soumillon M, Warnefors M *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505(7485), 635–640 (2014).
37. Chen J, Shishkin AA, Zhu X *et al.* Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* 17 (2016).
38. Darbellay F, Necseulea A. Comparative transcriptomics analyses across species, organs, and developmental stages reveal functionally constrained lncRNAs. *Mol. Biol. Evol.* 37(1), 240–259 (2020).
39. Engreitz JM, Haines JE, Perez EM *et al.* Local regulation of gene expression by lncRNA promoters, transcription, and splicing. *Nature* 539(7629), 452–455 (2016).
40. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147(7), 1537–1550 (2011).
41. Owens MC, Clark SC, Yankey A, Somarowthu S. Identifying structural domains and conserved regions in the long non-coding RNA lncTCF7. *Int. J. Mol. Sci.* 20(19), (2019).
42. Xue Z, Hennelly S, Doyle B *et al.* A G-rich motif in the lncRNA braveheart interacts with a Zinc-Finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* 64(1), 37–50 (2016).
43. Uroda T, Anastasakou E, Rossi A *et al.* Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* 75(5), 982–995.e9 (2019).
44. Matsumura K, Kawasaki Y, Miyamoto M *et al.* The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene* 36(9), 1191–1199 (2017).
45. Somarowthu S, Legiewicz M, Chillón I, Marcia M, Liu F, Pyle AM. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58(2), 353–361 (2015).
46. Pollard KS, Salama SR, King B *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics* 2(10), e168 (2006).
47. Pollard KS, Salama SR, Lambert N *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443(7108), 167–172 (2006).
- **This was the first study to identify the *HARI* loci.**
48. Beniaminov A, Westhof E, Krol A. Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 14(7), 1270–1275 (2008).
49. Ziegler M, Cevce M, Richter C, Schwalbe H. NMR studies of HAR1 RNA secondary structures reveal conformational dynamics in the human RNA. *Chem. Bio. Chem.* 13(14), 2100–2112 (2012).
- **Investigates the structure of the *HARI* region.**
50. Lares MR. Synthesis, purification and crystallization of a putative critical bulge of HAR1 RNA. *PLoS ONE* 14(11), e0225029 (2019).
51. Walter Costa MB, Höner Zu Siederdisen C, Tulpan D, Stadler PF, Nowick K. Temporal ordering of substitutions in RNA evolution: uncovering the structural evolution of the human accelerated region 1. *J. Theor. Biol.* 438, 143–150 (2018).
52. Fonouni-Farde C, Christ A, Blein T *et al.* Sequence-unrelated long noncoding RNAs converged to modulate the activity of conserved epigenetic machineries across kingdoms. *bioRxiv.* doi:2021.02.26.433017 (2021) (Epub ahead of print).
53. Frotscher M. Dual role of Cajal-Retzius cells and reelin in cortical development. *Cell Tissue Res.* 290(2), 315–322 (1997).
54. Hirota Y, Nakajima K. Control of neuronal migration and aggregation by Reelin signaling in the developing cerebral cortex. *Front. Cell Dev. Biol.* 5, 40 (2017).
55. Johnson R, Richter N, Jauch R *et al.* Human accelerated region 1 noncoding RNA is repressed by REST in Huntington's disease. *Physiol. Genomics* 41(3), 269–274 (2010).
- **Identified *REST* and *HARIA* expression in Huntington's disease as a possible mechanism of action.**
56. Roos RA. Huntington's disease: a clinical review. *Orphanet J. Rare Dis.* 5(1), 40 (2010).
57. Kay C, Hayden MR, Leavitt BR. Epidemiology of Huntington disease (Chapter 3). In: *Handbook of Clinical Neurology.* Feigin AS, Anderson KE (Eds). Elsevier, 31–46 (2017).
58. Gopalakrishnan V. REST and the RESTless: in stem cells and beyond. *Future Neurol.* 4(3), 317–329 (2009).
59. Myers SJ, Peters J, Huang Y, Comer MB, Barthel F, Dingledine R. Transcriptional regulation of the GluR2 gene: neural-specific expression, multiple promoters, and regulatory elements. *J. Neurosci.* 18(17), 6723–6739 (1998).
60. Rikani AA, Choudhry Z, Choudhry AM *et al.* The mechanism of degeneration of striatal neuronal subtypes in Huntington disease. *Ann. Neurosci.* 21(3), 112–114 (2014).
61. Chen X, Yan G-Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* 29(20), 2617–2624 (2013).

62. Chen X, Yan CC, Zhang X, You Z-H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinformatics* 18(4), 558–576 (2017).
63. Tolosa A, Sanjuán J, Leal C, Costas J, Moltó MD, de Frutos R. Rapid evolving RNA gene HAR1A and schizophrenia. *Schizophr. Res.* 99(1), 370–372 (2008).
64. Chanda K, Mukhopadhyay D. LncRNA Xist, X-chromosome instability and Alzheimer's disease. *Curr. Alzheimer Res.* 17(6), 499–507 (2020).
65. Simchovitz A, Hanan M, Yayon N *et al.* A lncRNA survey finds increases in neuroprotective LINC-PINT in Parkinson's disease substantia nigra. *Aging Cell* 19(3), e13115 (2020).
66. Sunwoo JS, Lee ST, Im W *et al.* Altered expression of the long noncoding RNA NEAT1 in Huntington's disease. *Mol. Neurobiol* 54(2), 1577–1586 (2017).
67. Zou H, Wu LX, Yang Y *et al.* lncRNAs PVT1 and HAR1A are prognosis biomarkers and indicate therapy outcome for diffuse glioma patients. *Oncotarget* 8(45), 78767–78780 (2017).
- **Identified low HAR1A expression results in worse patient survival for diffuse glioma patients.**
68. Chen Y, Guo Y, Chen H, Ma F. Long non-coding RNA expression profiling identifies a four-long non-coding RNA prognostic signature for isocitrate dehydrogenase mutant glioma. *Front. Neurol.* 11, 573264 (2020).
69. Lei B, Yu L, Jung TA *et al.* Prospective series of nine long noncoding RNAs associated with survival of patients with glioblastoma. *J. Neurol. Surg. A: Cent. Eur. Neurosurg.* 79(6), 471–478 (2018).
70. Yamada H, Takahashi M, Watanuki M *et al.* lncRNA HAR1B has potential to be a predictive marker for pazopanib therapy in patients with sarcoma. *Oncol. Lett.* 21(6), 455 (2021).
71. Liu H, Ye T, Yang X *et al.* A panel of four-lncRNA signature as a potential biomarker for predicting survival in clear cell renal cell carcinoma. *J. Cancer* 11(14), 4274–4283 (2020).
72. Verdelli C, Morotti A, Tavanti GS *et al.* The core stem genes SOX2, POU5F1/OCT4, and NANOG are expressed in human parathyroid tumors and modulated by MEN1, YAP1, and β -catenin pathways activation. *Biomedicines* 9(6), 637 (2021).
73. Liu H, Li J, Koirala P *et al.* Long non-coding RNAs as prognostic markers in human breast cancer. *Oncotarget* 7(15), 20584–20596 (2016).
74. Ma B, Liao T, Wen D *et al.* Long intergenic non-coding RNA 271 is predictive of a poorer prognosis of papillary thyroid cancer. *Sci. Rep.* 6, 36973 (2016).
75. Shi Z, Luo Y, Zhu M *et al.* Expression analysis of long non-coding RNA HAR1A and HAR1B in HBV-induced hepatocellular carcinoma in Chinese patients. *Lab. Med.* 50(2), 150–157 (2019).
76. Lee CP, Ko AMS, Nithiyantham S, Lai CH, Ko YC. Long noncoding RNA HAR1A regulates oral cancer progression through the alpha-kinase 1, bromodomain 7, and myosin IIA axis. *J. Mol. Med.* 99, 1323–1334 (2021).
- **Recent paper identifying the mechanism of action and functional roles of HAR1A in oral cancer, and the location of HAR1A in the cell nucleus.**
77. Guo B, Wu S, Zhu X *et al.* Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 39(1), e102190 (2020).
78. Hung T, Wang Y, Lin MF *et al.* Extensive and coordinated transcription of noncoding RNAs within cell cycle promoters. *Nat. Genet.* 43(7), 621–629 (2011).
79. Pandey GK, Mitra S, Subhash S *et al.* The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* 26(5), 722–737 (2014).
80. Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol. Med.* 24(3), 257–277 (2018).
81. Marín-Béjar O, Mas AM, González J *et al.* The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biology* 18(1), 202 (2017).
82. Sotrel A, Paskevich PA, Kiely DK, Bird ED, Williams RS, Myers RH. Morphometric analysis of the prefrontal cortex in Huntington's disease. *Neurology* 41(7), 1117–1123 (1991).
83. Mughal AA, Zhang L, Fayzullin A *et al.* Patterns of invasive growth in malignant gliomas – the hippocampus emerges as an invasion-spared brain region. *Neoplasia* 20(7), 643–656 (2018).
84. Bierhoff H. Analysis of lncRNA-protein interactions by RNA-protein pull-down assays and RNA immunoprecipitation (RIP). *Methods Mol. Biol.* 1686, 241–250 (2018).
85. Han H, Wang S, Meng J *et al.* Long noncoding RNA PART1 restrains aggressive gastric cancer through the epigenetic silencing of PDGFB via the PLZF-mediated recruitment of EZH2. *Oncogene* 39(42), 6513–6528 (2020).
86. Pintacuda G, Young AN, Cerase A. Function by structure: spotlights on Xist long non-coding RNA. *Front. Mol. Biosci.* 4 (2017).

