# Long non-coding RNAs: insights into functions

*Tim R. Mercer, Marcel E. Dinger and John S. Mattick*

Abstract | In mammals and other eukaryotes most of the genome is transcribed in a developmentally regulated manner to produce large numbers of long non-coding RNAs (ncRNAs). Here we review the rapidly advancing field of long ncRNAs, describing their conservation, their organization in the genome and their roles in gene regulation. We also consider the medical implications, and the emerging recognition that any transcript, regardless of coding potential, can have an intrinsic function as an RNA.

The RNA world hypothesis proposes that early life was based on RNA, which subsequently devolved the storage of information to more stable DNA, and catalytic functions to more versatile proteins. Consequently, despite crucial roles in the ancient processes of translation and splicing, RNA is assumed to have been largely relegated to an intermediate between gene and protein, encapsulated in the central dogma 'DNA makes RNA makes protein'. However, the finding that most of the genome in complex organisms is transcribed, apparently in a developmentally regulated fashion[1–6], and the discovery of new classes of regulatory non-coding RNAs (ncRNAs), challenges this assumption and suggests that RNA has continued to evolve and expand alongside proteins and DNA.

Although the current literature is dominated by short RNAs, there are an increasing number of reports describing long transcripts that, rather than encoding protein, act functionally as RNAs. Although we currently lack satisfactory classifications for these transcripts, long ncRNAs are arbitrarily considered to be longer than ~200 nucleotides, on the basis of a convenient practical cut-off in RNA purification protocols that excludes small RNAs[4].

## Identification of long ncRNAs

As a transcriptional class, long ncRNAs were first described during the large-scale sequencing of full-length cDNA libraries in the mouse[6]. Although distinguishing long ncRNAs from other protein-coding mRNAs is not a trivial process (BOX 1), it has nevertheless become apparent that a significant portion of the transcriptome has little or no protein-coding capacity. The increased sensitivity of genome tiling arrays provides an even more detailed view, revealing that the extent of non-coding sequence transcription is at least four times greater than coding sequence, and that an abundance of non-polyadenylated non-coding transcripts had been previously overlooked[4].

These studies also showed that the transcriptome is surprisingly complex, with long ncRNAs often overlapping with, or interspersed between, multiple coding and non-coding transcripts[1,5] (FIG. 1). This complexity has prompted a shift in our understanding of gene organization from a linear to a modular model, in which it is possible for a sequence to be transcribed into a range of sense and antisense, coding and non-coding transcripts. Attempts to untangle this complexity have led to crude classifications of ncRNAs based on their genomic proximity to protein-coding genes, including overlapping, *cis*-antisense, bidirectional or intronic ncRNAs. In reality many transcripts resist classification into any particular category, and instead exhibit a combination of these qualities. Other unusual species of long ncRNAs, such as *trans*-spliced transcripts, macroRNAs that encompass huge genomic distances and multigene transcripts that encompass several genes or even the whole chromosome, further confound efforts for systematic classification[2,3].

## Widespread functionality of long ncRNAs

Given their unexpected abundance, long ncRNAs were initially thought to be spurious transcriptional noise resulting from low RNA polymerase fidelity[7]. However, the expression of many long ncRNAs is restricted to particular developmental contexts[8], and large numbers of mouse ncRNAs are specifically expressed during embryonic stem cell differentiation[9] and in the brain, often exhibiting precise subcellular localization[10]. The binding of transcription factors to non-coding loci, together with evidence of purifying selection acting on ncRNA promoters, suggests that this type of expression is explicitly regulated[1,11,12].

Nevertheless, despite such signatures of functionality, the generally low sequence conservation of long ncRNAs has fuelled the assertion that they are not functional. However, this conclusion needs to be carefully considered. First, it ignores many examples that are conserved, and a recent study ascribes functional roles to a high proportion of such ncRNAs[13]. Second, long ncRNAs are likely to exhibit different patterns of conservation to protein-coding genes, which are subject to strict functional constraints and must preserve an ORF. By contrast, long ncRNAs can exhibit shorter stretches of sequence that are conserved to maintain functional domains and structures. Indeed, many long ncRNAs with a known function, such as *Xist*, only exhibit high conservation over short sections of their length[14]. Third, rather than being indicative of non-functionality, low sequence conservation can also be explained by high rates of primary sequence evolution if long ncRNAs have, like promoters and other regulatory elements, more plastic structure–function constraints than proteins[14]. Many conserved regions of the human genome that have been subject to recent and rapid evolutionary change are transcribed into long ncRNAs, including *HAR1*, a ncRNA expressed in Cajal–Retzius neurons in the developing neocortex[15]. Moreover, the adaptive radiation

of non-coding (that is, regulatory) sequences is likely to specify most of the phenotypic differences between, and within, species[16].

Low sequence conservation of long ncRNAs also prompted the alternative suggestion that it is the process rather than the product of transcription that is functional. For example, the cascading transcription of ncRNAs across the fructose bisphosphate *fbp1*$^{+}$ promoter in yeast is associated with the progressive opening of chromatin, thereby increasing access to transcriptional activators and RNA polymerase[17]. However, genome-wide evidence of conserved secondary structure[18], splicing patterns[12] and subcellular localization[10] suggest that a significant portion of ncRNAs fulfil functional roles beyond transcriptional remodelling. The diverse selection pressures acting on long ncRNAs probably reflect the wide range of their functions and their relative importance.

## Functions of long ncRNAs

Unlike microRNAs or proteins, ncRNA function cannot currently be inferred from sequence or structure, with the diversity of long ncRNAs described to date precluding simple generalizations. The broad functional repertoire of long ncRNAs includes roles in high-order chromosomal dynamics, telomere biology and subcellular structural organization[8]. One major emergent theme is the involvement of these ncRNAs in regulating the expression of neighbouring protein-coding genes. The importance of this localized regulation was foreshadowed by the phenomenon of 'transvection', in which non-coding loci affect the expression of nearby protein-coding genes in *trans*[19]. Additionally, the recent observation that human chromosome 21 largely recapitulates its native expression profile in mouse cells, despite interspecies differences in epigenetic machinery, cellular environment and transcription factors, suggests that most of the information required for gene regulation is embedded in the chromosome sequence[20]. In the following sections we focus on the ability of long ncRNAs to regulate gene expression at the level of chromatin modification, transcription and post-transcriptional processing.

***Chromatin modification.*** Long ncRNAs can mediate epigenetic changes by recruiting chromatin remodelling complexes to specific genomic loci. For example, hundreds of long ncRNAs are sequentially expressed along the temporal and spatial developmental axes of the human homeobox (Hox) loci, where they define chromatin domains of differential histone methylation and RNA polymerase accessibility[21]. One of these ncRNAs, Hox transcript antisense RNA (*HOTAIR*), originates from the *HOXC* locus and silences transcription across 40 kb of the *HOXD* locus in *trans* by inducing a repressive chromatin state, which is proposed to occur by recruitment of the Polycomb chromatin remodelling complex PRC2 by *HOTAIR*[21] (FIG. 2a). This model fits other chromatin modifying complexes, such as MLL, PcG, and G9a methyltransferase, which can be similarly directed by their associated ncRNAs[9,22–24]. Such a mechanism might resolve the paradox of how a small repertoire of chromatin remodelling complexes, which often have RNA binding domains but little DNA sequence specificity, are able to specify the complex array of chromatin modifications that are apparent throughout development.

Although the model of recruitment of chromatin modifying complexes by ncRNAs has been informative in our understanding of epigenetic phenomenon such as imprinting, it is probably only part of the story. X chromosome inactivation is mediated by the iconic long ncRNA, *Xist*. A small internal non-coding transcript from the *Xist* locus, *RepA*, recruits PRC2 to silence one X chromosome[25], whereas PRC2 is titrated from the remaining active X chromosome by the antisense transcript *Tsix*. However, another study describes an alternative mechanism whereby *Xist* and *Tsix* anneal to form an RNA duplex that is processed by Dicer to generate small interfering RNAs (siRNAs), which are required for the repressive chromatin modifications on the inactive X chromosome[26]. The contribution of these two different pathways to coordinate long and small RNAs in chromatin remodelling infers the existence of a global, integrated regulatory network based in RNA.

***Transcriptional regulation.*** The pervasive transcription of enhancers[27] and promoters[28] anticipates a core role for long ncRNA in regulating the process of transcription. The means by which such ncRNAs regulate transcription are expanding to encompass a diversity of mechanisms, as shown by the following examples.

Proximal promoters can be transcribed into long ncRNAs that recruit and integrate the functions of RNA binding proteins into the transcriptional programme, as exemplified by the repression of cyclin D1 transcription in human cell lines[29]. DNA damage signals induce the expression of long ncRNAs associated with the cyclin D1 gene promoter, where they act cooperatively to modulate the activities of the RNA binding protein TLS. TLS subsequently inhibits the histone acetyltransferase activities of CREB binding protein and p300 to silence cyclin D1 expression (FIG. 2b). The ability of

### Box 1 | Parsing coding and non-coding transcripts

Coding and non-coding RNAs (ncRNAs) can be difficult to distinguish. In eukaryotes, a protein-coding transcript is commonly defined by the presence of an ORF greater than 100 amino acids. However, a long ncRNA might contain such an ORF by chance alone, and many well-characterized long ncRNAs do indeed contain long ORFs. Reciprocally, proteins smaller than 100 amino acids might also be translated, with functional peptides as small as 11 amino acids being reported in *Drosophila* species[42]. The observation that selection favours synonymous over non-synonymous mutations to preserve codon usage has been exploited to help distinguish between transcripts with true rather than spurious ORFs[43]. Nevertheless, despite such improvements in the annotation of transcripts in recent years, we still lack a satisfactory definition of ncRNAs and there remain many ambiguous transcripts that exhibit both coding and non-coding traits. This might ultimately reflect the likelihood that the genome has evolved to encode a continuous spectrum of transcripts and information with little regard for our arbitrary definitions of coding and non-coding transcripts[44].

The extensive overlapping of alternatively spliced coding and non-coding isoforms further confounds the problem of distinguishing coding and non-coding transcripts, and indeed there might be a false dichotomy between them. One of the first and best characterized long ncRNAs, *SRA*, was later found to also encode a protein that acts antagonistically to the function of the ncRNA[45]. Reciprocally, many mRNAs can also function at an RNA level. For example, the *p53* mRNA acts intrinsically as an RNA regulator by binding the Mdm2 protein, which in turn induces *p53* expression and function[46]. A synonymous 'silent' mutation interferes with this process, and other similar silent mutations that affect protein translation are prevalent throughout the genome[47]. Moreover, 3′ UTRs of mRNAs can be expressed separately from the associated mRNA[48] and can impart functional information independently of the encoded protein[49,50]. Together these studies indicate that transcripts can potentially function both at an RNA level and to encode protein.

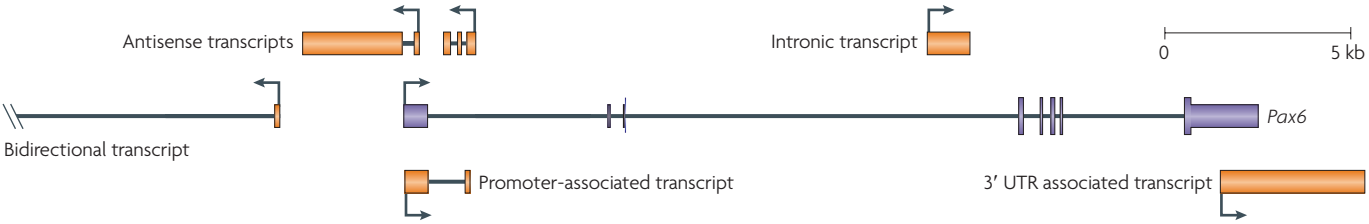


Figure 1 | **Genomic organization of coding and non-coding transcripts.** Schematic diagram illustrating the complexity of the interleaved networks of long non-coding transcripts (orange) that are associated with paired box gene 6 (*Pax6*; purple).

ncRNAs to recruit RNA binding proteins, one of the largest protein classes in the mammalian proteome, to gene promoters hugely expands the regulatory repertoire available to the transcriptional programme[29].

Long ncRNAs also act as co-factors to modulate transcription factor activity. For example, in mice, the ncRNA *Evf2* is transcribed from an ultraconserved distal enhancer and recruits the binding and action of the transcription factor DLX2 to this same enhancer to induce expression of adjacent protein-coding genes[30] (FIG. 2c). Many similar enhancers are transcribed in cells in which they are active — this could be a general strategy for regulating the expression of key developmental genes[27].

Long ncRNAs can regulate RNA polymerase (RNAP) II activity through other mechanisms, including by interaction with the initiation complex to influence promoter choice. For example, in humans, a ncRNA transcribed from an upstream region of the dihydrofolate reductase (*DHFR*) locus forms a triplex in the major promoter of *DHFR* to prevent the binding of the transcriptional co-factor TFIID[31] (FIG. 2d). This could be a widespread mechanism for controlling promoter usage as thousands of triplex structures exist in eukaryotic chromosomes[32].

Long ncRNAs can also effect global changes by interacting with basal components of the RNAP II-dependent transcription machinery. ncRNAs that interact with RNAP II machinery are typically transcribed by RNAP III, thereby decoupling their expression from the RNAP II-dependent transcription reaction they regulate. For example, *Alu* elements that are transcribed in response to heat shock bind tightly to RNAP II to preclude the formation of active preinitiation complexes[33]. *Alu* elements contain modular domains that can independently mediate polymerase binding and repression. In light of their abundance and distribution in the mammalian genome, these functional domains might have been co-opted into other ncRNAs during evolution; an observation supported by the finding that functional repeat sequence domains are a common characteristic of several known long ncRNAs[8].

***Post-transcriptional regulation.*** The ability of ncRNAs to recognize complementary sequences also allows highly specific interactions that are amenable to regulating various steps in the post-transcriptional processing of mRNAs, including their splicing, editing, transport, translation and degradation. Most mammalian genes express antisense transcripts, which might constitute a class of ncRNA that is particularly adept at regulating mRNA dynamics[34].

Antisense ncRNAs can mask key *cis*-elements in mRNA by the formation of RNA duplexes, as in the case of the *Zeb2* (also called *Sip1*) antisense RNA, which complements the 5′ splice site of an intron in the 5′ UTR of the zinc finger Hox mRNA *Zeb2* (REF. 35). Expression of the ncRNA prevents the splicing of an intron that contains an internal ribosome entry site required for efficient translation and expression of the ZEB2 protein (FIG. 2e). This sets a precedent for ncRNAs in directing the alternative splicing of mRNA isoforms. Indeed, a number of studies have noted the prevalence of ncRNAs antisense to introns, and they could similarly regulate splicing[34].

Alternatively, the annealing of ncRNA can target protein effector complexes to the sense mRNA transcript in a manner analogous to the targeting of the RNA-induced silencing complex (RISC) to mRNAs by siRNAs. RNA duplexes resulting from the annealing of complementary transcripts or even of long ncRNAs with extended internal hairpins can be processed into endogenous siRNAs to silence gene expression, raising the possibility that many long ncRNAs feed into RNA silencing pathways[26].

There are probably many other functions of long ncRNAs awaiting discovery. For example, the ncRNA *NRON* has been shown to regulate the nuclear trafficking of the transcription factor NFAT[36], and the observation that many long ncRNAs are located in the cytoplasm[4] suggests that they might have undiscovered roles in cell biology.

## Medical significance

There is increasing interest in the potential involvement of ncRNAs in disease aetiology, owing to aberrant function of ncRNAs in differentiation and developmental processes. The ability of ncRNAs to regulate associated protein-coding genes might contribute

### Glossary

**Adaptive radiation**
Evolution of new morphological or functional characteristics in lineages that diversify in response to environmental changes or to enable colonization of new ecological niches.

**Epigenetic**
Heritable changes in phenotype caused by mechanisms outside of the genomic sequence. Such changes might remain through cell divisions during, for example, cellular differentiation, or they might persist through subsequent generations. Epigenetic changes include chromatin modifications, such as histone acetylation, or chemical alterations to the DNA itself, such as DNA methylation.

**Long ncRNA**
Transcripts longer than 200 nucleotides that have little or no protein-coding capacity. Long ncRNAs can regulate gene expression through a diversity of mechanisms.

**MicroRNA**
Single-stranded RNAs of approximately 21–23 nucleotides that regulate gene expression by partial complementary base pairing to specific mRNAs. This annealing inhibits protein translation and can also facilitate degradation of the target mRNA.

**Transvection**
Apparent cross-talk between alleles on homologous chromosomes, in which complementation is observed between promoter mutations in one allele and structural mutations in the other. Transvection can cause either gene activation or repression.

**X chromosome inactivation**
A process in which one of the two copies of the X chromosomes in female mammals is inactivated. X inactivation occurs so that females produce the same dosage of gene products from the X chromosome as males.

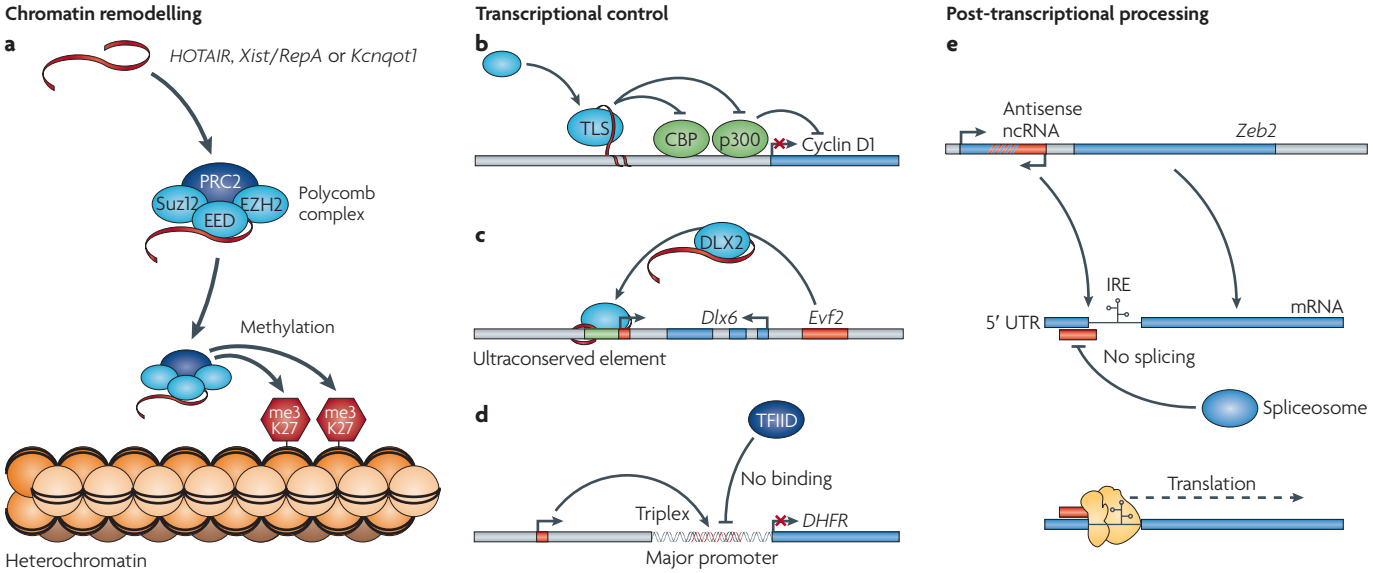


Figure 2 | **Functions of long non-coding RNAs (ncRNAs).** Illustrative mechanisms by which long ncRNAs regulate local protein-coding gene expression at the level of chromatin remodelling, transcriptional control and post-transcriptional processing. **a** | ncRNAs can recruit chromatin modifying complexes to specific genomic loci to impart their catalytic activity. In this case, the ncRNAs *HOTAIR*[21], *Xist* and *RepA* (the small internal non-coding transcript from the *Xist* locus)[25], or *Kcnqot1* (REF. 24) recruit the Polycomb complex to the *HoxD* locus, the X chromosome, or the *Kcnq1* domain, respectively, where they trimethylate lysine 27 residues (me3K27) of histone H3 to induce heterochromatin formation and repress gene expression. **b** | ncRNAs can regulate the transcriptional process through a range of mechanisms. ncRNAs tethered to the cyclin D1 gene recruit the RNA binding protein TLS to modulate the histone acetyltransferase activity of CREB binding protein (CBP) and p300 to repress gene transcription[29]. **c** | An ultraconserved enhancer is transcribed as a long ncRNA, *Evf2*, which subsequently acts as a co-activator to the transcription factor DLX2, to regulate the *Dlx6* gene transcription[30]. **d** | A ncRNA transcribed from the *DHFR* minor promoter in humans can form a triplex at the major promoter to occlude the binding of the general transcription factor TFIID, and thereby silence *DHFR* gene expression[31]. **e** | An antisense ncRNA can mask the 5′ splice site of the zinc finger homeobox mRNA *Zeb2* from the spliceosome, resulting in intron retention. The translation machinery can then recognize and bind an internal ribosome entry site (IRE) in the retained intron, resulting in efficient *Zeb2* translation and expression[35].

towards disease if their misexpression deregulates a gene of clinical significance. For example, an antisense ncRNA transcribed from the *p15* tumour suppressor locus induces changes to local heterochromatin and DNA methylation status, thereby regulating *p15* expression[37], and is potentially involved in oncogenesis as the antisense ncRNA and protein have inverse expression profiles in leukaemia. Many other tumour suppressor genes that are frequently silenced by epigenetic mechanisms in cancer also have antisense partners[37].

An appreciation of long ncRNAs might inform a reinterpretation of the functional basis of many disease-associated polymorphisms and chromosomal alterations that occur in non-coding regions[38]. For example, the translocation and induced expression of an antisense long non-coding transcript causes the epigenetic silencing of an adjacent α-globin gene, resulting in α-thalassaemia[39]. The role of ncRNAs in disease is likely to have been overlooked in genetic screens because of the subtlety of their effects and the emphasis to date, both intellectually and practically, on mutation scanning of protein-coding exons. Moreover, the complexity of non-coding transcription can make understanding the specific contribution of embedded polymorphisms a bewildering exercise. For example, a SNP that occurs both in the 3′ UTR of the zinc finger gene *ZFAT* and also in the promoter of an antisense transcript increases the expression of *ZFAT* — not through increasing mRNA stability, but by repressing the expression of the antisense transcript[40].

## Conclusion

Continuing advances in transcriptomics are resulting in ncRNA being recognized as an important functional expression of the genome. Rather than being reduced to a simple messenger role, it seems likely that the sophisticated structural and informational capacity of RNA has continued to be harnessed by evolution in a range of biological roles that interact with, but are distinct from, protein and DNA. The concomitant increase in non-coding content with organismal complexity supports the proposition that evolutionary innovations and expansion of regulatory RNAs were fundamental to the genetic programming of complex eukaryotes[41].

There are still huge gaps in our understanding of long ncRNAs, including the proportion that is functional and the range and mechanistic basis of their functions. What is required, and is currently developing, is the application of genome-wide techniques to reveal the full extent of ncRNA expression. Unbiased techniques, such as next-generation sequencing, have the benefit of not being constrained by current protein-centric annotations. Such data will progressively build a catalogue of ncRNAs with common characteristics that will aid in the identification and prediction of functional features, complemented by experimental analyses of individual examples to determine the mechanisms by which long ncRNAs act. This will increasingly involve the intersection of techniques from other fields, such as live cell RNA imaging, RNA proteomics (that is, the analysis of RNA-associated protein complexes) and RNA structural biology.

Long ncRNAs have the potential to rival the functional repertoire of the proteome, albeit with a different spectrum. It is already apparent that any RNA, regardless of protein-coding capacity, might have its own intrinsic messages to deliver (BOX 1),

suggesting that the genome might encompass an RNA-based information suite that is far more sophisticated than expected. Should the majority of ncRNAs prove to be functional, their characterization will have a considerable impact on our understanding of the genetic programming of complex organisms and will bring new answers to old questions in evolution, development and the understanding of disease.

*Tim R. Mercer, Marcel E. Dinger and John S. Mattick are at the Australian Research Council Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia.*

*T.R.M. and M.E.D. contributed equally to this work*

*Correspondence to J.S.M.*
*e-mail: j.mattick@imb.uq.edu.au*

doi:10.1038/nrg2521
Published online 3 February 2009

1. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
2. Denoeud, F. *et al.* Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.* **17**, 746–759 (2007).
3. Horiuchi, T. & Aigaki, T. Alternative *trans*-splicing: a novel mode of pre-mRNA processing. *Biol. Cell* **98**, 135–140 (2006).
4. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
5. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15**, 987–997 (2005).
6. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
7. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Struct. Mol. Biol.* **14**, 103–105 (2007).
8. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19**, 454–492 (2008).
9. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).
10. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **105**, 716–721 (2008).
11. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
12. Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).
13. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* (in the press).
14. Pang, K. C., Frith, M. C. & Mattick, J. S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* **22**, 1–5 (2006).
15. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
16. Carroll, S. B. Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
17. Hirota, K. *et al.* Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130–134 (2008).
18. Torarinsson, E., Sawera, M., Havgaard, J. H., Fredholm, M. & Gorodkin, J. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res.* **16**, 885–889 (2006).
19. Mattick, J. S. & Gagen, M. J. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **18**, 1611–1630 (2001).
20. Wilson, M. D. *et al.* Species-specific transcription in mice carrying human chromosome 21. *Science* **322**, 434–438 (2008).
21. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
22. Morris, K. V., Santoso, S., Turner, A. M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.* **4**, e1000258 (2008).
23. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
24. Pandey, R. R. *et al. Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
25. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
26. Ogawa, Y., Sun, B. K. & Lee, J. T. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336–1341 (2008).
27. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* **11**, 2494–2509 (1997).
28. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
29. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454**, 126–130 (2008).
30. Feng, J. *et al.* 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**, 1470–1484 (2006).
31. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
32. Ohno, M., Fukagawa, T., Lee, J. S. & Ikemura, T. Triplex-forming DNAs in the human interphase nucleus visualized *in situ* by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma* **111**, 201–213 (2002).
33. Mariner, P. D. *et al.* Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* **29**, 499–509 (2008).
34. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. The antisense transcriptomes of human cells. *Science* **322**, 1855–1857 (2008).
35. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769 (2008).
36. Willingham, A. T. *et al.* A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**, 1570–1573 (2005).
37. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–206 (2008).
38. Hindorff, L. A., Junkins, H. A. & Manolio, T. A. A catalog of published genome-wide association studies. *National Human Genome Research Institute* [online], <http://www.genome.gov/gwastudies> (2008).
39. Tufarelli, C. *et al.* Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature Genet.* **34**, 157–165 (2003).
40. Shirasawa, S. *et al.* SNPs in the promoter of a B cell-specific antisense transcript, *SAS-ZFAT*, determine susceptibility to autoimmune thyroid disease. *Hum. Mol. Genet.* **13**, 2221–2231 (2004).
41. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
42. Kondo, T. *et al.* Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biol.* **9**, 660–665 (2007).
43. Lin, R., Maeda, S., Liu, C., Karin, M. & Edgington, T. S. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* **26**, 851–858 (2007).
44. Dinger, M. E., Pang, K. C., Mercer, T. R. & Mattick, J. S. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* **4**, e1000176 (2008).
45. Chooniedass-Kothari, S. *et al.* The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.* **566**, 43–47 (2004).
46. Candeias, M. M. *et al. p53* mRNA controls p53 activity by managing Mdm2 functions. *Nature Cell Biol.* **10**, 1098–1105 (2008).
47. Komar, A. A. Silent SNPs: impact on gene function and phenotype. *Pharmacogenomics* **8**, 1075–1080 (2007).
48. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38**, 626–635 (2006).
49. Jenny, A. *et al.* A translation-independent role of *oskar* RNA in early *Drosophila* oogenesis. *Development* **133**, 2827–2833 (2006).
50. Rastinejad, F. & Blau, H. M. Genetic complementation reveals a novel regulatory role for 3′ untranslated regions in growth and differentiation. *Cell* **72**, 903–917 (1993).

**Acknowledgements**
We thank P. Amaral and other laboratory colleagues for many discussions related to this article, and the Australian Research Council for financial support. We apologize both to readers and colleagues for references that were omitted owing to editorial constraint.

**FURTHER INFORMATION**
**Mattick laboratory web site:**
http://jsm-research.imb.uq.edu.au
**CPC (coding potential calculator):** http://cpc.cbi.pku.edu.cn
**NRED (ncRNA expression database):**
http://jsm-research.imb.uq.edu.au/NRED
**RNAdb (RNA database):**
http://research.imb.uq.edu.au/RNAdb

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**