Seminars in Cell & Developmental Biology 23 (2012) 200–205

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Exploring long non-coding RNAs through sequencing

Sophie R. Atkinson[a,b], Samuel Marguerat[a], Jürg Bähler[a,*]

[a] University College London, Department of Genetics, Evolution & Environment and UCL Cancer Institute, Darwin Building, Gower Street, London WC1E 6BT, United Kingdom
[b] CoMPLEX, University College London, Gower Street, London WC1E 6BT, United Kingdom

ARTICLE INFO

Article history:
Available online 20 December 2011

Keywords:
Next-generation sequencing
RNA-seq
Non-coding RNA
Antisense transcript
Gene regulation

ABSTRACT

Long non-coding RNAs (lncRNAs) are emerging as an important class of regulatory transcripts that are implicated in a variety of biological functions. RNA-sequencing, along with other next-generation sequencing-based approaches, enables their study on a genome-wide scale, at maximal resolution, and across multiple conditions. This review discusses how sequencing-based studies are providing global insights into lncRNA transcription, post-transcriptional processing, expression regulation and sites of function. The next few years will deepen our insight into the overall contribution of lncRNAs to genome function and to the information flow from genotype to phenotype.

Contents

1. Introduction

RNA-sequencing (RNA-seq), based on next-generation (NGS) sequencing of cDNAs, is transforming the characterisation and quantification of transcriptomes. Unlike microarray-based approaches, RNA-seq does not depend on available genome annotation or sequence, can accurately detect expression levels over a wide dynamic range, and can reveal entire transcriptomes with high sensitivity and to nucleotide resolution (reviewed in [1–3]). A profusion of microarray and RNA-seq studies, on species ranging from simple microbes to humans, has revealed that the transcribed portions of genomes are much more pervasive and more complex than anticipated [4–8]. For example, less than 2% of the human genome encodes for proteins, yet as much as 80% of all DNA is transcribed [9,10].

Such pervasive transcription leads to a multitude of previously unknown non-coding RNAs (ncRNAs), many of which have an arbitrary minimal length cut-off of 200 nucleotides (due to RNA-seq library preparation protocols that exclude small RNAs), and are subsequently referred to as long ncRNAs (lncRNAs). These

Abbreviations: AS, antisense; ncRNAs, non-coding RNAs; ChIP-seq, chromatin immunoprecipation with sequencing; CAR, chromatin-associated ncRNA; CUT, cryptic unstable transcript; eRNAs, enhancer RNAs; GRO-seq, global run-on sequencing; GWAS, genome-wide association studies; lincRNAs, long intergenic ncRNAs; lncRNAs, long non-coding RNAs; NET-seq, native elongating transcript sequencing; NGS, next-generation-sequencing; Pol II, RNA polymerase II; RNA-seq, RNA sequencing; SUT, stable unannotated transcript.

* Corresponding author. Tel.: +44 0203 108 1602; fax: +44 0207 679 7096.
E-mail address: j.bahler@ucl.ac.uk (J. Bähler).

1084-9521 © 2011 Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.semcdb.2011.12.003

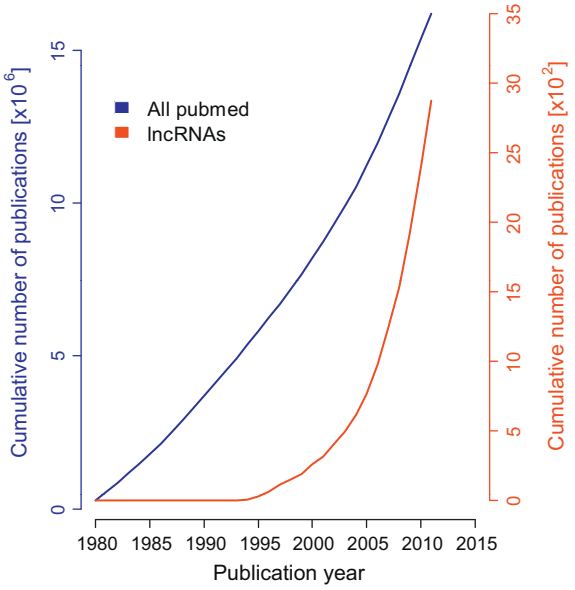

Fig. 1. Research on long non-coding RNAs is rapidly increasing. Cumulative plot of the total number of publication entries in PubMed (blue line and axis) and of entries related to long non-coding RNAs (red line and axis).

transcripts are distinct from small regulatory ncRNAs such as siRNAs, miRNAs and piRNAs, which are typically conserved, shorter than 30 nucleotides, and are involved in transcriptional and post-transcriptional gene silencing through specific base-pairing with their mRNA targets (reviewed in [11]). lncRNAs tend to be lowly expressed [7,12,13] and, although they are less conserved than protein-coding genes, there is evidence that they have been subject to purifying selection, hinting at their possible functionality [14,15]. Moreover, the evolutionary rates and expression levels of lncRNAs are anti-correlated, a characteristic they share with protein-coding genes [16]. While protein-coding genes are largely conserved in number and function between organisms of highly different complexities [10], the extent of lncRNAs increases much more with organismal complexity. Differences in genome regulation, possibly mediated to a substantial degree by lncRNAs, may contribute significantly to phenotypic differences between and within species (reviewed in [10]).

Determining the nature and possible biological functions of lncRNAs has been a rapidly developing field over the past decade (Fig. 1). The field has been pioneered by mechanistic studies of single genes, which have unravelled a variety of functions of lncRNAs in gene regulation (reviewed in [17,18]). Today, with its unrivalled sensitivity and resolution, RNA-seq allows the detection and quantification of lncRNAs on a genome-wide scale, revealing the global extent and complexity of non-coding transcriptomes. This review will provide an overview of how RNA-seq and other sequencing-based approaches are enlightening the genomic origins, expression regulation, and functions of lncRNAs.

2. Genomic origins of lncRNA transcription

RNA-seq studies have provided a genome-wide view of where lncRNAs are transcribed relative to coding regions. Several classes of lncRNAs can be distinguished, which are discussed in separate sections below: transcripts antisense to protein-coding genes, bidirectional promoter transcripts, transcripts associated with enhancers or repetitive regions, and other transcripts originating in intergenic regions (Fig. 2). Note that these different classes are not mutually exclusive, and it is not yet clear whether and how the classification reflects biological function.

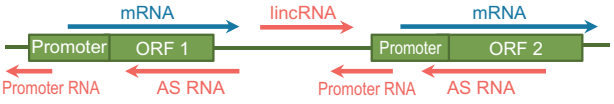

Fig. 2. Different types of long non-coding RNAs. Scheme showing different categories of transcripts. Green: DNA, with promoters and open reading frames (ORFs) depicted as boxes; blue arrows: messenger RNAs (mRNAs); red arrows: lncRNAs. Examples for different lncRNAs are antisense (AS) RNAs, promoter RNAs transcribed in opposite direction to mRNAs, and a long intergenic RNA (lincRNA). Some lncRNAs are associated with enhancer elements or with repeated regions. See main text for details.

2.1. Antisense transcripts

Standard RNA-seq libraries do not preserve information about transcriptional direction. However, several methods now exist for strand-specific RNA-seq (reviewed in [19]), revealing complex overlapping transcription in several genomes, with many lncRNAs being transcribed from the complementary strand of protein-coding genes. Such lncRNAs are referred to as antisense (AS) transcripts. Larger genomes have more complex transcriptomes, and therefore require a greater sequencing depth for adequate coverage. Given sequencing costs, simpler genomes are more amenable to high coverage to reveal their entire transcript repertoire. Therefore simple model organisms with smaller, more compact genomes, such as yeasts, have been instrumental in revealing the nature of often lowly expressed and complex AS transcripts.

Initial studies using whole-genome tiling arrays in both fission and budding yeast have revealed extensive AS transcription in rapidly proliferating cells [7,20–23]. Application of strand-specific RNA-seq to these eukaryotic microbes has now provided maps of AS transcription across different growth conditions at the highest possible resolution [24–26]. Such studies provide accurate detection of AS transcript boundaries, uncovering overlaps with variable portions of coding genes. In addition, AS transcript co- or anti-regulation (or lack thereof) with neighbouring genes suggest that they can arise from novel transcription start sites, but also from bidirectional transcription at promoters, or *via* transcriptional read-through in the case of convergent genes [24–26]. Several chromatin structure mutants also show higher prevalence of AS transcription [27,28]. The same studies have additionally reported that detectable AS transcripts do not originate randomly in the genome but are more likely to overlap genes involved in sexual differentiation or stress response [24–26], as well as genes with higher variability in transcript levels [23]. Notably, certain sense/AS transcript pairs and their co-regulation are conserved across several yeast species [25,26]. Taken together, these data suggest that AS transcription is a phenomenon of pervasive gene expression with diverse features and impacts on gene regulation. For instance, in the case of genes with highly variable gene expression, AS transcripts help to tightly shut off basal transcription [23]. Microarray- and NGS-based approaches have also revealed AS transcription in mammalian genomes [5,9,29], but to what degree and by what mechanisms AS transcripts control sense transcripts remains to be fully elucidated (reviewed in [30]).

2.2. Bidirectional promoter transcription

NGS-based approaches have helped to reveal an unanticipated property of RNA polymerase II (Pol II): transcriptional initiation can be bi-directional. For example, the cryptic unstable transcripts (CUTs) are Pol II-dependent transcripts defined in budding yeast can result from such bi-directional transcription. CUTs are degraded by the nuclear exosome shortly after synthesis [31]. Two recent studies have generated genome-wide maps of CUTs using NGS [32]

or high-density tiling arrays [22]. Both studies have revealed that CUTs are well-defined transcriptional units that tend to be transcribed from gene promoters in an opposite direction to the coding RNAs. This divergent transcription from promoters is not limited to unstable transcripts that are rapidly degraded by the nuclear exosome: stable unannotated transcripts (SUTs) can also arise from divergent transcription from known promoters [22].

Similar transcriptional patterns have been observed in other eukaryotes. Exosome depletion in human fibroblasts followed by tiling array analysis has revealed lncRNAs which map upstream of known protein-coding genes. Such promoter-upstream transcripts have been termed PROMPTs [33]. In addition, stable transcripts mapping to both strands of promoters are detected in metazoan cells. RNA-seq of short RNAs (~20–200 nucleotides) from mouse embryonic stem cells has shown that many of these transcripts originate from promoters (transcription start site-associated RNAs) and are transcribed in a non-random, divergent direction [34]. Furthermore, Core et al. [35] have employed global run-on sequencing (GRO-seq) to identify nascent RNAs associated with actively transcribing Pol II in human fibroblasts, independently of nascent transcript length or stability. This analysis has revealed that 77% of the transcriptionally active protein-coding genes display significant divergent transcription from promoters. Moreover, bidirectional transcripts have been reported to initiate from retrotransposons [36].

In summary, bidirectional transcription from promoters seems to be a widespread phenomenon conserved across evolution. There have been suggestions that such bidirectional transcription may facilitate protein-coding gene expression by promoting an open-chromatin structure at the promoter or by recruiting positive or negative transcriptional regulators. It is not known, however, whether most bidirectional transcripts are non-functional by-products of coding transcription, or whether they play regulatory roles (reviewed in [37]).

2.3. Enhancer associated lncRNAs

Enhancers spatially and temporally regulate protein-coding transcription from a distance and in an orientation-independent manner, relative to the regulated gene. Initial studies of the β-globin locus [38] have revealed that the hypersensitive site 2 (HS2) enhancer is transcribed into a lncRNA. Subsequently, several NGS-based studies have revealed widespread transcription from enhancers, resulting in a class of lncRNAs termed enhancer RNAs (eRNAs).

Using chromatin immunoprecipitation with sequencing (ChIP-seq) to map genomic binding sites of the enhancer protein CBP, Kim et al. [39] identified over 12,000 stimulus-dependent enhancers in mouse neurons. Furthermore, ChIP-seq showed Pol II to be present at over 25% of those enhancers. RNA-seq of total RNA from neurons has revealed lncRNAs that are produced in both directions from such Pol II-bound enhancers, and the expression levels of these eRNAs are correlated with mRNA synthesis from nearby genes [39].

In a related study, De Santa et al. [40] have applied ChIP-seq of Pol II in macrophages to reveal actively elongating Pol II bound to putative enhancer sites upstream of lipopolysaccharide-inducible genes. Many of these Pol II-bound enhancers are actively transcribed, and eRNA expression correlates with the expression of neighbouring genes. Notably, eRNA synthesis frequently precedes the induction of the adjacent coding gene [40]. Furthermore, a meta-analysis of RNA-seq data has shown that enhancer function associated with production of lncRNAs from mammalian ultra-conserved elements is much more widespread than previously thought [41].

Such NGS studies have raised the possibility that enhancer function may in fact be mediated through transcribed eRNAs. While it has been speculated that eRNAs may recruit enhancer-associated proteins, or perhaps facilitate chromatin looping to provide contact between the enhancer region and the promoter of the regulated gene, further studies are required to determine any biological functions and mechanisms of action of eRNAs (reviewed in [42]).

2.4. Long intergenic ncRNAs (lincRNAs)

In mammals, large projects such as those of the FANTOM and ENCODE consortiums [5,9] have uncovered widespread intergenic transcription which does not overlap with protein-coding genes. Guttman et al. [15] have developed a method to systematically identify such long intergenic ncRNAs (lincRNAs) using genome-wide chromatin-state maps, generated by ChIP-seq, to identify discrete transcriptional units occurring between protein-coding genes. Based on the observation that Pol II-transcribed genes display H3K4Me3 marks at their promoters and H3K36Me3 marks along their bodies, 'K4–K36' domains lying outside of known protein-coding genes were uncovered. This approach identified 1600 lincRNAs in four mouse cell types.

More recently, a comprehensive annotation of human lincRNAs has been achieved using transcriptome assembly of RNA-seq data from 24 human tissues and cell types [13]. 'Align-then-assemble' *ab initio* transcriptome assembly should theoretically be more sensitive than 'assemble-then-align' *de novo* approaches which are biased towards highly expressed transcripts, although a comprehensive comparison has yet to be made [43]. By selecting from the RNA-seq reconstructions those lincRNAs that are reliably expressed, a stringent set of 4662 lincRNAs has been produced [13]. Importantly, the authors have found that the identified lincRNAs are expressed in a highly tissue-specific manner, much more so than protein-coding genes. In addition, protein-coding genes proximal to lincRNAs are disproportionately associated with development and transcriptional regulation [14,15,29], hinting at possible functional roles of lincRNAs.

One of the most well-characterised lincRNAs is HOTAIR, which is transcribed within the HOXC cluster and represses genes in the HOXD cluster by binding and recruiting the chromatin-modifying complex PRC2 [44]. Khalil et al. [45] show that many lincRNAs are bound by PRC2 and other chromatin-modifying complexes. Moreover, RNAi-based depletion of various PRC2-associated lincRNAs results in activation of genes known to be repressed by PRC2 [45]. Such studies have led to the suggestion that lincRNAs guide chromatin-modifying complexes to specific genomic loci (reviewed in [46]).

2.5. Repetitive element-associated ncRNAs

Repetitive elements such as retrotransposons comprise 30–50% of mammalian genomes, and the advent of NGS technologies has uncovered transcriptional activity associated with such elements. Cross-hybridisation problems have hampered the use of array-based approaches to study genome-wide repetitive element transcription. In contrast, sequence-tag technologies can detect single base-pair differences between repetitive elements, enabling their discrimination.

Using a 'deep-CAGE' method to globally map transcription start sites (cleavage of ~20-nucleotide tags from extreme 5′ and 3′ ends of cDNAs, followed by sequencing), the FANTOM4 project has revealed extensive transcription of retrotransposons in human and mouse genomes [36]. Retrotransposons are expressed in a tissue-specific manner and proximal to protein-coding genes, suggesting roles of controlling alternative promoters or post-transcriptionally regulating protein-coding transcripts (reviewed in [47]).

Pseudogenes form another class of repetitive elements that can be transcribed into lncRNAs, as has recently been shown for the

PTEN and KRAS loci, which can regulate expression of their corresponding protein-coding genes by competing for regulatory miRNA binding [48,49].

3. Post-transcriptional processing of lncRNAs

Understanding the molecular pathways governing the production, function, and turnover of lncRNA is key to understanding their diversity and functions. For instance, the transcriptomes of budding yeast *rrp6* mutants – defective for the nuclear exosome complex involved in RNA degradation – have revealed a class of transcripts referred to as CUTs, which are rapidly degraded after synthesis (see Section 2.2) [22,32]. There is evidence that physiological conditions may affect the unstable nature of CUTs and render them stable. For example, loss of the Rrp6 protein in budding yeast leads to stabilisation of an AS transcript to the *PHO84* gene, and subsequent repression of *PHO84* transcription. Intriguingly, the same phenotype is observed during chronological aging, as the Rrp6 protein shows weaker association with the *PHO84* locus under these conditions [50]. In addition, a class of meiotic ncRNA is actively degraded during the mitotic cell-cycle by the nuclear exosome and becomes stabilised as the cell proceeds into meiotic differentiation [51]. These findings indicate that modulation of Rrp6 function by external, physiologically relevant cues could contribute to gene regulation, and that unstable transcripts could represent regulatory classes of lncRNAs.

Strand-specific RNA-seq of an *xrn1* exonuclease mutant has identified 1658 Xrn1p-sensitive unstable transcripts (XUTs) degraded by the 5′–3′ cytoplasmic RNA decay pathway in budding yeast [52]. More than 50% of the identified XUTs are AS transcripts, and they accumulate in lithium-containing media indicating a possible role in adaptive responses to changing growth conditions.

Native elongating transcript sequencing (NET-seq) – the deep sequencing of the 3′ ends of nascent transcripts associated with Pol II – follows RNAs as they are produced, regardless of their stability, making it ideally suited to analyze unstable transcripts. This technique has recently been used to profile nascent transcripts in budding yeast defective for the Rpd3S deacetylation complex, revealing a pervasive increase in unstable AS transcription produced from bidirectional promoters [53]. This study suggests that Rpd3S deacetylation enforces strong directionality to most promoters, suppressing bidirectional transcription under normal growth conditions. Similarly, it has also been shown in fission yeast that the variant histone H2A.Z, which localises at 5′ ends of genes, cooperates with heterochromatin and RNA interference factors to suppress AS transcription under normal growth conditions [27]. Finally, another study in yeast suggests an unexpected role of the Nrd1–Nab3–Sen1 complex in termination of ncRNA processing, showing its association not only with snoRNAs and CUTs but also with a class of AS transcripts [54].

Such transcriptomic studies demonstrate that lncRNAs are actively regulated transcriptionally and post-transcriptionally and illustrate that systematic RNA-seq analysis of mutants defective for chromatin maintenance or RNA processing pathways is a powerful approach to reveal novel, possibly regulatory, classes of lncRNAs.

4. Functional sites and expression regulation of lncRNAs

To determine biological roles of lncRNAs, it is important to establish the conditions and cell types they are expressed in and the subcellular locations where they function. Specific and dynamic expression patterns, are emerging for several lncRNAs, which intrinsically suggest functionality ([55], reviewed in [10]).

4.1. Subcellular locations of lncRNAs

A study in human cell lines suggests that ∼30% of lncRNAs are found exclusively in the nucleus, ∼15% are found exclusively in the cytoplasm, while ∼50% show both nuclear and cytoplasmic localisation [56]. Such a subcellular distribution implicates lncRNAs in gene regulation at both transcriptional and post-transcriptional levels. Indeed, these are two functional themes emerging from recent reports based on the analysis of individual lncRNAs.

At a transcriptional level, lncRNAs emerging from promoters or enhancers may act as scaffolds by binding, recruiting, or coordinating transcriptional activators and repressors [17,18]. Additionally, it has been suggested that the process of lncRNA transcription, rather than the lncRNA product itself, may be functional by facilitating an open chromatin structure at protein-coding promoters, thereby increasing access to transcriptional activators and to Pol II [57].

At a post-transcriptional level, the ability of lncRNAs to recognise complementary sequences may enable highly specific regulatory interactions with mRNAs. This provides opportunities for lncRNAs, in particular AS transcripts, to regulate the splicing, editing, transport, translation, or degradation of mRNAs.

Examples of different modes of action of lncRNAs in transcriptional or post-transcriptional gene regulation have been extensively reviewed elsewhere [17,18]. RNA-seq profiling of different cellular fractions will help to obtain a global view of lncRNAs potentially acting at transcriptional and/or post-transcriptional levels.

Two recent studies have employed ribosomal profiling – the deep sequencing of ribosome-protected RNAs – to determine whether any lncRNAs are found in association with ribosomes [58,59]. A key step in the annotation of lncRNAs is demonstrating the absence of a conventional open reading frame. Remarkably, however, these studies have revealed that over half of the annotated SUTs in budding yeast [58] and the majority of annotated lincRNAs in mouse embryonic stem cells [59] are exported to the cytoplasm, where they are engaged by the protein translation machinery. The extent to which lncRNAs, such as SUTs and lincRNAs, might act *via* proteins they encode rather than *via* RNA itself, is now an open question. It has been suggested that the low-level translation of some lncRNA transcripts may help provide the raw material for *de novo* birth of protein-coding genes [58].

Furthermore, deep-sequencing of the human mitochondrial transcriptome has revealed the presence of mitochondrial lncRNAs [60]. It will be of interest to determine whether such lncRNAs play any mitochondrial-specific roles.

4.2. Chromatin-related functions

As described above, an additional emerging theme of lncRNA function is in guiding chromatin-modifying complexes to specific genomic loci. Such a function helps to solve the apparent paradox of how chromatin-remodelling complexes with little DNA sequence specificity, but often with RNA-binding domains, are able to control complex chromatin modifications at specific genomic loci.

Many lncRNAs whose mechanism of action are well-characterised, such as HOTAIR [44] and Xist (reviewed in [61]), use chromatin as a substrate to execute their biological function. It has therefore been postulated that lncRNAs associated with chromatin may be more likely to have biological functions. By sequencing the RNAs of chromatin fractions, approximately 200 such chromatin-associated ncRNAs (CARs) have been detected in human fibroblast cells, implicating them in possible structural and functional roles in chromatin organisation [62].

To further probe the function of such CARs, it will be necessary to determine exactly where they bind to chromatin.

A recently described technique, Chromatin Isolation by RNA Purification(ChIRP)-seq, could enable such analyses on a genome-wide scale [63]. Chromatin is cross-linked to lncRNAs, and biotinylated oligonucelotide probes are then used to retrieve specific lncRNAs, together with bound DNA sequences which can then be interrogated by NGS. Just as ChIP-seq has greatly improved our understanding of protein–DNA interactions on a genomic scale, ChIRP-seq has the potential to map lncRNA:chromatin interactions *in vivo*, genome-wide and at high resolution. ChIRP-seq has been applied to three lncRNAs – TERC, HOTAIR and roX2 – revealing that lncRNA binding sites resemble transcription factor binding sites in being focal, numerous and sequence-specific [63].

4.3. Specific and dynamic expression of lncRNAs

In addition to determining where in the genome lncRNAs originate and where in the cell they function, determining under what conditions and in what cells they are expressed can provide key insights into their function. Initial studies in mouse and human have demonstrated that many lncRNAs are expressed in a cell- and tissue-specific manner during development and differentiation, suggesting that they might participate in the regulation of such processes ([64], reviewed in [65]).

RNA-seq is ideally suited to quantify transcripts, and expression levels can be easily compared across different conditions and tissues, without the need for complicated normalisation methods (reviewed in [2]). A recent RNA-seq study [13] has demonstrated that the vast majority of lincRNAs show tissue-specific expression patterns, with as much as 78% being categorised as tissue-specific compared to only ~19% of the coding genes. Notably, a large group of lincRNAs are specific to testes [13]. Another RNA-seq study [66] has analyzed the transcriptomes of 102 prostate cancer samples, defining 121 lncRNAs whose expression patterns distinguish two stages of cancer development. One such lncRNA is a prostate-specific regulator of cell proliferation [66]. Similar RNA-seq studies of dynamic lncRNA expression across different conditions and developmental stages in health and disease will provide comprehensive, sensitive and high-resolution data of lncRNA expression regulation.

5. Conclusions and outlook

RNA-seq is a powerful tool for the detection and quantification of lncRNAs. Several ongoing developments promise further advances in the insights that RNA-seq will be able to provide. For instance, it is increasingly appreciated that gene expression in individual cells deviates significantly from the average behaviour of cell populations [67]. Single-cell transcriptome profiling could revolutionise our understanding of genome regulation beyond population averages. Additionally, single-cell RNA-seq would enable transcriptome studies where only tiny amounts of cellular material are available, such as during early embryonic development or disease-associated samples. By exploiting microfluidics systems for single-cell delivery, technologies for single-cell transcriptome profiling are already emerging (reviewed in [3]).

Biases and artifacts in RNA-seq data are frequently introduced during cDNA synthesis and the subsequent manipulation steps in RNA library preparation. Direct high-throughput sequencing of RNA molecules, without prior cDNA conversion, offers the potential to mitigate many of the current problems and biases. Direct RNA sequencing technologies are already on the horizon [68].

Advances in algorithms used to analyse sequence data will also increase the power of RNA-seq to unravel transcriptomes. For example, methods which enable transcriptome reconstruction in the absence of a reference genome will help to reveal transcriptomic complexities in incompletely sequenced genomes, as well as uncovering the full complexity for those with known genome sequences. Such methods will also prove useful for cells with highly rearranged genomes such as cancer cells (reviewed in [43,69]).

A number of genome-wide association (GWA) studies have shown that variations associated with complex disorders often map to non-coding regions of the genome, implicating lncRNAs in disease (reviewed in [70]). While the mechanisms by which lncRNAs may contribute to pathogenesis are largely unknown, several lncRNAs have already been suggested to play a role in various diseases, through alterations in their expression levels or interactions with RNA-binding proteins [13]. A recently described technique, termed RNA CaptureSeq, employs tiling arrays of targeted genomic regions to capture cDNAs, followed by deep sequencing [71]. For targeted loci, such an approach enables a greater depth of coverage than can be achieved by conventional RNA-seq, and has revealed additional complexities in the human transcriptome. By enabling comprehensive interrogation of specific genomic regions, RNA CaptureSeq could be used to profile all transcripts produced from non-coding regions implicated in disease-susceptibility by GWA studies.

Being potentially more amenable drug targets than proteins, lncRNAs present a new frontier in biomedicine. No doubt, RNA-seq and other NGS-based approaches will continue to advance our understanding of the nature, extent and possible functions of lncRNAs in health and disease.

Acknowledgements

We thank Danny Bitton, Antonia Lock, Vincent Plagnol, Stephen Watt and Judith Zaugg for comments on the manuscript. Research in our laboratory is funded by the BBSRC and by a Wellcome Trust Senior Investigator Award.

References

[1] Marguerat S, Bähler J. RNA-seq: from technology to biology. Cell Mol Life Sci 2010;67:569–79.
[2] Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.
[3] Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 2011;12:87–98.
[4] Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 2002;420:563–73.
[5] Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science 2005;309:1559–63.
[6] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 2008;5:621–8.
[7] Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 2008;453:1239–43.
[8] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008;320:1344–9.
[9] Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799–816.
[10] Mattick JS. The central role of RNA in human development and cognition. FEBS Lett 2011;585:1600–16.
[11] Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet 2009;10:94–108.
[12] van Bakel H, Nislow C, Blencowe BJ, Hughes TR. Most "dark matter" transcripts are associated with known genes. PLoS Biol 2010;8:e1000371.
[13] Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25:1915–27.
[14] Ponjavic J, Oliver PL, Lunter G, Ponting CP. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet 2009;5:e1000617.
[15] Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009;458:223–7.

[16] Managadze D, Rogozin IB, Chernikova D, Shabalina SA, Koonin EV. Negative correlation between expression level and evolutionary rate of long intergenic non-coding RNAs. Genome Biol Evol 2011.
[17] Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10:155–9.
[18] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–14.
[19] Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 2010;7:709–15.
[20] David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, et al. A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA 2006;103:5320–5.
[21] Dutrow N, Nix DA, Holt D, Milash B, Dalley B, Westbroek E, et al. Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA–DNA hybrid mapping. Nat Genet 2008;40:977–86.
[22] Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature 2009;457:1033–7.
[23] Xu Z, Wei W, Gagneur J, Clauder-Munster S, Smolik M, Huber W, et al. Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol 2011;7:468.
[24] Ni T, Tu K, Wang Z, Song S, Wu H, Xie B, et al. The prevalence and regulation of antisense transcripts in *Schizosaccharomyces pombe*. PLoS One 2010;5:e15271.
[25] Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, et al. Comparative functional genomics of the fission yeasts. Science 2011;332:930–6.
[26] Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, et al. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol 2010;11:R87.
[27] Zofall M, Fischer T, Zhang K, Zhou M, Cui B, Veenstra TD, et al. Histone H2A.Z. cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature 2009;461:419–22.
[28] Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, et al. Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. PLoS Biol 2008;6:e277.
[29] Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 2010;28:503–10.
[30] Werner A, Swan D. What are natural antisense transcripts good for? Biochem Soc Trans 2010;38:1144–9.
[31] Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 2005;121:725–37.
[32] Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 2009;457:1038–42.
[33] Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, et al. RNA exosome depletion reveals transcription upstream of active human promoters. Science 2008;322:1851–4.
[34] Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, et al. Divergent transcription from active promoters. Science 2008;322:1849–51.
[35] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 2008;322:1845–8.
[36] Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 2009;41:563–71.
[37] Wei W, Pelechano V, Jarvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. Trends Genet 2011;27:267–76.
[38] Ling J, Baibakov B, Pi W, Emerson BM, Tuan D. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. J Mol Biol 2005;350:883–96.
[39] Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010;465:182–7.
[40] De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 2010;8:e1000384.
[41] Licastro D, Gennarino VA, Petrera F, Sanges R, Banfi S, Stupka E. Promiscuity of enhancer, coding and non-coding transcription functions in ultraconserved elements. BMC Genomics 2010;11:151.
[42] Orom UA, Shiekhattar R. Long non-coding RNAs and enhancers. Curr Opin Genet Dev 2011;21:194–8.
[43] Iyer MK, Chinnaiyan AM. RNA-seq unleashed. Nat Biotechnol 2011;29:599–600.
[44] Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007;129:1311–23.
[45] Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA 2009;106:11667–72.
[46] Koziol MJ, Rinn JL. RNA traffic control of chromatin complexes. Curr Opin Genet Dev 2010;20:142–8.
[47] Faulkner GJ, Carninci P. Altruistic functions for selfish DNA. Cell Cycle 2009;8:2895–900.
[48] Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465:1033–8.
[49] Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell 2011;147:344–57.
[50] Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. Cell 2007;131:706–17.
[51] Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, et al. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci USA 2011;108:1058–63.
[52] van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 2011;475:114–7.
[53] Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 2011;469:368–73.
[54] Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, et al. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet 2011;7:e1002329.
[55] Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci USA 2008;105:716–21.
[56] Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007;316:1484–8.
[57] Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 2008;456:130–4.
[58] Wilson BA, Masel J. Putatively noncoding transcripts show extensive association with ribosomes. Genome Biol Evol 2011;3:1245–52.
[59] Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 2011;147:789–802.
[60] Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AM, et al. The human mitochondrial transcriptome. Cell 2011;146:645–58.
[61] Pontier DB, Gribnau J. Xist regulation and function explored. Hum Genet 2011;130:223–36.
[62] Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. Characterization of the RNA content of chromatin. Genome Res 2010;20:899–907.
[63] Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. Mol Cell 2011.
[64] Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 2008;18:1433–45.
[65] Amaral PP, Mattick JS. Noncoding RNA in development. Mamm Genome 2008;19:454–92.
[66] Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol 2011;29:742–9.
[67] Kaufmann BB, van Oudenaarden A. Stochastic gene expression: from single molecules to the proteome. Curr Opin Genet Dev 2007;17:107–12.
[68] Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, et al. Direct RNA sequencing. Nature 2009;461:814–8.
[69] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 2011;29:644–52.
[70] Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol 2011;21:354–61.
[71] Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol 2011, doi:10.1038/nbt.2024.