View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Queensland eSpace

The evolution of RNAs with multiple functions

Marcel E. Dinger*, Dennis K. Gascoigne, John S. Mattick*

Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia

*Correspondence should be addressed to:

Marcel E. Dinger, Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4101, Australia; phone: +61 (7) 3346 2070; fax: +61 (7) 3346 2101; email: m.dinger@uq.edu.au.

John S. Mattick, Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4101, Australia; phone: +61 (7) 3346 2079; fax: +61 (7) 3346 2101; email: j.mattick@uq.edu.au.

Keywords: dual functional RNAs; long noncoding RNAs; evolution of genes; origin of new genes; complex loci

Abstract

Increasing numbers of transcripts have been reported to transmit both protein-coding and regulatory information. Apart from challenging our conception of the gene, this observation raises the question as to what extent this phenomenon occurs across the genome and how and why such dual encoding of function has evolved in the eukaryotic genome. To address this question, we consider the evolutionary path of genes in the earliest forms of life on Earth, where it is generally regarded that proteins evolved from a cellular machinery based entirely within RNA. This led to the domination of protein-coding genes in the genomes of microorganisms, although it is likely that RNA never lost its other capacities and functionalities, as evidenced by cis-acting riboswitches and UTRs. On the basis that the subsequent evolution of a more sophisticated regulatory architecture to provide higher levels of epigenetic control and accurate spatiotemporal expression in developmentally complex organisms is a complicated task, we hypothesize: (i) that mRNAs have been and remain subject to secondary selection to provide trans-acting regulatory capability in parallel with protein-coding functions; (ii) that some and perhaps many protein-coding loci, possibly as a consequence of gene duplication, have lost protein-coding functions en route to acquiring more sophisticated trans-regulatory functions; (iii) that many transcripts have become subject to secondary processing to release different products; and (iv) that novel proteins have emerged within loci that previously evolved functionality as regulatory RNAs. In support of the idea that there is a dynamic flux between different types of informational RNAs in both evolutionary and real time, we review recent observations that have arisen from transcriptomic surveys of complex eukaryotes and reconsider how these observations impact on the notion that apparently discrete loci may express transcripts with more than one function. In conclusion, we posit that many eukaryotic loci have evolved the capacity to transact a multitude of overlapping and potentially independent functions as both regulatory and protein-coding RNAs.

Introduction

The paradigm that "DNA is transcribed into RNA, which is translated into a protein that exerts a phenotype" [1] has instilled a common notion that assumes each component in this pathway maps neatly on a one-to-one basis from one stage to the next. Such paradigms fit comfortably in an anthropomorphic universe where such relationships are typically created by design (e.g., relational database tables), but seem to seldom apply in biological space where evolution has not been inhibited by equivalent organizational or logistical constraints. Indeed, despite the very concept of the gene as an independent functional unit becoming increasingly controversial and its utility correspondingly diminished [2-6], it has to a large extent persevered by virtue of the databases that are used to organize their information, which are predominantly designed around one-to-one relationships between identifiers, symbols, sequences, products and functions. The recognition that RNAs can exert functions as both regulatory molecules and as information carriers to encode protein [7, 8], serves as another reminder that not only is the generally understood concept of the gene faulty, but that biology has not evolved within the limitations imposed by linguistic semantics or human-designed information storage and retrieval architectures. In this perspective, we consider the possible scenarios by which new genes (which we define here simply as functional outputs of the genome) may have evolved, and rationalize why it is likely that many genes act not only as both regulatory and messenger RNAs, but embed a multitude of other functions that are transacted at both RNA and protein levels.

1. RNA - the ancestral gene

RNA is an extraordinarily versatile molecule, with the capacity for information encoding, sequence-specific interactions, three- and four dimensional structure, and catalytic activity. The widely accepted hypothesis that cellular life originated within an RNA world of RNA, where RNA fulfilled both informational and catalytic functions, defines life's earliest genes as noncoding RNAs [9, 10]. In light of the extraordinary repertoire of functions and mechanisms by which noncoding RNAs are know to act within extant life, we envisage that the first stages in the evolution of any new transcribed region that lacks translational signals can similarly be selected for on the basis of its function acting as an RNA. Indeed, given the vast amount of noncoding space in eukaryotic genomes and the lack of signals required to enable function compared to that required for translation (i.e. ribosome binding site, in-frame start and stop codons, and appropriate signals for transport to the ribosome), it might be expected that considerable functional exploration may still be largely initiated at the RNA level. Once such RNAs come under positive selection by conferring some advantage to the host, the appropriate regulatory signals can continue to evolve to optimize the functionality of the new gene. It is at this juncture, where a new gene is under some level of regulated expression, that it becomes enabled as a platform for selection of a novel translatable protein product. At this stage, as with RNAs that can be transcribed before evolving a specific function, small ORFs in ncRNAs could be translated before evolving a specific function and then expanded if it becomes the subject of positive selection. This scenario provides a model for the gradual evolution of a completely novel protein rather than just a reshuffling of existing domains into new architectures. In the event of positive selection for the protein product, conflicting selective pressures between the noncoding and protein-coding function may emerge. As with analogous scenarios in nature, diverse outcomes are possible in such relationships; e.g. the protein-coding function supplants the noncoding function or the two co-adapt in a symbiotic relationship.

It is widely accepted, as the favoured hypothesis, that early in the evolution of modern cellular life RNA devolved its information storage and inter-generational transmission functions largely to the more stable and easily replicable DNA, and its catalytic functions to the more chemically versatile proteins, which comprise the bulk of the analogue effectors of the system. Consequently, the genomes of microorganisms are dominated by protein-coding sequences, although it is also clear that RNA has retained regulatory capacities, at least in cis, as exemplified by riboswitches [11], as well as by untranslated regions (UTRs) in eukaryotic mRNAs (see below). By contrast, mammals and other complex organisms have only a minority of their genomes occupied by protein-coding sequences, and the majority of the RNA produced is non-protein-coding [12]. These noncoding RNAs are expressed in very precise patterns in different cells and tissues, and there is increasing evidence of their functionality, especially in the regulation of epigenetic processes, which are essential to differentiation and development [13]. Indeed, we have argued elsewhere that the major challenge for the emergence of developmentally complex and cognitively advanced organisms was regulatory, and that RNA allowed the separation of signal from consequent action, a more efficient framework [8, 14-16].

This scenario paints a picture where RNA functionality both precedes and follows protein-coding function and that the subsequent evolution of novel protein-coding function can co-exist with noncoding functions. For such a process to be evolutionary favourable, the advantage of developing the regulatory signals required for specific spatiotemporal expression would need to outweigh the constraints imposed by the co-evolution of coding and noncoding functions within the same transcript. In fact, irrespective for the expanded need for epigenetic regulatory RNAs, it may (also) be essential in complex organisms for efference regulatory information to be produced in parallel with protein-coding functions, either within the same RNA or intronically-derived RNA products. Given the extraordinary degree of sophistication involved in ensuring appropriate expression [17-22], this conjecture is plausible. Indeed, in an analogous situation in viruses where there is selective constraint against genome size, similar overlapping functionality is also observed in the same genomic sequence. The hypothesis that the development of regulatory architecture outstrips the expense of functionality evolving within limited sequence space fits with the observed structure of the genome, where genes commonly reside in clusters that incorporate complex overlapping transcriptional activity. As exemplified in a comparison of the *Dlx6* locus in mouse and human, lineage-specific coding and noncoding transcripts seem to have independently evolved within this homeotic locus, perhaps as a means to tune the developmental patterning related to this gene (Figure 1). Such clusters of transcription may represent genomic environments that evolved specialized architecture to facilitate appropriate spatiotemporal expression and/or the recruitment of specific post-transcriptional splicing or editing machinery and therefore become favoured sites for the evolution of new coding or noncoding genes.

This scenario can be likened to the development of a city, where the combined architecture and resources serve as leverage for new business ("genes") that would be less competitive or could not exist outside of these regions. To further this analogy, competition for space necessitates increased efficiencies in use of real estate, which in the case of cities is manifested in its extreme as skyscrapers in the central business district. Similarly, it can be posited that, despite established regulatory architecture serves as an attractant for evolutionary innovation resulting in a similar competition for genetic real estate and increased pressure for overlapping functions.

2. Long noncoding RNAs commonly contain open-reading frames

One of the key difficulties in distinguishing long noncoding RNAs (lncRNAs) from protein-coding RNAs lies in the fact that many lncRNAs contain substantial open reading frames (ORFs). We have shown previously that long ORFs can occur by chance alone [23], but examination of functionally annotated lncRNAs show that the ORFs present in these transcripts are actually much longer than expected by chance (Figure 2). Furthermore, it is expected that in cases where novel protein-capacity has evolved on the platform of a functional noncoding RNA, that such proteins would initially be constrained to be short in length. Indeed, given the reliance on protein-coding gene annotation on similarity to known proteins or overall ORF length, it is probable that there are many small and/or lineage-specific proteins and peptides that have yet to be discovered [24]. Furthermore, the absence of a detectable protein encoded by lncRNAs does not rule out its existence as it may only be translated at low levels and/or in particular developmental or cell-type specific contexts. Therefore, as demonstrated by instances such as SRA, which was initially described as a noncoding RNA and later identified to also encode a functional protein [25], it is conceivable that many lncRNAs can in fact encode proteins as either their primary or secondary function. On this basis, examples of functionally-validated lncRNAs such as H19 [26], and *TUG1* [27] that have large open reading frames (256 and 82 amino acids respectively) cannot be reasonably ruled to never be translated into functional proteins; it may simply be the case that they are translated in very specific contexts or at low levels.

Finally, in light of the growing number of examples of RNAs that function at both coding and noncoding levels, the corollary that many described messenger RNAs may also act as noncoding RNAs merits consideration. In the following section, evidence is presented that illustrates scenarios under which messenger RNAs may be expected to have retained or evolved additional functionality as noncoding RNAs.

3. Adoption of noncoding functionality into protein-coding genes and gene loci

As best illustrated by ancestral life in the RNA world, protein-coding function can emerge on the platform of noncoding RNA. However, it is similarly feasible for protein-coding RNAs to adopt noncoding functions. In cases such as *XIST*, it seems that the protein-coding function has been supplanted by a noncoding function [28]. However, as vividly illustrated in the case of snoRNAs and their host genes, noncoding RNAs can readily co-exist within protein-coding genes [29]. In addition to snoRNAs, many other stable RNAs have been identified that arise from intronic regions of the genome [30]. For example, we recently described an unspliced lncRNA that arises from the intron of the *SPRY4* protein-coding gene that has an important role in melanocyte proliferation and mobility [31]. Furthermore, messenger RNAs have been shown to act as regulatory molecules by competing for microRNA binding [32] and in the case of p53, the mRNA region encoding the Mdm2-binding site interacts directly with Mdm2, which in turn impairs the E3 ligase activity of Mdm2 and promotes p53 mRNA translation [33]. In another striking example, an alternative splicing variant of the nuclear receptor LXR acts as an RNA co-activator to LXR itself [34]. In terms of small RNAs, an intersection of protein-coding exons and miRNAs in human reveals at least 20 miRNAs that are encoded within protein-coding exons. In each of these examples, it is likely that the protein-coding function preceded the adoption of the noncoding function. Moreover, the respective functions of the coding and noncoding portions of the transcript in many cases appear to bear little or no relationship to one another, despite sharing the same regulatory architecture.

In light of recent transcriptomic studies revealing the regulated and conserved post-transcriptional processing of protein-coding transcripts into smaller RNAs [35, 36], it is possible that many transcripts harbour noncoding functionality. In some cases such functions overlap the protein-coding exons, but in others it occurs within untranslated regions (UTRs). Indeed, in a recent study, we showed that 3'UTRs can in some cases be expressed separately from the host transcript [37]. This phenomenon, which bears the hallmarks of biological function in terms of its tight regulation and conservation across diverse species, raises the hypothesis that UTRs harbour functions independent of the host protein-coding transcript. As with intronic regions, UTRs provide a practical substrate for evolutionary innovation in the sense that they do not necessarily interfere with the function of the protein-coding portion of the gene, yet already have the appropriate regulatory infrastructure to be expressed at some level within a particular spatiotemporal pattern. Once such a transcript confers a function, further post-transcriptional events can be adapted to fine-tune its expression or, as in the case of some of the 3'UTRs we examined, become expressed distinctly from the the host gene [37]. There are several examples of 3'UTRs conferring function independent to that of the encoded protein in the associated transcript, such as oskar, where the 3'UTR alone is sufficient to rescue the egg-less defect in an *oskar* null mutant. Similarly, the 3'UTRs of troponin I, tropomyosin, alpha-cardiac actin, ribonucleotide reductase, myotonic dystrophy (DM) protein kinase and prohibitin genes can act *in trans* to control cell proliferation and differentiation in the absence of associated coding-regions [38-42]. In addition, it is also possible for novel protein-coding function to emerge from within UTRs, such as identified for c-myc, where a novel independently translated protein was identified in the 5'UTR [43].

The possibility of widespread independent functionality of UTRs, cleavage of protein-coding exons into functional small RNAs, and overlapping regulatory functions within mRNAs, is supported by conservation of synonymous sites of codons [44, 45] and the enrichment of conserved RNA secondary structure within both coding exons and UTRs [46-49]. Growing evidence for functional selection challenges assumptions that synonymous sites evolve neutrally [50]. For example, comparisons between protein-coding and intergenic regions in human and chimp indicate that ~39% of synonymous sites are deleterious and subject to negative selection [51]. The additional conserved information embedded in mRNAs has been shown to effect splicing outcomes [52], co-translational protein folding [53] and other regulatory roles including co-repression [33], co-activation [34] and nuclear organization [54, 55].

Another mechanism by which existing regulatory infrastructure for an existing gene can be harnessed for the evolution of a new gene is through bi-directional promoters. Many promoters are symmetrical in nature and provide the opportunity for RNA polymerase to initiate transcription in either direction [56, 57], albeit that one direction may be favoured over the other. This bi-directional transcription affords the possibility for a transcript to be expressed coordinately with another gene, while having the benefit that it is not under any sequence constraints by the associated gene. This model for genomic evolution appears to be common in the genome, with evidence for thousands of bi-directional transcripts in vertebrate genomes [58, 59]. We recently explored such an example in the *Znfx1* loci, where a noncoding RNA (*Zfas1*) was expressed opposite to the *Znfx1* promoter [60]. Detailed expression profiling of the protein-coding and noncoding gene revealed that although the expression of these transcripts was coordinated in most tissues, there were some instances where their expression appeared to be independent. This suggests that the functions fulfilled by such transcripts may in fact be independent and that subsequent regulatory mechanisms have evolved to provide a means by which to decouple their expression in some contexts. Intriguingly, this noncoding gene, the spliced product of which we showed had a specific role in the proliferation of epithelial cells in mammary development, was also host to three intronic snoRNAs [60]. Amongst these snoRNAs, we also observed a degree of decoupling of expression levels amongst their expression and the host transcript, again suggestive of independent regulatory controls at a post-transcriptional level. At least some of the non-proportionality between the snoRNA expression could be rationalized by a highly thermodynamically stable structure for one of the snoRNAs that may have conferred it with a much longer half-life. The examination of the *Znfx1* loci provides a glimpse into the chain of evolutionary events that must have occurred to bring these disparate RNAs together and illustrates the co-evolution of coding and noncoding elements within a single locus. Cursory browsing through the transcriptional outputs of any of the well-annotated mammalian genomes indicates that such microcosms of gene evolution are relatively common and supports the notion that evolutionary innovation is favoured to occur within existing loci, presumably as a consequence of leveraging off the existing regulatory architecture.

Concluding remarks - the advent of multifunctional RNAs

A recent review on bifunctional RNAs bore the title "When one is better than two: RNA with dual functions" [7]. Here we propose that RNAs may potentially have numerous functions. This functionality may reside within introns, UTRs or overlap coding exons and can be manifested in both noncoding RNA function or translated products. Through the prism of our model of gene evolution, where regulatory architecture and recruitment of post-transcriptional processing machinery provides a favourable platform for genetic innovation, we predict that many loci will contain multiple layers of overlapping and potentially independent functions. This model is supported by those hotspots in the genome that contain a remarkable intersection of transcriptional and epigenetic complexity, whose products are processed through highly complex pathways involving splicing, editing and other modifications. Such a model of innovation can be likened to that of mobile phones, which were initially developed to fulfil the single function of allowing mobile voice calls. However, the platform required to necessitate this functionality, a mobile network, a display, a speaker, microphone, keyboard, and a battery, could be adapted to enable other functions, such as mobile internet, media playback, and games. Indeed, with the advent of downloadable applications, the mobile phone has been rapidly transformed into a system that has enabled the rapid innovation of thousands of novel functions. We envisage an analogous recruitment of complex regulatory architecture in the eukaryotic genome may similarly have been capitalized upon to generate diverse functions from a single locus. In light of the remarkable complexity observed at both the epigenetic and transcriptional levels in particular foci in eukaryotic genomes, a complexity that continues to grow with the technological means by which to examine these processes [61-63], we suggest that this possibility merits consideration in the interpretation of genetic studies, particularly those associated with complex loci, as well as more broadly in our conception of how information is stored in the genome.

Acknowledgments

We thank Paulo Amaral and Seth Cheetham for helpful discussions and critical reading of the manuscript and Tim Mercer for assisting in preparation of the figures. This work was supported by a National Health and Medical Research Council of Australia (NHMRC) Australia Fellowship (631668; JSM), a NHMRC Career Development Award (631542; MED), and a Queensland Government Department of Employment, Economic Development and Innovation Smart Futures Fellowship (MED).

References

[1] S. Brenner, F. Jacob, M. Meselson, An unstable intermediate carrying information from genes to ribosomes for protein synthesis, Nature 190 (1961) 576-581.

[2] M.E. Dinger, P.P. Amaral, T.R. Mercer, J.S. Mattick, Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications, Brief Funct Genomic Proteomic 8 (2009) 407-423.

[3] T.R. Gingeras, Origin of phenotypes: genes and transcripts, Genome Res 17 (2007) 682-690.

[4] G. Pesole, What is a gene? An updated operational definition, Gene 417 (2008) 1-4.

[5] M.B. Gerstein, C. Bruce, J.S. Rozowsky, D. Zheng, J. Du, J.O. Korbel, O. Emanuelsson, Z.D. Zhang, S. Weissman, M. Snyder, What is a gene, post-ENCODE? History and updated definition, Genome Res 17 (2007) 669-681.

[6] J.S. Mattick, Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms, Bioessays 25 (2003) 930-939.

[7] D. Ulveling, C. Francastel, F. Hube, When one is better than two: RNA with dual functions, Biochimie 93 (2011) 633-644.

[8] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long noncoding RNAs: insights into function, Nat Rev Genet 10 (2009) 155-159.

[9] L.E. Orgel, Prebiotic chemistry and the origin of the RNA world, Crit Rev Biochem Mol Biol 39 (2004) 99-123.

[10] G.F. Joyce, The antiquity of RNA-based evolution, Nature 418 (2002) 214-221.

[11] W.C. Winkler, R.R. Breaker, Regulation of bacterial gene expression by riboswitches, Annu Rev Microbiol 59 (2005) 487-517.

[12] R.J. Taft, M. Pheasant, J.S. Mattick, The relationship between non-protein-coding DNA and eukaryotic complexity, Bioessays 29 (2007) 288-299.

[13] J.S. Mattick, P.P. Amaral, M.E. Dinger, T.R. Mercer, M.F. Mehler, RNA regulation of epigenetic processes, Bioessays 31 (2009) 51-59.

[14] J.S. Mattick, RNA regulation: a new genetics?, Nat Rev Genet 5 (2004) 316-323.

[15] J.S. Mattick, A new paradigm for developmental biology, J Exp Biol 210 (2007) 1526-1547.

[16] J.S. Mattick, R.J. Taft, G.J. Faulkner, A global view of genomic information--moving beyond the gene and the master regulator, Trends Genet 26 (2010) 21-28.

[17] R.I. Kumaran, R. Thakar, D.L. Spector, Chromatin dynamics and gene positioning, Cell 132 (2008) 929-934.

[18] T. Misteli, Beyond the sequence: cellular organization of genome function, Cell 128 (2007) 787-800.

[19] A.D. Goldberg, C.D. Allis, E. Bernstein, Epigenetics: a landscape takes shape, Cell 128 (2007) 635-638.

[20] M. Bulger, M. Groudine, Functional and mechanistic diversity of distal transcription enhancers, Cell 144 (2011) 327-339.

[21] B. Li, M. Carey, J.L. Workman, The role of chromatin during transcription, Cell 128 (2007) 707-719.

[22] T. Kodadek, D. Sikder, K. Nalley, Keeping transcriptional activators under control, Cell 127 (2006) 261-264.

[23] M.E. Dinger, K.C. Pang, T.R. Mercer, J.S. Mattick, Differentiating protein-coding and noncoding RNA: challenges and ambiguities, PLoS Comput Biol 4 (2008) e1000176.

[24] M.C. Frith, A.R. Forrest, E. Nourbakhsh, K.C. Pang, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, T.L. Bailey, S.M. Grimmond, The abundance of short proteins in the mammalian proteome, PLoS Genet 2 (2006) e52.

[25] S. Chooniedass-Kothari, E. Emberley, M.K. Hamedani, S. Troup, X. Wang, A. Czosnek, F. Hube, M. Mutawe, P.H. Watson, E. Leygue, The steroid receptor RNA activator is the first functional RNA encoding a protein, FEBS Lett 566 (2004) 43-47.

[26] C.I. Brannan, E.C. Dees, R.S. Ingram, S.M. Tilghman, The product of the H19 gene may function as an RNA, Mol Cell Biol 10 (1990) 28-36.

[27] T.L. Young, T. Matsuda, C.L. Cepko, The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina, Curr Biol 15 (2005) 501-512.

[28] N. Brockdorff, A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, S. Rastan, The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus, Cell 71 (1992) 515-526.

[29] G. Dieci, M. Preti, B. Montanini, Eukaryotic snoRNAs: a paradigm for gene expression flexibility, Genomics 94 (2009) 83-88.

[30] H.I. Nakaya, P.P. Amaral, R. Louro, A. Lopes, A.A. Fachel, Y.B. Moreira, T.A. El-Jundi, A.M. da Silva, E.M. Reis, S. Verjovski-Almeida, Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription, Genome Biol 8 (2007) R43.

[31] D. Khaitan, M.E. Dinger, J. Mazar, J. Crawford, M.A. Smith, J.S. Mattick, R.J. Perera, The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion, Cancer Res (2011).

[32] L. Poliseno, L. Salmena, J. Zhang, B. Carver, W.J. Haveman, P.P. Pandolfi, A coding-independent function of gene and pseudogene mRNAs regulates tumour biology, Nature 465 (2010) 1033-1038.

[33] M.M. Candeias, L. Malbert-Colas, D.J. Powell, C. Daskalogianni, M.M. Maslon, N. Naski, K. Bourougaa, F. Calvo, R. Fahraeus, p53 mRNA controls p53 activity by managing Mdm2 functions, Nat Cell Biol (2008).

[34] K. Hashimoto, E. Ishida, S. Matsumoto, N. Shibusawa, S. Okada, T. Monden, T. Satoh, M. Yamada, M. Mori, A liver X receptor (LXR)-beta alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR-beta, Biochem Biophys Res Commun 390 (2009) 1260-1265.

[35] K. Fejes-Toth, V. Sotirova, R. Sachidanandam, G. Assaf, G.J. Hannon, P. Kapranov, S. Foissac, A.T. Willingham, R. Duttagupta, E. Dumais, et al., Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs, Nature 457 (2009) 1028-1032.

[36] T.R. Mercer, M.E. Dinger, C.P. Bracken, G. Kolle, J.M. Szubert, D.J. Korbie, M.E. Askarian-Amiri, B.B. Gardiner, G.J. Goodall, S.M. Grimmond, et al., Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome, Genome Res (2010).

[37] T.R. Mercer, D. Wilhelm, M.E. Dinger, G. Solda, D.J. Korbie, E.A. Glazov, V. Truong, M. Schwenke, C. Simons, K.I. Matthaei, et al., Expression of distinct RNAs from 3' untranslated regions, Nucleic Acids Res 39 (2011) 2393-2403.

[38] F. Rastinejad, H.M. Blau, Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation, Cell 72 (1993) 903-917.

[39] F. Rastinejad, M.J. Conboy, T.A. Rando, H.M. Blau, Tumor suppression by RNA from the 3' untranslated region of alpha-tropomyosin, Cell 75 (1993) 1107-1117.

[40] H. Fan, C. Villegas, A. Huang, J.A. Wright, Suppression of malignancy by the 3' untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs, Cancer Res 56 (1996) 4366-4369.

[41] E.R. Jupe, X.T. Liu, J.L. Kiehlbauch, J.K. McClung, R.T. Dell'Orco, The 3' untranslated region of prohibitin and cellular immortalization, Exp Cell Res 224 (1996) 128-135.

[42] J.D. Amack, A.P. Paguio, M.S. Mahadevan, Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model, Hum Mol Genet 8 (1999) 1975-1984.

[43] H. Choi, N.L. Jackson, D.R. Shaw, P.D. Emanuel, Y.L. Liu, A. Tousson, Z. Meng, S.W. Blume, mrtl-A translation/localization regulatory protein encoded within the human c-myc locus and distributed throughout the endoplasmic and nucleoplasmic reticular network, J Cell Biochem 105 (2008) 1092-1108.

[44] S. Itzkovitz, U. Alon, The genetic code is nearly optimal for allowing additional information within protein-coding sequences, Genome Res 17 (2007) 405-412.

[45] T. Bollenbach, K. Vetsigian, R. Kishony, Evolution and multilevel optimization of the genetic code, Genome Res 17 (2007) 401-404.

[46] S. Itzkovitz, E. Hodis, E. Segal, Overlapping codes within protein-coding sequences, Genome Res 20 (2010) 1582-1589.

[47] M. Kertesz, Y. Wan, E. Mazor, J.L. Rinn, R.C. Nutter, H.Y. Chang, E. Segal, Genome-wide measurement of RNA secondary structure in yeast, Nature 467 (2010) 103-107.

[48] G. Kudla, A.W. Murray, D. Tollervey, J.B. Plotkin, Coding-sequence determinants of gene expression in Escherichia coli, Science 324 (2009) 255-258.

[49] S. Steigele, W. Huber, C. Stocsits, P.F. Stadler, K. Nieselt, Comparative analysis of structured RNAs in S. cerevisiae indicates a multitude of different functions, BMC Biol 5 (2007) 25.

[50] J.V. Chamary, J.L. Parmley, L.D. Hurst, Hearing silence: non-neutral evolution at synonymous sites in mammals, Nat Rev Genet 7 (2006) 98-108.

[51] I. Hellmann, S. Zollner, W. Enard, I. Ebersberger, B. Nickel, S. Paabo, Selection on human genes as revealed by comparisons to chimpanzee cDNA, Genome Res 13 (2003) 831-837.

[52] A.A. Komar, Silent SNPs: impact on gene function and phenotype, Pharmacogenomics 8 (2007) 1075-1080.

[53] C. Kimchi-Sarfaty, J.M. Oh, I.W. Kim, Z.E. Sauna, A.M. Calcagno, S.V. Ambudkar, M.M. Gottesman, A "silent" polymorphism in the MDR1 gene changes substrate specificity, Science 315 (2007) 525-528.

[54] M. Kloc, K. Wilk, D. Vargas, Y. Shirato, S. Bilinski, L.D. Etkin, Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes, Development 132 (2005) 3445-3457.

[55] S.P. Shevtsov, M. Dundr, Nucleation of nuclear bodies by RNA, Nat Cell Biol 13 (2011) 167-173.

[56] N. Adachi, M.R. Lieber, Bidirectional gene organization: a common architectural feature of the human genome, Cell 109 (2002) 807-809.

[57] K.V. Morris, S. Santoso, A.M. Turner, C. Pastori, P.G. Hawkins, Bidirectional transcription directs both transcriptional gene activation and suppression in human cells, PLoS Genet 4 (2008) e1000258.

[58] P.G. Engstrom, H. Suzuki, N. Ninomiya, A. Akalin, L. Sessa, G. Lavorgna, A. Brozzi, L. Luzi, S.L. Tan, L. Yang, et al., Complex loci in human and mouse genomes, PLoS Genet 2 (2006) e47.

[59] N.D. Trinklein, S.F. Aldred, S.J. Hartman, D.I. Schroeder, R.P. Otillar, R.M. Myers, An abundance of bidirectional promoters in the human genome, Genome Res 14 (2004) 62-66.

[60] M.E. Askarian-Amiri, J. Crawford, J.D. French, C.E. Smart, M.A. Smith, M.B. Clark, K. Ru, T.R. Mercer, E.R. Thompson, S.R. Lakhani, et al., SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer, Rna 17 (2011) 878-891.

[61] S. Roy, J. Ernst, P.V. Kharchenko, P. Kheradpour, N. Negre, M.L. Eaton, J.M. Landolin, C.A. Bristow, L. Ma, M.F. Lin, et al., Identification of functional elements and regulatory circuits by Drosophila modENCODE, Science 330 (2010) 1787-1797.

[62] M.B. Gerstein, Z.J. Lu, E.L. Van Nostrand, C. Cheng, B.I. Arshinoff, T. Liu, K.Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, et al., Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project, Science 330 (2010) 1775-1787.

[63] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, et al., Identification and analysis of

functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799-816.

[64] P.P. Amaral, M.B. Clark, D.K. Gascoigne, M.E. Dinger, J.S. Mattick, lncRNAdb: a reference database for long noncoding RNAs, Nucleic Acids Res 39 (2011) D146-151.

[65] M.E. Dinger, P.P. Amaral, T.R. Mercer, K.C. Pang, S.J. Bruce, B.B. Gardiner, M.E. Askarian-Amiri, K. Ru, G. Solda, C. Simons, et al., Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation, Genome Res 18 (2008) 1433-1445.

[66] T.R. Mercer, M.E. Dinger, S.M. Sunkin, M.F. Mehler, J.S. Mattick, Specific expression of long noncoding RNAs in the mouse brain, Proc Natl Acad Sci U S A 105 (2008) 716-721.

[67] J. Feng, C. Bi, B.S. Clark, R. Mady, P. Shah, J.D. Kohtz, The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator, Genes Dev 20 (2006) 1470-1484.

[68] A.M. Bond, M.J. Vangompel, E.A. Sametsky, M.F. Clark, J.C. Savage, J.F. Disterhoft, J.D. Kohtz, Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry, Nat Neurosci 12 (2009) 1020-1027.

Figures

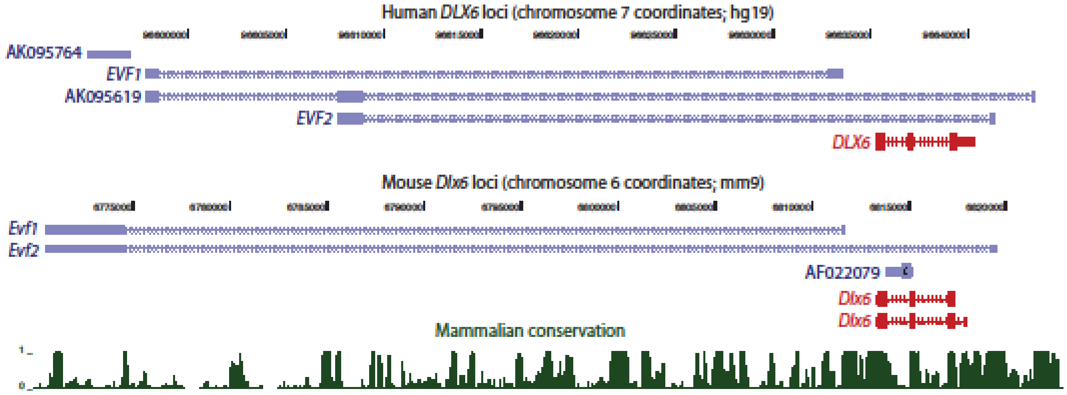
Human *DLX6* loci (chromosome 7 coordinates; hg19)
AK095764
EVF1
AK095619
EVF2
DLX6
Mouse *Dlx6* loci (chromosome 6 coordinates; mm9)
Evf1
Evf2
AF022079
Dlx6
Dlx6
Mammalian conservation
1
0

Figure 1. Are complex transcriptional loci hotbeds for evolutionary innovation? The genome representations compare long noncoding transcripts associated with the *DLX6* loci in human and mouse. The antisense *DLX6* transcript *EVF2* has been shown to recruit DLX transcription factors to important DNA regulatory elements in mouse [67] and human [68]. Despite the functional similarities, there are significant differences in the *EVF* variants between mouse and human, including a putative novel protein-coding antisense transcript within the intron of mouse *Dlx6*. Although the biological roles of the various *DLX6* antisense transcripts are known, it appears that the innovation of novel transcripts acting as either novel proteins or regulatory RNAs is much greater than the *DLX6* protein, which is relatively unchanged.

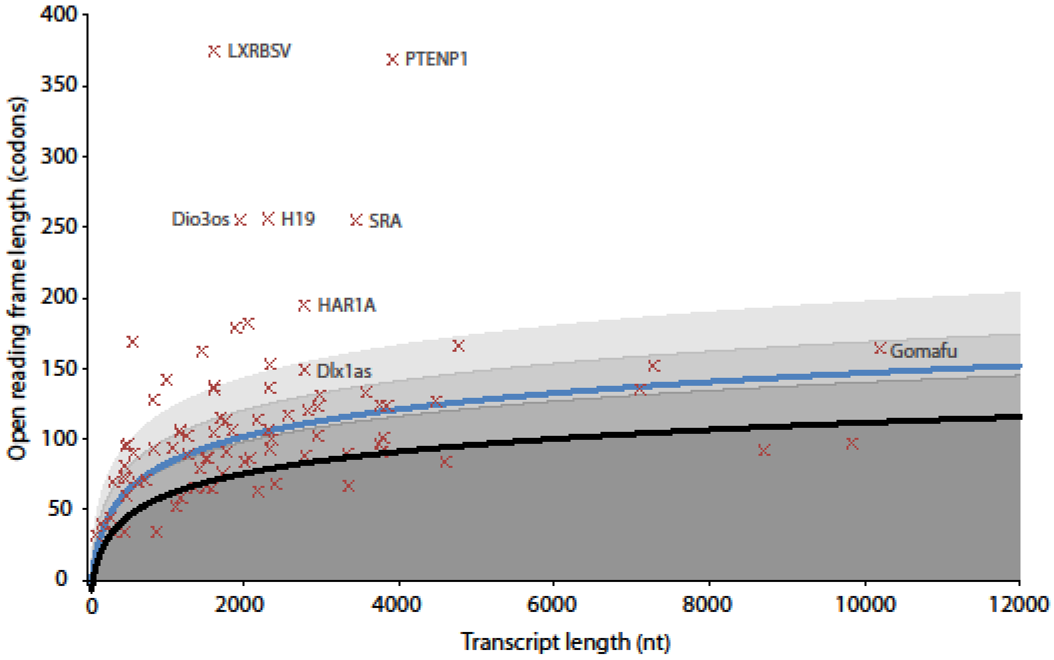

Figure 2. Do many long noncoding RNAs also encode proteins? The graph illustrates the incidence of open reading frames (ORFs) as a function of transcript length in characterized mouse and human long noncoding RNAs (lncRNAs) from lncRNAdb (http://lncrnadb.org) [64]. A number of well-described lncRNAs are labeled. The black line indicates the mean ORF length in a randomly generated set of 20,000 transcripts, with the shaded regions showing 1, 2, and 3 standard deviations above the mean [23]. The blue line represents a fitted curve from the maximum theoretical ORF and transcript lengths from a set of ~9,000 human lncRNAs annotated using a previously described approach [65, 66]. These lncRNAs have no overlap with any coding region annotated in either UCSC Genes or RefSeq.