Research Techniques Made Simple: Identification and Characterization of Long Noncoding RNA in Dermatological Research

Dario Antonini[1], Maria Rosaria Mollo[2] and Caterina Missero[2,3]

Long noncoding RNAs (lncRNAs) are a functionally heterogeneous and abundant class of RNAs acting in all cellular compartments that can form complexes with DNA, RNA, and proteins. Recent advances in high-throughput sequencing and techniques leading to the identification of DNA-RNA, RNA-RNA, and RNA-protein complexes have allowed the functional characterization of a small set of lncRNAs. However, characterization of the full repertoire of lncRNAs playing essential roles in a number of normal and dysfunctional cellular processes remains an important goal for future studies. Here we describe the most commonly used techniques to identify lncRNAs, and to characterize their biological functions. In addition, we provide examples of these techniques applied to cutaneous research in healthy skin, that is, epidermal differentiation, and in diseases such as cutaneous squamous cell carcinomas and psoriasis. As with protein-coding RNA transcripts, lncRNAs are differentially regulated in disease, and can serve as novel biomarkers for the diagnosis and prognosis of skin diseases.

Journal of Investigative Dermatology (2017) **137,** e21–e26; doi:10.1016/j.jid.2017.01.006

CME Activity Dates: February 22, 2017
Expiration Date: February 22, 2018
Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: All authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationship(s) with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:
- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of William Beaumont Hospital and the Society for Investigative Dermatology. William Beaumont Hospital is accredited by the ACCME to provide continuing medical education for physicians.
William Beaumont Hospital designates this enduring material for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the Journal of Investigative Dermatology website: http://www.jidonline.org/current. Tests for CME credits may only be submitted online at https://beaumont.cloud-cme.com/RTMS-Mar17 – click 'CME on Demand' and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

INTRODUCTION

The sequencing and functional analysis of the human genome has demonstrated that while well-characterized protein-coding genes account for only 2% of the genome, approximately 80% of the genome is transcribed from one or both strands, resulting in a large number of RNAs with little or no protein coding potential. Long noncoding RNAs (lncRNAs) are defined as RNA transcripts equal to or longer than 200 nucleotides that do not encode for proteins (Geisler and Coller, 2013). Regulation of their expression occurs by

[1]*IRCSS SDN, Napoli, Italy;* [2]*CEINGE Biotecnologie Avanzate, Center for Genetic Engineering, Napoli, Italy; and* [3]*Department of Biology, University of Naples Federico II, Napoli, Italy*

Correspondence: Caterina Missero, CEINGE Biotecnologie Avanzate, via G. Salvatore 486, 80145 Napoli, Italy. E-mail: missero@ceinge.unina.it

Abbreviations: cSCC, cutaneous squamous cell carcinoma; FISH, fluorescence in situ hybridization; lncRNA, long noncoding RNA; PICSAR, p38 inhibited cutaneous squamous cell carcinoma associated lincRNA; RNA-seq, RNA sequencing; TINCR, tissue differentiation-inducing non-protein coding RNA

SUMMARY

Advantages:

- RNA-seq coupled with advanced bioinformatics tools allow detection of even low abundance of known or previously unidentified lncRNAs, determining their primary structure and expression pattern.
- Well-established techniques such as quantitative reverse transcription polymerase chain reaction, Northern blots, and RNA FISH can be applied to validate expression, length, and to identify the localization of lncRNAs.
- Recent advances in genomics and proteomics can be applied to the study of lncRNAs by identifying in a high-throughput fashion lncRNA-associating DNA, RNA, and proteins.
- Because lncRNAs can serve as crucial components of large complexes, their functional characterization can be instrumental for therapeutic intervention.

Limitations:

- lncRNAs have highly heterogeneous functions requiring ad hoc studies for each single lncRNA.
- lncRNAs form complicated secondary and tertiary structures that can depend on their interacting molecules; therefore the function is not easily predictable by their primary structure.
- lncRNAs may be part of different complexes; therefore identification of protein, RNA, and DNA partners may not be entirely predictive of their functions.

mechanisms similar to those observed for coding genes. Similar to coding genes, lncRNA genes are often spliced, although with fewer exons, capped at their 5′ end, and polyadenylated at their 3′ end, although some lncRNAs do not contain a poly(A) tail.

LncRNAs have versatile functions due to their ability to pair with other nucleic acids and to form secondary and tertiary structures that can serve as scaffolds for multiple protein complexes. Several studies have revealed the functional relevance of lncRNAs in normal physiology, and their clinical implication in a number of malignancies as well as in other pathologies (Leucci et al., 2016; Schmitt and Chang, 2016; Yan et al., 2015).

IDENTIFICATION AND VALIDATION OF lncRNAs

To generate a comprehensive and global expression profile of lncRNAs in a given cell type or tissue, the most accurate and high-resolution method is RNA sequencing (RNA-seq) that takes advantage of high-throughput next-generation sequencing technologies (Figure 1a). The experimental design is crucial to obtain quantitative data and the appropriate comparison of different samples, especially when considering healthy and patient tissues. The first step is to create a cDNA library from RNA samples by depleting the highly abundant ribosomal RNA (for details see Whitley et al., 2016). After sequencing, basic bioinformatics analysis, and data extraction, the sequenced reads can be mapped to the corresponding genome or transcriptome reference, using several databanks including reference sequence (RefSeq; https://www.ncbi.nlm.nih.gov/refseq/), GENCODE (https://www.gencodegenes.org), lncrbadb (http://lncrnadb.com), and NONCODEv4 (http://noncode.org). An advantage of RNA-seq as opposed to microarrays is the potential of discovering novel transcribed regions and alternatively spliced forms of known genes by transcriptome reconstruction.

Using RNA-seq to compare undifferentiated and differentiated human keratinocytes, Kretz et al. (2013) identified the first lncRNA (terminal differentiation-induced ncRNA [TINCR]) that controls human epidermal differentiation by a post-transcriptional mechanism. PolyA-selected RNA was used to generate cDNA, and high-throughput transcriptome sequencing was undertaken using the Illumina HiSeq platform. Differential expression analysis was performed using human RefSeq transcripts as a reference transcriptome.

The largest expression dataset of lncRNAs in skin generated to date is from polyA+ RNA-derived cDNA from 216 samples of lesional, nonlesional psoriatic skin, and normal skin (Tsoi et al., 2015). More recently, Gupta et al. (2016) performed RNA-seq to compare the expression of lncRNAs in normal skin from healthy individuals and in lesional skin from patients with psoriasis before and after treatment with adalimumab, a humanized monoclonal antibody against tumor necrosis factor-alpha. cDNA was generated from ribosomal-depleted RNA from more than 15 individuals for each group, and sequences were obtained using the Illumina HiSeq. Interestingly, in this case, lncRNAs were mapped using a combined dataset derived from RefSeq, GENCODE, and a previously generated lncRNA catalog (Hangauer et al., 2013). Data were validated by reanalyzing previously published RNA-seq data obtained from an independent set of psoriatic skin (Li et al., 2014) with the combined database. This approach validated the top-scoring lncRNA identified by Gupta et al., underlying the crucial importance of publicly available RNA-seq raw data that can be reanalyzed in subsequent experiments by other groups.

Once specific lncRNAs of interest have been identified, a validation step is required (Figure 1b). Quantitative reverse transcription polymerase chain reaction approaches are useful to confirm lncRNA expression levels under different conditions. Although less quantitative and sensitive, Northern blot can provide reliable visual evidence for the abundance and length of the transcripts (Kretz et al., 2013). A useful technique to determine the subcellular localization of lncRNAs is single-molecule RNA fluorescence in situ hybridization (RNA FISH) analysis (Kretz et al., 2013; Piipponen et al., 2016) (Figure 2c). Visualization of the subcellular localization of a given lncRNA by FISH technology can shed light on its putative functions. This fluorescent method led to the demonstration that the lncRNA TINCR, a crucial regulator of keratinocyte differentiation, is present at low levels both in the nucleus and in the cytoplasm of undifferentiated human keratinocytes, whereas its expression