Toward a Systematic Understanding of mRNA 3' Untranslated Regions

Wenxue Zhao¹, Denitza Blagev¹, Joshua L. Pollack¹, and David J. Erle¹

¹Lung Biology Center, Department of Medicine, University of California, San Francisco, San Francisco, California

Messenger RNAs (mRNAs) contain prominent untranslated regions (UTRs) that are increasingly recognized to play roles in mRNA processing, transport, stability, and translation. 3' UTRs are believed to harbor recognition sites for a diverse set of RNA-binding proteins that regulate gene expression as well as most active microRNA target sites. Although the roles of 3' UTRs in the normal and diseased lung have not yet been studied extensively, available evidence suggests important roles for 3' UTRs in lung development, inflammation, asthma, pulmonary fibrosis, and cancer. Systematic, genome-wide approaches are beginning to catalog functional elements within 3' UTRs and identify the proteins and microRNAs that interact with these elements. Application of new data sets and experimental approaches should provide powerful insights into how 3' UTR-mediated regulatory events contribute to disease and may inspire novel therapeutic approaches.

Keywords: gene expression regulation; microRNAs; RNA, messenger; RNA-binding proteins; lung

The transcriptome (the full set of RNA transcripts) is a necessary intermediate between the genome and the proteome. In addition to translated (or coding) regions, messenger RNAs (mRNAs) contain 5' and 3' untranslated regions (UTRs). A rapidly growing body of evidence indicates that these noncoding mRNA sequences have diverse roles in regulating RNA processing, transport, stability, and translation. A comprehensive systems biology approach to gene regulation will require that we systematically identify active sequences within UTRs (*cis* elements), identify the cellular *trans* factors that interact with these *cis* elements, and determine the functional consequences of these interactions. Here we review available information about human 3' UTRs, known classes of 3' UTR *cis* elements, evidence that supports the importance of 3' UTRs in lung disease, and recent progress toward transcriptome-wide analysis of 3' UTRs. The 5' UTR, which plays a major role in regulating translation, has been reviewed elsewhere (1).

3' UTRs ARE PROMINENT FEATURES OF mRNAs

The 3' UTR begins at the stop codon. Polyadenylation signals (typically AAUAAA or a similar sequence) within immature mRNA 3' UTRs are recognized by an enormous complex of approximately 85 proteins that directs cleavage of the mRNA and addition of a poly(A) tail of approximately 250 nucleotides (nt) (2). Most human genes have more than one polyadenylation signal, and usage of alternative polyadenylation sites is one source of variability in 3' UTR length (3). For some genes,

alternative splicing (with use of an alternative final exon) produces variation in both 3' UTR length and sequence (4). For example, subjects with systemic lupus erythematosus had increased levels of an alternatively spliced form of the T cell receptor ζ transcript that was less stable than the predominant form seen in healthy control subjects (5). Our analysis of a large set of well-characterized human transcripts (Figure 1) indicates that the length of the 3' UTR varies widely, but on average 3' UTRs (not including the poly[A] tail) are 1,278 nt, or 36% of the total length of each mRNA. In contrast, yeast 3' UTRs average approximately 100 nt and very few are greater than 400 nt (6), suggesting that 3' UTRs serve additional regulatory functions in complex multicellular organisms. Approximately 15 to 20% of human 3' UTR sequence was found to be conserved in four other vertebrate species (7). Some conserved regions contain known or predicted *cis* regulatory elements, but others do not, suggesting that additional classes of *cis* elements remain to be discovered. The existence of large amounts of non-conserved or poorly conserved 3' UTR sequence (7) also suggests that 3' UTRs contribute to differences in gene regulation between species.

INTERACTIONS BETWEEN 3' UTRs AND RNA BINDING PROTEINS REGULATE mRNA PROCESSING, LOCALIZATION, STABILITY, AND TRANSLATION

Interactions between RNAs and proteins lead to formation of ribonucleoprotein (RNP) complexes (8). Many RNA-binding proteins have been identified and in some cases consensus RNA sequence motifs or structures required for protein binding have been defined (8, 9). Although RNA-binding proteins interact with all portions of mRNAs (5' UTRs, coding sequences, introns, and 3' UTRs), certain RNA-binding proteins interact preferentially with 3' UTRs (Table 1). This relates at least in part to the prevalence of specific recognition sequences in 3' UTRs. Individual 3' UTRs may contain multiple protein binding sites, which may allow for synergistic, antagonistic, or other combinatorial regulatory interactions (10). Systems biology approaches similar to those used to understand combinatorial interactions within DNA promoters (11) will be required to fully understand how 3' UTR signals are integrated.

RNP complex formation is critical for RNA processing and localization (8, 9). The best-known example of a 3' UTR *cis* element with a role in mRNA processing is the polyadenylation signal, which is present in virtually all eukaryotic mRNAs (with the exception of histone mRNAs) (12). Studies in neurons and other polarized cells have led to the identification of 3' UTR sequences important in targeting mRNAs to specific regions within the cytoplasm, allowing for localized protein synthesis (13). For example, the RNA binding protein zipcode binding protein-1 binds a 54-nucleotide "zipcode" sequence within the 3' UTR of β -actin mRNA and directs the mRNA to the leading edge of fibroblasts and other cells where actin filament assembly is required (14).

Interactions between 3' UTRs and RNA-binding proteins can affect protein production by altering mRNA stability and translation (15, 16). One of the best-studied 3' UTR *cis*-regulatory

(Received in original form July 30, 2010; accepted in final form December 16, 2010)

Supported by National Institute of Health grants 5R21HG004665, 5T32HL007185, and 1F32HL095338; UCSF Sandler Asthma Basic Research (SABRE) Center; and UCSF Program for Breakthrough Biomedical Research.

Correspondence and requests for reprints should be addressed to David J. Erle, M.D., UCSF Mail Code 2922, San Francisco, CA 94143. E-mail: david.erle@ucsf.edu

Proc Am Thorac Soc Vol 8, pp 163–166, 2011

DOI: 10.1513/pats.201007-054MS

Internet address: www.atsjournals.org

Figure 1. 3' Untranslated regions (UTRs) are prominent features of human messenger RNAs (mRNAs). We analyzed all RefSeq mRNA sequences available from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/projects/RefSeq/>, accessed on July 27, 2010) and determined lengths of (A) the 5' UTR, (B) the coding region, and (C) the 3' UTR for the 15,295 human RefSeq mRNA sequences that included annotations indicating positions of start and stop codons and one or more polyadenylation sites. For 5,672 of these transcripts, two or more polyadenylation sites were annotated and 3' UTR lengths were calculated for each of these variants. Mean lengths were 243 nucleotides (nt) for the 5' UTR, 1,766 nt for the coding region, and 1,278 nt for the 3' UTR.

elements is the AU-rich element (ARE) (17). AREs contain one or more repeats of an AUUUA motif or other AU-rich sequences. ARE-binding proteins, such as AUF1/hnRNP, HuR, TIA-1, KSRP, and tristetraprolin, recognize these sequences leading to changes in stability of mRNA transcripts (15, 16, 18). Some ARE-binding proteins tend to increase RNA stability, whereas others decrease stability, and effects of a given

ARE can depend on cellular context (18). For example, cytokine-induced changes in binding of KSRP and HuR to an ARE in the 3' UTR of human inducible nitric oxide synthase (iNOS) mRNA were found to contribute to regulation of iNOS, because KSRP destabilized this mRNA, whereas HuR had a stabilizing effect (19). Mechanisms underlying ARE-mediated changes in mRNA stability are complex (20) and incompletely understood. Certain AREs promote recruitment of mRNA to cytoplasmic processing bodies (P bodies) where enzymes involved mRNA decay are concentrated (21). Evidence suggests that AU-rich elements can affect protein production by altering mRNA stability, by sequestering mRNAs from ribosomes (e.g., by sequestration in P bodies), and by recruiting proteins that directly affect translation efficiency (22, 23). Sequences that are recognized by microRNAs (miRNAs) and other small (20–30 nt) RNAs, including small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) (24), also tend to be found largely in 3' UTRs. These small RNAs associate with members of the Argonaute family of small RNA-binding proteins within larger RNA-induced silencing complexes (RISCs) (25). Base-pairing interactions between a small RNA in the RISC and a partially complementary sequence in the 3' UTR lead to specific interactions between mRNAs and RISCs (24). RISCs are heterogeneous, but may contain Argonaute family members and other proteins that can cleave small RNA:mRNA heteroduplexes, block translational initiation or elongation, recruit poly(A) deadenylating and mRNA decapping complexes, and direct mRNAs to P bodies (24). In addition to AREs and small RNA/RISC targets, many other 3' UTR *cis* elements can also affect mRNA stability (13).

3' UTR IN DEVELOPMENT AND DISEASE

Given the fundamental regulatory role of 3' UTRs, it is not surprising that 3' UTRs and associated protein complexes make critical contributions to normal development (26). Levels of many miRNAs change substantially during development of the lung (27, 28) and other organs (29, 30). Mice with targeted deletions of specific miRNAs are beginning to reveal the functions of these miRNAs *in vivo*. For example, deletion of the muscle-specific miRNA miR1-2, which recognizes targets in the 3' UTR of the cardiac transcription factor *Irf5* mRNA and other cardiac mRNAs, showed a role for this miRNA in cardiac morphogenesis, electrical conduction, and cell cycle control (31). Although little is known about the *in vivo* effects of specific miRNAs in lung cells, miRNA overexpression experiments together with analyses of effects of concurrent changes in miRNA and mRNA levels suggest that miRNA interactions with miRNA targets in 3' UTRs also play important regulatory roles in lung development (27, 28).

AREs are frequently found in 3' UTRs of mRNAs encoding cytokines and these AREs are important in regulating inflammatory and immune responses (32). A striking demonstration of the importance of these elements is provided by studies of mice

TABLE I. SELECTED 3' UNTRANSLATED REGION FUNCTIONAL ELEMENTS

<i>Cis</i> Element	Motif	<i>Trans</i> Factor	Major Function	References
Polyadenylation signal	AAUAAA or similar	Cleavage/polyadenylation specificity factor complex	Marks site for addition of poly(A) tail	2
Zipcode	54-nt sequence	Zipcode binding protein-1	Localizes β -actin mRNA in cytoplasm	14
AU-rich element (ARE)	One or more AUUUA repeats	ARE-binding proteins	Control of mRNA stability	17
miRNA target	Complementary to miRNA sequence	miRNA in RISC complex	Inhibition of mRNA stability and/or translation	24, 42, 44

Definition of abbreviations: miRNA, micro RNA; mRNA, messenger RNA; nt, nucleotide; RISC, RNA-induced silencing complexes.

with targeted deletions of the ARE in the 3' UTR of tumor necrosis factor (TNF) mRNA (33). These mice have impaired regulation of TNF and develop chronic inflammatory arthritis and inflammatory bowel disease. Recent studies of experimental asthma indicate that many miRNAs, including miR-21 (34), miR-126 (35), and let-7 (36), are involved in establishment of Th2 responses and allergic inflammation. Some relevant target sites in 3' UTRs have been identified (e.g., the miR-21 target site in IL-12p35 mRNA [34]), but more work will be required to identify the full range of targets responsible for *in vivo* effects of miRNA inhibition in these asthma models. Alterations in 3' UTRs or in 3' UTR interacting proteins and miRNAs have also been implicated in several other lung diseases, including lung cancer (37, 38) and pulmonary fibrosis (39). A small number of rare disease-causing mutations in 3' UTRs have been identified (40), but the potential contributions of many common 3' UTR polymorphisms to disease susceptibility remain largely unknown.

SYSTEMS BIOLOGY APPROACHES TO THE 3' UTR

Although many valuable insights have come from traditional, low-throughput studies of individual 3' UTRs or 3' UTR *cis* elements and the *trans*-regulatory complexes that interact with them, a systematic understanding of the roles of 3' UTRs will require the use of multiple transcriptome-wide approaches. Computational approaches (based on sequence analysis) predict the existence of AREs in approximately 5 to 8% of human 3' UTRs (41) and of one or more miRNA targets in as many as 60% or more of human 3' UTRs (42). Experimental approaches have been used to define the complement of RNAs that associate with specific RNA-binding proteins ("ribonomics") (43). Recent improvements in high-throughput sequencing and cross-linking techniques have enabled fine mapping of protein binding sites within the transcriptome (44). High-throughput sequencing has recently been used together with massively parallel DNA synthesis for detailed analysis of DNA regulatory elements (45), and similar approaches are likely to be suitable for analyzing 3' UTR regulatory elements. These types of transcriptome-wide approaches should refine our understanding of known classes of *cis* elements and *trans*-regulatory factors and seem likely to identify novel classes of 3' UTR regulatory elements.

CONCLUSIONS

A systematic understanding of gene regulation will require that we understand in detail how each level of regulation works and how different levels are integrated. Ample evidence indicates that 3' UTRs of human mRNAs contain many *cis* regulatory elements that control RNA processing, localization, stability, and/or translation. These elements are important in normal development and in disease in the lung and other organs. Much more work is required to identify the complete set of 3' UTR *cis* elements and the *trans*-regulatory elements that interact with them and to determine the functional consequences of these interactions. Fortunately, powerful transcriptome-wide computational and experimental methods are now being used to address these issues. Together with lower-throughput reductionist approaches, these transcriptome-wide approaches should move us closer to a systems biology understanding of how 3' UTRs contribute to gene regulation in health and disease.

Author Disclosure: W.Z. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. D.B. received grant support from the National Heart, Lung, and Blood Institute. J.L.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. D.J.E. receives royalties from one or more companies paying through the University of California San Francisco.

References

- Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol* 2005;16:39–47.
- Shi Y, Di Giandomartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 2009;33:365–376.
- Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 2000;10:1001–1010.
- Hughes TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet* 2006;22:119–122.
- Chowdhury B, Tsokos CG, Krishnan S, Robertson J, Fisher CU, Warke RG, Warke VG, Nambiar MP, Tsokos GC. Decreased stability and translation of T cell receptor zeta mRNA with an alternatively spliced 3'-untranslated region contribute to zeta chain down-regulation in patients with systemic lupus erythematosus. *J Biol Chem* 2005;280:18959–18966.
- Tuller T, Ruppert E, Kupiec M. Properties of untranslated regions of the *S. cerevisiae* genome. *BMC Genomics* 2009;10:391.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–1050.
- McKee AE, Silver PA. Systems perspectives on mRNA processing. *Cell Res* 2007;17:581–590.
- Morris AR, Mukherjee N, Keene JD. Systematic analysis of post-transcriptional gene expression. *Wiley Interdiscip Rev Syst Biol Med* 2010;2:162–180.
- Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007;8:533–543.
- Gertz J, Siggia ED, Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* 2009;457:215–218.
- Shatkin AJ, Manley JL. The ends of the affair: capping and polyadenylation. *Nat Struct Biol* 2000;7:838–842.
- Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, *et al*. UTRdb and UTRsite (release 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 2010;38:D75–D80.
- Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. Characterization of a beta-actin mRNA zipcode-binding protein. *Mol Cell Biol* 1997;17:2158–2165.
- Bag J, Bhattacharjee RB. Multiple levels of post-transcriptional control of expression of the poly (a)-binding protein. *RNA Biol* 2010;7:5–12.
- Chen CY, Gherzi R, Ong SE, Chan EL, Rajmakers R, Puijn GJ, Stoecklin G, Moroni C, Mann M, Karin M. AU binding proteins recruit the exosome to degrade AU-rich mRNAs. *Cell* 2001;107:451–464.
- Xu N, Chen CY, Shyu AB. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* 1997;17:4611–4621.
- Barreau C, Paillard L, Osborne HB. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 2005;33:7138–7150.
- Linker K, Pautz A, Fehrer M, Hubrich T, Greeve J, Kleinert H. Involvement of KSRP in the post-transcriptional regulation of human iNOS expression-complex interplay of KSRP with TTP and HuR. *Nucleic Acids Res* 2005;33:4813–4827.
- Tebo J, Der S, Frevel M, Khabar KS, Williams BR, Hamilton TA. Heterogeneity in control of mRNA stability by AU-rich elements. *J Biol Chem* 2003;278:12085–12093.
- Franks TM, Lykke-Andersen J. TTP and BRF proteins nucleate processing body formation to silence mRNAs with AU-rich elements. *Genes Dev* 2007;21:719–735.
- Espel E. The role of the AU-rich elements of mRNAs in controlling translation. *Semin Cell Dev Biol* 2005;16:59–67.
- Vasudevan S, Steitz JA. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 2007;128:1105–1118.
- Pratt AJ, MacRae IJ. The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem* 2009;284:17897–17901.
- Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 2005;123:631–640.
- Mendell JT, Dietz HC. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* 2001;107:411–414.

27. Bhaskaran M, Wang Y, Zhang H, Weng T, Baviskar P, Guo Y, Gou D, Liu L. MicroRNA-127 modulates fetal lung development. *Physiol Genomics* 2009;37:268–278.
28. Dong J, Jiang G, Asmann YW, Tomaszek S, Jen J, Kislinger T, Wigle DA. MicroRNA networks in mouse lung organogenesis. *PLoS ONE* 2010;5:e10854.
29. Baltimore D, Boldin MP, O'Connell RM, Rao DS, Taganov KD. MicroRNAs: new regulators of immune cell development and function. *Nat Immunol* 2008;9:839–845.
30. Blakaj A, Lin H. Piecing together the mosaic of early mammalian development through microRNAs. *J Biol Chem* 2008;283:9505–9508.
31. Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, Muth AN, Tsuchihashi T, McManus MT, Schwartz RJ, Srivastava D. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking mirna-1-2. *Cell* 2007;129:303–317.
32. Khabar KS. Post-transcriptional control during chronic inflammation and cancer: a focus on AU-rich elements. *Cell Mol Life Sci* 2010;67: 2937–2955.
33. Kontoyiannis D, Pasparakis M, Pizarro TT, Cominelli F, Kollias G. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* 1999;10:387–398.
34. Lu TX, Munitz A, Rothenberg ME. MicroRNA-21 is up-regulated in allergic airway inflammation and regulates il-12p35 expression. *J Immunol* 2009;182:4994–5002.
35. Mattes J, Collison A, Plank M, Phipps S, Foster PS. Antagonism of microRNA-126 suppresses the effector function of TH2 cells and the development of allergic airways disease. *Proc Natl Acad Sci USA* 2009;106:18704–18709.
36. Polikepahad S, Knight JM, Naghavi AO, Oplt T, Creighton CJ, Shaw C, Benham AL, Kim J, Soibam B, Harris RA, et al. Proinflammatory role for let-7 microRNAs in experimental asthma. *J Biol Chem* 2010; 285:30139–30149.
37. Wu X, Piper-Hunter MG, Crawford M, Nuovo GJ, Marsh CB, Otterson GA, Nana-Sinkam SP. MicroRNAs in the pathogenesis of lung cancer. *J Thorac Oncol* 2009;4:1028–1034.
38. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009;138:673–684.
39. Liu G, Friggeri A, Yang Y, Milosevic J, Ding Q, Thannickal VJ, Kaminski N, Abraham E. miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. *J Exp Med* 2010;207:1589–1597.
40. Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat Med* 2000;6:637–641.
41. Bakheet T, Williams BR, Khabar KS. ARED 3.0: The large and diverse AU-rich transcriptome. *Nucleic Acids Res* 2006;34:D111–D114.
42. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009;19:92–105.
43. Tenenbaum SA, Lager PJ, Carson CC, Keene JD. Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* 2002;26:191–198.
44. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010;141:129–141.
45. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* 2009;27:1173–1175.