Advances in Experimental Medicine and Biology 722

Lesley J. Collins *Editor*

RNA Infrastructure and Networks

LANDES BIOSCIENCE

Springer

RNA Infrastructure and Networks

ADVANCES IN EXPERIMENTAL MEDICINE AND BIOLOGY

Recent Volumes in this Series

Volume 713
CELL FUSION IN HEALTH AND DISEASE, I: CELL FUSION IN HEALTH
Thomas Dittmar

Volume 714
CELL FUSION IN HEALTH AND DISEASE, II: CELL FUSION IN DISEASE
Thomas Dittmar

Volume 715
BACTERIAL ADHESION: BIOLOGY, CHEMISTRY AND PHYSICS
Dirk Linke and Adrian Goldman

Volume 716
MAST CELL BIOLOGY: CONTEMPORARY AND EMERGING TOPICS
Alasdair M. Gilfillan and Dean D. Metcalfe

Volume 717
KAINATE RECEPTORS: NOVEL SIGNALING INSIGHTS
Antonio Rodriguez-Moreno and Talvinder S. Sihra

Volume 718
BRAIN INSPIRED COGNITIVE SYSTEMS 2010
Ricardo Sanz, Jaime Gomez and Carlos Hernandez

Volume 719
HOT TOPICS IN INFECTION AND IMMUNITY IN CHILDREN VIII
Nigel Curtis

Volume 720
HUMAN CELL TRANSFORMATION
Johng S. Rhim and Richard Kremer

Volume 721
SPHINGOLIPIDS AND METABOLIC DISEASE
L. Ashley Cowart

Volume 722
RNA INFRASTRUCTURE AND NETWORKS
Lesley J. Collins

RNA Infrastructure and Networks

Edited by

Lesley J. Collins
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Springer Science+Business Media, LLC

Landes Bioscience

Springer Science+Business Media, LLC
Landes Bioscience

Printed in the USA.

Springer Science+Business Media, LLC, 233 Spring Street, New York, New York 10013, USA
http://www.springer.com

Please address all inquiries to the publishers:
Landes Bioscience, 1806 Rio Grande, Austin, Texas 78701, USA
Phone: 512/ 637 6050; FAX: 512/ 637 6079
http://www.landesbioscience.com

The chapters in this book are available in the Madame Curie Bioscience Database.
http://www.landesbioscience.com/curie

RNA Infrastructure and Networks, edited by Lesley J. Collins. Landes Bioscience / Springer Science+Business Media, LLC dual imprint / Springer series: Advances in Experimental Medicine and Biology.

ISBN: 978-1-4614-0331-9

Library of Congress Cataloging-in-Publication Data

RNA infrastructure and networks / edited by Lesley J. Collins.
p. ; cm. -- (Advances in experimental medicine and biology ; v. 722)
Includes bibliographical references and index.
ISBN 978-1-4614-0331-9
1. RNA. I. Collins, Lesley J., 1968- II. Series: Advances in experimental medicine and biology ; v. 722. 0065-2598
[DNLM: 1. RNA. W1 AD559 v.722 2011 / QU 58.7]
QP623.R5727 2011
579.2'5--dc23

2011018033

DEDICATION

This book is dedicated to all my family, especially Maurice and Shannen, the central hub of my RNA world.

PREFACE

Previously, RNA was investigated merely as an intermediate between DNA and proteins. Studies of regulatory pathways led to the discovery that small RNAs are major regulators in animal cells, and other non-coding RNAs responsible for RNA cleavage and modification. Now studies reveal that small RNA regulation exists throughout eukaryotes and prokaryotes, and long non-coding RNAs exert a large influence over epigenetics. We can also see that the RNA components of the cell are not merely 'fixed accidents' of an ancient RNA world, but instead are continuing to evolve and affect the metabolism of all life.

RNAs form complexes with proteins and other RNAs. The RNA-infrastructure represents the spatiotemporal interaction of these proteins and RNAs in a cell-wide network. This volume brings together these ideas to illustrate the scope of RNA-based biology, and how connecting RNA mechanisms is a powerful tool to investigate regulatory pathways. The first chapter is an introduction to the RNA-infrastructure and how RNAs and proteins interact in networks. Following this is a chapter on RNA interactions, explaining how RNA folds upon itself, and then interacts with other RNAs as well as DNAs and proteins.

The second section focuses on RNA interference (RNAi), where siRNA is shown to be important in defending plants against viruses. We then explore how miRNAs act in the regulation of many gene networks, then the spatiotemporal aspects of miRNA-mediated gene regulation. This networking is further investigated with two specific examples; (i) spliceosomal RNA infrastructure and (ii) RNA-binding proteins, expanding on traditional analysis of protein-protein interaction networks to include regulatory RNAs as well as other interacting RNAs.

The third section describes some of the lesser understood RNA mechanisms. First we examine how post-transcriptional control is reinforced by RNA protein complexes (RNPs), responsible for controlling the abundance of gene expression during development. The next chapter summarizes how non-random tRNA fragments can guide mRNA cleavage, inhibit translation and promote morphological changes. A chapter on programmed DNA elimination in the protozoan ciliate *Tetrahymena* then describes a mechanism for RNAi-directed heterochromatin formation. Long non-coding RNAs are next introduced

in their context of epigenomics, and a promoter-associated long non-coding RNA that binds to the TLS protein, is examined in more detail.

RNA networks are also found throughout prokaryotes, and the fourth section reviews prokaryotic viral defence (CRISPRs), regulation (riboswitches and small RNAs) and other RNAs surrounding tRNA processing, as well as the spatiotemporal nature of the prokaryotic RNA infrastructure.

The last section reviews some of the latest technology that has greatly increased our knowledge of RNAs including small RNA discovery using high-throughput approaches. We conclude by discussing evolutionary aspects of RNA networking and examining the mechanisms we see today, how such RNA-mechanisms evolved and whether some date back to the ancient RNA world.

This work is but a taste of the wide range of RNA-based mechanisms that connect in the RNA infrastructure. We have only begun to untangle this complex cellular web, and have still an awful lot to learn about the cell.

Lesley J. Collins

ABOUT THE EDITOR...

Photo Credit : Robin Atherton, New Zealand.

LESLEY J. COLLINS is a Senior Research Fellow in the Institute of Fundamental Sciences at Massey University, New Zealand. With an early background in molecular biology she now uses that knowledge in evolutionary genomics and bioinformatics. Her research interests focus on the evolution of RNA networks, especially on how they relate to the evolution of ancient eukaryotes. With the rise of new sequencing technology she was involved in the establishment of next generation sequencing at Massey University. In her spare time she is a bioinformatics consultant, a practitioner of Tai Chi, a knitter and an avid reader.

PARTICIPANTS

Patrick J. Biggs
Institute of Veterinary, Animal
and Biomedical Sciences
Massey University
Palmerston North
New Zealand

Věra Čapková
Laboratory of Pollen Biology
Institute of Experimental Botany AS CR
Rozvojová
Czech Republic

X. Sylvia Chen
Department of Biochemistry
University of Otago
Dunedin
New Zealand

Lesley J. Collins
Institute of Fundamental Sciences
Massey University
Palmerston North
New Zealand

Tamas Dalmay
School of Computing Sciences
University of East Anglia
Norwich
UK

Toni Daly
Allan Wilson Centre for Molecular
Ecology and Evolution
Massey University
Palmerston North
New Zealand

Said Hafidh
Laboratory of Pollen Biology
Institute of Experimental Botany AS CR
Rozvojová
Czech Republic

David Honys
Laboratory of Pollen Biology
Institute of Experimental Botany AS CR
Rozvojová
and
Department of Plant Experimental Biology
Faculty of Science
Charles University in Prague
Viničná
Czech Republic

Rebecca L. Hurto
Department of Molecular Genetics
The Ohio State University
Columbus, Ohio
USA

Sarath Chandra Janga
MRC Laboratory of Molecular Biology
Cambridge
UK
and
Institute of Genomic Biology
University of Illiniois at Urbana-Champaign
Urbana, Illinois
USA

Chandrasekhar Kanduri
Department of Genetics and Pathology
Uppsala University
Uppsala
Sweden

Kensuke Kataoka
Institute of Molecular Biotechnology of the Austrian Academy of Sciences
Vienna
Austria

Kenneth C. Keiler
Department of Biochemistry and Molecular Biology
Penn State University
Althouse Laboratory
University Park, Pennslyvania
USA

Riki Kurokawa
Division of Gene Structure and Function
Research Center for Genomic Medicine
Saitama Medical University
Yamane, Hidaka-shi, Saitama-Ken
Japan

Manja Marz
Bioinformatics Group
Department of Computer Science and Interdisciplinary Center for Bioinformatics
University of Leipzig
Leipzig
and
RNA-Bioinformatics Group
Institut für Pharmazeutische Chemie
Philipps-Universität Marburg
Marburg
Germany

Nitish Mittal
Department of Biotechnology
National Institute of Pharmaceutical Education and Research
SAS Nagar
Punjab
India

Kazufumi Mochizuki
Institute of Molecular Biotechnology of the Austrian Academy of Sciences
Vienna
Austria

Vincent Moulton
School of Computing Sciences
University of East Anglia
Norwich
UK

Simon Moxon
School of Computing Sciences
University of East Anglia
Norwich
UK

Helio Pais
School of Computing Sciences
University of East Anglia
Norwich
UK

Vitantonio Pantaleo
Istituto di Virologia Vegetale del CNR
Torino
Italy

David Penny
Institute of Molecular BioSciences
Massey University
Palmerston North
New Zealand

Joris Pothof
Department of Genetics
Erasmus MC
University Medical Center
Netherlands

Peter F. Stadler
Bioinformatics Group
Department of Computer Science
and Interdisciplinary Center
for Bioinformatics
University of Leipzig
Leipzig
Germany;
Institute for Theoretical Chemistry
University of Vienna
Wien
Austria;
Fraunhofer Institute for Cell Therapy
and Immunology
Leipzig
Germany;
and
Santa Fe Institute
Santa Fe, New Mexico
USA

Swathi Vallabhaneni
Institute of Genomic Biology
University of Illiniois at Urbana
Champaign
Urbana, Illinois
USA

Dik C. van Gent
Department of Genetics
Erasmus MC
University Medical Center
Rotterdam
Netherlands

CONTENTS

1. THE RNA INFRASTRUCTURE: AN INTRODUCTION TO ncRNA NETWORKS 1

Lesley J. Collins

Abstract 1
Introduction 1
RNAs Processing Other RNAs 2
Spatial Regulation of Eukaryotic RNA Processing 7
RNA Regulation, Connecting Components of the RNA-Infrastructure 9
Conclusion 15

2. RNA INTERACTIONS 20

Manja Marz and Peter F. Stadler

Abstract 20
Introduction 20
ncRNA-DNA Interactions 21
RNA-RNA Interactions 23
ncRNA-Protein Interactions 31
Conclusion 33

3. PLANT RNA SILENCING IN VIRAL DEFENCE 39

Vitantonio Pantaleo

Abstract 39
Introduction 39
Generation of Viral Small Interfering RNAs: Viruses, vsiRNAs and Dicers 40
Post-Dicing Modification of vsiRNAs 47
Antiviral Effector Complexes 47
Amplification and Spread of Antiviral Defence 49
Movement of Antiviral Silencing 51
Viral Suppressors of RNA Silencing 52
Conclusion 53

4. MicroRNAS AS POST-TRANSCRIPTIONAL MACHINES AND THEIR INTERPLAY WITH CELLULAR NETWORKS 59

Sarath Chandra Janga and Swathi Vallabhaneni

Abstract 59
Introduction 60
Biogenesis, Genomic Organization and Evolution of miRNAs 61
Methods to Identify miRNAs and Their Targets 63
MicroRNA Interplay with Other Cellular Networks 65
Expression Dynamics and Crosstalk of miRNAs with Other Regulatory Factors 69
Conclusion 70

5. SPATIOTEMPORAL ASPECTS OF MicroRNA-MEDIATED GENE REGULATION 75

Joris Pothof and Dik C. van Gent

Abstract 75
Introduction 75
DNA Damage Response as an Example of Multilayered Regulation 77
MicroRNA-Mediated Gene Silencing in the DDR 78
Mechanism of Fast MicroRNA Regulation 81
Localization of the MicroRNA-Mediated Gene Silencing Machinery 82
Conclusion 84

6. SPLICEOSOMAL RNA INFRASTRUCTURE: THE NETWORK OF SPLICING COMPONENTS AND THEIR REGULATION BY miRNAS 86

Lesley J. Collins

Abstract 86
Introduction 86
The Spliceosome as an Example of the RNA-Infrastructure 87
RNA-Protein Interaction Networks 89
Regulatory (Expression Modulated) Networks 93
Spatiotemporal Dynamic RNA Networks 98
Conclusion 100

7. CONSTRUCTION, STRUCTURE AND DYNAMICS OF POST-TRANSCRIPTIONAL REGULATORY NETWORK DIRECTED BY RNA-BINDING PROTEINS 103

Sarath Chandra Janga and Nitish Mittal

Abstract 103
Introduction 104
RNA Binding Proteins and Post-Transcriptional Regulation 104
Methods to Identify RBPs and Their Targets 107
RBPs and Post-Transcriptional Operons 109
Post-Transcriptional Network Formed by RBPs 110

Expression Dynamics of RBPs in Post-Transcriptional Networks **112**
Conclusion **114**

8. SAFE KEEPING THE MESSAGE: mRNP COMPLEXES TWEAKING AFTER TRANSCRIPTION **118**

Said Hafidh, Věra Čapková and David Honys

Abstract **118**
Introduction **119**
RNP-Infrastructure; Post-Transcriptional Modulator of Gene Expression **119**
Pre-mRNA Processing and the Initiation of mRNP Formation **122**
Mobilization to the Site of Storage and mRNA Localization **125**
Cytoskeleton: Actin and Microtubule Dynamics as a Floating Raft for mRNPs Subcellular Localization **126**
Intra- and Intercellular Translocation of mRNPs **126**
Storage of mRNPs and mRNA Turnover **127**
Systematic Translation Repression by Small RNAs **131**
Reinitiating Translation of the Localized, Translationally Sequestered mRNAs **132**
Conclusion **133**

9. UNEXPECTED FUNCTIONS OF tRNA AND tRNA PROCESSING ENZYMES **137**

Rebecca L. Hurto

Abstract **137**
Introduction **137**
tRNAs and Transcription **138**
tRNA Fragments **140**
Aminoacyl-tRNA Functions Beyond Translation **147**
tRNA Processing Enzymes Affect More Than tRNA **148**
Aminoacyl-tRNA Synthetases, tRNA and Disease **149**
Conclusion **151**

10. PROGRAMMED DNA ELIMINATION IN *TETRAHYMENA*: A SMALL RNA-MEDIATED GENOME SURVEILLANCE MECHANISM **156**

Kensuke Kataoka and Kazufumi Mochizuki

Abstract **156**
Introduction **157**
Nuclear Dimorphism in *Tetrahymena* **159**
Epigenetic Regulation of DNA Elimination by Small RNAS **163**
Biogenesis of scnRNA **163**
scnRNA Selection **165**
scnRNA-Induced Heterochromatin Formation and DNA Elimination **166**
Evolutionary Considerations **167**
Future Prospects **169**
Conclusion **170**

11. LONG NONCODING RNA AND EPIGENOMICS 174

Chandrasekhar Kanduri

Abstract 174
Introduction 174
Long ncRNAs and Tumor Suppressor Genes 176
Long ncRNAs and Homeobox Genes 177
Long ncRNAs and Dosage Compensation 179
Long ncRNAs and Pluripotency 184
Long ncRNAs and Genomic Imprinting 186
Long ncRNAs and Centromeric Heterochromatin 189
Conclusion 191

12. PROMOTER-ASSOCIATED LONG NONCODING RNAs REPRESS TRANSCRIPTION THROUGH A RNA BINDING PROTEIN TLS 196

Riki Kurokawa

Abstract 196
Introduction 197
The Target RNA Sequences Recognized by TLS 197
Domain Structure of TLS 200
Transcription-Regulatory Long ncRNA, Cyclin D1 pncRNA 203
Conclusion 206

13. RNA NETWORKS IN PROKARYOTES I: CRISPRs AND RIBOSWITCHES 209

Patrick J. Biggs and Lesley J. Collins

Abstract 209
Introduction 209
Riboswitches Control Gene and Operon Expression Via Metabolite Binding 210
The CRISPR/Cas System is Analogous, but Not Homologous, to Eukaryotic RNAi 213
Conclusion 218

14. RNA NETWORKS IN PROKARYOTES II: tRNA PROCESSING AND SMALL RNAs 221

Lesley J. Collins and Patrick J. Biggs

Abstract 221
Introduction 221
RNase P and tRNA Processing in Bacteria and Archaea 222
Bacterial Small RNAs and Sigma Factors 224
Transfer Messenger RNA 227
Conclusion 228

15. LOCALIZATION OF THE BACTERIAL RNA INFRASTRUCTURE 231

Kenneth C. Keiler

Abstract..... 231
Introduction..... 231
Localized Transcription and mRNA 232
Spatial Control of RNA Degradation and Processing..... 233
EF-Tu is a Component of the Cytoskeleton 235
Localization of *Trans*-Translation Components..... 235
Spatial Control of Gene Expression 236
Conclusion 236

16. SMALL RNA DISCOVERY AND CHARACTERISATION IN EUKARYOTES USING HIGH-THROUGHPUT APPROACHES.....239

Helio Pais, Simon Moxon, Tamas Dalmay and Vincent Moulton

Abstract..... 239
Introduction..... 239
The Mechanisms of RNA Silencing 240
Using Microarrays and DNA Sequencing to Measure RNA 243
High-Throughput Approaches for the Discovery and Functional Classification of sRNAs 244
Conclusion 250

17. HOW OLD ARE RNA NETWORKS? 255

Toni Daly, X. Sylvia Chen and David Penny

Abstract..... 255
Introduction..... 255
Regulatory Networks of Small RNAs..... 256
RNA Regulation and Defence Against the Dark Arts..... 259
Other Regulatory RNAs..... 266
How Old are the Different Interactions of RNA? 267
Conclusion 269

INDEX..... 275

ACKNOWLEDGEMENTS

I would like to acknowledge the wonderful assistance from David Penny for mentoring, highly stimulating conversation and funding to delve deep into the Pandora's Box we called the RNA infrastructure. This work was also assisted greatly by the Institute of Fundamental Sciences at Massey University. Not in the very least, my greatest thanks to all the colleagues who contributed, reviewed and revised chapters to help me bring together this book that highlights our modern RNA World.

CHAPTER 1

THE RNA INFRASTRUCTURE: An Introduction to ncRNA Networks

Lesley J. Collins
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
Email: l.j.collins@massey.ac.nz

Abstract: The RNA infrastructure connects RNA-based functions. With transcription-to-translation processing forming the core of the network, we can visualise how RNA-based regulation, cleavage and modification are the backbone of cellular function. The key to interpreting the RNA-infrastructure is in understanding how core RNAs (tRNA, mRNA and rRNA) and other ncRNAs operate in a spatial-temporal manner, moving around the nucleus, cytoplasm and organelles during processing, or in response to environmental cues. This chapter summarises the concept of the RNA-infrastructure, and highlights examples of RNA-based networking within prokaryotes and eukaryotes. It describes how transcription-to-translation processes are tightly connected, and explores some similarities and differences between prokaryotic and eukaryotic RNA networking.

INTRODUCTION

RNA biology in both eukaryotes and prokaryotes exists in a spatiotemporal network of processes we call the RNA-infrastructure. In eukaryotes, there are numerous subtypes of noncoding (nc) RNA genes involved including rRNA, mRNA, tRNA, snRNA, snoRNAs, several classes of regulatory RNAs (RNAi) and many long ncRNAs. In prokaryotes, in addition to tRNAs, mRNAs and rRNAs, we can have small RNAs, CRISPRs and tmRNAs, and even viruses can contain small RNAs. ncRNAs are generally involved in the transcription-to-translation processes surrounding the conversion and regulation of information from DNA to protein, implicated in viral defence mechanisms, or are involved in gene regulation (e.g., RNA interference; RNAi). What we are only beginning

RNA Infrastructure and Networks, edited by Lesley J. Collins.

to understand is how these processes are integrated, and how RNA plays a previously understated role in the overall regulation of the cell.

There are some key differences in cells that are differentiated (i.e., from multicellular eukaryotes), single celled eukaryotes and prokaryotes, but there are also striking similarities in how RNA processing and regulation works in different types of cells, giving us clues to their evolution. Although we are more familiar with RNA networks from eukaryotes, prokaryotic noncoding RNA research is using concepts developed from eukaryotic work to discover new RNA-based systems in bacteria and archaea. Although finding RNA genes is becoming a standard step in genomic investigations, it is now clear that discovering connections between these genes, and their associated proteins is just as important. Once we add in regulatory and epigenetic elements (such as methylation and histone modification) our regulatory networks can become very complex, but these complex networks have the ability to indicate linkages between cellular machineries not previously observed. The examples in this chapter will show how RNA-based processes within both prokaryotic and eukaryotic cells interact in networks in both a spatial and temporal manner.

RNAs PROCESSING OTHER RNAs

A good example in how RNA-processes are connected comes from examining the transcription-to-translation processes which form the core of the RNA-infrastructure (Fig. 1).[1] The processing of the three core-RNAs in eukaryotes (mRNA, tRNA and rRNA) includes the RNA-based mechanisms of RNA cleavage and modification (Fig. 1A). In eukaryotes these are: rRNA by RNase MRP and snoRNAs; tRNA by RNaseP; and mRNAs spliced by snRNAs within the spliceosome. We can then expand this idea to include spatial movement and regulation during RNA-processing (Fig. 1B). In prokaryotes we still have tRNAs and rRNAs being processed either directly by RNAs (e.g., RNase P of tRNA and tmRNA) or indirectly (where rRNAs are released by tRNA processing) (Fig. 1C).

Examining the connections between these processes in more detail we see networking between different mRNA machineries. For example, transcription by RNA Polymerase II (Pol II) and mRNA splicing in mammals are carried out in close proximity,[2] and this coupling may protect the newly synthesised RNA from degradation[3] before the termination of transcription.[4,5] Some splicing may occur cotransciptionally and this significantly improves processing efficiency (reviewed in ref. 6). At the other end of the transcript, 3′-end cleavage and polyadenylation of mRNA can be promoted by splicing proteins (e.g., U2AF65 reviewed in ref. 7). It is clear that splicing (the processing of mRNA with snRNAs) connects to other mRNA-processes including RNA localisation, translational yield and mRNA decay.[8] In another example, the Exon Junction Complex (EJC) is a set of proteins deposited on 5′ end of the exon during the second step of the splicing cycle, and remain bound to the spliced mRNA as it is exported to the cytoplasm.[9] This complex interacts transiently with many factors that connect the mRNA to the downstream RNA processing network,[10,11] as it is a major link between mRNA-splicing and mRNA export, as well as having a potential role in RNA degradation. The EJC appears to relay the previous location of introns,[8] and thus detects incorrect splicing that introduces premature stop-codons. It has been shown that in mammals at least, spliceosomal proteins and especially those involved in exon-definition, remain associated with the pre-mRNA to be available for the splicing of the next introns. This allows for efficient splice site recognition for subsequent introns since splice site recognition only needs to be carried out once for a site.[6] With splicing central to downstream

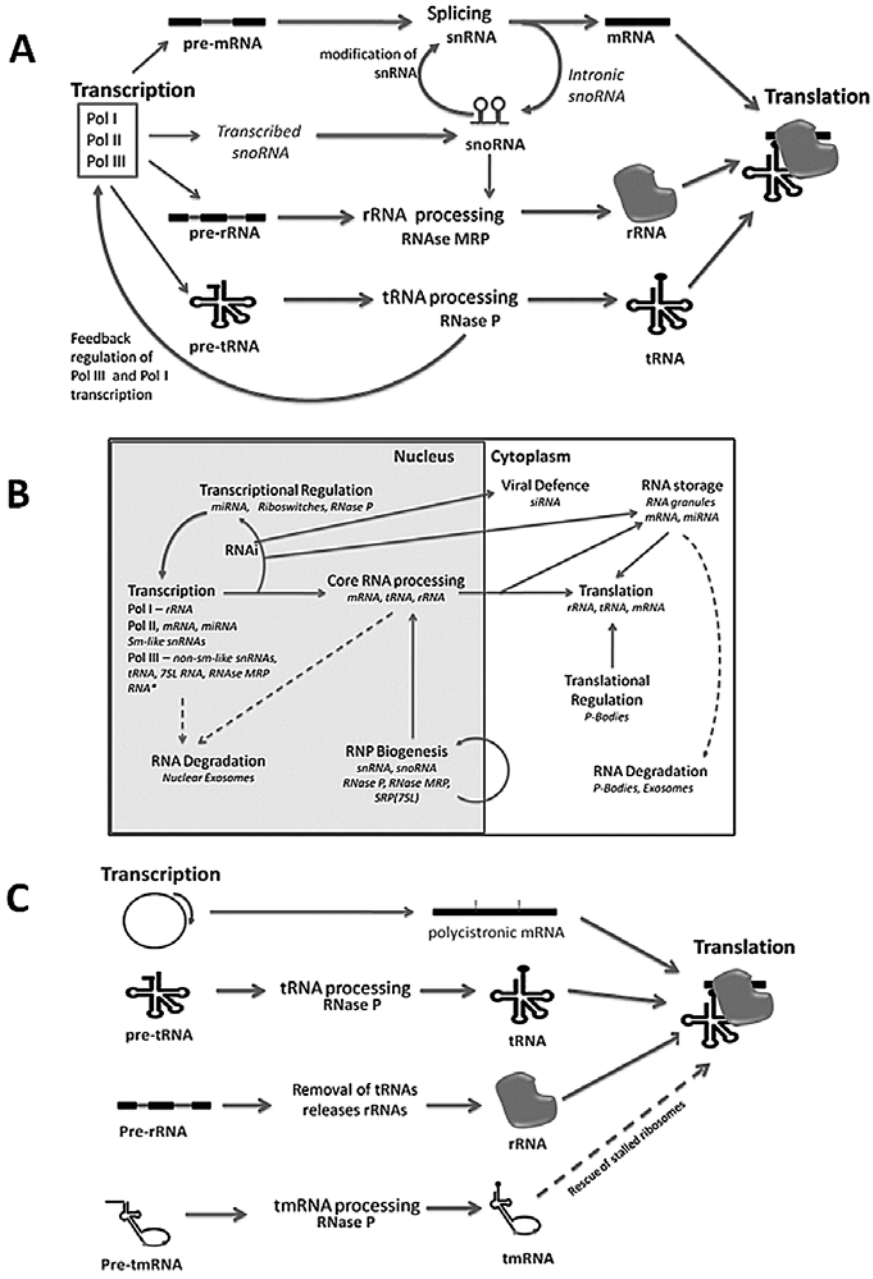

Figure 1. RNAs processing other RNAs are the key feature of the RNA-infrastructure. A) In transcription-to-translation machineries in eukaryotes, ncRNAs are involved in the processing of mRNA, tRNA and rRNA. B) In eukaryotes, compartmentalisation and biogenesis pathways permit regulation of these processes in the RNA infrastructure. *MRP RNA may not be in all eukaryotes. C) In prokaryotes in general, there is still processing of tRNAs and rRNAs but less mRNA processing. A and B adapted from[1] with permission from authors. A color version of this figure is available at www.landesbioscience.com/curie.

RNA processing it is not surprising that many proteins are now seen as having roles in splicing as well as their own function (e.g., transcription or capping). However, it remains to be seen whether these proteins actually influence catalysis in the spliceosome, or are detected due to the close proximity of these RNA processing complexes.

Similarly, transcribed pre-tRNAs require processing before being able to function as amino acid transfer molecules for translation. Leader sequences at the 5′ and 3′ ends of the pre-tRNAs require cleaving, introns within the tRNA may need to be removed and in some cases a 3′ CCA tail needs to be added.[12] In addition, certain nucleotides within the tRNA require modification by aminoacylation. The ribonucleoprotein RNase P is responsible for cleaving the 5′ leader sequence of pre-tRNAs in all cells, although the overall structure of this protein-RNA complex differs in eukaryotes, bacteria and archaea. In bacteria there is one small protein that plays diverse roles such as enhancing substrate binding, altering substrate recognition, stabilising RNA conformation, and aiding catalysis by discriminating between the substrate and product by binding to the 5′ leader sequence of the pre-tRNA.[13,14] Eukaryotes have 9-10 proteins in the complex with a single RNA. Archaeal RNase P also has multiple proteins (five including the ribosomal protein L7Ae) which do show some homology to some of the eukaryotic RNase P proteins. The RNase P RNA from some representatives from each kingdom can be induced to perform weak catalysis without its accompanying proteins, but only with high salt and high cation conditions in vitro (summarized in ref. 15).

RNase P plays key networking roles in both the eukaryote's and prokaryote's RNA infrastructure, resulting in the cleavage of additional substrates and the repression of transcription (Fig. 2).[16,17] In bacteria, as well as cleaving the 5′ leader sequence of tRNAs, it cleaves a similar leader sequence for tmRNA. tmRNA (transfer-messenger RNA) is a specialised tRNA molecule that together with the SmpB protein (small protein B) rescues stalled ribosomes in a process called trans-translation (reviewed in ref. 18).With a structure partly a tRNA molecule and partly an mRNA molecule,[19] the tRNA part binds to the stalled ribosome, allows the translation to proceed along the mRNA part which encodes a distinctive degradation signal and a translation stop signal. When the mistranslated protein is released after the stop signal it is targeted for degradation. This process of trans-translation is conserved throughout bacteria and is also present in some mitochondria and chloroplasts.[20,21] In prokaryotes other cleavage products by RNase P include some riboswitches and some viral RNAs as well as the 4.5S rRNA which is part of the Signal Recognition Particle involved in post-translational transport (reviewed in ref. 22).The eukaryotic counterpart of the 4.5S rRNA (7SL RNA) is also cleaved by RNase P (reviewed in ref. 13). In yeast the HRA1RNA and some C/D box snoRNAs are processed by RNase P although whether cleavage is the exact mechanism is yet to be completely determined. In humans MALAT1, another long ncRNA, is cleaved by RNase P. tRNAs in organelles within eukaryotes (in some species) either encode their own RNase P RNA (e.g., the yeast *S. cerevisiae*), use the nuclear counterpart (e.g., humans) or occasionally do without the RNA component altogether.[23] Additionally, in the archaeans *Nanoarchaeum equitans and Pyrobaculum aerophilum* and the hyperthermophilic bacterium *Aquifex aeolicus*, there does not appear to be any RNase P-like RNA sequence in their genomes.[24] In *N. equitans* the requirement for RNase P has been replaced by a strict placement of the promoter 26 nucleotides upstream of the mature tRNA sequence allowing transcription of leaderless tRNAs.[25]

There is a feedback affect of RNase P on its own polymerase RNA Pol III[16] and the polymerase affecting rRNA transcription, RNA Pol I.[17] Thus, RNase P in eukaryotes has a large effect on other aspects of RNA processing including splicing (U6 snRNA is transcribed by RNA Pol III), and RNA modification (by some yeast C/D box snoRNAs).

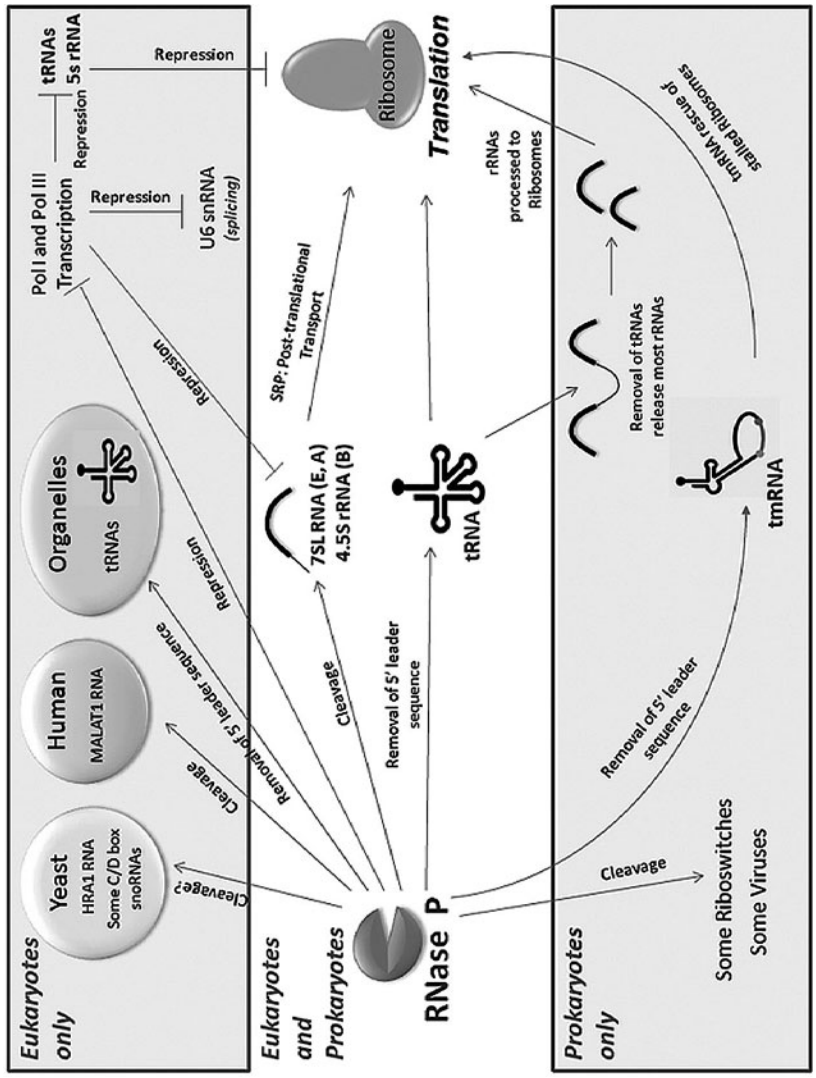

Figure 2. RNase P is central to RNA processing in eukaryotes and prokaryotes. In both eukaryotes and prokaryotes, RNase P cleaves tRNA leader sequences and also the SRP RNA (7SL RNA, called 4.5S rRNA in bacteria). Within eukaryotes RNase P also interacts with other RNAs in repressing Pol I and Pol III transcription which affects splicing and rRNA maturation. Within prokarytoes, RNase P also affects the processing of tmRNA which rescues stalled ribosomes, as well as affecting rRNA processing and cleaving some riboswitches and viruses. A color version of this figure is available at www.landesbioscience.com/curie.

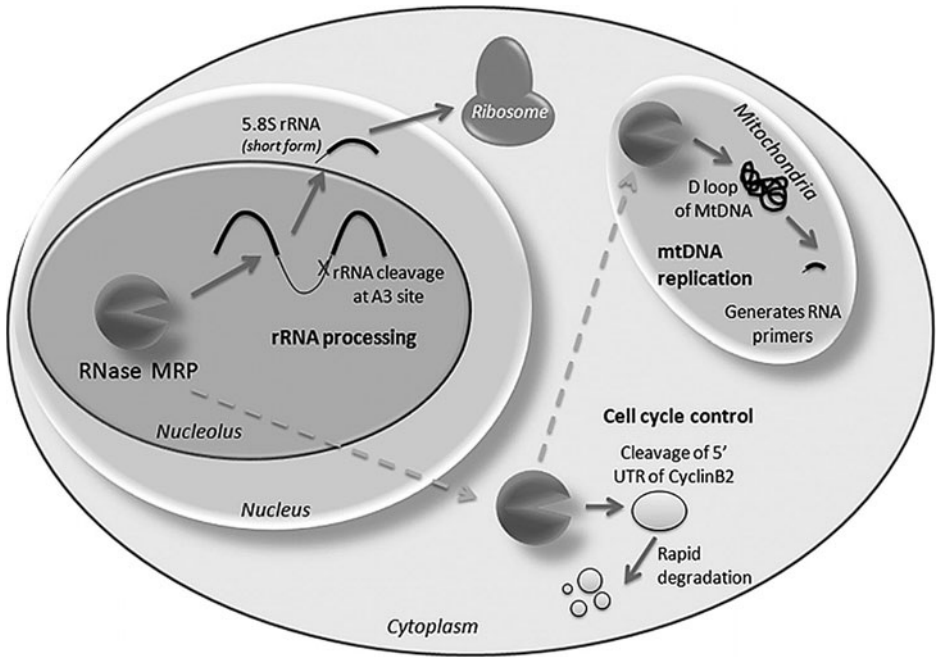

Figure 3. RNase MRP has different functions in different cellular compartments. Within the nucleolus it is involved in rRNA processing. However, in the cytoplasm it cleaves the leader sequence of the cell cycle control protein CyclinB2, and in the mitochondria it is crucial for mtRNA replication where it cleaves the D loop of the mtDNA to generate RNA primers. Whether this macromolecule consisting of one catalytic RNA and ~9 proteins moves as a whole through the cell, or disassembles and re-assembles at the different areas in which it functions, is not yet known. A color version of this figure is available at www.landesbioscience.com/curie.

It clearly plays a central role in the RNA infrastructure of both eukaryotes and prokaryotes and it is likely that other substrates and processing connections, especially in prokaryotes, are still to be uncovered.

RNase MRP is a ribonucleoprotein found only in eukaryotes, but closely related and sharing many of the same proteins with RNase P (for a review see ref. 13). It too has multiple roles (Fig. 3), processing the A3 site of rRNA in the nucleolus, a critical cell cycle control protein (Cyclin B2) in the cytoplasm, and the D-loop of mitochondrial DNA (MtDNA) in the mitochondria to generate RNA primers for Mt DNA replication. This is a good illustration of the spatial nature of RNA-Protein complexes that have different roles in different cellular compartments. RNase MRP is transcribed by RNA Pol III and thus is affected by the RNase P feedback on the polymerase.

Other aspects of rRNA processing in eukaryotes are linked to transcription and downstream rRNA maturation. Extensive modification of the pre-rRNAs includes methylation of 2′ hydroxyl groups of ribose (guided by C/D box snoRNAs [small nucleolar RNAs]) and pseudouridine formation from uracil (guided by H/ACA snoRNAs).[26] In vertebrates, these snoRNAs are mostly found within introns, and are spliced out by snRNAs, illustrating the strong network of RNA biogenesis and splicing machineries. Yeast models (primarily in *S. cerevisiae*) indicate that RNA Pol I, elongation factors and

rRNA sequence elements appear to optimize transcription elongation and co-ordinate interactions (including those with snoRNAs) with the pre-rRNA for correct rRNA processing and ribosome assembly.[27] In addition, a protein complex of three transcription factors (the CURI complex comprising of Rap1, Fhl1 and Ifh1) links ribosomal protein production and pre-rRNA processing.[28] Thus, rRNA processing also uses feedback from the later stages of processing to regulate transcription.

SPATIAL REGULATION OF EUKARYOTIC RNA PROCESSING

Spatial placement of both RNA and protein macromolecular components plays an important part in the regulation of RNA-processing. In eukaryotes, this is clearly demonstrated by how RNAs move through nuclear bodies (such as Cajal bodies, Gems and nucleoli) and for some of them, into cytoplasmic bodies such as P-bodies and RNA granules. As an example, Figure 4 illustrates the biogenesis of snRNAs and snoRNAs in

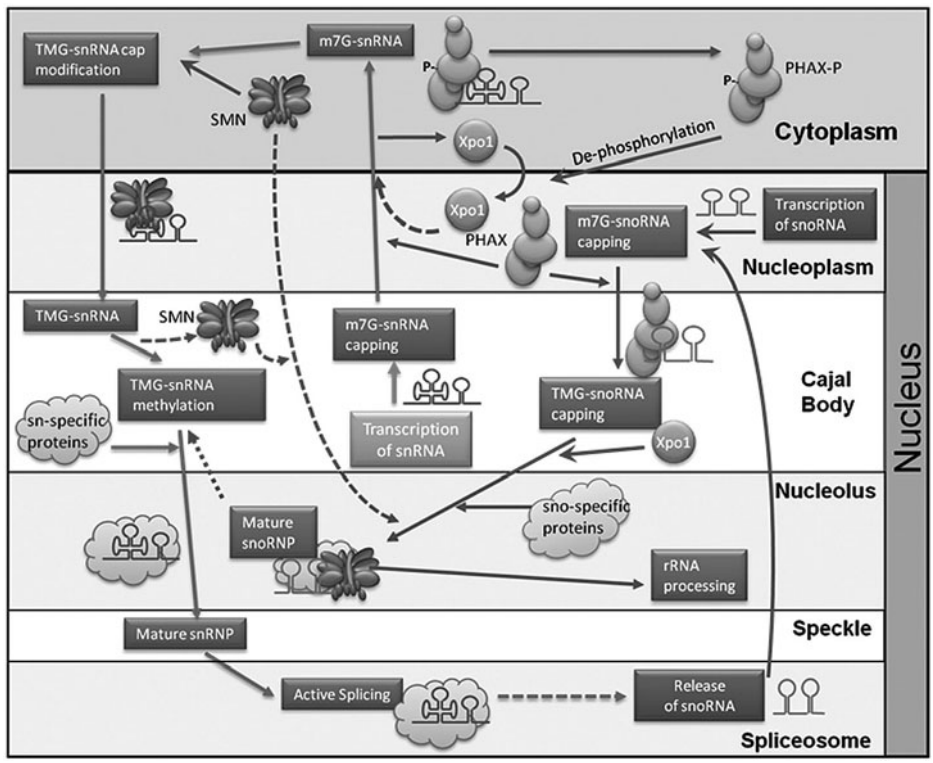

Figure 4. The network of Sm-class snRNA and snoRNA biogenesis pathways connected by the SMN complex and the PHAX complex. Transcribed snRNAs move through nuclear compartments during initial processing then into the cytoplasm using the PHAX complex where they gain the SMN complex. After this, the snRNA/SMN macromolecule moves back into the nucleus for further maturation before being used for active splicing. In contrast snoRNAs do not enter the cytoplasm but instead use the PHAX complex for intranuclear transport and the SMN complex for macromolecule maturation. Figure adapted with permission of authors. A color version of this figure is available at www.landesbioscience.com/curie.

humans. Typically there are different stages of RNA-processing taking place within different nuclear sub-compartments, but for the Sm-class-snRNAs, the processing moves to the cytoplasm, before the re-import of the snRNP-complexes back to the nucleus. In contrast, the Lsm-class snRNAs (U6 and $U6_{atac}$ snRNAs) in humans never leave the nucleus,[29] although in yeast there may be some nuclear export and re-import of U6 snRNA.[29] Cajal Bodies in particular appear to be important sub-nuclear compartments for RNPs since they are not only repositories for the biogenesis of RNPs. Mature snRNPs travel through Cajal Bodies, sometimes moving from one Cajal body to another suggesting that the Cajal Body is being used as a 'recycling center', enabling the re-assembly of the tri-snRNPs.[30] In contrast, the assembly of C/D box snoRNPs appears to occur cotranscriptionally, but much of the intra-nuclear and intra-cellular trafficking of snoRNPs remain to be characterised.[31]

A feature of intra-cellular RNP trafficking is how some proteins assist these different RNPs in different manners. One such group of proteins linking snRNA and snoRNA biogenesis (Fig. 4) is the PHAX complex (consisting of PHAX, Cap Binding Protein (CBC), CRM1 and RanGTP) which in humans at least, transports snRNAs from the nucleus to the cytoplasm as well as transport of some snoRNAs (especially U3, U8, U13) around the nucleus to speckles, Cajal bodies and nucleoli.[32] Although PHAX is a metazoan protein there has been a similar protein characterised in the protist *Cryptosporidium parvum*.[32] Another important RNA-escorting macromolecule is the SMN protein complex, which is found in the nucleoplasm and nuclear bodies called Gems.[33] The SMN complex scrutinizes cellular RNAs to ensure that Sm cores (of highly reactive RNA-binding Sm proteins) are only assembled on proper snRNAs,[34] and the Gemin5 protein of this SMN complex can distinguish snRNAs from other cellular RNAs for snRNP biogenesis.[34] The SMN complex also plays a role in other biogenesis pathways including those for hnRNPs and microRNPs.[33] The above pathways for snRNP and snoRNP biogenesis have been largely characterised for mammalian and yeast systems, and although there is now some plant information,[35] there is little known about how these RNPs complexes form in the many different groups of protists. As with plants we expect some different proteins to be involved and there will likely be different pathways.

After nuclear export some mRNAs are translated immediately, but many mRNAs are recruited to RNA granules (See Chapter 8, Table 1 pg 124 for cellular component definitions) in the cytoplasm until developmental or environmental cues signal their translation.[36] Cytoplasmic RNA granules (reviewed in refs. 36,37)include Processing-bodies and Stress Granules as well as compartments found in germ cells (polar and germinal granules) and neurons (neuronal granules). Processing-bodies (P-bodies or GW bodies) are involved with post-transcriptional processes, including mRNA degradation, nonsense-mediated mRNA decay (NMD), translational repression and RNA-mediated gene silencing (reviewed in ref. 38). mRNA degradation is initiated by the deadenylation (shortening) of the 3′ polyA-tail followed by decapping.[39] Stress Granules are a cytoplasmic RNA granule that typically forms during stress response (whereas P-bodies are present continuously).[40] Stress Granules contain polyadenylated transcripts and are not degraded, making them available for rapid re-initiation after stress recovery,[40] whereas mRNAs recruited to P-bodies are largely deadenylated.[37] mRNAs within stress granules P-bodies and Stress Granules constantly exchange RNAs and proteins with the cytosol[37] and mRNAs can move from one to the other. P-bodies have been investigated in yeasts, plants, trypanosomatids, insects and vertebrates (reviewed in ref. 37) and thus are likely to be important eukaryotic RNA-based cellular features. Evidence is suggesting that Stress Granules are the consequence, not the cause

of the shut off of translation during stress, and the formation of critical macromolecules may be linked to the sequestering of key components (reviewed in ref. 37).

Although RNAs such as miRNAs, siRNAs and tRNAs typically act in the cytoplasm, there are some miRNAs that may re-enter the nucleus, possibly playing a role in modification or nuclear component assembly processes (summarized in refs. 41,42). tRNAs in particular show interesting nuclear-cytoplasmic dynamics. tRNA transcription and 5′ processing is typically in the nucleolus, however 3′ processing has been found mostly in the nuceloplasm; tRNA modification is usually in the cytoplasm but in some species, tRNA splicing takes place on the mitochondrial cytoplasmic surface (reviewed in ref. 42). Subsequently however, a retrograde pathway exists where the tRNAs are imported from the cytoplasm back into the nucleus[43] but can then be re-exported to the cytoplasm in response to nutrient availability.[44] Although major studies of tRNA cellular dynamics has been mainly in yeast, the retrograde process (moving cytoplasm to nucleus) at least appears conserved in vertebrates (summarized in ref. 42).

Spatiotemporal movement of RNAs or key components of RNA-based machineries, is not restricted to eukaryotes. The Gram-negative bacteria *Caulobacter crescentus* is dimorphic in that it has a stalked form that adheres to surfaces (with a holdfast and stalk), and a swarmer form that is mobile with a flagellum. Often used as a model for bacterial cell cycle and cell differentiation studies, *C. crescentus* shows substructure localisation and temporal timing of trans-translation.[45-47] tmRNA and its small protein SmpB are colocalised to a helix-like pattern in swarmer and predivisional cells but they are delocalised in stalked cells.[46] However, the protein RNase R which interacts with the tmRNA is localised separately to another helix-like pattern that is out of phase with the tmRNA-SmpB pattern. Trans-translation requires that the individual tmRNA-SmpB molecules would have to disassociate from the helix-like structure in order to pass through the ribosome, and it is feasible that these structures facilitate the regulation of trans-translation.[46] In a possible feedback mechanism, the tmRNA of *C. crescentus* is regulated in the cell cycle by temporally controlled transcription and translation.[45] With trans-translation required for many functions across bacteria, including sporolation in *Bacillus subtilis*,[48] symbiosis in *Bradyrhizobium japonicum*[109], and pathogenecity in *Salmonella enteria* (summarized in ref. 47), the tight regulation of the RNA component is not unexpected. It is even possible that these helix-like structures seen in *C. crescentus* are analogous to the P-bodies we find in eukaryotes.

RNA REGULATION, CONNECTING COMPONENTS OF THE RNA-INFRASTRUCTURE

RNAi Networks

RNA regulation (including RNAi, riboswitches and RNA-editing), storage and degradation are linked to the processes discussed in the earlier sections. RNAi (RNA interference involving miRNAs, siRNAs and piRNAs are reviewed in refs. 49, 50). Although best known for roles in regulating mRNA levels, RNAi is also directly involved in many cellular processes including chromatin-mediated gene silencing and DNA re-arrangements. It is also not a matter of one target to one regulator. It has been shown[51] that a single miRNA can directly or indirectly down-regulate the production of thousands of genes. Although RNAi as a mechanism appears general in eukaryotes, the

timing and location of miRNA expression varies even between vertebrates due to changes in miRNA copy number, genomic context (either exclusively intergenic, or intronic and intergenic) or both.[52] There can also be expression differences when there is conservation of the miRNA sequence.[52]

Other forms of RNA-based transcriptional regulation include regulation by RNase P which has a positive effect on Pol III promoter activity.[16,53] RNase P associates with the chromatin of tRNA and 5S rRNA genes which contain the Type-1 Pol III promoter sequences, but not with the U6 snRNA and 7SL-RNA that have Type-3 Pol III promoter sequences. Transcription of these Pol III transcribed ncRNAs declines sharply in extracts depleted of active RNase P.[16] RNase P may also have a role in the splicing-independent maturation of snoRNAs as recently demonstrated in yeast[54] linking the production of tRNAs and rRNA.

A number of longer ncRNAs directly target transcription (reviewed in ref. 55) including SRA (a transcriptional co-activator for several steroid-hormone receptors), NRSE (Neuron-restrictive silencer element dsRNA), HSR1 (heat shock RNA-1) and 7SK RNA. This latter ncRNA is transcribed by Pol III and represses transcript elongation by Pol II (also reviewed in ref. 55). Another instance is a regulatory transcript from a minor promoter interfering with the expression of the main transcript.[56] With the ongoing discovery[57] of new ncRNAs in a wider range of eukaryotes we certainly expect the identification of other direct transcriptional regulators.

An interesting trend is the discovery that small regulatory RNAs can be derived from other ncRNAs. A number of studies have characterised miRNA-like RNAs derived from snoRNA-derived RNAs[58] and RNAs derived from the Vault RNA.[59] tRNA-derived RNAs are thought to be involved in translational repression.[60,61] Studied in mammals, plants, fungi, and the protists *Giardia* and *Tetrahymena*, tRNAs are cleaved by members of the Ribonuclease A or T2 protein families in the anticodon loop forming 5′ and 3′ tRNA halves. Although how these different tRNA halves regulate translation inhibition is still very much under investigation, in mammals it has been shown that 5′ tRNA halves induces Stress Granule formation[62] and that the original cleavage is enhanced by stress. In Trypanosomes, granules are formed that are distinct from Stress Granules.[37,63] Other translational-inhibition small RNAs include qiRNAs (QDE-2 associated RNAs) from the fungus Neurospora which inhibit protein translation during DNA damage response.[64] With mass RNA sequencing still in its early days, these may represent only a fraction of the real amount of derived regulatory RNAs.

Transcription-initiation RNAs* are typically transcribed from a repeat motif called a 'spanion cluster'.[65] These RNAs have a strong preference towards transcription initiation sites. Other small regulatory RNAs of note are the tiny RNAs of mammals, which are 17-18 bp in length and have a connection to splicing in that their 3′ ends map precisely to the splice donor site of internal exons.[66] A subgroup of these splice-site RNAs are seen to be associated with highly expressed genes.[66] How widespread these types of regulatory RNAs are remains to be investigated, but high throughput sequencing technology has enabled researchers to uncover these types and more.

RNAi is seen thus as a typical eukaryotic feature but there are some lineages that have lost their RNAi proteins but some still maintain some form of ncRNA-based regulation. The yeast *S. cerevisiae* does not have the 'standard' RNAi system since it lacks Dicer-like RNases, Argonaute or Piwi-like proteins, but it does have ncRNAs

* In a wave of confusion we are also seeing different types of regulatory RNAs given the same prefix. Two types of small RNA, transcription-initiation RNAs,[65] and tiny RNAs[66] have both been given the name tiRNAs. To avoid confusion the expanded name rather than tiRNA will be used for both cases.

that act in the regulation of its genes.[67] These ncRNAs including the 'cryptic unstable transripts' (CUTS), tend to stem from bidirectional transcription and may be passive by-products of transcriptional noise rather than any specific mechanism.[67] In single celled protists, some species of Trypanosomes have lost their RNAi systems while others have retained them.[68] Additionally, for some Trypanosomes the retention of associated viruses is necessary and it has been suggested that the loss of RNAi has facilitated viral retention. However genome plastidity is also a potential effect of this loss.[68] Whether these RNAi-less Trypanosomes have evolved a different type of RNA-regulation system to compensate, is as yet unknown.

RNA Networks and Epigenetics

ncRNAs play a major role in epigenetics[69-74] and include networks consisting of long ncRNAs (such as XIST and HOTAIR), and short ncRNAs[69,74,75] such as miRNAs, siRNAs and piRNAs. miRNAs in particular have been shown to be important in RNA networks behind stem-cell self-renewal and differentiation (reviewed in ref. 76). In general, there are two types of stem-cell, tissue stem cells (which include somatic and germline cells which develop, maintain and repair tissues in developing and adult organisms), and embryonic stem cells (ES) which develop from an embryo to give rise to the foetus. In one example of a miRNA-epigenetic network, the expression of the miR-290-295 miRNA cluster (a group of miRNAs that share a 5′ proximal AAGUGC motif) increases during pre-implantation development and remains high in undifferentiated ES cells, but then decreases after ES cell differentiation.[77] These miRNAs act as post-transcriptional regulators of retinoblastoma-like 2 (Rbl2) which in turn acts as a transcriptional repressor of DNA methyl transferases (DNMTs), Dnmt3a and Dnmt3b. DNMTs epigenetically silence OCT4, a key transcription factor of ES cell renewal and differentiation.[77,78] Alternatively if Dicer is knocked out, miRNAs are depleted and the methylation of the Oct4 promoter is severely impaired during differentiation. Many other candidate targets of the AAGUGC seed-containing miRNAs have been identified as well as many indirectly regulated targets,[77] but it remains to be seen how other aspects of self-renewal and differentiation are affected by the miR-290 cluster.

Networks involving multiple long ncRNAs (defined generally as having a length > 200 nt) are also known, with a classic example being the long ncRNA control of X-chromosome inactivation (reviewed in refs. 74,79). In mammals the potential double dosage of gene expression from the X chromosome in XX females (when compared to XY males) is controlled by inactivating one of the X chromosomes. In mice there are two forms of X Chromosome inactivation (XCI)[79] where XCI is imprinted in extra-embryonic tissues and the paternal X (Xp) is inactivated. Further along in development just before the embryo proper, the Xp is re-activated after which random XCI is initiated during early embryonic development. In humans XCI is randomly activated but it is not clear if the imprinted form is present.[79] During random XCI in humans (Fig. 5), the long ncRNA Xist is repressed on the future active X chromosome Xa by another long ncRNA Tsix,[80] and activated on the future inactive X chromosome by a third long ncRNA Jpx.[81]

This complex network of long ncRNA and methylation processes can be seen if we examine this system in more detail (Fig. 5). In mammals, in pre-XCI embryonic stem cells (ES) Tsix is transcribed at a higher level than Xist, and triggers H3-K4 dimethylation along both the *Xist* and *Tsix* genes. Xist becomes elevated when the major pluripotency factors Nanog, Oct3/4 and Sox2 dissociate from intron 1 within *Xist* initiating XCI. One

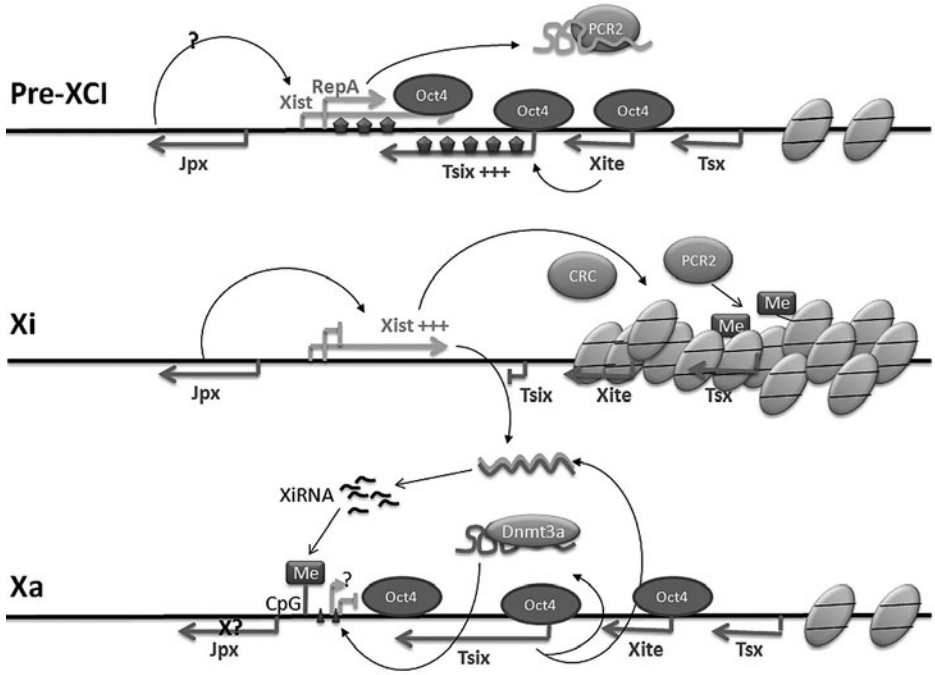

Figure 5. RNA networking during mammalian random X chromosome inactivation (XCI). During Pre-XCI, Tsix and RepA, compete for binding to the polycomb repressive complex PRC2. Tsix is expressed at a high level upon XCI and triggers dimethylation (wide vertical arrows) along both the *Xist* and *Tsix* genes, leading to active transcription of Xist and Tsix. On the future inactive X chromosome Xi, Oct4 binding is lost so Tsix is downregulated and Xist induced and is further enhanced by Jpx. A coating of Xist RNA forms a chromatin compartment and recruits the chromatin repressive complex (CRC) to the Xi. The inactive status of this chromosome is maintained by PCR2. On the future active X chromosome Xa, Oct4 is retained and maintains Tsix expression. Tsix, associated with methyltransferase Dnm3a, directs the methylation on the *Xist* promoter and Xist is repressed. Dicer-dependent XiRNAs are possibly produced from Xist and Tsix ncRNA duplexes and could direct methylation along the future Xi, and also direct methylation of CpG islands of the Xist promoter region in the Xa. Gene distances are not to scale. Figure adapted from[74] with permission from authors. A color version of this figure is available at www.landesbioscience.com/curie.

of these proteins Oct4 is known to active Tsix and another RNA region Xite which is an activator of Tsix. Oct4 also acts as a repressor of Xist aiding in the control of the Xist:Tsix balance in XCI.[82] During XCI different events occur upon the future active X chromosome (Xa) and the future inactive X chromosome (Xi).

On the future inactive X chromosome (Xi), Oct4 binding is lost so Tsix is downregulated and Xist is induced.[82] A coating of Xist RNA forms a silent chromatin compartment where X-linked genes become 'localised' through binding to the Xist RNA.[83] Xist RNA also recruits the chromatin repressive complex (CRC) to Xi. The inactive status is maintained by the polycomb repressive complex PCR2. In pre-XCI Tsix and another RNA, Repeat A (RepA), compete for PRC2 binding[84] but upon XCI, RepA recruits PRC2 to Xist and PRC2 methylates Xist at H3K27 to upregulate Xist[84]. RepA also collaborates with the long ncRNA Jpx in an as yet unknown mechanism to transcriptionally activate Xist.[81] On the future active X chromosome Xa, Oct4 is retained and maintains Tsix expression. Tsix,