The dark matter rises: the expanding world of regulatory RNAs

Michael B. Clark*†[1], Anupma Choudhary*, Martin A. Smith*†, Ryan J. Taft* and John S. Mattick†[1]

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

†*Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2010, Australia*

Abstract

The ability to sequence genomes and characterize their products has begun to reveal the central role for regulatory RNAs in biology, especially in complex organisms. It is now evident that the human genome contains not only protein-coding genes, but also tens of thousands of non–protein coding genes that express small and long ncRNAs (non-coding RNAs). Rapid progress in characterizing these ncRNAs has identified a diverse range of subclasses, which vary widely in size, sequence and mechanism-of-action, but share a common functional theme of regulating gene expression. ncRNAs play a crucial role in many cellular pathways, including the differentiation and development of cells and organs and, when misregulated, in a number of diseases. Increasing evidence suggests that these RNAs are a major area of evolutionary innovation and play an important role in determining phenotypic diversity in animals.

[1]*Correspondence may be addressed to either of these authors (email m.clark3@uq.edu.au or j.mattick@garvan.org.au).*

Introduction

Recent advances in molecular biology, led by the large-scale sequencing of genomes and the characterization of transcriptomes, have revealed that animal genomes are far more complex and intricate than previously anticipated, containing a great diversity of sequences that can be effectors of genetic information, i.e. genes. Central to this has been the discovery that many genes do not encode proteins, but rather produce ncRNAs (non-coding RNAs), sometimes referred to as genomic 'dark matter'. Unlike in prokaryotes and most unicellular eukaryotes, such as yeast, only a small fraction of the mammalian genome encodes protein, yet the vast majority of the mammalian genome is transcribed, producing tens of thousands of small and long ncRNAs [1,2].

The properties of RNA molecules, including their ability to form higher-order structures, to specifically hybridize with other RNAs or DNA, and to assemble RNA–protein complexes, makes them effective and versatile regulatory molecules that can direct relatively generic effector proteins to sequence-specific targets [3]. The functional versatility of RNA has led to the 'RNA world hypothesis', which postulates that the dual catalytic and informational storage properties of RNA provided the molecular platform for the evolution of early life [4]. While proteins and DNA now fulfil most catalytic and information storage roles respectively, RNA continues to have a variety of functional roles, including as a regulator, in all kingdoms of life [4]. The emergence of multicellular life, however, has required increasingly complex regulatory circuits to orchestrate the development and organization of specialized tissues and organs. Hence, although prokaryotes and unicellular eukaryotes do contain regulatory ncRNAs, the role of ncRNAs as regulators of gene expression has seemingly expanded in the genomes of multicellular organisms [5].

The explosion of ncRNA research has revealed that there is an abundance of small and long ncRNAs involved in regulating almost all steps of gene expression, including, but not limited to, chromatin modification, transcriptional control, mRNA degradation, translational efficiency and splicing [6,7]. Hence, ncRNAs function in a wide range of cellular processes, play crucial roles in development and disease, and may even play a central role in the evolution of different species and complex organisms [8–10].

The animal genome and the non-coding universe

For half a century, protein-centric convictions predicated on the central dogma of molecular biology dominated the discipline. However, the completion of multiple genome sequencing projects has revealed that protein-coding sequences encompass only a small fraction of animal genomes and less than 2% of the genome in humans and other mammals [11,12]. Thus either animal genomes are largely composed of 'junk' DNA, or they contain another form of genetic information that has thus far been overlooked. Several lines of evidence support the latter, including the positive correlation between the proportion of the genome that is 'junk'/non-protein-coding and developmental complexity [5], the presence of extensive conserved non-coding sequences outside protein-coding regions [13], and the pervasive differential transcription of the vast majority of the genome [1,14].

Characterization of the mammalian transcriptome has revealed that RNA transcripts are produced from a considerably greater proportion of the genome than the ~40% covered (including introns and exons) by known genes [1,15,16]. This pervasive transcription of the genome, defined as occurring when "the majority of (the genome's) bases are associated with at least one primary transcript" [17] has been identified by a number of independent techniques, including genome-wide tiling arrays, large-scale cloning and sequencing of cDNAs, and next-generation RNA sequencing. For example, results from the ENCODE project, which aimed to identify all functional elements within the human genome, demonstrated that at least 75% of bases were transcribed [1]. A number of studies have shown that although the vast majority of the genome is transcribed at some level, most transcription, including novel unannotated transcription, clusters around known genes [16,18–20]. Such analyses led Kapranov et al. [19] to propose "a model of genome organization where protein–coding genes are at the center of a complex network of overlapping sense and antisense (long) RNA transcription, with interleaved (small) RNAs" (Figure 1).

The net result of pervasive transcription is a complex and interleaved transcriptome producing approximately 20000 coding genes, along with at least as many, and possibly a much greater number of, transcripts that do not encode proteins and could function at the RNA level [2]. These can be separated into two major groups (the small and long ncRNAs) on the basis of size and mechanism of synthesis.

Small ncRNAs

Small RNAs are generally defined as ncRNAs shorter than 200 nt in length, and are usually produced by the post–transcriptional processing of longer transcripts by endogenous
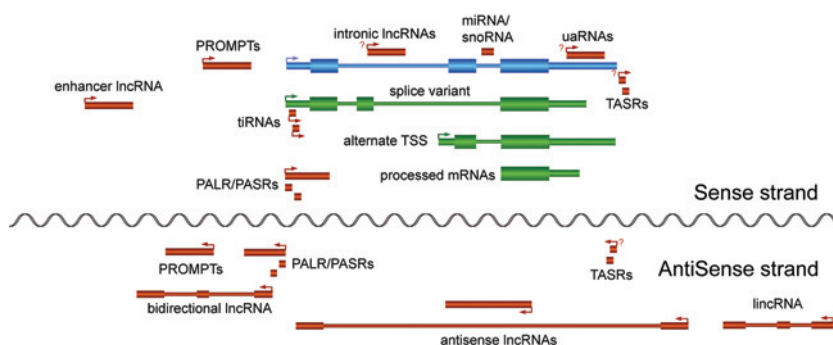

Figure 1. Pervasive transcription around a hypothetical protein coding gene
A standard mRNA transcript is shown in blue, other coding or potentially coding transcripts are shown in green, and non-coding transcripts are in red. Introns are represented by thin lines, non-coding exons are indicated by medium thickness lines and coding exons are indicated by the thickest lines. Arrows represent transcription start sites. The absence of an arrow indicates that the transcripts are generated by processing. An arrow plus a question mark refers to transcripts where the origin is often unclear, and could involve transcription initiation or processing. PALR, promoter-associated long RNA [19]; PASR, promoter-associated small RNA [19]; PROMPT, promoter upstream transcript [51]; snoRNA, small nucleolar RNA; TASR, termini-associated small RNA [19]; tiRNA, transcription initiation (tiny) RNA [49]; TSS, transcription start site; uaRNA, 3′-UTR-derived RNAs [118].

RNases (i.e. RNA cleavage enzymes). Based primarily on their size, mode of biogenesis and function, small RNAs can be divided into various subclasses. The three subclasses of small RNAs that have thus far received the most attention are those that participate in RNAi (RNA interference) pathways, namely, miRNAs (microRNAs), siRNAs (small interfering RNAs) and piRNAs (piwi-interacting RNAs), which produce mature RNAs ~20–30 nt in length. Nucleotide sequence complementarity lies at the heart of the widespread and potent regulatory control that these small RNAs exert on their targets. In all known RNAi pathways, this regulatory control is mediated by binding of the small RNA to a complex of proteins, chief among them being the Argonaute proteins. There are also slightly larger small RNA species (~100 nt) that have important cellular roles, including snoRNAs (small nucleolar RNAs), which guide RNA base modification [21] [and can also be processed into other classes of regulatory small RNA, including miRNAs and sdRNAs (sno-derived RNAs)] [22]; snRNAs (small nuclear RNAs), which mediate RNA splicing and are important components of the spliceosome [23]; Y RNAs, which appear to regulate the Ro autoantigen [24]; and vault RNAs, components of the vault ribonucleoprotein complex [25]. Most small ncRNAs function as part of RNA–protein complexes to regulate gene expression, with the small RNA often acting to specify the target for regulation through nucleotide sequence complementarity (Figures 2A and 2B) [26].

miRNAs are ~22 nt long and bind to short regions of complementary sequence, usually located in the 3′ UTR (untranslated region), of target mRNAs [27]. miRNAs can bring about the translational repression or degradation of target transcripts, depending on the extent of complementarity between them (Figure 2A) [26]. The ability of these small RNAs to modulate gene expression was first identified 20 years ago with the discovery of *lin–4* in the nematode worm *Caenorhabditis elegans* [28,29]. Since then, the miRNA field has evolved rapidly and there are now over 1500 miRNAs annotated in the human genome [30]. Mature miRNAs are produced as part of a two-step enzymatic process from a longer pri-miRNA (primary miRNA) that is processed into a pre-miRNA (precursor miRNA) in the nucleus. The pre-miRNA is then transported to the cytoplasm where it is cleaved to release mature miRNAs [31]. The mature miRNA is then loaded on to the RISC (RNA-induced silencing complex), which is composed of the Argonaute2 protein, the miRNA and other auxiliary proteins [26]. The significance of miRNAs in biological processes is highlighted by the fact that one miRNA can potentially target, and hence control the expression of, many hundreds of mRNAs [32,33]. miRNAs can also regulate transcription through mechanisms that are not yet fully understood in mammalian cells [34]. However, it seems likely that there is more to miRNA function than just the inhibition of translation. There are increasing reports describing the presence of mature miRNAs in the nucleus [35,36], suggesting that they may also be directly or indirectly involved in transcriptional gene silencing.

siRNAs are ~21 nt long and function mainly by degrading transcripts they have perfect sequence complementarity to. The precursor transcripts of endogenous siRNAs include dsRNAs (double-stranded RNAs), pseudogenes [37] or transcripts with very long stem-loop structures [38,39]. Endogenous siRNAs have been proposed to protect eukaryotic cells from dsRNA viruses and are also important for silencing transposons and other 'selfish' genomic elements [40,41]. Since siRNAs use the same enzymatic machinery as miRNAs to function, synthetic siRNAs can be introduced into cells to 'knock-down' any given gene, a feature that has been exploited in scientific research and in a number of recently developed therapeutics [42].
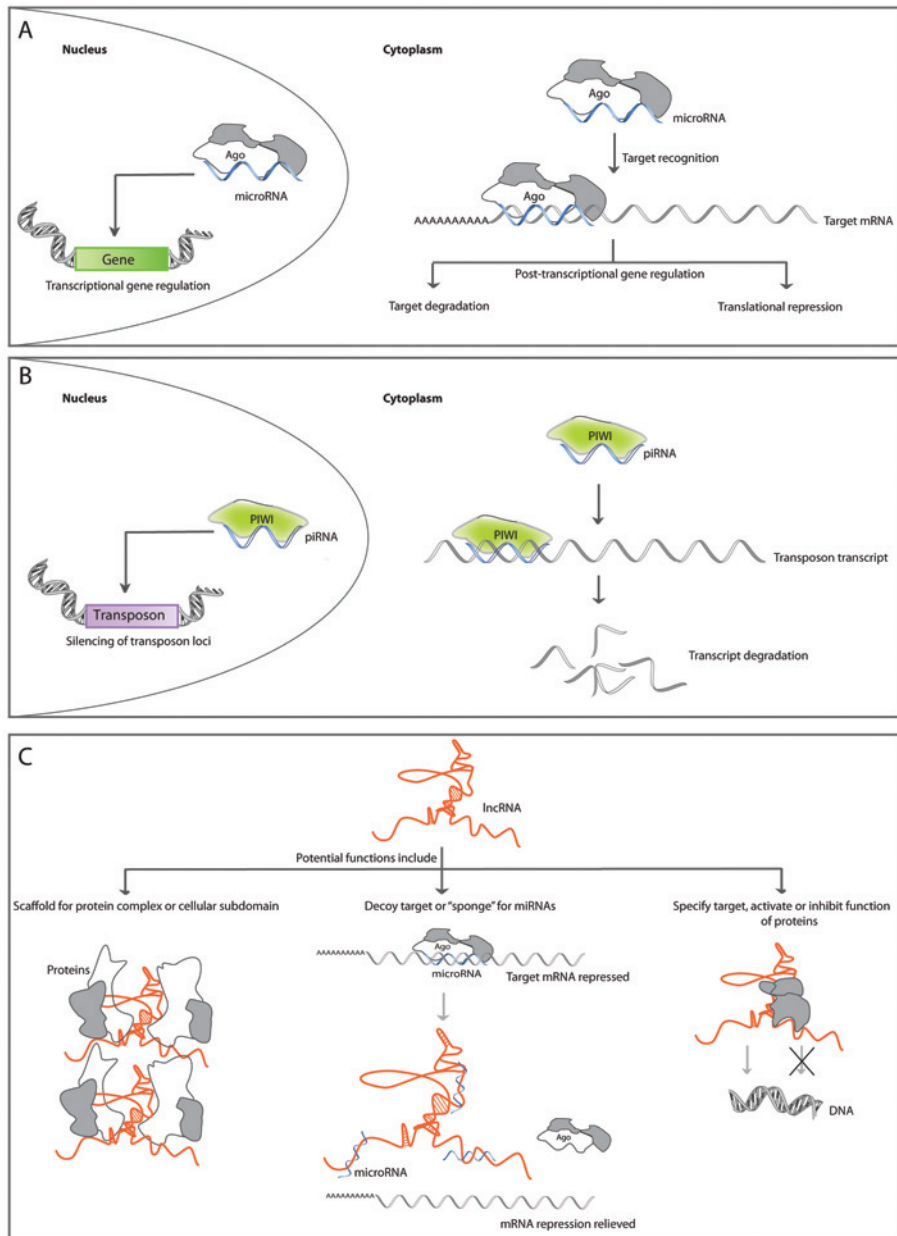
Figure 2. Examples of ncRNA function

(**A**) Transcriptional and translational regulation by miRNAs (shown in blue), which are expressed in nearly every tissue and cell type in complex animals, most simple animals, plants and fungi. Mature miRNAs are loaded into an Argonaute (Ago) protein-containing RISC complex in the cytoplasm. Argonaute proteins have also been reported to function in the nucleus of various organisms. (**B**) piRNA silencing of transposons in germ cells. piRNAs (shown in blue and white) bind to PIWI proteins (shown in green) in the germ line and cause the degradation of transposon transcripts. (**C**) Some potential functions of a lncRNA. Folded lncRNA is shown in red. Proteins are shown in white or grey. miRNAs are in blue. The series of A on the target mRNAs represents the polyA tail. DNA is shown as a double helix.

piRNAs are a related class of small (28–32 nt) RNAs that are found only in animals and principally expressed in spermatids [43]. piRNAs are found in clusters in the genome, and appear to arise from long single-stranded precursor transcripts [44,45]. Unlike miRNAs and siRNAs, which bind to the Ago subclade of the Argonaute proteins, piRNAs associate with the Piwi subclass of Argonaute proteins. They have an important role in suppressing the expression of repetitive elements by guiding DNA methylation, and have been shown to be involved in gametogenesis (Figure 2B) [46,47], as well as in neuronal plasticity in the sea hare, *Aplysia* [48].

Apart from these major categories, other small ncRNAs have been described that originate from regions adjacent to transcription start-sites, such as tiRNAs (transcription initiation RNAs). These RNAs are ~17–18 nt long and are abundant at active promoters as well as at loci with evidence of bidirectional transcription [49], and have been shown to influence the epigenetic state of the genomic region from which they are derived [50]. RNAs of similar size [spliRNAs (splice site RNAs)] are also associated with splice sites [36]. Other, but less well-defined, small RNAs found at or near transcription start-sites that have been reported include PASRs (promoter-associated small RNAs) [19], PROMPTs (promoter upstream transcripts) [51], TSSa-RNAs (transcription start-site-associated RNAs) [52] and TASRs (gene termini-associated small RNAs) [19].

Consistent with their role in gene regulation, small RNAs are involved in many cellular processes and their dysfunction is implicated in a number of physiological and developmental defects [53]. For example, aberrant expression of miRNAs is implicated in a wide variety of diseases ranging from disorders of the heart to immune diseases and cancer [54]. Additionally, a snoRNA has been demonstrated to play a central role in Prader–Willi syndrome [55], whereas another has been associated with autism [56].

lncRNAs (long ncRNAs)

lncRNAs are generally defined as ranging in size from ~200 nt to over 100 kb in length [57,58]. Although the 200 nt cut-off for lncRNAs is quite arbitrary, it has the advantage of excluding most transcripts accepted to be members of small RNA classes. Unlike small ncRNAs, lncRNAs cannot be easily divided into different subclasses on the basis of sequence characteristics and mode-of-action, and this inability to classify lncRNAs into different subtypes has contributed to their current arbitrary definition. We have previously suggested a more flexible definition that lncRNAs are "noncoding RNAs that may have a function as either primary or spliced transcripts, which are independent of processing into known classes of small RNAs, such as miRNAs, piRNAs and snoRNAs, while also excluding structural RNAs from classical housekeeping families" [59], such as rRNAs.

The presence of lncRNAs with important functions has been known for some time, with the characterization of lncRNAs such as *Xist* (X-inactive specific transcript, which controls the silencing of one X chromosome in female mammals) in the 1990s [60,61]. However, the first database of eukaryotic lncRNAs had less than 12 entries by the end of the millennium [62]. The identification of thousands of putative lncRNA transcripts from genome-wide transcriptome analysis [15,63,64], along with prominent examples of functional lncRNAs [65–67], demonstrated that lncRNAs such as *Xist* were not rare genomic oddities, but were instead the first characterized examples of a large class of novel genes. By the end of 2010 over 100

lncRNAs had been functionally characterized as part of a surge of research into the lncRNA world [59]. The subset of lncRNAs transcribed from intergenic regions (i.e. genomic loci some distance from and not overlapping protein-coding genes) have received the most recent attention and are known as lincRNAs (long/large intergenic ncRNAs). Along with those from intergenic loci, lncRNAs are transcribed from many other regions of the genome including promoters, enhancers, introns, UTRs, as overlapping or non-coding isoforms of coding genes, antisense to other transcripts and from pseudogenes [15,68–70].

Although often of similar length to mRNAs, there are a number of differences between mRNAs and lncRNAs beyond the absence of a functional ORF (open reading frame) in the latter. Analysis of lncRNA expression has shown they have lower expression levels and are more likely to be expressed in highly tissue- and cell-specific patterns [63,71–73]. Unlike most mRNAs and many small ncRNAs, lncRNAs are not as highly conserved, although they do show evidence of conservation in their promoters, primary sequences and splice sites [15,71,73–75]. Furthermore, many lncRNAs consist of a single exon and those that are spliced have fewer exons than protein-coding genes [63,72,73]. lncRNAs also commonly contain transposable elements and other repeats [73]. Sequences from such genetic elements can be 'domesticated' during evolution and contribute to lncRNA function by promoting their expression and providing functional motifs [76].

lncRNAs carry out a diverse range of functions in the cell (Figure 2C). Although few are reported to function catalytically, many carry out RNA–protein, RNA–DNA and RNA–RNA interactions. Similar to many small ncRNAs, lncRNAs can regulate gene expression via RNA–protein (ribonucleoprotein) complexes (Figure 2C) [77]. A common function of lncRNAs appears to be directing the activity of chromatin-modifying complexes and transcription factors by specifying their genomic DNA targets and activating or inhibiting their function [67,78–83]. In these and other contexts, lncRNAs have the ability to act as scaffolds, nucleating the assembly of larger complexes or cellular structures [84–86]. Other reported lncRNA functions include acting as miRNA sponges to 'soak up' miRNAs, relieving the repression of mRNAs and so controlling mRNA expression levels and mRNA translation [87,88].

lncRNAs can function both locally and *in trans*. An example of the former is *Airn*, which silences the expression of neighbouring genes to regulate imprinting of the *Igf2r* (insulin-like growth factor 2 receptor) locus [57,65]. *Trans*-acting lncRNAs include *HOTAIR*, which is expressed from the *HOXC* locus and acts to silence gene expression at many genomic locations, including the *HOXD* locus, by recruiting repressive chromatin modification complexes [67,89].

Given their range of functions, a number of lncRNAs are also implicated or involved in disease states, including functioning as oncogenes or tumour suppressors [87,90,91], as well as being linked to other complex diseases such as myocardial infarction [92] and Alzheimer's disease [93].

Regulatory RNAs in prokaryotes

The versatility of RNA as a regulator has also been used by prokaryotes, which contain numerous ncRNAs. Similar to eukaryotes, prokaryotic ncRNAs are being found to play an increasingly important role in regulating gene expression [94]. Despite these similarities, few ncRNA classes are shared between prokaryotes and eukaryotes, with the exception of snoRNAs, which are present in archaea (although not in bacteria) [95].

Many prokaryotic sRNAs (small RNAs; abbreviation only used in prokaryotes) and anti-sense RNAs function as ncRNAs. sRNAs are generally defined as transcripts <500 nt and can be expressed from any region of the genome. Antisense RNAs include both sRNAs and longer RNAs that are antisense to coding genes, creating some overlap between the sRNA and anti-sense RNA classes. ncRNAs are common in prokaryotic genomes, with approximately 170 non-coding sRNAs predicted in the archaea *Methanosarcina mazei* [96], ~50 in the bac-teria *Listeria monocytogenes* [97] and 165 in *Pseudomonas aeruginosa* [98]. In comparison, *Pseudomonas* was found to contain 384 antisense RNAs [98], whereas *Helicobacter pylori* was reported to have antisense transcription across most of the genome, covering 46% of ORFs [99].

Prokaryotic regulatory RNAs function by a variety of mechanisms (reviewed in [94,100]). *Cis*-antisense ncRNAs commonly act to repress the expression of the sense coding gene. Repression can occur either at the level of transcription, RNA turnover or by inhibiting translation, with the extensive nucleotide complementarity between the two important for many of these mechanisms. Examples include ncRNAs regulating the copy number of mobile elements such as plasmids and repressing the translation of toxic proteins, such as the SymR antisense sRNA in *Escherichia coli* that represses the synthesis of the SymE toxin protein [101,102].

Trans-encoded sRNAs generally act to repress translation or destabilize target RNA(s), demonstrating some functional similarity to eukaryotic miRNAs. With more limited comple-mentarity than *cis*-antisense RNAs, *trans* RNAs often require the RNA chaperone protein Hfq to bind their targets [94]. An example is the association of four sRNAs with Hfq in *Vibrio chol-erae* to control quorum sensing by destabilizing the mRNA of the quorum-sensing master regu-lator [103]. Some sRNAs also bind directly to proteins by mimicking other nucleic acid sequences. For example, the 6S sRNA mimics the structure of an open promoter to bind RNA polymerase and regulate transcription [94,104].

Another important class of prokaryotic ncRNAs are CRISPRs (clustered regularly inter-spersed short palindromic repeats). First discovered in *E. coli* [105], CRISPRs are now known to exist in most bacteria and archaea [106]. CRISPR loci contain short direct repeats inter-spersed with spacer regions derived from invading mobile elements. CRISPRs are transcribed and processed to generate small crRNAs (CRISPR RNAs), which function to protect the cell from invading bacteriophages and conjugative plasmids [106,107].

The role of ncRNAs in evolutionary innovation

The sequencing and initial annotation of mammalian genomes provided two large surprises: the large fraction of the genome comprised of sequences derived from transposable elements and the much lower than expected number of protein-coding genes [11,12]. In fact, the num-ber of recognized human protein-coding genes (20687) [2] is similar to that in the nematode worm *C. elegans* (20517) [108] and that in a basal metazoan, the sponge *Amphimedon queens-landica* (18500–30000) [109]. Furthermore, much of the protein-coding 'toolkit' that controls multicellular processes in more complex animals is also present in the sponge [109]. There is widespread use of alternative splicing in human genes [110], which can diversify the proteome

without an increase in gene number, suggesting that it is one mechanism to explain differences in complexity. However, alternative splicing itself requires regulation and hence it has been hypothesized that increases in gene regulatory complexity underlie much of morphological complexity [5,8,10].

Moreover, given the relatively stable protein-coding complement, it is clear that most evolutionary adaptation occurs in regulatory sequences, which are fast evolving and show little conservation over long evolutionary distances [111–113]. The discovery that most of this non-coding DNA is dynamically transcribed to generate tens of thousands of ncRNAs [1,15,16] provides a hitherto unexpected mechanism to explain this increase in regulatory complexity.

ncRNAs, with their potential to bind DNA, RNA and protein in a sequence- or structure-specific manner, are versatile and effective regulatory molecules. By providing specificity to generic protein complexes [3], ncRNAs can act as guides to selectively target effector proteins to different loci and thereby regulate the transcription or translation of many genes [90,114].

Lastly the pervasive transcription of different ncRNAs in the genome provides a large dynamic pool of transcripts for selection to act upon, as most ncRNAs are subject to more flexible structure–function constraints than protein-coding RNAs [17,111,115]. For instance, many ncRNAs function via the formation of stable secondary and tertiary structures, which can accommodate compensatory nucleotide substitutions, e.g. A:U base pairs to G:U or G:C, without disrupting their structural (and thus functional) integrity [116]. Moreover, regulatory sequences are also subject to positive selection for adaptive radiation [117]. The overarching conclusion is that regulatory ncRNAs represent a vast hidden layer of evolutionarily plastic *cis*- and *trans*-acting regulatory information that directs the epigenetic pathways that underpin animal development and diversity [8–10].

Conclusions

The last decade has revolutionized our understanding of genomes and what constitutes a gene. It has become increasingly apparent that many cellular functions are mediated by RNA, a realization that has far-reaching implications for understanding human biology and treating human disease. The following chapters outline the state-of-the-art in the characterization of various types of ncRNAs, although the continuing rapid pace of discovery and unknown function of so many ncRNAs makes it clear that much remains to be done before this poorly charted sphere of biology is fully explored.

Summary
- Non-coding RNA genes are abundant in the genome, with similar numbers of protein-coding and non-coding genes in humans.
- Non-coding RNAs are structurally diverse, ranging from less than 20 nt to over 100 kb in length.
- The properties of RNA molecules allow them to function through both sequence complementarity to other RNAs or DNA, as well as forming structures that can interact with proteins and/or nucleic acids.

- Many well-characterized subclasses of small non-coding RNAs are known, with each member of a subclass having a similar functional mechanism, whereas the subclasses and mechanism-of-action of long non-coding RNAs are much less well understood.
- Most functionally characterized non-coding RNAs (whether small or long) function in the regulation of gene expression.
- Non-coding RNAs play essential roles in many biological processes and are crucial for development and disease, and perhaps even the evolution of organisms.

References

1. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F. et al. (2012) Landscape of transcription in human cells. Nature **489**, 101–108
2. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. **22**, 1760–1774
3. Huttenhofer, A. and Schattner, P. (2006) The principles of guiding by RNA: chimeric RNA-protein enzymes. Nat. Rev. Genet. **7**, 475–482
4. Atkins, J.F., Gesteland, R.F. and Cech, T. (2011) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
5. Taft, R.J., Pheasant, M. and Mattick, J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays **29**, 288–299
6. Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mattick, J.S. (2008) The eukaryotic genome as an RNA machine. Science **319**, 1787–1789
7. Brosnan, C.A. and Voinnet, O. (2009) The long and the short of noncoding RNAs. Curr. Opin. Cell Biol. **21**, 416–425
8. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. Hum. Mol. Genet. **15**, R17–R29
9. Prasanth, K.V. and Spector, D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev. **21**, 11–42
10. Mattick, J.S. (2011) The central role of RNA in human development and cognition. FEBS Lett. **585**, 1600–1616
11. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. Nature **409**, 860–921
12. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature **420**, 520–562
13. Stephen, S., Pheasant, M., Makunin, I.V. and Mattick, J.S. (2008) Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. Mol. Biol. Evol. **25**, 402–408

14. Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A. et al. (2011) The reality of pervasive transcription. PLoS Biol. **9**, e1000625

15. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. et al. (2005) The transcriptional landscape of the mammalian genome. Science **309**, 1559–1563

16. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. et al. (2005) Transcriptional maps of 10 human chromosomes at 5–nucleotide resolution. Science **308**, 1149–1154

17. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447**, 799–816

18. Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G. et al. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. **14**, 331–342

19. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L. et al. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science **316**, 1484–1488

20. Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G.J., Kapranov, P., Foissac, S., Willingham, A.T., Duttagupta, R., Dumais, R. and Gingeras, T.R. (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature **457**, 1028–1032

21. Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. Biochimie **84**, 775–790

22. Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P. and Mattick, J.S. (2009) Small RNAs derived from snoRNAs. RNA **15**, 1233–1240

23. Wachtel, C. and Manley, J.L. (2009) Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Mol. Biosyst. **5**, 311–316

24. Sim, S., Weinberg, D.E., Fuchs, G., Choi, K., Chung, J. and Wolin, S.L. (2009) The subcellular distribution of an RNA quality control protein, the Ro autoantigen, is regulated by noncoding Y RNA binding. Mol. Biol. Cell **20**, 1555–1564

25. Berger, W., Steiner, E., Grusch, M., Elbling, L. and Micksche, M. (2009) Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. Cell. Mol. Life Sci. **66**, 43–61

26. Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. Nat. Rev. Genet. **10**, 94–108

27. Lai, E.C. (2002) Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. **30**, 363–364

28. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell **75**, 855–862

29. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**, 843–854

30. Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res **39**, D152–D157

31. Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. **21**, 4663–4670

32. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell **129**, 1401–1414

33. Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. **19**, 92–105

34. Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R. and Frank, W. (2010) Transcriptional control of gene expression by microRNAs. Cell **140**, 111–122

35. Hwang, H.W., Wentzel, E.A. and Mendell, J.T. (2007) A hexanucleotide element directs microRNA nuclear import. Science **315**, 97–100

36. Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J., Rasko, J.E. et al. (2010) Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat. Struct. Mol. Biol. **17**, 1030–1034

37. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T. et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature **453**, 539–543

38. Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R. et al. (2008) An endogenous small interfering RNA pathway in *Drosophila*. Nature **453**, 798–802

39. Okamura, K., Balla, S., Martin, R., Liu, N. and Lai, E.C. (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. Nat. Struct. Mol. Biol. **15**, 581–590

40. Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. **10**, 126–139

41. Chung, W.J., Okamura, K., Martin, R. and Lai, E.C. (2008) Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. Curr. Biol. **18**, 795–802

42. Li, S.D., Chono, S. and Huang, L. (2008) Efficient oncogene silencing and metastasis inhibition via systemic delivery of siRNA. Mol. Ther. **16**, 942–946

43. Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T. et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. Nature **442**, 203–207

44. Thomson, T. and Lin, H. (2009) The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu. Rev. Cell Dev. Biol. **25**, 355–376

45. Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell **128**, 1089–1103

46. Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B. et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell **129**, 69–82

47. Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T. and Hannon, G.J. (2008) A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. Mol. Cell **31**, 785–799

48. Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T. and Kandel, E.R. (2012) A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell **149**, 693–707

49. Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K. et al. (2009) Tiny RNAs associated with transcription start sites in animals. Nat. Genet. **41**, 572–578

50. Taft, R.J., Hawkins, P.G., Mattick, J.S. and Morris, K.V. (2011) The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization. Epigenetics Chromatin **4**, 13

51. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H. and Jensen, T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. Science **322**, 1851–1854

52. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. and Sharp, P.A. (2008) Divergent transcription from active promoters. Science **322**, 1849–1851

53. Esteller, M. (2011) Non-coding RNAs in human disease. Nat. Rev. Genet. **12**, 861–874

54. Boyd, S.D. (2008) Everything you wanted to know about small RNA but were afraid to ask. Lab. Invest. **88**, 569–578

55. Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W. and Beaudet, A.L. (2008) Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat. Genet. **40**, 719–721

56. Nakatani, J., Tamada, K., Hatanaka, F., Ise, S., Ohta, H., Inoue, K., Tomonaga, S., Watanabe, Y., Chung, Y.J., Banerjee, R. et al. (2009) Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. Cell **137**, 1235–1246

57. Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C. and Barlow, D.P. (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat. Genet. **25**, 19–21

58. Furuno, M., Pang, K.C., Ninomiya, N., Fukuda, S., Frith, M.C., Bult, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. et al. (2006) Clusters of internally primed transcripts reveal novel long noncoding RNAs. PLoS Genet. **2**, e37

59. Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E. and Mattick, J.S. (2011) lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. **39**, D146–D151

60. Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J. and Willard, H.F. (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell **71**, 527–542

61. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. and Brockdorff, N. (1996) Requirement for Xist in X chromosome inactivation. Nature **379**, 131–137

62. Erdmann, V.A., Szymanski, M., Hochberg, A., de Groot, N. and Barciszewski, J. (1999) Collection of mRNA-like non-coding RNAs. Nucleic Acids Res. **27**, 192–195

63. Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M. et al. (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res **16**, 11–19

64. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458**, 223–227

65. Sleutels, F., Zwart, R. and Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature **415**, 810–813

66. Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E. et al. (2003) MALAT–1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early–stage non–small cell lung cancer. Oncogene **22**, 8031–8041

67. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell **129**, 1311–1323

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

69. Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El–Jundi, T.A., da Silva, A.M., Reis, E.M. and Verjovski-Almeida, S. (2007) 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**, R43

70. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature **465**, 182–187

71. Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C. et al. (2010) *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. **28**, 503–510

72. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. **25**, 1915–1927

73. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G. et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. **22**, 1775–1789

74. Marques, A.C. and Ponting, C.P. (2009) Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol. **10**, R124

75. Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnar, Z. and Ponting, C.P. (2010) Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. **11**, R72

76. Dinger, M.E., Amaral, P.P., Mercer, T.R. and Mattick, J.S. (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Briefings Funct. Genomics Proteomics **8**, 407–423

77. Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. **81**, 145–166

78. Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson–Smith, A.C., Feil, R. and Fraser, P. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science **322**, 1717–1720

79. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. and Lee, J.T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science **322**, 750–756

80. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M. et al. (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell **142**, 409–419

81. Kino, T., Hurt, D.E., Ichijo, T., Nader, N. and Chrousos, G.P. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci. Signaling **3**, ra8

82. Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L. et al. (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature **477**, 295–300

83. Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC–1 complex. Cell **97**, 17–27

84. Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A. and Lawrence, J.B. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol. Cell **33**, 717–726

85. Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S. and Spector, D.L. (2009) MENε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res. **19**, 347–359

86. Shevtsov, S.P. and Dundr, M. (2011) Nucleation of nuclear bodies by RNA. Nat. Cell Biol. **13**, 167–173

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

88. Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A. and Bozzoni, I. (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell **147**, 358–369

89. Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. and Chang, H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science **329**, 689–693

90. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L. et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature **464**, 1071–1076

91. Zhang, X., Gejman, R., Mahta, A., Zhong, Y., Rice, K.A., Zhou, Y., Cheunsuchon, P., Louis, D.N. and Klibanski, A. (2010) Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. Cancer Res. **70**, 2350–2358

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

93. Mus, E., Hof, P.R. and Tiedge, H. (2007) Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc. Natl. Acad. Sci. U.S.A. **104**, 10679–10684

94. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. Cell **136**, 615–628

95. Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R. and Dennis, P.P. (2000) Homologs of small nucleolar RNAs in Archaea. Science **288**, 517–522

96. Jager, D., Sharma, C.M., Thomsen, J., Ehlers, C., Vogel, J. and Schmitz, R.A. (2009) Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. Proc. Natl. Acad. Sci. U.S.A. **106**, 21878–21882

97. Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K. et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. Nature **459**, 950–956

98. Wurtzel, O., Yoder-Himes, D.R., Han, K., Dandekar, A.A., Edelheit, S., Greenberg, E.P., Sorek, R. and Lory, S. (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLoS Pathog. **8**, e1002945

99. Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R. et al. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464**, 250–255

100. Thomason, M.K. and Storz, G. (2010) Bacterial antisense RNAs: how many are there, and what are they doing? Annu. Rev. Genet. **44**, 167–188

101. Tomizawa, J. and Itoh, T. (1981) Plasmid ColE1 incompatibility determined by interaction of RNA I with primer transcript. Proc. Natl. Acad. Sci. U.S.A. **78**, 6096–6100

102. Kawano, M., Aravind, L. and Storz, G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol. Microbiol. **64**, 738–754

103. Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S. and Bassler, B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell **118**, 69–82

104. Wassarman, K.M. and Storz, G. (2000) 6S RNA regulates *E. coli* RNA polymerase activity. Cell **101**, 613–623

105. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. **169**, 5429–5433

106. Jore, M.M., Brouns, S.J. and van der Oost, J. (2012) RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. Cold Spring Harbor Perspect. Biol. **4**, a003657

107. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science **315**, 1709–1712

108. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S. et al. (2011) Ensembl 2011. Nucleic Acids Res. **39**, D800-D806

109. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U. et al. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature **466**, 720–726

110. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature **456**, 470–476

111. Meader, S., Ponting, C.P. and Lunter, G. (2010) Massive turnover of functional sequence in human and other mammalian genomes. Genome Res. **20**, 1335–1343

112. Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T. and Marques, A.C. (2012) Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet. **8**, e1002841

113. Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S. et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science **328**, 1036–1040

114. Guo, H., Ingolia, N.T., Weissman, J.S. and Bartel, D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature **466**, 835–840

115. Heinen, T.J., Staubach, F., Haming, D. and Tautz, D. (2009) Emergence of a new gene from an intergenic region. Curr. Biol. **19**, 1527–1531

116. Smit, S., Knight, R. and Heringa, J. (2009) RNA structure prediction from evolutionary patterns of nucleotide composition. Nucleic Acids Res. **37**, 1378–1386

117. Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C. and Peterson, K.J. (2008) MicroRNAs and the advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. U.S.A. **105**, 2946–2950

118. Mercer, T.R., Wilhelm, D., Dinger, M.E., Solda, G., Korbie, D.J., Glazov, E.A., Truong, V., Schwenke, M., Simons, C., Matthaei, K.I. et al. (2011) Expression of distinct RNAs from 3′ untranslated regions. Nucleic Acids Res. **39**, 2393–2403