

Long Non-Coding RNA

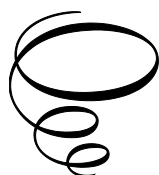
Long Non-Coding RNA:

*From Disease Biomarkers to
Targeted Therapeutics*

Edited by

Syed Mudasar Ahmad
and Nazir Ahmad Ganai

**Cambridge
Scholars
Publishing**



Long Non-Coding RNA:
From Disease Biomarkers to Targeted Therapeutics

Edited by Syed Mudasir Ahmad and Nazir Ahmad Ganai

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2020 by Syed Mudasir Ahmad, Nazir Ahmad Ganai
and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-5821-5

ISBN (13): 978-1-5275-5821-2

This book is dedicated to



Dr. A. P. J. Abdul Kalam
(1931–2015)

The whole of science is nothing more than a refinement of everyday thinking.

—Albert Einstein (1879–1955)

TABLE OF CONTENTS

Preface	ix
List of Abbreviations	xii
Chapter 1	1
Functional Aspects of Long Non-Coding RNAs: A Review Peerzada Tajamul Mumtaz, Qamar Taban, Sameena Ismail, Divya Srivastava, Mashooq Ahmad Dar, Riaz Ahmad Shah, and Syed Mudasir Ahmad*	
Chapter 2	22
Cancer Development and Therapy Resistance: Spotlight on the Dark Side of the Genome Eleonora Leucci*	
Chapter 3	45
Long Non-coding RNAs in Multiple Sclerosis: An Update Insha Zahoor* Shailendra Giri, Mir Khurshid Iqbal, Ashaq Hussain Mir, Amrina Shafi Qamar Taban and Ehtishamul Haq	
Chapter 4	64
Long Non-coding RNAs in Infection and Immunity Qamar Taban, Peerzada Tajamul Mumtaz, Sameena Ismail, Ehtishamul Haq, Mashooq Dar, Insha Zahoor, and Syed Mudasir Ahmad*	
Chapter 5	91
LncRNAs as Important Mediators for Diagnosis, Prognosis, and Potential Therapeutics in Leukemias, Hepatocellular Carcinoma, and Breast Cancer Aashiq Hussain Bhat* Khalid Bashir Dar, and Showkat Ahmad Ganie	
Chapter 6	116
Long Non-coding RNA in Neurological Diseases: A Mechanistic Insight Khalid Bashir Dar, Aashiq Hussain Bhat, Shajrul Amin, and Showkat Ahmad Ganie*	

Chapter 7	140
Long Non-coding RNAs in Hormonal Disorders with a Focus on Polycystic Ovary Syndrome Mudasar Nabi, Shayaq Ul Abeer Rasool, Sairish Ashraf, and Shajrul Amin*	
Chapter 8	160
Diverse Functions of Long Non-coding RNAs in Cardiovascular Disease Bilal Ahmad Tantry* Qamar Taban, Mudasar Nabi, Peerzada Tajamul Mumtaz, Shaik Rahiman, and Mohammad Abdul Hafeez	
Chapter 9	178
Regulatory Mechanisms of Long Non-coding RNAs in Host–Pathogen Interactions Syed Mudasir Ahmad*	

PREFACE

In RNA biology, there has been a profound shift in our understanding of the role of non-coding RNA molecules. They were once regarded as mere templates for generating proteins, but growing evidence is proving they are becoming as large and diverse as their protein coding mRNA counterparts. LncRNAs have been found to be involved in a range of developmental processes and diseases, but details of the mechanisms by which they act are still surprising. It was suggested they play emerging roles in chromatin states and epigenetic and post-transcriptional gene regulations. The aberrant expression of lncRNAs and changes in their structure are associated with a wide spectrum of diseases. Some lncRNAs play a fundamental role in cellular identity, cell development, and disease progression. They are now widely studied in diverse organisms to understand their function. Thus, understanding the role of lncRNAs has become fundamental in biology, keeping in mind the fact of their emerging profile. An increasing number of experimental studies are providing evidence that lncRNAs mediate disease pathogenesis, thereby challenging the concept that protein-coding genes are the sole contributors to the development of human diseases.

LncRNAs are regulatory non-coding RNAs that are greater than 200 nucleotides in length; they have emerged as another critical layer of gene regulation. Given the significant role of lncRNAs in bimolecular regulatory interactions within cells and their prospects as drug targets in numerous diseases, it becomes highly imperative to understand their functionalities in disease pathogenesis. With this fascination, this book focuses on various types of disease to illustrate the potential impacts of lncRNAs on diverse biological processes and their therapeutic significance. This book is primarily centered on a few fundamental biological questions: How are lncRNAs involved in disease pathogenesis? How do lncRNAs regulate gene expression and contribute to cellular identities in physiological and pathological contexts, and how can they be targeted as potential biomarkers in drug therapeutics? This book is thus essential reading for biologists as well as interested researchers and students working in the domain of RNA biology. The book contains nine chapters.

Chapter 1 by Peerzada Tajamul Mumtaz and associates introduces the theme of the book, beginning with biogenesis and the mechanism of

action. This chapter lays the foundation for the book by describing the types of lncRNAs and their characteristic features.

Chapter 2 by Eleonora Leucci summarizes current knowledge of the role of lncRNAs in cancer, and their potential clinical applications as biomarkers and therapeutic targets. It also provides an insight into the role of lncRNAs as major regulators of cancer cell plasticity and drug resistance.

Chapter 3 by Insha Zahoor et al. aims to provide a recent and cohesive but comprehensive update on the involvement of lncRNAs in MS pathogenesis, which will indeed lay down a much needed foundation for exploiting their potential as biomarkers in the early diagnosis of this complex disease and in predicting disease activity and its course, serving as novel targets for the development of alternative therapeutic interventions for its complete cure.

Chapter 4 by Qamar Taban and associates discusses the regulatory role of lncRNAs in both innate and adaptive branches of immunity and also enlightens us about their role in hematopoietic development, myeloid differentiation, inflammation, and host defense against pathogens.

Chapter 5 by Aashiq Hussain Bhat and associates details the lncRNAs involved in gene expression regulation and as mediators for diagnosis and prognosis, and as potential therapeutics in leukemias, hepatocellular carcinoma, and breast cancer.

Chapter 6 by Khalid Bashir Dar and associates gives mechanistic insights into the role of lncRNAs in brain function and more particularly in the pathogenesis of neurological diseases like Alzheimer's disease, Parkinson's disease, Angelman syndrome, and amyotrophic lateral sclerosis.

Chapter 7 by Mudasir Nabi et al. gives a detailed introduction to non-coding RNAs with special reference to lncRNAs and their biological role in various cellular pathways that contribute to the pathogenesis of PCOS. It also gives an insight into the future prospective of lncRNAs as non-invasive biomarkers and potential therapeutic agents in PCOS.

Chapter 8 by Dr. Bilal et al. gives us a brief insight into the role of lncRNAs in heart-related disorders and cardiovascular diseases (CVDs).

Chapter 9 by Dr. Syed Mudasir Ahmad gives an idea about the regulatory role of long non-coding RNAs and their role in host–pathogen interactions in general.

It is hoped that readers will find each chapter stimulating and thought provoking, which will add new dimensions to future RNA research. As editors of the book, we are indebted to the authors for the time and energy they spent in shaping the book, and the advancement of knowledge in the field that it generates. We are highly thankful to Prof. Khalid Rehman Hakeem, King Abdulaziz University, Jeddah, Saudi Arabia, for encouraging

us to write a book in the field of LncRNA research. We are also grateful to Prof. Riaz Ahmad Shah, Head of the Department of Animal Biotechnology, FVSc and AH SKUAST-K, India, for his valuable suggestions in completing this assignment. Finally, we extend equal appreciation to our associate editors, Peerzada Tajamul Mumtaz and Qamar Taban, who wholeheartedly supported us in compiling the book. Gratitude is expressed to all the authors who have contributed chapters to this book.

This book is a part of the National Agricultural Higher Education Project (NAHEP) of SKUAST, Kashmir, India.

Dr. Syed Mudasir Ahmad
Division of Animal Biotechnology
Faculty of Veterinary Sciences
Husbandry, SKUAST-Kashmir, India

LIST OF ABBREVIATIONS

AD: Alzheimer's disease
AE-PCOS: Androgen Excess and PCOS Society
AGO2: Argonaute 2
Airm: antisense of insulin-like growth factor 2 non-coding RNA
ALL: acute lymphoblastic leukemia
ALS: amyotrophic lateral sclerosis
AML: acute myeloid leukemia
AMPK: AMP-activated protein kinase
ANRIL: antisense non-coding RNA in the INK4 locus
APCs: antigen-presenting cells
ASD: autism spectrum disorder
ASO: antisense oligonucleotides
BACE1-AS: B-secretase-1-AS lncRNA
BANCR: BRAF-activated non-coding RNA
BC: brain cytoplasmic lncRNA
BDNF: brain derived neurotrophic factor
BDNF-AS: brain-derived neuro-trophic factor antisense RNA
BGL3: beta globin locus 3
Bvht: braveheart
CAA: chronic aplastic anemia
CAGE: cap analysis of gene expression
CARMEN: cardiac mesoderm enhancer-associated non-coding RNA
CCAT1: colon cancer-associated transcript 1
CCAT1-L: colorectal cancer associated transcript 1-long isoform
cDCs: classical DCs
CDKN2A: cyclin dependent kinase inhibitor 2A
cET: constrained EThyl
CETPH: chronic thromboembolic pulmonary hypertension
Chaer: cardiac hypertrophy-associated epigenetic regulator
CHAST: cardiac hypertrophy-associated transcript
CHRF: cardiac hypertrophy-related factor
CIS: clinically isolated syndrome
CNS: central nervous system
COX2: cyclooxygenase 2 gene
CPC: cardiac precursor cell

CTBP1-AS: C-Terminal binding protein 1 antisense
CTNNB1: oncogene β catenin
CVD: cardiovascular diseases
DDIT4: DNA damage inducible transcript 4
DYNLL1: dynein light chain 1
DYRK: dual-specificity tyrosine phosphorylation-regulated kinase
EBV: Epstein Barr virus
EMT: epithelial to mesenchymal transition
ERV: endogenous retroviral sequence
ESCs: embryonic stem cells
FAS-AS1: fas cell surface death receptor-antisense 1
FDA: food and drug administration
Fendrr: fetal-lethal non-coding developmental regulatory RNA
FGF19: fibroblast growth factor 19
FTLD: frontotemporal lobar degeneration
FUS/TLS: fused in sarcoma/translocated in sarcoma
GABA: g-aminobutyric acid
GAS5: growth arrest-specific transcript 5
GO: gene Ontology
H3K27ac: histone H3 Lys27 acetylation
HAR1: human accelerated region 1 gene
HAT: histone acetyltransferase
HCC: hepatocellular Carcinomas
HD: Huntington's disease
HLA: human leukocyte antigen
hnRNPs: heterogeneous nuclear ribonucleoproteins
HOST2: human ovarian cancer-specific transcript
HOTAIR: HOX transcript antisense intergenic RNA
HOTAIRM1: HOXA transcript, antisense RNA myeloid-specific 1
HOTTIP: HOXA transcript at the distal tip
HOXC9: Homeobox C9
HSCs: hematopoietic stem cells
HSV: herpes simplex virus
HUVECs: human umbilical endothelial cell
IAV: influenza A
IGF2: insulin-like growth factor 2
IIM: idiopathic inflammatory myopathy
IL2RA: interleukin-2 receptor alpha
ISGs: interferon stimulated genes
JAK1: Janus kinase 1
KD: knock-down

KEGG: Kyoto Encyclopedia of Genes and Genomes analysis
KLF2: kruppel like factor 2
LINK-A: long intergenic NcRNA for Kinase Activation
Lmx1a: LIM homeobox transcription factor 1, alpha
LNA: locked nucleic acids
LncRNA: long non-coding RNAs
lncRNA-ATB: long non-coding RNA activated by TGF beta
LSD1: lysine specific demethylase 1
LTR: long terminal repeat
LUNAR1: Leukemia-induced non-coding activator rna-1
MALAT1: metastasis-associated lung adenocarcinoma transcript 1
MDRV: Muscovy Duck Reovirus
MeCP2: methyl CpG binding protein 2
MEG3: maternally expressed gene 3
MITF: microphthalmia-associated transcription factor
MLL: myeloid-lymphoid leukemia
Morbid: myeloid RNA regulator of Bim-induced death
MS: multiple sclerosis
MSA: multiple system atrophy
MVIH: microvascular invasion
MYC: myelocytomatosis homologs
MyD: myeloid differentiation
NBAT-1: lncRNA neuroblastoma associated transcript-1
ncRNAs: non-coding RNAs
NEAT1: nuclear paraspeckles assembly transcript
NEIGHBOR: National Eye Institute Glaucoma Human Genetics
Collaboration consortium
NeST: Nettoie Salmonella pas Theilers's
NFAT: Nuclear Factor of Activated T cells
NRAV: negative regulator of antiviral response
p53: tumor suppressor gene
PABP1: poly(A)-binding protein
PACER: p50-associated Cox2 extragenic RNA
PAN RNA: polyadenylated nuclear RNA
PANDAR: promoter of CDKN1A antisense DNA damage activated RNA
PCOS: Polycystic ovary syndrome
PD: Parkinson's disease
pDCs: Plasmacytoid DCs
PDX: patient derived xenograft
PGK1: Phosphoglycerate kinase 1
PINK1: PTEN -induced kinase 1

PMO: phosphorodiamidate morpholino oligomer
PNA: peptide nucleic acids
POAG: primary open-angle glaucoma
PPAR: proliferator-activated receptor
PPMO: peptide-conjugated PMO
PS: phosphorothioate
PTEN: phosphatase tensin homolog
PVT1: plasmacytoma variant translocation 1
RBPs: RNA-binding proteins
RISC: RNA-induced silencing complex
RMST: Rhabdomyosarcoma 2 associated transcript
RNA: ribonucleic acid
RNCR3: Retinal non-coding RNA 3
SAMMSON: survival-associated mitochondria melanoma-specific
oncogenic ncRNA
SCA8: spinocerebellar ataxia type 8
SeV: Sendai virus
siRNA: short interfering RNA
SMILR: smooth muscle-induced lncRNA
Sor11: sortilin-related receptor 1
SRA: steroid receptor RNA activator
STAT3: signal transducer and activator of transcription 3
SYNCRIP: synaptotagmin-binding cytoplasmic RNA interacting protein
TDP-43: transactive response DNA binding protein 43 kDa
TERT: telomerase reverse transcriptase
TGF-beta: transforming grow factor beta
THRIL: TNF and heterogeneous nuclear ribonucleoprotein L related
immunoregulatory lincRNA
TLRs: toll-like Receptors
TSS: transcriptional start site
TUG1: taurine up-regulated gene 1
UBE3A: Ubiquitin protein ligase E3A
UCA1: urothelial cancer associated 1
Uchl1: ubiquitin carboxy-terminal hydrolase L1
XIST: X chromosome inactive-specific transcript
ZEB: zinc finger E-box-binding homeobox
Zeb2NAT: Zeb2 Natural antisense transcript

CHAPTER 1

FUNCTIONAL ASPECTS OF LONG NON-CODING RNAs: A REVIEW

PEERZADA TAJAMUL MUMTAZ,^{1,2}
QAMAR TABAN,^{1,3} SAMEENA ISMAIL,⁴
DIVYA SRIVASTAVA,²
MASHOOQ AHMAD DAR,¹
RIAZ AHMAD SHAH,¹
AND SYED MUDASIR AHMAD^{*1}

¹ Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Shuhama, SKUAST-K-190006, India

² Department of Biochemistry, School of Life Sciences, Jaipur National University, Jaipur, Rajasthan-302017, India

³ Drug Therapeutics and Neurobiology Lab, Department of Biotechnology, School of Biological Sciences, University of Kashmir, Hazratbal, Srinagar, Jammu and Kashmir-190006, India

⁴ Department of Biotechnology, Government Degree College Anantnag, Jammu and Kashmir-192101

*Corresponding author:

Syed Mudasir Ahmad,

Department of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-K

Tel: 9419514131; E-mail: mudasirbio@gmail.com

Abstract

Recent reports on RNA sequencing studies have revealed that most of the genome is transcribed, but very little of it encodes proteins, while the rest of the transcripts are known as non-coding transcripts. Long non-coding RNAs (lncRNAs) are non-coding transcripts longer than 200 nucleotides. Members of the non-coding genome include microRNA (miRNA), small

regulatory RNAs, and other short RNAs. Most long non-coding RNAs (lncRNAs) are poorly annotated. Recent recognition of lncRNAs highlights their effects in many biological and pathological processes. LncRNAs are dysfunctional in a variety of human diseases varying from cancerous to non-cancerous diseases. Characterization of these lncRNA genes and their modes of action may allow their use for diagnosis, monitoring of progression, and targeted therapies in various diseases. In this review, we summarize the functional perspectives as well as the mechanism of action of lncRNAs.

Keywords: LncRNA, biogenesis, cellular localization, stability, functional utility

Introduction

As the complete transcriptional panorama of all organisms is more complex than was originally thought, the majority of the genomic sequence is arranged into a distinct array of protein-coding RNAs and non-coding RNAs (ncRNAs). Our index of genetic elements is brimming with long non-coding RNAs (lncRNAs), as in the genomic and transcriptomic view. Long non-coding RNAs are an enigmatic group of long RNA transcripts with no accepted protein-coding role. LncRNAs are usually found in every aspect of life; both the diversity and the size of non-coding RNA expression stocks rather than of protein-coding genes are associated with organismal complexity. Because of the diverse number of lncRNAs in humans, it is estimated that the number of lncRNAs is greater than the number of protein-coding genes. The total number of lncRNAs is continuously increasing with the advent of high-throughput next-generation sequencing (NGS) technology, the latest epigenomic technologies, and computational prediction techniques. Thus, understanding the role of lncRNAs has become fundamental in biology, keeping in view its emerging ranks. LncRNAs are regulatory non-coding RNAs that are greater than 200 nucleotides in length. On the basis of their roles, lncRNAs can be generally divided into three broad categories: the first category includes lncRNAs that are non-functional and likely to be the result of transcriptional noise; the second category includes lncRNAs for which the act of transcription alone is sufficient for their function but the transcript itself is not necessary; and the third category includes functional lncRNAs that act in cis and/or in trans ways. LncRNA transcripts have emerged as another critical layer of gene regulation.

Nomenclature of lncRNAs

lncRNAs are named after their function and expression pattern like XIST, HOTAIR. But this rule of nomenclature is inefficient due to the poor understanding of the functional utility of the majority of lncRNAs. To address such issues in the nomenclature of lncRNAs, the HUGO Gene Nomenclature Committee (HGNC) has drafted certain basic outlines to name unknown lncRNAs. The committee has decided to use the symbol “-AS” for anti-sense lncRNAs, as in “BOK-AS1” for “BOK antisense RNA 1,” located in the opposite strand of BOK. Likewise, “-IT” is used for “intronic lncRNA,” as in “MAGI2- IT1” for “MAGI2 intronic lncRNA transcript 1.” The prefix “LINC” is followed by numbers used for intergenic lncRNAs (Volders et al. 2013; Wright and Bruford 2011).

lncRNA stability and turnover

After the study of the genome, it was found that the stability of about 7200 lncRNAs along with almost 20,000 protein-coding transcripts have a shorter turnover with a median of 3.5 hours compared with 5.1 hours for mRNAs. In this study, the half-lives of the lncRNAs ranged from 30 minutes to 48 hours and that of the protein-coding transcripts ranged from 2 hours to 12 hours (Niazi and Valadkhan 2012). This wide range of stability for lncRNA transcripts reflects the complexity of their metabolism and boundless functions (Clark et al. 2012; Wu et al. 2013). It was found that protein-coding genes with enzymatic and housekeeping functions were more stable as they did not involve any dynamic regulatory pathways. Since the regulatory elements are required for a short period, the proteins with regulatory activity have low stability. Correspondingly, lncRNAs involved in regulatory functions have shorter lifespans compared with housekeeping lncRNAs (Clark et al. 2012; Tani et al. 2012; Wu et al. 2013). A large number of lncRNAs are involved in the regulation of chromatin remodeling and gene expression in response to the external stimulus. The instability of lncRNAs with a regulatory function is paramount to their rapid dynamic regulation (Clark et al. 2012; Tani et al. 2012). Other factors such as subcellular localization, splicing, and genomic location also affects the stability of lncRNAs; this explains the greater stability of spliced lncRNAs than the single exon transcripts and lower stability of nuclear transcripts in line with their regulatory functions (Clark et al. 2012).

Protein coding potential of lncRNAs

To determine the protein-coding capability of lncRNAs, a number of experimental methods have been used (Ma et al. 2012). Further, different computational algorithms have been used to predict the protein-coding potential of lncRNAs and to determine the length of their ORFs. Another motive of these studies was to define the evolutionary conservation of lncRNA transcripts (Rinn and Chang 2012). These studies enlisted the shorter RNA transcripts (<300 nt) into the lncRNA group, since the ORF length of most of the eukaryotic protein-coding RNAs are longer than 300 nt or 100 amino acids. Studies have shown a significant difference between lncRNAs and mRNAs in their ORF length, with a median of 250 nt versus 1200 nt, respectively (Niazi and Valadkhan 2012), and also revealed that the start and stop codons in lncRNAs are distributed randomly among their genes (Ma et al. 2012). However, in eukaryotic cells some protein coding genes with an ORF shorter than 300 nt and lncRNAs with longer ORFs (>300 nt) have also been found (Ma et al. 2012). Therefore, other criteria like codon content, sequence context at -3 to +4 of the start codon, other protein domains, and so on were also employed to define the protein-coding capacity of these lncRNAs. The codon content of the lncRNA ORFs was found to be similar to the 3'UTRs of the protein-coding RNAs and to the intron codon alignment, indicating that lncRNAs and protein-coding RNAs have undergone evolutionary pressures through different pathways (Niazi and Valadkhan 2012). Further, it was shown that the ORFs of the longest lncRNAs lack the sequence context at -3 to +4 of the start codon indicating its translational efficiency (Niazi and Valadkhan 2012). lncRNA ORFs were also examined to detect known protein domains, which could be a sign for the protein-coding sequences (Volders et al. 2013). The details are reported on LNCipedia, a database for human long noncoding RNAs (<http://www.lncipedia.org>). Some other experimental methods such as ribosome profiling were used to identify the association of lncRNAs with ribosomes as well as their protein-coding potentials (Rinn and Chang 2012). Since all the susceptible translatable lncRNAs, which are associated with ribosomes, lack the evolutionary conservation of their proposed coding regions, such methods must be employed precisely (Rinn and Chang 2012). As a result, any interaction between ribosomes and lncRNAs does not necessarily result in translation. Further, it was shown that some of the ribosome-associated lncRNAs modulate the translation of protein-coding genes (Yoon et al. 2012). In conclusion, it is yet to be revealed whether lncRNAs possess any protein coding potential.

Genomic and cellular location of lncRNAs

lncRNAs can be found in either DNA strand or in both (Wang et al. 2011a). On the basis of genomic diversification with respect to protein-coding genes, lncRNAs can be classified into broad categories: for example, intergenic, intragenic, overlapping, and bidirectional lncRNA (figure 1). Sense lncRNAs are transcribed from loci, which overlap with a protein-coding gene; antisense lncRNAs are transcribed from loci that overlap with the antisense strand of the protein-coding gene; bidirectional lncRNAs are located on the antisense strand of a protein-coding gene whose start codon is <1000 base pairs away; intronic lncRNAs are transcribed from introns of protein-coding genes; and long intergenic noncoding RNAs (or intervening lncRNAs) loci are not in the proximity of a protein-coding gene at all (Da Sacco et al. 2012; Hauptman and Glavac 2013; Shahandeh 2013).

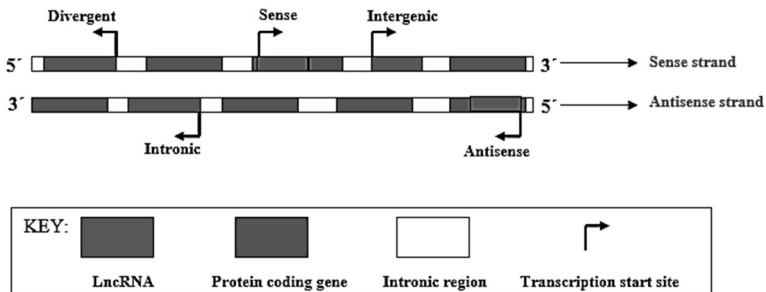


Figure 1. Different classes of long non-coding RNAs (lncRNAs). lncRNA genes can be classified into divergent, intronic, sense, antisense, and intergenic groups according to their location relative to the nearby protein-coding genes.

Additionally, these categories can be further classified into several subtypes. Prepared by the GENCODE project, among 14,880 lncRNAs, 9518 belong to intergenic lncRNAs and 5362 are from intragenic lncRNAs (Derrien et al. 2012). Intergenic lncRNAs are localized in the gene deserts, at least 5 kb away from protein-coding genes (Shahandeh 2013; Hauptman and Glavac 2013) and are also known as “large or long intergenic RNAs” (lincRNAs) (Knauss and Sun 2013). Enhancer RNA (eRNA) is a subclass of lincRNAs that is transcribed from enhancers and regulates the expression of its adjacent protein-coding genes (Wang and Chang 2011; Prensner and Chinnaiyan 2011). Beside, many other lincRNAs are transcribed from the gene-free chromosomal regions and regulate gene expression, both in cis

and trans (Knauss and Sun 2013). Intragenic lncRNAs reside within a protein-coding gene and can be further sub-classified into four groups: (i) Sense-exonic lncRNAs, which overlap with one or more exons of protein-coding genes in the same strand. (ii) Antisense exonic lncRNAs, also known as “natural antisense transcripts” (NATs), which overlap with the antisense strand of a protein-coding gene and that can be transcribed from a host gene or from a different genomic locus with the same sequence complementarity (Da Sacco et al. 2012; Hauptman and Glavač 2013; Papait et al. 2013). NATs regulate gene expression through multiple mechanisms, for example through forming endo-siRNAs (Knauss and Sun 2013). (iii) Intronic sense and antisense lncRNAs are derived from an intron of a protein-coding transcript from the same or the opposite strand. (iv) 3'UTRs and 5'UTRs associated with RNAs or uaRNAs are derived from untranslated regions of the protein-coding transcripts (Nie et al. 2012; Wu et al. 2013). Overlapping lncRNA transcripts contain a protein-coding gene within an intron on the same strand (Derrien et al. 2012; Cheng et al. 2013). These lncRNAs may regulate the promoter of the overlapping genes. To date, 52 overlapping-sense lncRNAs have been identified in the human genome (Knauss and Sun 2013). Bidirectional or divergent lncRNAs are located in the opposite strand of an adjacent promoter and reside in fewer than approximately 1000 bp of a protein-coding gene (Wu et al. 2013; Knauss and Sun 2013; Shahandeh 2013). These lncRNAs usually have the same CpG island promoter as adjacent protein-coding genes and they also regulate the expression of their neighboring protein-coding genes (Sun et al. 2013). Figure 1 provides a schematic representation of the locus and transcription direction of lncRNAs. By identifying the subcellular localization of lncRNAs, scientists could be partially able to determine their functions (Lipovich et al. 2010). Analysis of the ENCODE RNAseq data for nucleus and cytoplasm revealed that the nucleus is highly enriched with lncRNA and that the lncRNAs are significantly located in the chromatin fraction indicating their regulatory roles (Derrien et al. 2012). Some lncRNAs are located only in the cytoplasm, and others in both the nucleus and the cytoplasm (Wang et al. 2011a). Thus, lncRNAs possess extensive functional capacity, especially in gene and chromatin regulation (Lipovich et al. 2010). Moreover, these findings verified Paul and Duerksen’s research to some extent, in which it was reported that some parts of chromatin include RNAs twice as much as in DNAs. Figure 2 shows the cellular localization of lncRNAs.

Origin of lncRNAs

In eukaryotes, noncoding DNA has often been called “junk DNA,” because with no evolutionary restriction, it experiences natural selection (Ponting et al. 2009). Each group, based on the differences between the nature of coding and non-protein-coding genes is differently disposed to evolutionary constraints. In contrast to protein-coding genes, non-coding genes are under secondary structure conservation (Gingeras 2007; Kutter et al. 2012). Since non-coding DNAs and protein-coding DNAs reveal different evolutionary pathways, they possess different conservations (Kaessmann 2010). Compared with protein-coding genes, we know little about the origin and evolution of lncRNAs. They show a low sequence conservation and rapid evolution among mammals (Kutter et al. 2012). In addition to this, various evolutionary assumptions have been considered for the origin of lncRNAs. The first hypothesis is the metamorphosis of the protein coding gene generated through a gene-duplication process (Ponting et al. 2009; Kaessmann 2010; Kapusta et al. 2013). During evolution, one copy of a protein-coding gene is exposed to accumulate mutations and eventually loses its protein-coding capacity. Then, a new functional lncRNA gene is designed that caters to some of the previous coding fractions including regulatory elements, splicing signals, exon sequences, and polyadenylation sequences (Ponting et al. 2009; Kaessmann 2010). An X inactive-specific transcript (XIST) is essential for sex-chromosome dosage compensation in mammals and was originated from the chicken protein-coding *Lnx3* gene (Kaessmann 2010; Gribnau and Grootegoed 2012; Kapusta et al. 2013) and was then classified as a pseudo gene (Gerstein et al. 2007; Ponting et al. 2009). The XIST promoter region was originated from 5- UTR of *Lnx3* exons 1 and 2. Further, exons 4 and 5 of the human XIST gene were generated from *Lnx3* exons 4 and 11 (Elisaphenko et al. 2008). lncRNAs may be produced from segmental or whole-gene duplication of other lncRNA genes, which also include small and long non-coding RNAs (Ponting et al. 2009; Kapusta et al. 2013). The genomic examination of large homologous protein-coding gene families has shown that protein-coding gene duplication is a very common phenomenon; however, unlike protein-coding genes, there is very little evidence for the whole duplication of lncRNAs. This may be due to the rapid sequence divergence of lncRNAs (Ponting et al. 2009; Kaessmann 2010). Mouse nuclear-enriched abundant transcript 2 (*Neat2*) is an example of duplicated lncRNA that is paralogous to nonexonic sequences of the mouse genome (Ponting et al. 2009). Segmental gene duplication within lncRNA also seems to play a role in the generation of lncRNAs. Further, local and tandem duplications have been

observed in the 5 regions of *Kcnq1ot1* (Ponting et al. 2009; Kaessmann 2010). De novo generation is another possibility of the origins of lncRNAs (Kapusta et al. 2013). In this case, changes in the genomes such as chromosomal rearrangement and the generation of (proto-) splice sites and (proto-) promoters have transformed nonfunctional genomic sequences to functional lncRNAs (Kaessmann 2010). The last assumption considers the insertion of transposable elements (TEs) as the origin of lncRNAs (Ponting et al. 2009; Rinn and Chang 2012; Kapusta et al. 2013). In fact, TEs have given a lot, not only to the origin of new lncRNAs but also in terms of diversification and the regulation of vertebrate lncRNAs (Kapusta et al. 2013). Comparing lncRNAs with other genes like pseudo genes, small lncRNAs, and protein-coding genes, the majority of human lncRNAs have TE segments in their body. For instance, TEs are present within the internal part of exons, transcription start sites, polyadenylation (polyA) sites, or in the combination of all these parts. Studies have shown that 75% of human lncRNAs have at least one exon with partial TE origin. Most TE-based lncRNAs tend to have a few amounts of these sequences (Kapusta et al. 2013). Compared with other lncRNAs, intergenic lncRNAs are mostly covered by TEs (Rinn and Chang 2012). Four main classes of TEs are attributed to lncRNAs. However, LTR/ERV elements are the most abundant TE families among lncRNAs (Kapusta et al. 2013). TEs exert many important functions in lncRNA-related processes including the origin, diversification, and regulation of lncRNAs.

Role of TEs in lncRNA generation

The issue of generation of lncRNAs from transposable elements (TEs) is currently highly debated, thereby asking whether it is “lncRNA first” or “TE first”: divergence or emergence? However, the scientific studies confirm both models (Kapusta et al. 2013). Some studies revealed that about 56.5% of TE-derived lncRNAs show some signatures of TEs to their last exons thereby supporting the “lncRNA-first” model. There is also some evidence that supports the “TE-first” model of generation of lncRNAs including the existence of lncRNAs that are found to mostly or entirely consist of TEs, as it seems to be improbable to consider the insertion of all these TEs into nonTE ancestral lncRNAs. Also there are some lncRNAs that are found within the start region and the first exon of transcripts derived from TEs. Human Gencode v13 catalogued about 4404 lncRNAs with TE-derived TSS and 2213 lncRNAs, whose first exons are covered by TEs and derived from primate-specific TE families (Kapusta et al. 2013). Findings supported both models and also unveiled that 40% of TEs-derived human lncRNAs are

specific to primates and some are even restricted to hominoids (Kapusta et al. 2013). Therefore, these lncRNAs are considered to be young genes. Xist, which is developed from the Lnx3 pseudo gene and a set of diverse transposable elements, is one of the genes supporting both models. In this gene at least, a short fragment of the first exon, as well as exons 2, 3, 6, 7, 8, is reported to contain different types of TEs (Elisaphenko et al. 2008).

Role of TEs in the lineage-specific diversification of lncRNAs

Recently a large amount of evidence has evolved claiming a correlation between TEs' insertion in a regulatory region of genes and the emergence of lineage-specific lncRNAs. According to one study, about 36.3% of lncRNAs derived from TEs are found to be related to primate-specific lncRNAs (Kapusta et al. 2013). XIST, an lncRNA that evolved during the evolution of eutherians is found to have lineage-specific TEs like FLAM_C, lncRNA-RoR, and LTR7/HERVH, elements that cause varied expression of this lncRNA among certain species (Elisaphenko et al. 2008; Loh et al. 2011; Rinn and Chang 2012).

The role of TEs in the regulation of lncRNAs

Concordant TE insertion sites on lncRNA genes contribute to many aspects of lncRNA function and regulation. For instance, TEs contribute to transcription regulation of lncRNAs, their processing (e.g., splicing, alternative splicing, and polyadenylation) and the formation of lncRNAs' secondary structure formation (Kapusta et al. 2013; Hadjiargyrou and Delihis 2013). TEs that overlap with mapped DNaseI hypersensitive sites (DHS) indicate the active chromatin site and uncover the significant roles of TEs in lncRNA regulation. LTRs are revealed to be TEs that reside in the vicinity of the upstream region of lncRNAs. However, in protein-coding genes, SINES are the TEs located in the proximal upstream region of lncRNAs (Kapusta et al. 2013). The distribution of these TEs in the upstream regulator sites of protein and noncoding genes is not random. TEs take part in cell type-specific regulations of specific lncRNAs, since different TEs-derived transcription regulatory motifs on DNA recruit different types of transcription factors. Consequently, this results in different spatio-temporal expression patterns (Hadjiargyrou and Delihis 2013; Kapusta et al. 2013). TEs located in lncRNA transcripts are involved in their functional regulation, such as acting as cis-regulators or trans-regulators of protein-coding gene expressions. ANRIL is one of these

lncRNAs that regulates gene expression epigenetically. ANRIL binds to chromatin modifier factors and interacts with Alu supplementary motifs in promoters of ANRIL trans-regulating genes via its Alu motif (Holdt et al. 2013). One of the notable examples of gene expression transregulation is Alu-containing lncRNA originated and from chromosome 11. The Alu-containing lncRNAs bind to Alu sequences in 3'UTR of the target mRNAs forming specific secondary structures, recruiting Stau1 and subsequently UPF1. Their recruitments trigger target mRNA decay (Wang and Chang 2011; Hadjiargyrou and Delihis 2013).

Mechanism of action: lncRNAs as chromatin regulators

The number of lncRNAs with described functions is steadily increasing, and most of the reports revolve around their regulatory capacity. LncRNAs often function as important cis- and trans-acting modulators for the expression of protein-coding genes (Mao et al. 2011; Fejes-Toth et al. 2009). LncRNAs can mediate epigenetic modification by recruiting a chromatin-remodeling complex to a specific chromatin locus (figure 2A). At least 38% of the lncRNAs present in several tissues bind to the polycomb repressive complex 2 or the chromatin-modifying proteins, CoREST and SMCX (Fejes-Toth et al. 2009). Others bind to trithorax chromatin-activating complexes and/or activated chromatin (Kapranov et al. 2010). The well-characterized lncRNAs, ANRIL, XIST, HOTAIR, and KCNQ1OT1 are able to recruit epigenetic modifiers to specific loci for reprogramming the chromatin state. For example, KCNQ1OT1 binds to PRC2 and the methyltransferase G9A (also known as EHMT2), whereas ANRIL binds to PRC1 and PRC2 (Lee et al. 2011; Mak et al. 2004). HOTAIR and other lncRNAs function as scaffolds (figure 2B) coordinating the targeting of specific repressive histone-modifying complexes to target loci (Lee et al. 2011). However, within this framework, the detailed mechanism of how specific DNA regions are targeted by lncRNAs remains unclear. An example of such principle approaches working together is the XIST locus that controls the X-chromosome dosage compensation. XIST, in many ways can be considered as a model example of lncRNA biology, and the dissection of its mechanism of action can serve as a guide for deeper and more thorough future lncRNA studies. Principally, X chromosome inactivation is due to the ablation of the DNA and the promoter of XIST (Mak et al. 2004; Brown et al. 1991); nevertheless, such models cannot undermine the effect of the local act of transcription.

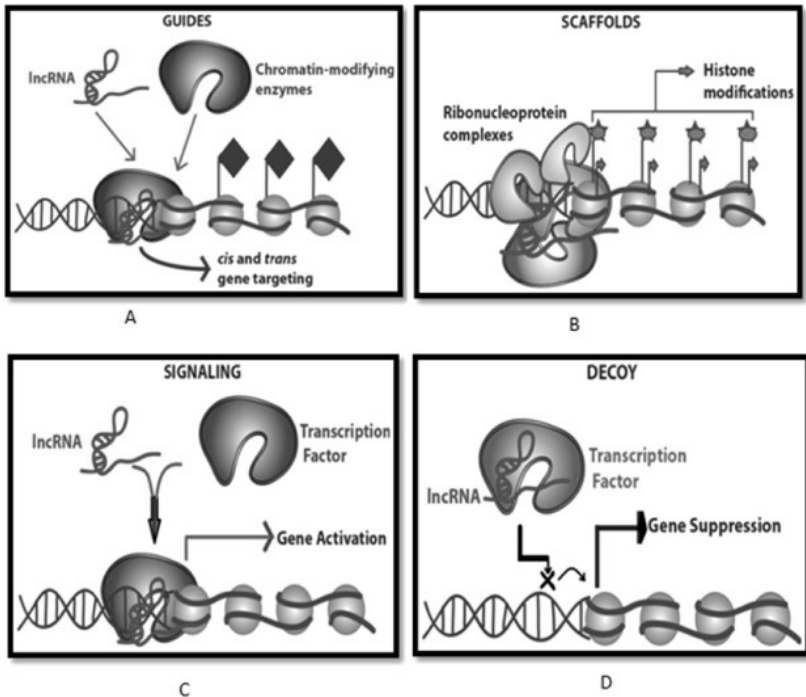


Figure 2. LncRNA-mediated transcriptional regulation. (Reproduced from our publication Bhat et al. 2016, *Noncoding RNA Research*).

Long non-coding RNAs (lncRNAs; red solid lines) regulate gene transcription through three main mechanisms: (A) Interaction with and recruitment of chromatin-modifying enzymes (e.g., histone methylases, acetylases, and deacetylases) to the target gene locus. Modulation of the chromatin state by these enzymes leads to activation or repression of local genes. (B) Interaction with other RNA-binding factors such as hnRNPs forms RNA-protein complexes (RNPs). RNPs can either promote transcription by recruiting key proteins to the target gene promoters or repress gene transcription by binding to existing gene repressors. (C) LncRNAs also have enhancer functions and help change the chromatin architecture and recruit transcriptional machinery proteins to the adjacent target gene locus to promote its transcription. (D) LncRNAs are also involved in the repression of certain pro-apoptotic genes, such as FAS and BIK, by acting as a decoy for the transcription factor (NF- κ B).

Transcriptional regulation

Long non-coding RNAs act as co-factors to modify the activity of the transcriptional factor. For example, the ncRNA *Evf2* is transcribed from the conserved distal enhancer and recruits the transcription factor (figure 2C) *DIX2* to the same enhancer to induce expression of adjacent protein-coding genes (Zhao et al. 2008). Various developmental genes are regulated in a similar fashion by transcribing the enhancers in the cells in which they are active (Clemson et al. 1996). LncRNAs can modify RNA polymerase (RNAP) II activity by interplaying with the initiation complex to steer promoter choice; for example, it inhibits the binding of the transcriptional co-factor TFIID31 in human transcription of a ncRNA from an upstream region of the dihydrofolate reductase (*DHFR*) locus to form a triplex in the major promoter of *DHFR*. This could be a general mechanism for controlling the usage of the promoter, as thousands of these triplex structures exist in eukaryotic chromosomes (Sado et al. 2005). Long lncRNAs may also affect global changes by interacting with some basic components of the RNAP II-dependent transcription machinery. LncRNAs interacting with RNAP II machinery are primarily transcribed by RNAP III, thereby delinking their expression from the RNAP II-dependent transcription reaction they regulate. For example, in the transcription of Alu elements in response to heat shock, they bind tightly to RNAP II to rule out the formation of active pre-initiation complexes (Wan et al. 2008). Alu elements can independently mediate polymerase binding and repression by their respective domain interaction. Due to their abundance and distribution in the mammalian genome, these functional domains might have been co-operated into other ncRNAs during evolution (Edwards et al. 2007).

Post-transcriptional regulation

The ability of ncRNAs to identify complementary sequences allows some specific interactions capable of regulating post-transcriptional processing of mRNAs like capping, splicing, editing, transport, translation, degradation, and stability at various control sites. For example, *MALAT1* affects alternate splicing by interacting with splicing factors. Another lncRNA, *Gomafu/MIAT*, which is localized to a nuclear domain and has a neuron-specific expression, may block spliceosome formation and affect mRNA splicing by sequestering splicing factor 1 (SF1) (He et al. 2014; Pandey et al. 2008). The NAT lncRNAs also play an essential role in regulating mRNA dynamics (figure 2D). About 61–72% of all transcribed regions possess lncRNAs in an antisense orientation (NATs) (Kanduri et al. 2006). Unlike

classical NAT lncRNAs in the imprinting genes (Tsix, Air, HOTAIR, and Evf-2) that are responsible for recruiting repressor complexes like PRC2 to the target site, certain NATs result in RNA duplexes to inhibit cis-regulatory elements, leading to an alternate splicing pattern of the paired gene. For example, the Zeb2/Sip1 NAT complementary binds to the 5' splice site of an intron in the 5'-UTR of the zinc finger Hox mRNA Zeb2, which plays a crucial role in epithelial-mesenchymal transition (EMT). Expression of the Zeb2 NAT upon EMT masks the splice site, hence preventing spliceosome action. Subsequently, the translation machinery can then recognize and bind to an internal ribosome entry site (IRES), resulting in more efficient Zeb2 translation.

The structural and functional diversity of lncRNAs, as well as their sheer number, make their characterization challenging: the NONCODE database, version 4, contains 95,135 human lncRNA transcripts (Mercer et al. 2013). The functional array of mechanisms of lncRNAs is diverse, ranging from scaffolds to decoys or signals (Xie et al 2014) that act either locally to regulate nearby genes in cis or remotely, that is, over one megabase away or on a different chromosome, in trans, to control the transcription of target genes (Klattenhoff et al. 2013). LncRNAs could directly silence or activate gene expression, or indirectly regulate the chromatin states of their target genes. Indeed, many lncRNAs are very much like mRNAs: they are transcribed by RNA polymerase II (Pol II) from genomic loci with similar chromatin states to mRNAs; they are often 5'-capped, spliced, and polyadenylated; and in most instances, they lack any biochemical distinction from mRNAs beside the absence of a translated ORF. As a class, lncRNAs run the gamut from mRNA-like to truly exotic, such as circular RNAs (He et al. 2013), lncRNAs spanning 100 kb (Bell et al. 2014), and lncRNAs with a restricted subnuclear localization (Wang et al. 2014a). When compared with mRNAs, lncRNAs are more enriched in the nucleus relative to the cytoplasm and within the nucleus; they occupy the chromatin fraction (17% of lncRNAs versus 15% of mRNAs are enriched in the nucleus, whereas 4% versus 26%, respectively, are enriched in the cytoplasm). This observation agrees with a report that lncRNA-read density is higher in the nuclear fraction. Nuclear lncRNAs can recruit sequestering proteins, such as transcription factors or histone-modifying enzymes, to modify the transcription from target loci; and cytoplasmic lncRNAs regulate protein localization, mRNA translation, and their stability in the cytoplasm. Many lncRNA species have very well-defined subcellular localizations, including XIST (on the inactive X), Gomafu (also known as MIAT; subnuclear domains), BORG (restricted to the nucleus), and GAS5 (exported to the cytoplasm). Several degradation pathways preferentially

act on lncRNAs, including subsequent 5'-to-3' exonuclear degradation via Xrn1, or by nonsense-mediated decay (Wang et al. 2014b)

lncRNAs are involved in numerous important biological phenomena such as imprinting genomic loci, shaping chromosome conformation, and allosterically regulating enzymatic activity and pathological mechanisms. The over-expression, deficiency, or mutation of lncRNA genes has been implicated in numerous diseases. By now, although the function of most lncRNAs is yet to be discovered, several lncRNAs have been clearly shown to play critical roles through their molecular mechanisms in various cellular processes, which include differentiation, growth, apoptosis, development, and tumorigenesis (Zangrando et al. 2014). Increasing evidence firmly supports the fact that lncRNAs may act as diagnostic markers or therapeutic targets in many diseases. For example, lncRNA H19 acts as a carcinogenic gene and was found to be involved in gastric cancer, colorectal cancer, and glioma. HOTAIR mediates epigenetic silencing of another HOX cluster on a different chromosome via binding PRC2. Its over-expression in breast cancer, hepatocellular carcinoma, colorectal cancer, and pancreatic cancer is associated with poor prognosis and tumor metastasis as well as invasion ability in gastrointestinal stromal tumors. It has been elucidated that there is a relationship between ANRIL and increased susceptibility to several types of cancer such as acute lymphoblastic leukemia, glioma, basal cell carcinoma, nasopharyngeal carcinoma, breast cancer, and plexiform neurofibromas (Ounzain et al. 2014). Another strong study proves that MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) might be a screening biomarker for lung cancer, uterine endometrial stromal sarcoma, cervical cancer, and hepatocellular carcinoma because of its association with metastasis. It promotes the motility of lung cancer cells through transcriptional or post-transcriptional regulation of motility-related genes. The lncRNA SLINKY has been found to be a robust prognostic biomarker in clear cell renal cell carcinomas (ccRCC), where it has been shown to function together with HNRNPK in cancer cell proliferation (Yang et al. 2014). A study also suggested that the dysregulation of lncRNA SNHG6 is highly cancer-type specific (Kumarswamy et al. 2014).

The lncRNA BACE1AS has been reported to play a vital role in the etiology of Alzheimer's disease (AD). Various lncRNAs including the Gas5 transcript is involved in host-pathogen interactions (Vausort et al. 2014). Although research on lncRNAs in endocrine diseases remains limited, the important roles of several genes in the metabolism and endocrine system have been reported, such as the PTEN-induced gene PINK1, which has been shown to be associated with diabetic status. The