Biochimica et Biophysica Acta 1859 (2016) 16–22

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

The specificity of long noncoding RNA expression☆

Brian S. Gloss, Marcel E. Dinger *

Division of Genomics and Epigenetics, Garvan Institute of Medical Research, Sydney, Australia
St Vincent's Clinical School, Faculty of Medicine, UNSW Australia

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form 12 August 2015
Accepted 12 August 2015
Available online 19 August 2015

Keywords:
Long noncoding RNA
Expression
Technology
Transcriptomics

ABSTRACT

Over the last decade, long noncoding RNAs (lncRNAs) have emerged as a fundamental molecular class whose members play pivotal roles in the regulation of the genome. The observation of pervasive transcription of mammalian genomes in the early 2000s sparked a revolution in the understanding of information flow in eukaryotic cells and the incredible flexibility and dynamic nature of the transcriptome. As a molecular class, distinct loci yielding lncRNAs are set to outnumber those yielding mRNAs. However, like many important discoveries, the road leading to uncovering this diverse class of molecules that act through a remarkable repertoire of mechanisms, was not a straight one. The same characteristic that most distinguishes lncRNAs from mRNAs, i.e. their developmental-stage, tissue-, and cell-specific expression, was one of the major impediments to their discovery and recognition as potentially functional regulatory molecules. With growing numbers of lncRNAs being assigned to biological functions, the specificity of lncRNA expression is now increasingly recognized as a characteristic that imbues lncRNAs with great potential as biomarkers and for the development of highly targeted therapeutics. Here we review the history of lncRNA research and how technological advances and insight into biological complexity have gone hand-in-hand in shaping this revolution. We anticipate that as increasing numbers of these molecules, often described as the dark matter of the genome, are characterized and the structure–function relationship of lncRNAs becomes better understood, it may ultimately be feasible to decipher what these non-(protein)-coding genes encode. This article is part of a Special Issue entitled: Clues to long noncoding RNA taxonomy1, edited by Dr. Tetsuro Hirose and Dr. Shinichi Nakagawa.

1. Technological history of RNA biology

The ability to measure gene transcription provided access to a fundamental molecular layer in the transaction of genomic information into biological function. As with many other sciences, the emerging understanding of these processes has been driven by technological advances that built on and addressed limitations of precursors (Table 1). The Northern blot first allowed researchers to identify the size and approximate expression levels of RNA species in the 1970s [1]. Advances in cloning and the advent of the polymerase chain reaction in the 1980s [2–4] facilitated an explosion in the characterization of RNAs. These candidate approaches were eclipsed in the late 1990s with the development of cDNA arrays, which for the first time allowed researchers a glimpse into the expression of thousands of RNAs in a single experiment. Researchers using arrays quickly realized major limitations — particularly imprecision at low transcript copy numbers and dependence on *a priori* knowledge of transcript sequence. Massively parallel sequencing technology fundamentally addressed this latter limitation and drastically improved the former. Two landmark projects (using arrays [5] and then sequencing [6]) made a number of surprising observations, including (i) that most of the genome was transcribed into RNA, the majority with little or no protein coding potential, (ii) that alternate transcript isoforms arising from alternate splicing and alternate promoters was prevalent throughout the genome and (iii) that this previously undetected transcriptional complexity often occurred rarely or at low levels. Together, these observations prompted a re-evaluation of our understanding of the information encoded in the genome, how transcription was regulated, and challenged existing concepts of the definition of the gene [7,8]. Indeed, later evidence of a continuum of protein coding/transcript-driven function within the transcriptome [9] further challenged these definitions [10]. Despite continued controversy over the biological significance of these new perspectives on the transcriptome, the advent of next generation sequencing has challenged numerous long-standing assumptions in the role of RNA across all life and has led to profound new understandings of the regulatory functions of RNA, including the discovery of entirely new functional classes of RNA.

2. Biological history of RNA technology

The main role of RNA in cells has historically been divided into noncoding and coding functions. The most ubiquitous RNA species in the cell are noncoding, mostly comprising transfer and ribosomal RNA.

Abbreviations: lncRNA, long noncoding RNA.

☆ This article is part of a Special Issue entitled: Clues to long noncoding RNA taxonomy1, edited by Dr. Tetsuro Hirose and Dr. Shinichi Nakagawa.

* Corresponding author at: Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst, NSW 2010, Australia.
E-mail address: m.dinger@garvan.org.au (M.E. Dinger).

http://dx.doi.org/10.1016/j.bbagrm.2015.08.005
1874-9399/

Table 1
Comparison of technologies described in this review.

Technique	Transcript and transcriptome coverage	Dependence on gene annotation	Sensitivity	Comments
Northern blot	Single region of gene but gives transcript size, not multiplexed	Yes	High	Low throughput
RT-PCR	Single region of gene, easily multiplexed	Yes	Down to single copy	Moderate throughput, high precision
Expression arrays	Multiple regions of tens of thousands of genes	Yes	Poor at low copy	Comprehensive but subject to evolving annotation, low precision
Tiling arrays	Full transcript, up to the whole transcriptome	No	Poor at low copy	Comprehensive but low precision
Short-read RNA-seq	Full transcript, up to whole transcriptome	No; limitations in de-novo assembly	Dependent on depth, typically (~30E6 reads) superior to arrays	Current standard, flexible and improved precision
Short read RNA CaptureSeq	Full or partial transcript, defined by allowable capture space	No; limitations in de-novo assembly	Orders of magnitude greater than RNA-seq, also depends on depth	Excellent for identifying and defining low-copy transcripts
HITS-CLIP	Captures small area of transcript binding specific protein	Partial; small fragment gives little insight into transcript structure	Greater than RNA-seq, depends on # copies bound to protein	Superior for characterizing RNA-protein interactions
Long read SMRT-sequencing	Up to full transcript in one read, up to whole transcriptome	No; excellent for transcript assembly, improving read accuracy	Currently dependent on depth and cost-benefit	Offers full-length transcripts
Single cell sequencing	Full transcript, up to whole transcriptome	No; limitations in de-novo assembly	Currently poor at low copy	Captures cell specific transcripts
RNA FISH	Multiple regions of single genes	Yes	Down to single copy	High precision, low throughput; provides subcellular localization

These “structural” RNAs are highly evolutionarily conserved, and occur in all known forms of life. Since the discovery of genes as a principle unit of heritability, the concept of RNA as a signaling molecule has typically been dominated by understanding that messenger RNA (mRNA) transduced information from DNA for translation into protein (thus protein coding). This understanding of RNA biology has dominated molecular biology for decades, with specific coding RNA providing a cellular identity through protein, and ubiquitous noncoding RNA providing a structural basis for basic cellular functions [11]. Discoveries of other ubiquitous noncoding RNA such as XIST (stabilizing X chromosome inactivation) reinforced this picture.

The ancient origins and central role of RNA in the evolution of life [12] prompt the idea that RNA continued to evolve functions in parallel to its more obvious and ubiquitous roles as core components of the ribosome (as opposed to purely messenger functions). This hypothesis predicts diverse functions for noncoding RNAs utilizing the extensive repertoire of biochemical processes that RNA can participate in to influence cellular behavior. Indeed early characterization of small noncoding RNAs known as microRNAs (miRs), which modulate gene expression by a variety of common pathways, showed that the expression of microRNAs more accurately defined the tissue of origin in cancer than coding RNAs [13]. The implication of noncoding RNA as a major determining factor of cellular phenotypic identity challenged the central dogma of molecular biology, where cell phenotype was driven by the process of DNA → RNA → protein [14]. These findings hinted at a higher order regulatory architecture for controlling gene expression at the transcript level, and that RNA was able to play the role of message, target and regulator.

3. The convergence of technology and biology

Historically, technology has driven scientific endeavor as much as understanding has shaped technology (Box 1). Much like revolutions in optical technology drove revolutions in the understanding of the very small (microbiology) and the very distant (astronomy), technological developments in molecular biology are driving our view of the sophistication and prevalence of DNA transcription.

Akin to the incredulity that traditionally faced new technologies in other scientific domains, early evidence of pervasive transcription of the mammalian genome as noncoding RNA led to comparisons with evidence of pervasive dark matter in the universe [15–17]. Despite extensive subsequent experimental evidence [5,6,18,19], the observation of pervasive transcription is still met with dismissal of the phenomena as a technical artifact, transcriptional noise or leaky transcription of nonfunctional genome regions [20]. The main reasons for this can be attributed to observed low levels of transcription (Fig. 1A); sometimes calculated at less than one copy per cell (e.g. from the ENCODE studies [21]), appearing stochastic by some measures, and an apparent lack of sequence conservation of these regions [11]. However, in parallel, growing numbers of studies showed that such conclusions were poorly substantiated [22]. These investigations revealed that long noncoding RNA (lncRNA; operationally defined as >200 nt transcripts lacking long open reading frames) species showed highly specific spatial and temporal expression patterns [23,24] and were often regulated in the same fashion at the (epi)genome [25,26] and RNA processing levels [27,28] as their protein coding cousins.

The apparent lack of sequence conservation for lncRNAs remains a heated area of debate. Some researchers question whether this lack argues against functionality [29], while others argue for structural conservation [30,31] and highlight how noncoding transcription scales with organismal complexity [32] and that lack of conservation does not necessarily imply lack of functionality [33]. Indeed analyses of RNA structural conservation present evidence for an ancient and plastic evolutionary process at the RNA level [34]. Regardless of the controversy surrounding lncRNA as a class (including long intergenic ncRNA (lincRNA), intronic, antisense & processed transcripts, emerging types such as enhancer associated RNAs [35] as well as repressive [36] & activating [37] promoter transcripts; with other types likely to still be described [11]) and the relative few that have been functionally studied and annotated (Fig. 1B) [38], the pioneering work in the discovery and classification of lncRNAs has resulted in a veritable explosion of research into – and recognition of – these molecules [38] (Fig. 2). Much of this work is allowing researchers to redefine molecular biology paradigms that have dominated the zeitgeist for decades [39].

An area of controversy and experimental challenge facing researchers studying lncRNA revolves around ascertaining the biological relevance of low expression levels and infrequent splicing events [26,40]. The observation of low expression of lncRNAs is commonly oversimplified by assuming that all cells in a population, even in complex tissues, possess uniform transcriptomes. Although this assumption is practical for differential expression analyses, it is biologically unlikely to be the case. There are a variety of scenarios that could yield the observed average low expression levels of lncRNAs, such as small numbers of cells with high expression or transient bursts of high expression (Fig. 3). With the applications of new sequencing technologies and

Box 1
Technology driven discovery.

Technological and Philosophical Period	The Universe	The Microcosm	The Noncoding Transcriptome
Medieval *(Historical)*	By eye: ~6,000 visible stars	By eye: circumstantial evidence of microbes	Northern hybridization: highly expressed ubiquitous ncRNAs
Renaissance *(Transformative but mostly observational)*	Binoculars: 30,000 stars	Microscope: near-ubiquity of diverse single-celled organisms	Arrays: pervasive noncoding transcription, active chromatin marks in intergenic space
Enlightenment *(Allows better understanding but highlights limitations)*	Telescopes: millions of stars, Galaxies, red-shift	Optical advances: remarkable diversity and subcellular structure of microbes	RNA-Seq: much higher resolution of gene-structure
Industrial Revolution *(Addresses limitations, hints at greater complexity)*	Mt Wilson observatory, the cosmic microwave background	Electron microscopy: viruses, mitochondria X-ray crystallography /atomic force microscopy: macro/micromolecular structures	RNA-Capture-Seq: complexity of gene structure, specificity of expression, and prevalence of transcription
Information Revolution *(Transformative and observational again)*	Hubble deep field, square kilometer array, planck cosmology probe	Metagenomics: millions of previously uncultured/unstudied microbes.	Future: real-time, full-length single-molecule-resolution transcriptomics

applications, it is becoming increasingly apparent that the lncRNA transcriptome is at least as large and complex as the coding transcriptome and that gene-splicing plays a huge role in the cell (transcript numbers in Fig. 2) [41]. Thus, as understanding and technology are progressing, lncRNAs are taking a center stage in understanding the molecular basis underlying the complexity of life, particularly complex life.

4. lncRNA expression is often tissue- and cell-type specific

Much of the pioneering work in characterizing lncRNAs was focused on differential expression studies. These studies revealed a surprising degree of tissue-type and developmental-stage specificity of lncRNA transcription [6,23,24,42,43]. And in recent years, the specificity of lncRNA expression has surpassed that of protein coding transcripts [44,45]. This characteristic of lncRNA specificity can be illustrated using the Illumina Body Atlas RNA sequencing data (Fig. 1B). Half of lncRNA genes are expressed in fewer than half of the tissue types tested, whereas protein-coding genes are much more frequently expressed in different tissue types. Indeed, the specific cell type specificity of lncRNA has been shown most elegantly in complex (brain [24], testes [44,46]) and primordial (stem cells [23]) systems (see Table 2). These observations suggest that lncRNA transcription is intimately linked with complex and plastic cell phenotypes [47].

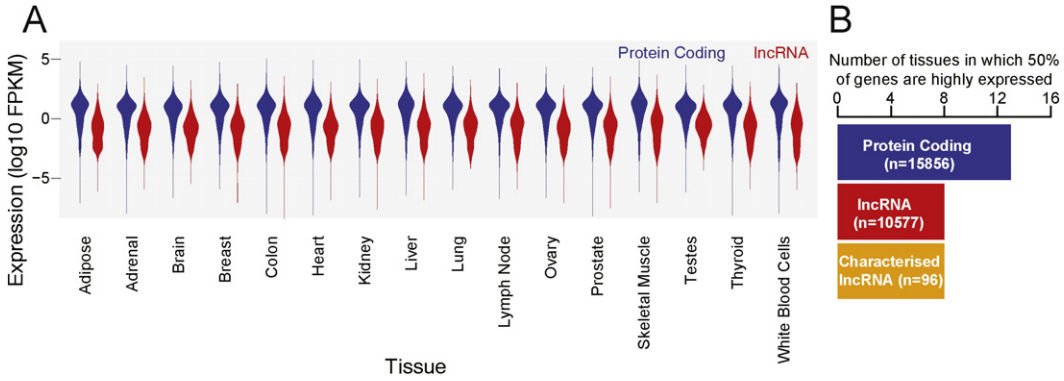

Fig. 1. lncRNAs are more lowly expressed (A) and more tissue specific (B) than protein coding genes. (A) Illumina body atlas data was aligned to hg19 with GENCODE 18 gene annotations. For each tissue, FPKM values for protein coding and lncRNAs (lincRNA and antisense biotypes) classes were compared; genes with FPKMs of 0 were omitted. The corresponding median expression of protein coding genes for each tissue is, on average, ~40-fold (range: 22 to 59-fold) higher than lncRNAs. (B) For each class, median expression for each tissue was calculated and genes were binned as highly (above median) or lowly expressed in each tissue. Genes were then ranked by number of tissues in which they were highly expressed and the median was plotted for protein coding (blue) and lncRNA (red) genes, compared to annotated functional lncRNAs (yellow; from lncRNAdb v2).

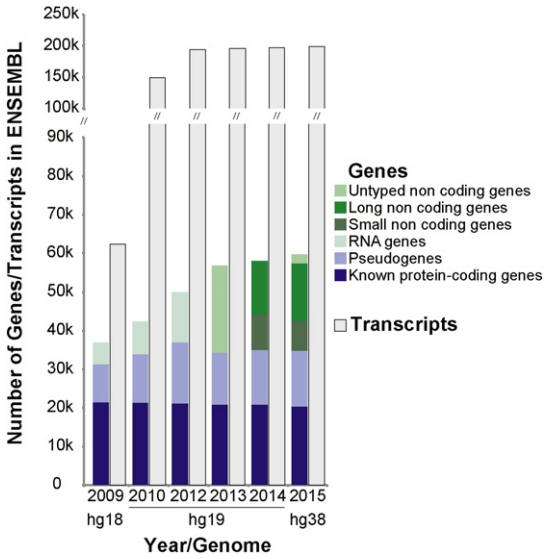

Fig. 2. The rise of noncoding RNA: ENSEMBL gene annotations from 2009 (version 54) to present (version 79) are presented broken down by designated gene type and compared to the observed number of unique transcripts with each compilation.

5. lncRNA expression is highly dynamic

Not only do lncRNAs exhibit cell type specificity, but also it is becoming clear that their expression can be highly context and time dependent. That transcriptomes are subject to large-scale variation in even well controlled systems is not a new idea and the impact of handling and conditions on cellular transcriptomes can be far reaching [48]. This highlights the dynamic nature of transcriptomes and argues against assumptions of steady state gene expression; a typical, yet flawed assumption (Fig. 3) commonly made when interpreting results from these technologies. Temporal and cell cycle specific processes have long been studied in specific contexts, but the potential impact in broader contexts have, until recently, been difficult to access and not well considered. Certain lncRNAs, such as HOTTIP [49], have been shown to exert an extremely short-lived function (providing a scaffold to initiate assembly of a chromatic remodeling complex) necessitating only a brief burst of expression. More broadly, lncRNAs, including those with known cellular functions, have on average significantly shorter transcript half-lives than protein coding genes [50–52] (Table 3). Together, these observations highlight how much of lncRNA expression patterns remain to be elucidated. Interestingly, the interaction of U1snRNP and promoter-proximal poly(A) sites of a transcript has been shown to be a critical determinant of transcript stability in preventing spurious bidirectional transcription [53]; this may provide valuable insight into the longevity of transcripts, particularly in the nucleus.

6. lncRNAs are commonly lowly expressed

Despite advances in the understanding and identification of lncRNAs in various systems, lncRNAs remain difficult to detect with conventional sequencing approaches [54]. This has two outcomes; that existing studies may not necessarily be considered comprehensive and that identification of novel transcripts is an ongoing and improving process. The collateral of these limitations is that despite the growing interest in lncRNA expression patterns, most published RNA-seq investigations remain focused on protein coding genes.

In addressing issues of comprehensiveness, observed low levels of transcription are currently considered to be due to the cellular specificity that is becoming a hallmark of these molecules, and due to the highly dynamic nature of their expression and action, most existing techniques of transcriptomic enquiry lack the power to effectively elucidate their expression. Indeed, current whole transcriptome sequencing approaches still lack the breadth in dynamic range to probe lncRNA expression [54], and researchers are turning to ever deeper sequencing or to emerging complementary techniques (Table 1) to address this limitation.

A promising technique for overcoming the high dynamic range of RNA expression, especially in the characterization of the extent of lncRNA expression, is RNA-capture sequencing (RNA CaptureSeq) [41,55,56]. Using a similar approach to exon capture sequencing, RNA CaptureSeq employs oligomer baits to enrich a sequencing library (in this case cDNA) for sequences of interest before sequencing. The technique yields, on average, a 2–3 $\log_{10}$ increase in signal over traditional sequencing, thereby providing a vital tool for the lowly and specifically expressed lncRNAs present in more common bulk RNA samples. RNA CaptureSeq studies have shown that the complexity of lncRNA and even coding RNA expression, especially in RNA processing patterns, is only starting to be appreciated. Indeed a single RNA CaptureSeq experiment in 20 tissue samples reported an increase of 20.6% in observed long-noncoding exons as well as a 13.5% increase in splice junctions than had previously been annotated — with similar improvements on existing coding

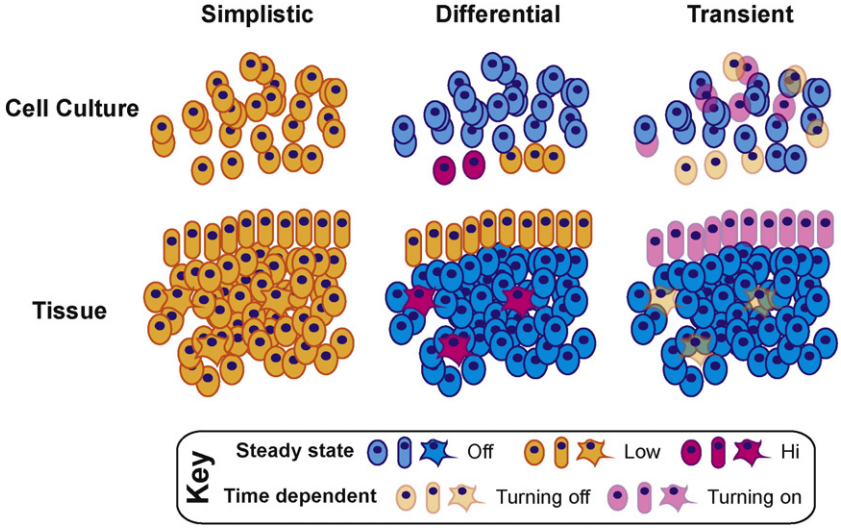

Fig. 3. Comparison of models of transcriptional activity in cell culture and tissue. For any given gene, a low (orange) signal is often interpreted as low transcription in all cells contained in that sample (simplistic). However, the same value would be measured if the gene were off in most but low or high in others (differential) or in the process of switching on or off in some (transient). This is even more likely in tissue samples where more complex cell type mixtures are likely to be present.

Table 2
Specificity of lncRNA expression.

Study	Tissues	Specificity	Other Observations
Mercer 2008	Brain	64% (849/1428) lncRNAs are expressed • 60 Ubiquitous • 513 Specific • 276 Diffuse	RNA structural conservation observed In Purkinje Cells (88) • 29% Nuclear • 61% Speckle • 10% Diffuse
Dinger 2008	Mouse ES cells	984 ncRNAs expressed during preimplantation differentiation 174 Differentially expressed • 12: pluripotency • 7: primitive streak • 31: mesoderm	More positively correlated to neighbouring protein coding genes • 338 Intronic • 61 bidirectional 36 cis-antisense (not more correlated)
Cabili 2011	24 Tissues	8000 lincRNAs More are tissue specific (78%) than protein coding (~19%) Enriched in testes	Co-expressed with neighbouring protein coding genes Some have potential coding portions
Washietl 2014	9 Tissues	1898 lncRNAs expressed Enriched in testes and brain	Small number with potential coding potential Decreasing fraction of orthologous regions with evolutionary divergence Orthologues retain tissue specificity Conserved promoter domains

annotations. Significantly, this study doubled the observed alternative splicing per locus from 1.8 to 3.6 [41], highlighting the transformative nature of this technology. That "bait" oligonucleotides can be tiled over large genomic regions gives researchers significant power to resolve lncRNA expression and sequence. However, the total capture space allowed and design and analytic techniques still present limitations owing to the relative early days of the technology. Complementary techniques such as HITS-CLIP [57] and RIP-Seq [58] address some of these limitations by employing the protein-binding nature of some lncRNAs to enrich transcript-sequencing pools for transcripts interacting with specific protein complexes and thus also can overcome the low copy-number limitation for studying lncRNAs. These methodologies, despite being limited to the specific protein being assayed, also provide crucial insight into how expressed lncRNAs may exert a biological effect.

Table 3
Dynamics of lncRNA expression.

Study	Method	Observations on stability
Clark 2012	Actinomycin D RNA polymerase inhibition Custom array analysis Time course	12,670 transcripts (823 lncRNAs)- lncRNA half-life is less than protein coding (median 3.5 vs. 5.1 h)- 240 lncRNAs with half life <2 h (including Neat1)- 51 lncRNAs with high stability (including Zfas1)- lncRNAs not overlapping protein coding genes are more stable- unstable lncRNAs are enriched in the nucleus- Decay rates are not consistent between human and mouse
Tani 2012	5′-bromo-uridine pulse Immunoprecipitation seq Time course	11,052 mRNAs + 1418 ncRNA (highly expressed)- Identification of 795 short lived noncoding transcripts based on observed short half life of regulatory lncRNAs- Unstable lncRNAs enriched in nucleus Some validated in Wang et al. 2014 (actD validation)- 10 lncRNAs half-life > 4 h- 7 lncRNAs half-life <4 h

This increasing complexity highlights another active area of research, namely the accuracy of transcript identification. The vast majority of lncRNA research has used short read sequencing, and the most common tool for assembling novel transcripts is the cufflinks [59] or *trinity* [60] tools. Researchers familiar with these tools will be aware that confidence in *de novo* assembled transcripts drops alarmingly at low levels [61]. Recent advances in the computational tools available have improved this level of confidence but the central limitation remains. Short read sequencing of low abundance transcripts does not appear to adequately capture or define these transcripts. More comprehensive transcript characterization methods such as 3′ & 5′ RACE and cDNA sequencing lack the throughput for comprehensive coverage, especially for lowly expressed genes. It is still widely considered that even with large-scale studies such as ENCODE [62,63], the extent and complexity of lncRNA transcription remain to be defined [55]. Excitingly, long-read sequencing such as the PacBio platform [64] offers great promise for resolving this shortcoming and comprehensively cataloging full-length coding and noncoding transcripts.

7. Resolving lncRNA specificity

Returning to the notion of technology driving innovation, it is clear that, while great inroads have been made into understanding the role of lncRNAs in determining cellular phenotype, much remains to be done. Indeed, it is apparent that the very hallmarks of lncRNA expression – cell and subcellular specificity, rapid action, short duration of existence and complex splicing patterns – demand that greater precision and comprehensiveness of investigations will be required for its adequate characterization.

Future research will require substantial effort to elucidate the precise molecular contexts in which lncRNA expression occurs. The Roadmap Epigenome [65] will provide an invaluable chromatin map to begin to investigate this in different cell states. Elucidation of the mechanism of how lncRNA interacts with and guides chromatin states [66] will refine the understanding of this roadmap. Fine temporal resolution will be required to capture short lived and cell-state/cell-cycle specific transcripts. The complementary RNA-Seq and RNA CaptureSeq, possibly in combination with long read sequencing techniques, will continue to bring the lower expressed transcripts to light. Single molecule RNA FISH, allowing subcellular localization of single RNA molecules can [67] and will continue to be required to probe very lowly expressed and cell or subcellular specific transcripts. Interestingly, this technique demonstrated – in cell culture – that observed low levels of lncRNA expression can indeed be due to uniformly lower numbers of RNA molecules in cells. Thus the characterization of the nature of lncRNA specificity will likely enable an improved understanding of the functional importance of these molecules.

8. Resolving lncRNA function

Currently, a significant challenge in understanding the specificity of lncRNA transcription is that the vast majority of lncRNAs remain to be functionally characterized. Indeed, at the time of writing, of the 15,900 annotated human lncRNA genes in GENCODE version 22 [40,68], less than 1% have been individually characterized [38]. Those that have been examined display the same signatures of low primary sequence conservation and tissue- and cell-type-specificity observed in the broader pool of lncRNAs (Fig. 1B). This suggests that lncRNA function is intimately tied to the cellular or tissue context in which they are expressed [69]. Therefore, high-throughput screens that have enabled large-scale inroads in understanding coding transcript function, may not be as fruitful in understanding lncRNA function because these methods are dependent on the cell-line being used and the assay for function being robust. Moreover, as lncRNAs are typically cell-type-specific and often exert functions with a subtle phenotype that may be short-lived and highly condition-specific, existing high throughput methods may be limited in defining lncRNA function. However, ongoing improvements in molecular tools for lncRNAs (such as the Dharmacon Lincode siRNA library) as well as advances in high content imaging and software will bring more analytical power to these investigations and the functional characterization of lncRNAs will continue. A central challenge to overcome will lie in the designs required to biochemically identify lncRNA function, which can be transacted in many and varied ways.

Furthermore, it is vitally important that the structure–function relationship of lncRNAs is elucidated fully in order to facilitate the functional characterization of their biological roles. It is known that lncRNAs display conservation of structural domains [30,33] but how these relate to function is unclear. Some lncRNAs act as “guides” to direct or provide specificity to protein–protein and DNA–protein interactions, whereas others acts as more generic scaffolds for such interactions [70]. It is also important to note that the coding–noncoding distinction is not yet fully resolved, with identification of functional short peptides [71], evidence of designated lncRNAs encoding peptides [72] and observation of transcript-mediated activity of protein coding genes [73]. Understanding the distinction between transcript- and protein-mediated effects [9,10] and elucidating the nature of RNA structures and interactions will be essential in defining how lncRNAs exert this function [11], which will ultimately enable prediction of lncRNA function based on sequence. Insight into this relationship remains elusive and will likely depend on further technological advances to resolve [74].

The field of lncRNA biology is still in its infancy and research to date has been predominantly observational, focusing largely on mapping the landscape and cherry-picking only those lncRNAs with the most distinct expression patterns. In trying to unravel insights into lncRNA biology, screening studies employ complex biological systems (such as brain or development) to maximize the pool of potentially functional lncRNAs and seek to assign putative functions using expression with genes of known function. This approach simultaneously identifies candidate molecules and a system in which to examine their role. This approach is valuable because it raises the potential for moderate throughput functional screening, a process more amenable to dissecting the specific functions of the transcript. The CRISPR-Cas9 system [75] promises unprecedented power in the systematic functional testing of lncRNAs [76] and will likely gain more traction as the tool of choice for characterizing the function of lncRNAs [77].

9. The future

As our understanding of the specificity and complexity of the lncRNA transcriptome increases, new technologies will be required to appropriately probe their patterns of expression. Spatial-transcriptomes, where whole tissue mounts can be applied to high-throughput RNA sequencing in a space-dependent fashion [78] offers an unprecedented chance to delve into the complexity of tissue-specific transcription. Single cell RNA sequencing is already beginning to unravel differences between individual cells in complex populations (e.g. cancer [79]), but thus far lacks the precision at low expression levels [80,81] to adequately probe lncRNA expression. Single cell qPCR arrays may aid this process, but current annotations and continuing identification of novel noncoding transcripts suggests that this technique lacks the throughput to systematically probe the complex and widespread transcription patterns beginning to emerge. Understanding the regulatory architecture surrounding lncRNA transcription, processing, actions and turnover is likely to shed light on how this vast transcriptional network is orchestrated in cells. Perhaps, most excitingly, these new technologies will hint at yet more elegant and complex processes underlying biology. It is likely that, as it has before, technology and understanding will converge and a wider survey of lncRNA function and regulation across an organism's tremendous repertoire of cellular phenotypes will be feasible. When such studies become possible, we should again be prepared to challenge and reevaluate existing assumptions and dogmas of genome biology.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

BG is supported by Cancer Institute NSW Early Career Fellowship 13/ECF/1-45.

References

[1] J.C. Alwine, et al., Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5350–5354.
[2] R.K. Saiki, et al., Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science (New York, N.Y.) 230 (1985) 1350–1354.
[3] K.B. Mullis, Target amplification for DNA analysis by the polymerase chain reaction, Ann. Biol. Clin. 48 (1990) 579–582.

[4] P.M. Holland, et al., Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 7276–7280.
[5] P. Kapranov, et al., Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays, Genome Res. 15 (2005) 987–997.
[6] P. Carninci, et al., The transcriptional landscape of the mammalian genome, Science (New York, N.Y.) 309 (2005) 1559–1563.
[7] M.B. Gerstein, et al., What is a gene, post-ENCODE? History and updated definition, Genome Res. 17 (2007) 669–681.
[8] P. Kapranov, et al., Genome-wide transcription and the implications for genomic organization, Nat. Rev. Genet. 8 (2007) 413–423.
[9] M.E. Dinger, et al., The evolution of RNAs with multiple functions, Biochimie 93 (2011) 2013–2018.
[10] M.E. Dinger, et al., Differentiating protein-coding and noncoding RNA: challenges and ambiguities, PLoS Comput. Biol. 4 (2008), e1000176.
[11] K.V. Morris, J.S. Mattick, The rise of regulatory RNA, Nat. Rev. Genet. 15 (2014) 423–437.
[12] P.G. Higgs, N. Lehman, The RNA World: molecular cooperation at the origins of life, Nat. Rev. Genet. 16 (2015) 7–17.
[13] J. Lu, et al., MicroRNA expression profiles classify human cancers, Nature 435 (2005) 834–838.
[14] J.S. Mattick, I.V. Makunin, Non-coding RNA, Hum. Mol. Genet. 15 (Spec No 1) (2006) R17–R29.
[15] P. Kapranov, et al., Large-scale transcriptional activity in chromosomes 21 and 22, Science (New York, N.Y.) 296 (2002) 916–919.
[16] Y. Okazaki, et al., Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs, Nature 420 (2002) 563–573.
[17] J.L. Rinn, et al., The transcriptional activity of human chromosome 22, Genes Dev. 17 (2003) 529–540.
[18] J. Cheng, et al., Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, Science (New York, N.Y.) 308 (2005) 1149–1154.
[19] S. Katayama, et al., Antisense transcription in the mammalian transcriptome, Science (New York, N.Y.) 309 (2005) 1564–1566.
[20] H. van Bakel, et al., Most "dark matter" transcripts are associated with known genes, PLoS Biol. 8 (2010), e1000371.
[21] A.F. Palazzo, T.R. Gregory, The case for junk DNA, PLoS Genet. 10 (2014), e1004351.
[22] M.B. Clark, et al., The reality of pervasive transcription, PLoS Biol. 9 (2011) e1000625 (discussion e1001102).
[23] M.E. Dinger, et al., Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation, Genome Res. 18 (2008) 1433–1445.
[24] T.R. Mercer, et al., Specific expression of long noncoding RNAs in the mouse brain, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 716–721.
[25] J.L. Rinn, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell 129 (2007) 1311–1323.
[26] M. Guttman, et al., Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, Nature 458 (2009) 223–227.
[27] S. Cawley, et al., Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs, Cell 116 (2004) 499–509.
[28] P. Johnsson, et al., A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells, Nat. Struct. Mol. Biol. 20 (2013) 440–446.
[29] C.M. Rands, et al., 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage, PLoS Genet. 10 (2014) e1004525.
[30] M.A. Smith, et al., Widespread purifying selection on RNA structure in mammals, Nucleic Acids Res. 41 (2013) 8220–8236.
[31] P. Johnsson, et al., Evolutionary conservation of long non-coding RNAs; sequence, structure, function, Biochim. Biophys. Acta 1840 (2014) 1063–1071.
[32] R.J. Taft, et al., The relationship between non-protein-coding DNA and eukaryotic complexity, BioEssays 29 (2007) 288–299.
[33] I. Ulitsky, et al., Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution, Cell 147 (2011) 1537–1550.
[34] A. Nitsche, et al., Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved, RNA 21 (5) (2015) 801–812.
[35] E.M. Darrow, B.P. Chadwick, Boosting transcription by transcription: enhancer-associated transcripts, Chromosom. Res. 21 (2013) 713–724.
[36] J. Han, et al., Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 12422–12427.
[37] N. Hamazaki, et al., Gene activation-associated long noncoding RNAs function in mouse preimplantation development, Development 142 (2015) 910–920.
[38] X.C. Quek, et al., lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs, Nucleic Acids Res. 43 (2015) D168–D173.
[39] M.E. Dinger, et al., Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications, Brief. Funct. Genomic. Proteomic. 8 (2009) 407–423.
[40] T. Derrien, et al., The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Res. 22 (2012) 1775–1789.
[41] M.B. Clark, et al., Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing, Nat. Methods 12 (4) (2015) 339–342.
[42] T. Ravasi, et al., Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome, Genome Res. 16 (2006) 11–19.
[43] K.C. Pang, et al., Genome-wide identification of long noncoding RNAs in CD8+ T cells, J. Immunol. 182 (2009) 7738–7748.
[44] M.N. Cabili, et al., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses, Genes Dev. 25 (2011) 1915–1927.
[45] S. Djebali, et al., Landscape of transcription in human cells, Nature 489 (2012) 101–108.
[46] S. Washietl, et al., Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals, Genome Res. 24 (2014) 616–628.
[47] T.R. Mercer, et al., Noncoding RNAs in long-term memory formation, Neuroscientist 14 (2008) 434–445.
[48] H. Dvinge, et al., Sample processing obscures cancer-specific alterations in leukemic transcriptomes, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 16802–16807.
[49] K.C. Wang, et al., A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression, Nature 472 (2011) 120–124.
[50] M.B. Clark, et al., Genome-wide analysis of long noncoding RNA stability, Genome Res. 22 (2012) 885–898.
[51] H. Tani, et al., Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals, Genome Res. 22 (2012) 947–956.
[52] L. Wang, et al., Exploring the stability of long intergenic non-coding RNA in K562 cells by comparative studies of RNA-Seq datasets, Biol. Direct 9 (2014) 15.
[53] A.E. Almada, et al., Promoter directionality is controlled by U1 snRNP and polyadenylation signals, Nature 499 (2013) 360–363.
[54] D. Sims, et al., Sequencing depth and coverage: key considerations in genomic analyses, Nat. Rev. Genet. 15 (2014) 121–132.
[55] T.R. Mercer, et al., Targeted RNA sequencing reveals the deep complexity of the human transcriptome, Nat. Biotechnol. 30 (2012) 99–104.
[56] A. Roberts, L. Pachter, RNA-Seq and find: entering the RNA deep field, Genome Med. 3 (2011) 74.
[57] D.D. Licatalosi, et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing, Nature 456 (2008) 464–469.
[58] J. Zhao, et al., Genome-wide identification of polycomb-associated RNAs by RIP-seq, Mol. Cell 40 (2010) 939–953.
[59] C. Trapnell, et al., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515.
[60] M.G. Grabherr, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.
[61] H. Inamdar, et al., Rule-based integration of RNA-Seq analyses tools for identification of novel transcripts, J. Bioinforma. Comput. Biol. 12 (2014) 1450026.
[62] K.R. Rosenbloom, et al., ENCODE whole-genome data in the UCSC Genome Browser: update 2012, Nucleic Acids Res. 40 (2012) D912–D917.
[63] E.P. Consortium, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816.
[64] J. Eid, et al., Real-time DNA sequencing from single polymerase molecules, Science (New York, N.Y.) 323 (2009) 133–138.
[65] B.E. Bernstein, et al., The NIH Roadmap Epigenomics Mapping Consortium, Nat. Biotechnol. 28 (2010) 1045–1048.
[66] F.P. Marchese, M. Huarte, Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code, Epigenetics 9 (2014) 21–26.
[67] M.N. Cabili, et al., Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution, Genome Biol. 16 (2015) 20.
[68] J. Harrow, et al., GENCODE: the reference human genome annotation for The ENCODE Project, Genome Res. 22 (2012) 1760–1774.
[69] J.S. Mattick, et al., A global view of genomic information—moving beyond the gene and the master regulator, Trends Genet. 26 (2010) 21–28.
[70] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, Annu. Rev. Biochem. 81 (2012) 145–166.
[71] A. Pauli, et al., Toddler: an embryonic signal that promotes cell movement via Apelin receptors, Science (New York, N.Y.) 343 (2014) 1248636.
[72] D.K. Gascoigne, et al., Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes, Bioinformatics 28 (2012) 3042–3050.
[73] L. Poliseno, et al., A coding-independent function of gene and pseudogene mRNAs regulates tumour biology, Nature 465 (2010) 1033–1038.
[74] C.K. Kwok, et al., The RNA structurome: transcriptome-wide structure probing with next-generation sequencing, Trends Biochem. Sci. 40 (2015) 221–232.
[75] L. Cong, et al., Multiplex genome engineering using CRISPR/Cas systems, Science (New York, N.Y.) 339 (2013) 819–823.
[76] T.T. Ho, et al., Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines, Nucleic Acids Res. 43 (2015), e17.
[77] J. Peng, et al., High-throughput screens in mammalian cells using the CRISPR-Cas9 system, FEBS J. 282 (11) (2015) 2089–2096.
[78] J.H. Lee, et al., Highly multiplexed subcellular RNA sequencing in situ, Science (New York, N.Y.) 343 (2014) 1360–1363.
[79] A.P. Patel, et al., Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma, Science (New York, N.Y.) 344 (2014) 1396–1401.
[80] P.V. Kharchenko, et al., Bayesian approach to single-cell differential expression analysis, Nat. Methods 11 (2014) 740–742.
[81] J. Eberwine, et al., The promise of single-cell sequencing, Nat. Methods 11 (2014) 25–27.